Graph Contrastive Learning with Low-Rank Regularization and Low-Rank Attention for Noisy Node Classification

Yancheng Wang[1] **Ping Li**[2] **Yingzhen Yang**[1]

[1]School of Computing and Augmented Intelligence, Arizona State University, Tempe, AZ 85281, USA, yancheng.wang@asu.edu, yingzhen.yang@asu.edu
[2]VecML Inc., Bellevue, WA 98004, USA, pingli98@gmail.com

Abstract

Graph Neural Networks (GNNs) have shown strong performance in learning node representations and node classification, but their effectiveness can be substantially degraded by noise in real-world graph data. To address this challenge, we introduce a robust and innovative node representation learning method named Graph Contrastive Learning with Low-Rank Regularization, or GCL-LRR, which follows a two-stage transductive learning framework for node classification. In the first stage, the GCL-LRR encoder is trained via prototypical contrastive learning with a low-rank regularization objective, and in the second stage, the learned representations are used by a linear transductive classifier to predict labels for unlabeled nodes. Our GCL-LRR is inspired by the Low Frequency Property (LFP) of the graph data and its labels, and it is also theoretically motivated by our sharp generalization bound for transductive learning. Our theoretical result is among the first to theoretically demonstrate the advantage of low-rank regularization in transductive learning, which is also supported by strong empirical results. To further enhance the performance of GCL-LRR, we present an improved model named GCL-LR-Attention, which incorporates a novel LR-Attention layer into GCL-LRR. GCL-LR-Attention reduces the kernel complexity of GCL-LRR and contributes to a tighter generalization bound, leading to improved performance. Extensive evaluations on standard benchmark datasets evidence the effectiveness and robustness of both GCL-LRR and GCL-LR-Attention. The code of GCL-LRR and GCL-LR-Attention is available at https://github.com/Statistical-Deep-Learning/GCL-LR-Attention.

1 INTRODUCTION

Graph Neural Networks (GNNs) are widely recognized as effective tools for node representation learning via message passing over graph structures [Kipf and Welling, 2017, Bruna et al., 2014, Hamilton et al., 2017, Xu et al., 2019b, Zhu and Koniusz, 2021]. However, most existing GNN methods inadequately handle noise in graph data [Zhu et al., 2024, Zhong et al., 2019], where attribute or label noise can significantly degrade generalization [Patrini et al., 2017]. Noise from a few nodes can propagate through the structure and affect others [Dai et al., 2021, Wang et al., 2023, 2024c], highlighting the need for GNN models that can learn effectively under noisy inputs. In this paper, we study the problem of learning robust node representations for semi-supervised node classification on noisy graphs affected by noise in both node labels and node attributes. Specifically, symmetric and asymmetric label noise are considered under standard label-noise modeling settings widely adopted in existing robust graph learning methods [Dai et al., 2022, Qian et al., 2023, Zhuang and Al Hasan, 2022, Dai et al., 2021, Wang et al., 2023, 2024c], where symmetric noise corresponds to uniformly random label flips across classes, while asymmetric noise involves class-dependent mislabeling patterns. In addition, attribute noise is also considered, where noise is introduced by randomly perturbing the input node features, following the settings in prior works [Ding et al., 2022, Li et al., 2024a, Ju et al., 2025]. Detailed settings for the label noise and attribute noise studied in this paper are presented in Section A of the supplementary.

Limitations of Existing Robust Graph Learning Methods on Noisy Graph Data. Traditional strategies for robust learning generally fall into two categories. The first modifies the loss function to accommodate corrupted data [Patrini et al., 2017, Goldberger and Ben-Reuven, 2017], while the second removes samples suspected to be noisy [Malach and Shalev-Shwartz, 2017, Jiang et al., 2018, Yu et al., 2019, Li et al., 2020, Han et al., 2018]. Although some of these ideas have been adapted for graph settings [Dai et al., 2021,

Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence (UAI 2026), PMLR 337:7200–7228.

Qian et al., 2023, Zhuang and Al Hasan, 2022], they often depend on heuristics and lack theoretical backing in semi-supervised scenarios. To this end, we introduce Graph Contrastive Learning with Low-Rank Regularization, abbreviated as GCL-LRR, which is a new encoder aimed at improving both robustness and generalization in node representation learning. The design of our GCL-LRR model is inspired by the low-frequency nature of graph signals, and it is also theoretically supported by a new generalization bound developed for transductive learning. To the best of our knowledge, this paper is among the first to provide a principled justification for the advantage of low-rank representation learning in graph contrastive settings. Experimental evaluations conducted on widely used benchmarks demonstrate that GCL-LRR consistently achieves strong robustness and high performance.

Although GNNs inherently act as low-pass filters, they do not explicitly model low-frequency signals, limiting their ability to leverage the Low-Frequency Property (LFP) under noisy labels. As shown in Figure 1, clean label information is concentrated in the low-rank structure of the observed label space. Unlike conventional GNNs, the GCL-LRR encoder explicitly captures this structure by learning low-rank representations. Prior work [Cheng et al., 2021] demonstrates that low-rank learning mitigates attribute noise, while recent graph attention studies emphasize balancing low- and high-frequency components [Choi et al., 2024a, Zhang et al., 2024a]. In comparison, GCL-LRR offers an improved balance by promoting low-frequency information through minimization of the Truncated Nuclear Norm (TNN), which aligns with the LFP. Our GCL-LRR method is also grounded in theory through our novel generalization bound in the transductive setting. To further boost the performance of GCL-LRR, we introduce an enhanced model termed GCL-LR-Attention, which integrates a novel LR-Attention layer into GCL-LRR. Existing works on GCL [Li et al., 2024b, Yuan et al., 2023, Jin et al., 2021, Mo et al., 2022, Zhang et al., 2023] suggest that node representations learned by GCL are usually more robust to input noise than GCN. Compared to GCN, GCL enforces semantic consistency across multiple stochastically augmented views generated by injecting noise into the input graph. This is also evidenced by the ablation study in Section B.14, where GCL-LR-Attention significantly outperforms GCN-LR-Attention, which is the GCN with low-rank regularization and LR-Attention, thereby justifying the use of the GCL encoder in this paper.

GCL-LR-Attention reduces the kernel complexity of GCL-LRR by the LR-Attention layer and yields a tighter generalization error bound than GCL-LRR, leading to even better performance than GCL-LRR. Performance comparisons in Table 1 of Section 5.2 demonstrate that GCL-LRR and GCL-LR-Attention outperform state-of-the-art methods based on attention and transformers, such as GFSA [Choi et al., 2024a] and HONGAT [Zhang et al., 2024a]. Furthermore, the trade-off between low- and high-frequency learning is evaluated through kernel complexity, a metric detailed in Section 4.2. As shown in Table 2 of Section 5.3, GCL-LRR achieves lower kernel complexity and a tighter upper bound on test loss than competing graph contrastive and attention-based methods, and GCL-LR-Attention further reduces the kernel complexity and the generalization bound via the LR-Attention layer, resulting in superior node classification performance.

Contributions. Our contributions are as follows.

First, we introduce Graph Contrastive Learning with Low-Rank Regularization (GCL-LRR), a novel and theoretically grounded GCL encoder. The design of GCL-LRR is motivated by the LFP illustrated in Figure 1, which shows that a low-rank projection of the clean label matrix captures most of its informative content. In contrast, label noise tends to spread uniformly across all eigenvectors of the classification kernel matrix. Drawing from this insight, GCL-LRR incorporates the TNN as a low-rank regularization component into the loss function of standard prototypical graph contrastive learning. This regularization encourages the model to produce inherently low-rank representations, which are subsequently used in a linear transductive classifier.

Second, we establish a strong theoretical foundation that guarantees the generalization performance of the linear transductive algorithm when applied to the low-rank features generated by GCL-LRR. In particular, we derive a novel generalization bound on the test loss for unlabeled nodes. To the best of our knowledge, this theoretical result is among the first demonstrating the benefit of learning low-rank representations for robust transductive classification under label noise. Motivated by this theoretical result and in pursuit of further performance gains, we introduce an enhanced variant of our model that incorporates a novel LR-Attention layer, termed the GCL-LR-Attention. GCL-LR-Attention achieves a further reduction in kernel complexity compared to GCL-LRR, which leads to a tighter bound on the generalization error in the transductive setting and improved empirical performance. As shown in Table 2 of Section 5.3, GCL-LRR achieves a lower generalization upper bound than existing methods, and GCL-LR-Attention further tightens this bound beyond GCL-LRR. Comprehensive experiments demonstrate the superiority of both GCL-LRR and GCL-LR-Attention over existing methods in node classification tasks on noisy graph data.

2 RELATED WORKS

2.1 ROBUST GRAPH NEURAL NETWORKS

Existing GNNs learn node representations through message passing over graph neighborhoods [Bruna et al., 2014, Kipf

and Welling, 2017, Hamilton et al., 2017, Veličković et al., 2017, Xu et al., 2019b]. In recent years, contrastive learning has become prominent on graphs [Feng et al., 2022a, Zhang et al., 2023, Lin et al., 2023], where most graph contrastive learning methods [Jiao et al., 2020, Peng et al., 2020, You et al., 2021, Jin et al., 2021, Mo et al., 2022] generate multiple augmented views and align the corresponding embeddings. Recent works [Xu et al., 2021, Guo et al., 2022, Li et al., 2021] further improve GCL by introducing semantic prototypes [Snell et al., 2017, Arik and Pfister, 2020, Allen et al., 2019, Xu et al., 2020]. Deep neural networks, including GNNs, are vulnerable to noisy inputs [Zhang et al., 2021], and existing robust learning methods mostly follow two strategies: loss correction, which modifies training objectives to handle corrupted data [Patrini et al., 2017, Goldberger and Ben-Reuven, 2017], and sample selection, which identifies clean samples for training [Malach and Shalev-Shwartz, 2017, Jiang et al., 2018, Yu et al., 2019, Li et al., 2020, Han et al., 2018]. Graph learning methods improve robustness via structural prediction, label denoising, and self-supervised objectives [Dai et al., 2021, Qian et al., 2023, Zhuang and Al Hasan, 2022, Li et al., 2024b, Yuan et al., 2023].

2.2 LEARNING LOW-FREQUENCY SIGNAL IN GRAPHS

Traditional GNNs, such as GCN [Kipf and Welling, 2017], aggregate information from neighboring nodes and thus act as low-pass filters, emphasizing low-frequency signals in graph structures and node features [NT and Maehara, 2019, Xu et al., 2019a, Wu et al., 2019, Yu and Qin, 2020]. However, excessive reliance on low-frequency components can lead to over-smoothing [Bo et al., 2021, Zhang et al., 2024b, Dong et al., 2025, Sun et al., 2022]. To mitigate this issue, recent works [Bo et al., 2021, Dong et al., 2021, Ju et al., 2022] propose adaptive mechanisms that balance low- and high-frequency signals [Tang et al., 2025]. Meanwhile, learning from low-rank graph features and structures has been shown to enhance robustness under noise [Tang et al., 2024, Yang et al., 2023]. Graph attention models such as GAT [Veličković et al., 2017] also emphasize low-frequency patterns [Zhang et al., 2024a, Choi et al., 2024b], with methods like HONGAT [Zhang et al., 2024a] alleviating over-smoothing via high-order dependencies and sparse attention. Integrating spectral filtering with attention has also emerged as a promising approach for adaptive, frequency-aware node representation learning [Chang et al., 2021, Sun et al., 2024, Wang et al., 2024a].

3 PROBLEM SETUP

Notations. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$ denote an attributed graph with N nodes. The node set is given by $\mathcal{V} = \{v_1, v_2, \ldots, v_N\}$ and the edge set satisfies $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. The matrix $\mathbf{X} \in \mathbb{R}^{N \times D}$ contains the attribute information of all nodes, where D corresponds to the dimensionality of each node's attributes. The adjacency matrix associated with $\mathcal{G}$ is denoted by $\mathbf{A} \in \{0,1\}^{N \times N}$ and satisfies $\mathbf{A}_{ij} = 1$ whenever there is an edge (v_i, v_j) in $\mathcal{E}$. When self-loops are added to the graph, the resulting adjacency matrix becomes $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$, and the corresponding degree matrix is defined as $\tilde{\mathbf{D}}$, which is diagonal. The notation $[N]$ refers to the set of integers from 1 to N inclusive. A subset $\mathcal{L} \subseteq [N]$ contains m labeled nodes, and its complement $\mathcal{U} = [N] \setminus \mathcal{L}$ has cardinality u with $u = N - m$. The sets $\mathcal{V}_\mathcal{L}$ and $\mathcal{V}_\mathcal{U}$ represent the collections of labeled and unlabeled nodes respectively, with $|\mathcal{V}_\mathcal{L}| = m$ and $|\mathcal{V}_\mathcal{U}| = u$. For any vector $\mathbf{u} \in \mathbb{R}^N$, the notation $[\mathbf{u}]_\mathcal{A}$ refers to the subvector composed of entries indexed by $\mathcal{A} \subseteq [N]$. In the case where $\mathbf{u}$ is a matrix, $[\mathbf{u}]_\mathcal{A}$ denotes the submatrix consisting of the rows indexed by $\mathcal{A}$. The Frobenius norm of a matrix is denoted by $\|\cdot\|_\mathrm{F}$, and the p-norm of a vector is expressed as $\|\cdot\|_p$.

Problem Description. In real-world graph data, noise frequently appears either in the node attributes or in the labels. Such noise can significantly deteriorate the quality of node representations produced by standard GCL encoders, which in turn impairs the performance of classifiers trained on these representations. Our objective is to develop node embeddings that remain robust under two specific scenarios. The first involves noise present in the labels of $\mathcal{V}_\mathcal{L}$, while the second concerns noise within the input node attributes $\mathbf{X}$. These two cases correspond to noisy labels and noisy features, respectively. The GCL-LRR model is designed to learn node representations that are low-rank and resilient to these types of noise. The representations are defined as $\mathbf{H} = g(\mathbf{X}, \mathbf{A})$, where $g(\cdot)$ denotes the GCL-LRR encoder. In this work, the encoder g is implemented as a standard GCN Kipf and Welling [2017]. The output $\mathbf{H} = \{\mathbf{H}_1; \mathbf{H}_2; \ldots; \mathbf{H}_N\} \in \mathbb{R}^{N \times d}$ consists of low-rank embeddings for all nodes. These representations are then used for transductive node classification. During this process, a linear classifier is trained using the labeled node set $\mathcal{V}_\mathcal{L}$, and the trained classifier is subsequently applied to predict labels for the unlabeled test nodes in $\mathcal{V}_\mathcal{U}$.

4 METHODS

4.1 GCL-LRR: GRAPH CONTRASTIVE LEARNING WITH LOW-RANK REGULARIZATION

Preliminary of Prototypical GCL. We adopt a contrastive learning framework to optimize the GCL encoder $g(\cdot)$, a standard GCN [Kipf and Welling, 2017]. Two augmented graph views, denoted as $G^1 = (\mathbf{X}^1, \mathbf{A}^1)$ and $G^2 = (\mathbf{X}^2, \mathbf{A}^2)$, are created. The resulting node representations are given by $\mathbf{H}^1 = g(\mathbf{X}^1, \mathbf{A}^1)$ and $\mathbf{H}^2 = g(\mathbf{X}^2, \mathbf{A}^2)$. We

enhance mutual information between $\mathbf{H}^1$ and $\mathbf{H}^2$ using the InfoNCE loss [Li et al., 2021], and incorporate prototypical contrastive learning [Li et al., 2021, Snell et al., 2017] by aligning node embeddings with K-means-derived cluster prototypes, computed as $\mathbf{c}_k = \frac{1}{|S_k|}\sum_{\mathbf{H}_i \in S_k} \mathbf{H}_i$ for every k in $[K]$. The training loss combines node-level and prototype-level objectives, $\mathcal{L}_{\text{node}}$ and $\mathcal{L}_{\text{proto}}$, which are computed as $\mathcal{L}_{\text{node}} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{s(\mathbf{H}_i^1, \mathbf{H}_i^2)}{s(\mathbf{H}_i^1, \mathbf{H}_i^2) + \sum_{j=1}^{N} s(\mathbf{H}_i^1, \mathbf{H}_j^2)}$ and $\mathcal{L}_{\text{proto}} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{\exp(\mathbf{H}_i \cdot \mathbf{c}_k)}{\sum_{k=1}^{K} \exp(\mathbf{H}_i \cdot \mathbf{c}_k)}$, where $s(\mathbf{H}_i^1, \mathbf{H}_i^2)$ is the cosine similarity between $\mathbf{H}_i^1$ and $\mathbf{H}_i^2$. The overall loss function of the PGCL is $\mathcal{L}_{\text{GCL}} = \mathcal{L}_{\text{node}} + \mathcal{L}_{\text{proto}}$.

GCL-LRR: Graph Contrastive Learning with Low-Rank Regularization. GCL-LRR enhances the robustness and generalization of node representations derived from Prototypical GCL by promoting a low-rank learned feature kernel. The gram matrix $\mathbf{K}$ of the node representations $\mathbf{H} \in \mathbb{R}^{N \times d}$ is calculated by $\mathbf{K} = \mathbf{H}\mathbf{H}^\top / N \in \mathbb{R}^{N \times N}$. Let $\left\{\widehat{\lambda}_i\right\}_{i=1}^{N}$ with $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \ldots \geq \widehat{\lambda}_{\min\{N,d\}} \geq \widehat{\lambda}_{\min\{N,d\}+1} = \ldots = 0$ be the eigenvalues of $\mathbf{K}$. In order to encourage the features $\mathbf{H}$ or the gram matrix $\mathbf{K}$ to be low-rank, we explicitly add the TNN $\|\mathbf{K}\|_{r_0} := \sum_{i=r_0+1}^{N} \widehat{\lambda}_i$ to the loss function of prototypical GCL. The starting rank $r_0 < \min(N, d)$ is the rank of the gram matrix of the features we aim to obtain with the GCL-LRR encoder, that is, if $\|\mathbf{K}\|_{r_0} = 0$, then $\text{rank}(\mathbf{K}) = r_0$. Therefore, the overall loss function of GCL-LRR is

$$\mathcal{L}_{\text{GCL-LRR}} = \mathcal{L}_{\text{node}} + \mathcal{L}_{\text{proto}} + \tau \|\mathbf{K}\|_{r_0}, \tag{1}$$

where $\tau > 0$ is the weighting parameter for the TNN $\|\mathbf{K}\|_{r_0}$. In practice, the TNN regularizer can be evaluated without explicitly constructing the full kernel matrix $\mathbf{K} = \mathbf{H}\mathbf{H}^\top / N \in \mathbb{R}^{N \times N}$. Let $\lambda_i(\mathbf{K})$ be the i-th largest eigenvalue of $\mathbf{K}$ and let $\sigma_i(\mathbf{H})$ be the i-th largest singular value of $\mathbf{H}$. Since the nonzero eigenvalues of $\mathbf{K}$ are determined by the singular values of $\mathbf{H}$, we have

$$\|\mathbf{K}\|_{r_0} = \sum_{i=r_0+1}^{\min\{N,d\}} \lambda_i(\mathbf{K}) = \frac{1}{N} \sum_{i=r_0+1}^{\min\{N,d\}} \sigma_i^2(\mathbf{H}).$$

Therefore, the low-rank regularization term can be computed from the singular values of $\mathbf{H}$ in $O(Nd^2)$ time when $N \geq d$, avoiding the explicit $O(N^2)$ construction of $\mathbf{K}$. We summarize the training algorithm for the GCL-LRR encoder in Algorithm 1. After the training is completed, we compute the low-rank node feature by $\mathbf{H} = g(\mathbf{X}, \mathbf{A})$. Algorithm 1 outlines the training procedure for Graph Contrastive Learning with Low-Rank Regularization (GCL-LRR). The characteristics of the datasets employed in our experimental evaluations are summarized in Table 3.

Motivation of Learning Low-Rank Features. We investigate how the information from the ground-truth clean labels and the label noise is distributed across different

Algorithm 1 Graph Contrastive Learning with Low-Rank Regularization

Input: Attribute matrix $\mathbf{X}$, adjacency matrix $\mathbf{A}$, training epochs $t_{\max}$
Output: Parameters of GCL-LRR encoder g

Initialize parameters of encoder g
for $t = 1$ to $t_{\max}$ **do**
  Compute node representations: $\mathbf{H} = g(\mathbf{X}, \mathbf{A})$
  Generate augmented views G^1, G^2 with $\mathbf{X}^1, \mathbf{A}^1$ and $\mathbf{X}^2, \mathbf{A}^2$
  Compute $\mathbf{H}^1 = g(\mathbf{X}^1, \mathbf{A}^1)$, $\mathbf{H}^2 = g(\mathbf{X}^2, \mathbf{A}^2)$
  Cluster $\{\mathbf{H}_i\}_{i=1}^{N}$ into K clusters $\{S_k\}_{k=1}^{K}$ using K-means
  for $k = 1$ to K **do**
    Compute prototype: $\mathbf{c}_k = \frac{1}{|S_k|}\sum_{\mathbf{H}_i \in S_k} \mathbf{H}_i$
  end for
  Compute eigenvalues $\{\widehat{\lambda}_i\}_{i=1}^{N}$ of kernel matrix $\mathbf{K} = \mathbf{H}\mathbf{H}^\top / N$
  Update parameters of g using gradient descent on $\mathcal{L}_{\text{GCL-LRR}}$
end for
return Encoder g

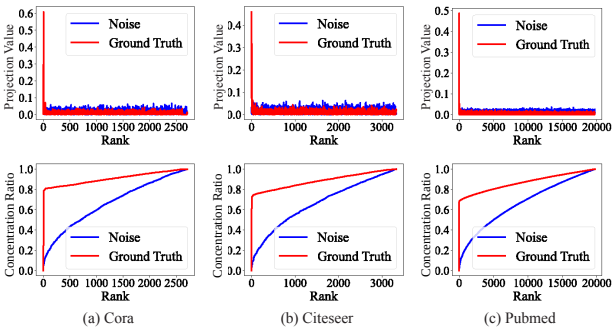


Figure 1: Eigen-projection (first row) and concentration ratios (second row) on Cora, Citeseer, and Pubmed. In the second row, the red curves denote the clean-label concentration ratios, whereas the blue curves denote the noise concentration ratios. The study in this figure is performed for asymmetric label noise with a noise level of 60%. By the rank $r = 0.2 \min\{N, d\}$, the clean-label concentration ratios of $\tilde{\mathbf{Y}}$ for Cora, Citeseer, and Pubmed are 0.844, 0.809, and 0.784 respectively. Figure 4 in Section B.7 further illustrates the eigen-projection and concentration ratios on more datasets.

eigenvectors of the feature gram matrix $\mathbf{K}$ through an eigen-projection analysis. Let $\tilde{\mathbf{Y}} \in \mathbb{R}^{N \times C}$ denote the clean label matrix without noise. We begin by computing the eigenvectors $\mathbf{U}$ of the gram matrix $\mathbf{K}$. The eigen-projection score for the r-th eigenvector is then given by $p_r = \frac{1}{C}\sum_{c=1}^{C} \left\|\mathbf{U}^{(r)\top} \tilde{\mathbf{Y}}^{(c)}\right\|_2^2 / \left\|\tilde{\mathbf{Y}}^{(c)}\right\|_2^2$ for $r \in [N]$, where C is the number of classes, and $\tilde{\mathbf{Y}} \in 0, 1^{N \times C}$ con-

sists of one-hot encoded clean labels. Here $\tilde{\mathbf{Y}}^{(c)}$ refers to the c-th column of $\tilde{\mathbf{Y}}$. We define $\mathbf{p} = [p_1, \ldots, p_N] \in \mathbb{R}^N$ as the vector of projection values. In the presence of label noise $\mathbf{N} \in \mathbb{R}^{N\times C}$, the observed label matrix becomes $\mathbf{Y} = \tilde{\mathbf{Y}} + \mathbf{N}$. The projection value p_r quantifies the proportion of the signal aligned with the r-th eigenvector of $\mathbf{K}$, while the clean-label concentration ratio at rank r of the ground truth class label is defined as $\left\|\mathbf{p}^{(1:r)}\right\|_1$, representing the cumulative contribution of the top r eigenvectors. Similarly, for each fixed rank r, the noise eigen-projection score is the scalar $q_r = \frac{1}{C}\sum_{c=1}^{C}\left\|\mathbf{U}^{(r)\top}\mathbf{N}^{(c)}\right\|_2^2 / \left\|\mathbf{N}^{(c)}\right\|_2^2 \in \mathbb{R}$. Stacking these scalar values over all ranks gives $\mathbf{q} = [q_1, \ldots, q_N] \in \mathbb{R}^N$, and the noise concentration ratio at rank r is defined as $\left\|\mathbf{q}^{(1:r)}\right\|_1$. Empirical results, shown as red curves in Figure 1, indicate that the clean label signals are primarily concentrated on the leading eigenvectors of $\mathbf{K}$. In contrast, the projection of the label noise appears more uniformly distributed across all eigenvectors, as demonstrated by the blue curves in the same figure. The above observation motivates low-rank features $\mathbf{H}$ or, equivalently, the low-rank gram matrix $\mathbf{K}$. This is because the low-rank part of the feature matrix $\mathbf{H}$ or the gram matrix $\mathbf{K}$ covers the dominant information in the ground truth label $\tilde{\mathbf{Y}}$ while learning only a small portion of the label noise. We refer to such property as the Low Frequency Property (**LFP**), which has been widely studied in deep learning [Rahaman et al., 2019, Arora et al., 2019, Cao et al., 2021, Choraria et al., 2022, Wang et al., 2024b]. Moreover, we remark that the regularization term $\|\mathbf{K}\|_{r_0}$ in the loss function (1) of GCL-LRR is also theoretically motivated by the sharp upper bound for the test loss using a linear transductive classifier, presented as (3) in Theorem 4.1. A smaller $\|\mathbf{K}\|_{r_0}$ renders a smaller upper bound for the test loss, which ensures better generalization capability of the linear transductive classifier to be introduced in the next subsection.

4.2 TRANSDUCTIVE NODE CLASSIFICATION

In this section, we present a standard yet straightforward linear transductive node classification method based on the low-rank node representations $\mathbf{H} \in \mathbb{R}^{N\times d}$ obtained from the GCL-LRR encoder. Furthermore, we establish a novel generalization bound on the test loss of this transductive classifier in the presence of label noise.

We begin by introducing the fundamental notations for the proposed algorithm. For each node v_i where $i \in [N]$, let $\mathbf{y}_i \in \mathbb{R}^C$ denote its observed one-hot class label vector. We then define the full observed label matrix as $\mathbf{Y} \coloneqq [\mathbf{y}_1; \mathbf{y}_2; \ldots; \mathbf{y}_N] \in \mathbb{R}^{N\times C}$, which may include label noise, modeled as $\mathbf{N} \in \mathbb{R}^{N\times C}$. The classifier's linear prediction is defined by $\mathbf{F}(\mathbf{W}) = \mathbf{HW}$, where $\mathbf{W} \in \mathbb{R}^{d\times C}$ is the learnable weight matrix. Final predictions are made using the softmax transformation $\text{softmax}(\mathbf{F}(\mathbf{W})) \in \mathbb{R}^{N\times C}$ to estimate class probabilities for the test nodes. We train the transductive classifier by minimizing the regular cross-entropy on the labeled nodes through the following optimization problem,

$$\min_{\mathbf{W}} L(\mathbf{W}) = \frac{1}{m}\sum_{v_i\in\mathcal{V}_\mathcal{L}} \text{KL}\left(\mathbf{y}_i, [\text{softmax}\left(\mathbf{HW}\right)]_i\right), \quad (2)$$

where KL is the KL divergence between the label $\mathbf{y}_i$ and the softmax of the classifier output at node v_i. We use a regular gradient descent to optimize (2) with a learning rate $\eta \in (0, \frac{1}{\widehat{\lambda}_1})$. $\mathbf{W}$ is initialized by $\mathbf{W}^{(0)} = \mathbf{0}$, and at the t-th iteration of gradient descent for $t \geq 1$, $\mathbf{W}$ is updated by $\mathbf{W}^{(t)} = \mathbf{W}^{(t-1)} - \eta\nabla_{\mathbf{W}} L(\mathbf{W})|_{\mathbf{W}=\mathbf{W}^{(t-1)}}$. We define $\mathbf{F}(\mathbf{W}, t) \coloneqq \mathbf{HW}^{(t)}$ as the output of the classifier after the t-th iteration of gradient descent for $t \geq 1$. We have the following theoretical result, Theorem 4.1, on the Mean Squared Error (MSE) loss of the unlabeled test nodes $\mathcal{V}_\mathcal{U}$ measured by the gap between $[\mathbf{F}(\mathbf{W}, t)]_\mathcal{U}$ and $\left[\tilde{\mathbf{Y}}\right]_\mathcal{U}$ when using the low-rank feature $\mathbf{H}$ with $r_0 \in [N]$, which is the generalization error bound for the linear transductive classifier using $\mathbf{F}(\mathbf{W}) = \mathbf{HW}$ to predict the labels of the unlabeled nodes. Similar to existing work [Kothapalli et al., 2023] that use the MSE to analyze the optimization and the generalization of GNNs, we employ the MSE loss to provide the generalization error of the node classifier in the following theorem. It is remarked that the MSE loss is necessary for the generalization analysis of transductive learning using the transductive local Rademacher complexity [Tolstikhin et al., 2014, Yang, 2025a,b].

Theorem 4.1. Suppose that a set $\mathcal{L}$ with $|\mathcal{L}| = m$ is sampled uniformly without replacement from $[N]$, and the remaining nodes $\mathcal{V}_\mathcal{U} = \mathcal{V} \setminus \mathcal{V}_\mathcal{L}$ are the test nodes with $|\mathcal{V}_\mathcal{U}| = u$. Then after the t-th iteration of gradient descent for all $t \geq 1$, for every $x > 0$ and every $\delta \in (0, 1)$, with probability at least $1 - \exp(-x) - \delta$, we have

$$\begin{aligned}\mathcal{U}_{\text{test}}(t) &\coloneqq \frac{1}{u}\left\|\left[\mathbf{F}(\mathbf{W}, t) - \tilde{\mathbf{Y}}\right]_\mathcal{U}\right\|_\text{F}^2 \\ &\leq \frac{2}{m}\left(L_1(\mathbf{K}, \tilde{\mathbf{Y}}, t) + L_2(\mathbf{K}, \mathbf{N}, t)\right) + c_0\text{KC}(\mathbf{K}) + \frac{c_0 x}{N_{u,m,\delta}},\end{aligned} \quad (3)$$

where $N_{u,m,\delta} \coloneqq \frac{\min\{u,m\}}{\log_2(4\min\{u,m\}/\delta)}$, c_0 is a positive number depending on $\mathbf{U}$, $\left\{\widehat{\lambda}_i\right\}_{i=1}^{r_0}$, and τ_0 with $\tau_0^2 = \max_{i\in[N]}\mathbf{K}_{ii}$. Moreover, $L_1(\mathbf{K}, \tilde{\mathbf{Y}}, t) \coloneqq \left\|\left(\mathbf{I}_m - \eta\mathbf{K}_{\mathcal{L},\mathcal{L}}\right)^t\left[\tilde{\mathbf{Y}}\right]_\mathcal{L}\right\|_\text{F}^2$, $L_2(\mathbf{K}, \mathbf{N}, t) \coloneqq \left\|\eta\mathbf{K}_{\mathcal{L},\mathcal{L}}\sum_{t'=0}^{t-1}\left(\mathbf{I}_m - \eta\mathbf{K}_{\mathcal{L},\mathcal{L}}\right)^{t'}[\mathbf{N}]_\mathcal{L}\right\|_\text{F}^2$, $\mathbf{K}_{\mathcal{L},\mathcal{L}} \coloneqq [\mathbf{H}]_\mathcal{L}[\mathbf{H}]_\mathcal{L}^\top$, and KC is the kernel complexity of the gram matrix $\mathbf{K} = \mathbf{HH}^\top/N$ defined as $\text{KC}(\mathbf{K}) \coloneqq \min_{r_0\in[N]} r_0\left(\frac{1}{u} + \frac{1}{m}\right) + \sqrt{\|\mathbf{K}\|_{r_0}}\left(\frac{1}{\sqrt{u}} + \frac{1}{\sqrt{m}}\right)$.

This theorem is proved in Section C. Specifically, $\mathcal{U}_{\text{test}}(t)$ denotes the test loss over unlabeled nodes, quantified by the discrepancy between the classifier output $\mathbf{F}(\mathbf{W}, t)$ and

the clean label matrix $\tilde{\mathbf{Y}}$. The upper bound on the test loss in (3) consists of three components: $L_1(\mathbf{K}, \tilde{\mathbf{Y}}, t)$, $L_2(\mathbf{K}, \mathbf{N}, t)$, and $\mathrm{KC}(\mathbf{K})$, each serving a distinct role. The term $L_1(\mathbf{K}, \tilde{\mathbf{Y}}, t)$ reflects the training loss of the classifier when using clean labels. $L_2(\mathbf{K}, \mathbf{N}, t)$ captures the impact of label noise on the classification loss. When there is no label noise, i.e., $\mathbf{N} = \mathbf{0}$, the term $L_2(\mathbf{K}, \mathbf{N}, t)$ vanishes, and the bound reduces to a clean-data generalization bound controlled by $L_1(\mathbf{K}, \tilde{\mathbf{Y}}, t)$ and $\mathrm{KC}(\mathbf{K})$. Thus, the strong performance of GCL-LRR in clean or low-noise settings is consistent with our theory, since low-rank regularization reduces kernel complexity even when the noise-dependent term is absent. The term $\mathrm{KC}(\mathbf{K})$ denotes the kernel complexity (KC), which assesses the structural complexity of the gram matrix $\mathbf{K}$ derived from node representations $\mathbf{H}$ produced by the GCL-LRR encoder. Importantly, the TNN $\|\mathbf{K}\|_{r_0}$ appears on the right-hand side of the upper bound in (3), thereby providing theoretical justification for incorporating the TNN regularizer $\|\mathbf{K}\|_{r_0}$ into the GCL-LRR objective function (1) to promote low-rank feature learning. Furthermore, under the LFP, which is consistently supported by the empirical evidence shown in Figure 1 and Figure 4, $L_1(\mathbf{K}, \tilde{\mathbf{Y}}, t)$ diminishes as the number of training iterations t increases. Simultaneously, $L_2(\mathbf{K}, \mathbf{N}, t)$ remains small due to the approximately uniform eigen-projection of label noise, while $\mathbf{K}$ remains close to a rank-r_0 matrix, as the TNN is effectively minimized through the GCL-LRR training objective (1). In our experiments, described in the subsequent section, we select the TNN rank r_0 via standard cross-validation across all graph datasets. As reported in Table 4, the optimal rank r_0 consistently falls within the range of $0.1 \min\{N, d\}$ to $0.3 \min\{N, d\}$.

4.3 GCL-LR-ATTENTION: IMPROVING GCL-LRR BY LOW RANK ATTENTION

To further enhance the performance of GCL-LRR, we introduce an improved model termed GCL-LR-Attention, which incorporates a novel LR-Attention layer into GCL-LRR. GCL-LR-Attention applies self-attention to the low-rank node representations $\mathbf{H} \in \mathbb{R}^{N \times d}$ produced by the GCL-LRR encoder via the optimization of (1). Specifically, the output of the LR-Attention layer of the GCL-LR-Attention model is computed as $\mathbf{F} = \mathbf{BH}$, where $\mathbf{F}$ denotes the attention-transformed features, and $\mathbf{B} \in \mathbb{R}^{N \times N}$ is the attention weight matrix. We recall that the gram matrix of the node features is given by $\mathbf{K} = \mathbf{H}\mathbf{H}^\top / N$. In our LR-Attention formulation, the attention matrix is defined as $\mathbf{B} = \mathbf{K}/\widehat{\lambda}_1$, where $\widehat{\lambda}_1$ is the largest eigenvalue of $\mathbf{K}$.

Efficiency of GCL-LR-Attention with the low-rank attention weight matrix B. The attention matrix in GCL-LR-Attention is computed by $\mathbf{B} = \mathbf{K}/\widehat{\lambda}_1$, which implies $\mathrm{rank}(\mathbf{B}) \leq \min\{N, d\}$. In practice, the embedding dimension d is significantly smaller than the number of nodes N. In all our experiments, we set $d = 256$, following standard settings in existing graph contrastive learning methods [Li et al., 2024b, Yuan et al., 2023]. As a result, $\mathbf{B}$ is inherently low-rank, and the LR-Attention operation $\mathbf{BH} = \mathbf{H}\mathbf{H}^\top\mathbf{H}/\widehat{\lambda}_1$ can be computed in $\mathcal{O}(Nd^2)$ time by first forming the $d \times d$ matrix $\mathbf{H}^\top\mathbf{H}$ and then multiplying it with $\mathbf{H}$, without explicitly constructing the dense $N \times N$ attention matrix.

The resulting gram matrix of the transformed features $\mathbf{F} \in \mathbb{R}^{N \times d}$ is $\mathbf{K_F} = \mathbf{F}\mathbf{F}^\top/N = \mathbf{K}^3/\widehat{\lambda}_1^2$. Let $\{\lambda_i\}_{i=1}^N$ denote the eigenvalues of $\mathbf{K_F}$, ordered as $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N \geq 0$. Then, for each $i \in [N]$, it holds that $\lambda_i = \widehat{\lambda}_i^3/\widehat{\lambda}_1^2$. Noting that $\lambda_i = \widehat{\lambda}_i \cdot \widehat{\lambda}_i^2/\widehat{\lambda}_1^2 \leq \widehat{\lambda}_i$ due to the fact that $\widehat{\lambda}_1 \geq \widehat{\lambda}_i$ for all $i \in [N]$, it follows that the LR-Attention layer reduces the KC of the original gram matrix $\mathbf{K}$. Therefore, the KC of $\mathbf{K_F}$ is guaranteed to be no greater than that of $\mathbf{K}$, demonstrating the regularization effect introduced by the LR-Attention mechanism. We then train a transductive classifier on top of $\mathbf{F}$ similar to Section 4.2 by minimizing the following loss function,

$$\min_{\mathbf{W}} L(\mathbf{W}) = \frac{1}{m} \sum_{v_i \in \mathcal{V}_\mathcal{L}} \mathrm{KL}\left(\mathbf{y}_i, [\mathrm{softmax}\left(\mathbf{FW}\right)]_i\right), \quad (4)$$

where $\mathbf{W}$ is the weight matrix for the classifier. The linear classifier trained with the LR-Attention layer via the optimization objective in (4) is referred to as GCL-LR-Attention. Based on the preceding analysis and the test loss upper bound given in Equation (3) in Theorem 4.1, GCL-LR-Attention exhibits a reduced KC compared to GCL-LRR. Consequently, the test loss $\mathcal{U}_{\text{test}}(t)$ for GCL-LR-Attention can be lower than that of GCL-LRR, implying improved predictive performance. This theoretical advantage is empirically supported by the results in Table 8 and Table 2, where GCL-LR-Attention consistently achieves both a lower kernel complexity and a tighter upper bound on the test loss compared to GCL-LRR.

5 EXPERIMENTS

In this section, we comprehensively evaluate the performance of GCL-LRR and GCL-LR-Attention on several public graph datasets. Section 5.1 outlines the experimental setup and implementation details for both models. In Section 5.2, we report the results of GCL-LRR and GCL-LR-Attention on semi-supervised node classification under asymmetric label noise and attribute noise. Section 5.3 investigates the KC and the theoretical upper bound of the test loss for both GCL-LRR and GCL-LR-Attention. Additional experiment results are presented in Section B of the supplementary. The effectiveness of GCL-LRR and GCL-LR-Attention on heterophilic graph datasets is studied in Section B.1. Additional evaluations on larger heterophilic graph datasets are reported in Section B.2, and experiments

on social network, protein-protein interaction, and business-review graphs are reported in Section B.3. Additional analysis of KC across additional datasets is given in Section B.4. Additional node classification results under symmetric label noise, asymmetric label noise, and attribute noise are reported in Section B.5. Comparisons with recent robust GCL methods are presented in Section B.6. Complementary results on eigen-projection and concentration ratios are presented in Section B.7. Section B.8 compares our models with existing graph contrastive learning methods that employ various classifier architectures. The statistical significance of improvements observed in Sections 5.2, B.5, and B.1 is assessed using the Student's t-test, detailed in Section B.9. Section B.10 presents a sensitivity analysis of the key hyperparameter. Section B.11 further evaluates how this hyperparameter affects L_1, L_2, KC, and the theoretical upper bound. In Section B.12, we conduct an ablation study to examine the effect of the rank parameter r_0 in the TNN. Additional rank ablations on Coauthor-CS and ogbn-arxiv are reported in Section B.13. The ablation study in Section B.14 demonstrates that the LR-Attention layer with low-rank regularization, as in GCL-LR-Attention, is also applicable to GCN so as to enhance the robustness of standard GCN [Kipf and Welling, 2017], and the resultant GCN-based model is termed GCN-LR-Attention, as detailed in Section B.14. However, GCL-LR-Attention substantially outperforms GCN-LR-Attention, justifying the use of the GCL encoder in this paper. The extended component ablations in Section B.15 further compare GCL-LR-Attention with vanilla GCL, GCL-LRR, GCN-LR-Attention, and GCL+Standard Attention on all eight benchmark datasets. Section B.16 presents a training comparison between GCL-LRR, GCL-LR-Attention, and the competing baseline methods.

5.1 EXPERIMENTAL SETTINGS

In our experiments, we evaluate the proposed methods using eight widely adopted graph benchmark datasets: Cora, Citeseer, and PubMed [Sen et al., 2008], Coauthor CS, ogbn-arxiv [Hu et al., 2020], Wiki-CS [Mernyei and Cangea, 2020], and Amazon-Computers and Amazon-Photos [Shchur et al., 2018]. Following existing GCL methods Li et al. [2024b], Yuan et al. [2023], we employ a three-layer GCN as the GCL encoder, with hidden dimensions set to 256. Detailed settings for the generation of the label noise, the cross-validation for tuning r_0 and τ, and the datasets are presented in Section A of the supplementary.

5.2 NODE CLASSIFICATION

Compared Methods. We conduct a comprehensive comparison of GCL-LRR and GCL-LR-Attention with representative semi-supervised node representation learning methods, including GCN [Kipf and Welling, 2017], GCE [Zhang and Sabuncu, 2018], S^2GC [Zhu and Koniusz, 2021], and GRAND+ [Feng et al., 2022b], as well as label-noise–robust baselines NRGNN [Dai et al., 2021] and RTGNN [Qian et al., 2023]. We further benchmark against leading GCL approaches, including MERIT [Jin et al., 2021], SUGRL [Mo et al., 2022], and SFA [Zhang et al., 2023]. In addition, we include attention-based GNNs that integrate low- and high-frequency information, namely GFSA [Choi et al., 2024a] and HONGAT [Zhang et al., 2024a], as well as contrastive methods designed for noisy labels, JoSRC [Yao et al., 2021], Sel-CL [Li et al., 2022], CRGNN [Li et al., 2024b], and CGNN [Yuan et al., 2023]. We further compare with recent robust GCL baselines, including DKGCCL [Chen et al., 2026], GRANCE [Zhuang et al., 2025], and ArnoldiGCL [Coskun et al., 2025], in Section B.6.

Experimental Results. To evaluate the robustness of GCL-LRR and its variant GCL-LR-Attention, we perform a series of experiments on graphs subjected to both symmetric and asymmetric label noise, with corruption levels varying from 40% to 80% in increments of 20%. In parallel, we investigate the effect of attribute noise under the same range and step size. Unlike prior GCL methods such as MERIT, SUGRL, and SFA, which adopt an inductive classification framework by training a linear classifier on node embeddings obtained via contrastive learning, GCL-LRR is evaluated under a transductive setting that incorporates test data during training. To ensure fair comparisons, all baseline models are retrained using the transductive classifier employed in GCL-LRR, along with a standard standard GCN transductive classifier. Details on the performance under different classifier configurations are provided in Section B.8. For methods initially designed for inductive learning, we report their best outcomes among the default inductive classifier and the two transductive classifiers. Table 1 summarizes the average accuracy and standard deviation across 10 runs for GCL-LRR, GCL-LR-Attention, and competing baselines for node classification on Cora, PubMed, and ogbn-arxiv with asymmetric label noise and attribute noise. Comparison between GCL-LRR, GCL-LR-Attention, and all baseline methods on Cora, Citeseer, PubMed, and ogbn-arxiv with asymmetric label noise, symmetric label noise, and attribute noise is deferred to Table 9 in Section B.5 of the supplementary. Table 10 in Section B.5 of the supplementary presents additional results on Coauthor-CS, Wiki-CS, Amazon-Computers, and Amazon-Photos. Across all datasets and noise conditions, GCL-LRR consistently achieves superior performance over all baselines. Moreover, the introduction of low-rank attention in GCL-LR-Attention leads to further gains. For instance, under 80% asymmetric label noise on PubMed, GCL-LRR outperforms the strongest baseline, CRGNN, by 2.1% in classification accuracy, while GCL-LR-Attention achieves an additional 2.3% improvement. These results underscore the effectiveness of both low-rank regularization and low-rank attention

Table 1: Performance comparison for node classification on Cora, PubMed, and ogbn-arxiv with asymmetric label noise and attribute noise. The highest values for each dataset under each setting in the table are bold. Comparison between GCL-LRR, GCL-LR-Attention, and all baseline methods on Cora, Citeseer, PubMed, and ogbn-arxiv with asymmetric label noise, symmetric label noise, and attribute noise is deferred to Table 9 in Section B.5 of the supplementary. Table 10 in Section B.5 of the supplementary presents additional results on Coauthor-CS, Wiki-CS, Amazon-Computers, and Amazon-Photos.

Dataset	Methods	Noise Type						
		0	40		60		80	
		-	Asymmetric	Attribute	Asymmetric	Attribute	Asymmetric	Attribute
Cora	GCN	0.815±0.005	0.547±0.015	0.639±0.008	0.405±0.014	0.439±0.012	0.265±0.012	0.317±0.013
	NRGNN	0.822±0.006	0.571±0.019	0.645±0.012	0.472±0.014	0.451±0.011	0.282±0.022	0.326±0.010
	RTGNN	0.828±0.003	0.570±0.010	0.678±0.011	0.474±0.011	0.457±0.009	0.280±0.011	0.342±0.016
	SUGRL	0.834±0.005	0.564±0.011	0.675±0.009	0.468±0.011	0.452±0.012	0.280±0.012	0.338±0.014
	MERIT	0.831±0.005	0.560±0.008	0.671±0.009	0.467±0.013	0.450±0.014	0.277±0.013	0.335±0.009
	SFA	0.839±0.010	0.564±0.011	0.676±0.015	0.473±0.014	0.457±0.014	0.282±0.016	0.344±0.017
	Sel-Cl	0.828±0.002	0.570±0.010	0.676±0.009	0.472±0.013	0.455±0.011	0.282±0.017	0.341±0.015
	Jo-SRC	0.825±0.005	0.571±0.006	0.679±0.007	0.473±0.011	0.458±0.012	0.285±0.013	0.345±0.018
	GRAND+	0.858±0.006	0.570±0.009	0.678±0.011	0.472±0.010	0.456±0.012	0.284±0.015	0.345±0.013
	GFSA	0.837±0.006	0.568±0.012	0.672±0.009	0.466±0.012	0.451±0.012	0.279±0.012	0.336±0.013
	HONGAT	0.833±0.004	0.566±0.011	0.667±0.010	0.464±0.010	0.449±0.010	0.278±0.013	0.334±0.014
	CRGNN	0.842±0.005	0.572±0.010	0.674±0.010	0.472±0.012	0.454±0.013	0.283±0.014	0.341±0.015
	CGNN	0.835±0.006	0.567±0.009	0.669±0.011	0.462±0.013	0.450±0.013	0.281±0.012	0.337±0.014
	GCL-LRR	0.858±0.006	0.589±0.011	0.695±0.011	0.492±0.011	0.477±0.012	0.306±0.012	0.363±0.011
	GCL-LR-Attention	**0.861±0.006**	**0.602±0.011**	**0.708±0.011**	**0.510±0.011**	**0.492±0.012**	**0.329±0.012**	**0.382±0.011**
PubMed	GCN	0.790±0.007	0.584±0.022	0.595±0.012	0.405±0.025	0.488±0.013	0.305±0.022	0.423±0.013
	NRGNN	0.797±0.008	0.602±0.022	0.603±0.008	0.443±0.012	0.499±0.009	0.330±0.023	0.433±0.011
	RTGNN	0.797±0.004	0.610±0.008	0.614±0.012	0.455±0.010	0.501±0.011	0.335±0.013	0.452±0.013
	SUGRL	0.819±0.005	0.603±0.013	0.615±0.010	0.445±0.011	0.501±0.007	0.321±0.009	0.446±0.010
	MERIT	0.801±0.004	0.593±0.011	0.613±0.011	0.447±0.012	0.497±0.009	0.328±0.011	0.445±0.009
	SFA	0.804±0.010	0.596±0.011	0.609±0.011	0.447±0.014	0.499±0.014	0.330±0.011	0.447±0.014
	Sel-Cl	0.799±0.005	0.605±0.014	0.614±0.012	0.455±0.014	0.502±0.008	0.334±0.021	0.456±0.014
	Jo-SRC	0.801±0.005	0.613±0.010	0.617±0.013	0.453±0.008	0.504±0.013	0.330±0.015	0.459±0.018
	GRAND+	0.845±0.006	0.610±0.011	0.617±0.013	0.453±0.008	0.503±0.010	0.331±0.014	0.458±0.014
	GFSA	0.823±0.005	0.608±0.012	0.616±0.009	0.450±0.013	0.500±0.010	0.333±0.013	0.455±0.012
	HONGAT	0.818±0.006	0.606±0.011	0.613±0.010	0.448±0.014	0.498±0.012	0.328±0.012	0.450±0.011
	CRGNN	0.829±0.005	0.612±0.010	0.618±0.011	0.452±0.011	0.503±0.009	0.335±0.013	0.457±0.012
	CGNN	0.822±0.006	0.607±0.013	0.615±0.010	0.449±0.012	0.499±0.010	0.332±0.014	0.454±0.013
	GCL-LRR	0.845±0.009	0.637±0.014	0.637±0.011	0.479±0.011	0.526±0.011	0.356±0.011	0.482±0.014
	GCL-LR-Attention	**0.846±0.009**	**0.652±0.014**	**0.655±0.011**	**0.498±0.011**	**0.544±0.011**	**0.379±0.011**	**0.498±0.014**
ogbn-arxiv	GCN	0.717±0.003	0.401±0.014	0.478±0.010	0.336±0.011	0.339±0.012	0.286±0.022	0.294±0.013
	NRGNN	0.721±0.006	0.449±0.014	0.485±0.012	0.371±0.020	0.342±0.011	0.330±0.018	0.300±0.010
	RTGNN	0.718±0.004	0.443±0.012	0.484±0.014	0.380±0.011	0.340±0.017	0.335±0.011	0.301±0.006
	SUGRL	0.693±0.002	0.439±0.010	0.480±0.012	0.365±0.013	0.341±0.009	0.327±0.011	0.295±0.011
	MERIT	0.717±0.004	0.442±0.009	0.483±0.010	0.368±0.011	0.341±0.012	0.324±0.012	0.304±0.009
	SFA	0.718±0.009	0.445±0.012	0.486±0.012	0.368±0.011	0.338±0.015	0.325±0.014	0.302±0.013
	Sel-Cl	0.719±0.002	0.447±0.007	0.486±0.010	0.375±0.008	0.344±0.013	0.331±0.008	0.304±0.012
	Jo-SRC	0.715±0.005	0.445±0.011	0.481±0.010	0.377±0.013	0.340±0.013	0.333±0.013	0.297±0.009
	GRAND+	0.725±0.004	0.445±0.008	0.481±0.011	0.378±0.010	0.344±0.010	0.332±0.010	0.303±0.009
	GFSA	0.719±0.004	0.443±0.012	0.482±0.011	0.370±0.012	0.342±0.011	0.328±0.012	0.299±0.011
	HONGAT	0.716±0.005	0.440±0.011	0.480±0.012	0.366±0.013	0.339±0.012	0.324±0.014	0.296±0.012
	CRGNN	0.721±0.003	0.446±0.010	0.483±0.009	0.372±0.010	0.343±0.010	0.330±0.012	0.302±0.010
	CGNN	0.717±0.006	0.441±0.013	0.481±0.010	0.368±0.014	0.340±0.011	0.326±0.015	0.298±0.012
	GCL-LRR	0.728±0.006	0.472±0.013	0.508±0.014	0.405±0.014	0.405±0.012	0.359±0.015	0.335±0.013
	GCL-LR-Attention	**0.731±0.006**	**0.487±0.013**	**0.523±0.014**	**0.423±0.014**	**0.423±0.012**	**0.374±0.015**	**0.350±0.013**

in enhancing the robustness of node classification models.

5.3 STUDY IN THE KERNEL COMPLEXITY AND THE UPPER BOUND OF THE TEST LOSS

This section provides a comparative analysis of each component in the upper bound of the test loss presented in Equation (3), specifically $L_1(\mathbf{K}, \tilde{\mathbf{Y}}, t)$, $L_2(\mathbf{K}, \mathbf{N}, t)$, and $\mathrm{KC}(\mathbf{K})$, for node representations produced by different methods. The corresponding results are reported in Table 2. Experiments are conducted on Cora, Citeseer, and PubMed under symmetric label noise at a corruption level of 40%. The findings indicate that GCL-LRR and GCL-LR-Attention consistently yield substantially lower values across all evaluated terms when compared to baseline approaches. This suggests superior generalization capacity of these methods in semi-supervised node classification, even in the presence of noisy labels. Additionally, we evaluate the KC of the gram matrix derived from node representations generated by GCL-LRR, GCL-LR-Attention, and other competitive GCL methods across additional datasets. The results in Table 8 in Section B.4 show that the representations from GCL-LRR exhibit notably lower complexity. This implies that transductive classifiers trained on such representations tend to incur smaller generalization errors

Table 2: Comparisons on $L_1(\mathbf{K}, \tilde{\mathbf{Y}}, t)$, $L_2(\mathbf{K}, \mathbf{N}, t)$, KC($\mathbf{K}$) and the value of the upper bound of the test loss from Theorem 4.1. The evaluation is performed on semi-supervised node classification with 40% of symmetric label noise. The lowest values for each dataset in the table are bold, and the second-lowest values are underlined. The results represent the mean values computed over 10 independent runs, with the standard deviation reported after $\pm$.

Datasets		MERIT	SFA	Jo-SRC	GCN	GFSA	HONGAT	GCL-LRR	GCL-LR-Attention
Cora	L_1	5.24 ±0.49	6.04 ±0.23	6.50 ±0.34	7.38 ±0.12	6.44 ±0.01	6.38 ± 0.13	<u>3.72</u> ± 0.38	**3.65** ± 0.38
	L_2	4.92 ±0.14	4.95 ±0.35	5.05 ±0.13	5.24 ±0.01	3.80 ±0.24	4.25 ± 0.26	<u>2.97</u> ± 0.45	**2.72** ± 0.42
	KC	0.37 ±0.29	0.42 ±0.09	0.48 ±0.39	0.44 ±0.40	0.35 ±0.31	0.40 ± 0.08	<u>0.20</u> ± 0.02	**0.18** ± 0.26
	Upper Bound	10.68±0.14	11.59±0.15	12.18 ±0.46	13.22±0.11	10.80±0.22	11.25 ± 0.02	<u>7.05</u> ± 0.43	**6.74** ± 0.32
Citeseer	L_1	4.72 ±0.42	4.85 ±0.28	4.92 ±0.23	5.10 ±0.40	4.54 ±0.46	4.69 ± 0.19	<u>4.02</u> ± 0.34	**3.95** ± 0.21
	L_2	4.33 ±0.04	4.69 ±0.07	4.42 ±0.15	5.08 ±0.25	4.20 ±0.00	4.42 ± 0.03	<u>3.75</u> ± 0.17	**3.60** ± 0.22
	KC	0.47 ±0.27	0.45 ±0.18	0.55 ±0.08	0.64 ±0.42	0.47 ±0.10	0.50 ± 0.42	<u>0.24</u> ± 0.18	**0.21** ± 0.16
	Upper Bound	9.77 ±0.14	10.21±0.28	10.17 ±0.34	11.07±0.24	9.40 ±0.25	9.84 ± 0.14	<u>8.20</u> ± 0.04	**7.97** ± 0.33
PubMed	L_1	3.97 ±0.29	4.02 ±0.08	4.11 ±0.14	4.35 ±0.06	4.26 ±0.12	3.95 ± 0.23	<u>3.38</u> ± 0.40	**3.40** ± 0.38
	L_2	2.69 ±0.20	2.54 ±0.28	2.60 ±0.32	2.88 ±0.08	2.98 ±0.09	2.85 ± 0.03	<u>2.32</u> ± 0.10	**2.26** ± 0.45
	KC	0.54 ±0.49	0.50 ±0.27	0.62 ±0.17	0.71 ±0.23	0.52 ±0.17	0.66 ± 0.16	<u>0.30</u> ± 0.29	**0.28** ± 0.37
	Upper Bound	7.44 ±0.22	7.28 ±0.03	7.59 ±0.37	8.15 ±0.26	7.99 ±0.23	7.63 ± 0.14	<u>6.25</u> ± 0.30	**6.16** ± 0.40

on unlabeled nodes.

6 LIMITATIONS

The effectiveness of GCL-LRR and GCL-LR-Attention depends on the presence of a meaningful separation between the clean-label signal and noise in the eigenspace of the learned feature gram matrix. In particular, the clean-label signal should be sufficiently concentrated in the leading eigenspace, while the noise should not dominate those leading components. When the noise energy becomes so large that its projection on the leading eigenvectors is larger than the projection of the clean-label signal, low-rank representation learning may fail to recover useful task information. This limitation is consistent with detectability results in noisy graph models, where the signal from the ground-truth labels can become unrecoverable below a signal-to-noise threshold [Nadakuditi and Newman, 2012, Abbe, 2018]. Another limitation is that the current empirical validation and theory focus on node classification. Extending low-rank regularization and LR-Attention to graph classification and link prediction is an important direction for future work.

7 CONCLUSIONS

This paper first introduces a novel graph contrastive learning encoder, Graph Contrastive Learning with Low-Rank Regularization (GCL-LRR), designed to enhance the robustness in node representation learning. GCL-LRR generates low-rank features, drawing motivation from the low-frequency characteristics commonly observed in real-world graph datasets and the tight generalization bound for transductive learning. The GCL-LRR encoder is trained under a prototypical GCL framework, incorporating the TNN as a regularization term. To further strengthen its performance, we extend GCL-LRR with a novel low-rank attention mechanism, leading to the GCL-LR-Attention model. GCL-LR-Attention further reduces the kernel complexity of GCL-LRR, thereby improving the generalization of GCL-LRR. Empirical results across a wide range of benchmarks confirm that both GCL-LRR and GCL-LR-Attention achieve superior robustness and outperform state-of-the-art approaches in learning effective node representations for noisy node classification tasks.

ACKNOWLEDGEMENTS

This work was supported by National Institute of Health 1OT2OD037955-01. This work was also supported by the 2023 Mayo Clinic and Arizona State University Alliance for Health Care Collaborative Research Seed Grant Program under Award No. AWD00038846.

References

Emmanuel Abbe. Community detection and stochastic block models: Recent developments. *Journal of Machine Learning Research*, 18(177):1–86, 2018.

Kelsey Allen, Evan Shelhamer, Hanul Shin, and Joshua Tenenbaum. Infinite mixture prototypes for few-shot learning. In *International Conference on Machine Learning*, pages 232–241. PMLR, 2019.

Sercan Ö Arik and Tomas Pfister. Protoattend: Attention-based prototypical learning. *The Journal of Machine Learning Research*, 21(1):8691–8725, 2020.

Sanjeev Arora, Simon S. Du, Wei Hu, Zhiyuan Li, and Ruosong Wang. Fine-grained analysis of optimization and generalization for overparameterized two-layer neural networks. In *International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 322–332, 2019.

Deyu Bo, Xiao Wang, Chuan Shi, and Huawei Shen. Beyond low-frequency information in graph convolutional

networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 3950–3957, 2021.

Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. Spectral networks and locally connected networks on graphs. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014.

Yuan Cao, Zhiying Fang, Yue Wu, Ding-Xuan Zhou, and Quanquan Gu. Towards understanding the spectral bias of deep learning. In Zhi-Hua Zhou, editor, *International Joint Conference on Artificial Intelligence*, pages 2205–2211, 2021.

Heng Chang, Yu Rong, Tingyang Xu, Wenbing Huang, Somayeh Sojoudi, Junzhou Huang, and Wenwu Zhu. Spectral graph attention network with fast eigen-approximation. In *Proceedings of the 30th ACM international conference on information & knowledge management*, pages 2905–2909, 2021.

Xiang Chen, Kun Yue, Wenjie Liu, Zhenyu Zhang, and Liang Duan. Dual-kernel graph community contrastive learning. In *Fortieth AAAI Conference on Artificial Intelligence, Thirty-Eighth Conference on Innovative Applications of Artificial Intelligence, Sixteenth Symposium on Educational Advances in Artificial Intelligence, AAAI 2026, Singapore, January 20-27, 2026*, pages 14583–14591. AAAI Press, 2026.

Xiuyuan Cheng, Zichen Miao, and Qiang Qiu. Graph convolution with low-rank learnable local filters. In *International Conference on Learning Representations*, 2021.

Jeongwhan Choi, Hyowon Wi, Jayoung Kim, Yehjin Shin, Kookjin Lee, Nathaniel Trask, and Noseong Park. Graph convolutions enrich the self-attention in transformers! *Advances in Neural Information Processing Systems*, 37: 52891–52936, 2024a.

Jeongwhan Choi, Hyowon Wi, Jayoung Kim, Yehjin Shin, Kookjin Lee, Nathaniel Trask, and Noseong Park. Graph convolutions enrich the self-attention in transformers! In *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, 2024b.

Moulik Choraria, Leello Tadesse Dadi, Grigorios Chrysos, Julien Mairal, and Volkan Cevher. The spectral bias of polynomial neural networks. In *International Conference on Learning Representations*, 2022.

Coskun et al. Arnoldigcl: Graph contrastive learning via learnable arnoldi-based guided spectral chebyshev polynomial filters. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2025.

Enyan Dai, Charu Aggarwal, and Suhang Wang. Nrgnn: Learning a label noise-resistant graph neural network on sparsely and noisily labeled graphs. *SIGKDD*, 2021.

Enyan Dai, Wei Jin, Hui Liu, and Suhang Wang. Towards robust graph neural networks for noisy graphs with sparse labels. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, pages 181–191, 2022.

Kaize Ding, Zhe Xu, Hanghang Tong, and Huan Liu. Data augmentation for deep graph learning: A survey. *arXiv preprint arXiv:2202.08235*, 2022.

Yushun Dong, Kaize Ding, Brian Jalaian, Shuiwang Ji, and Jundong Li. Adagnn: Graph neural networks with adaptive frequency response filter. In *Proceedings of the 30th ACM international conference on information & knowledge management*, pages 392–401, 2021.

Yushun Dong, Yinhan He, Patrick Soga, Song Wang, and Jundong Li. Graph neural networks are more than filters: Revisiting and benchmarking from a spectral perspective. In *The Thirteenth International Conference on Learning Representations*, 2025.

Shengyu Feng, Baoyu Jing, Yada Zhu, and Hanghang Tong. Adversarial graph contrastive learning with information regularization. In *Proceedings of the ACM Web Conference 2022*, pages 1362–1371, 2022a.

Wenzheng Feng, Yuxiao Dong, Tinglin Huang, Ziqi Yin, Xu Cheng, Evgeny Kharlamov, and Jie Tang. Grand+: Scalable graph random neural networks. In *Proceedings of the ACM Web Conference 2022*, pages 3248–3258, 2022b.

Jacob Goldberger and Ehud Ben-Reuven. Training deep neural-networks using a noise adaptation layer. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*, 2017.

Yuanfan Guo, Minghao Xu, Jiawen Li, Bingbing Ni, Xuanyu Zhu, Zhenbang Sun, and Yi Xu. Hcsc: hierarchical contrastive selective coding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9706–9715, 2022.

William L. Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 1024–1034, 2017.

Bo Han, Quanming Yao, Xingrui Yu, Gang Niu, Miao Xu, Weihua Hu, Ivor W. Tsang, and Masashi Sugiyama. Co-teaching: Robust training of deep neural networks with extremely noisy labels. In Samy Bengio, Hanna M. Wallach,

Hugo Larochelle, Kristen Grauman, Nicolò Cesa-Bianchi, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pages 8536–8546, 2018.

Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. Open graph benchmark: Datasets for machine learning on graphs. In *NeurIPS*, 2020.

Lu Jiang, Zhengyuan Zhou, Thomas Leung, Li-Jia Li, and Li Fei-Fei. Mentornet: Learning data-driven curriculum for very deep neural networks on corrupted labels. In *International Conference on Machine Learning*, pages 2304–2313. PMLR, 2018.

Yizhu Jiao, Yun Xiong, Jiawei Zhang, Yao Zhang, Tianqi Zhang, and Yangyong Zhu. Sub-graph contrast for scalable self-supervised graph representation learning. In *2020 IEEE international conference on data mining (ICDM)*, pages 222–231. IEEE, 2020.

Ming Jin, Yizhen Zheng, Yuan-Fang Li, Chen Gong, Chuan Zhou, and Shirui Pan. Multi-scale contrastive siamese networks for self-supervised graph representation learning. In *The 30th International Joint Conference on Artificial Intelligence (IJCAI)*, 2021.

Mingxuan Ju, Shifu Hou, Yujie Fan, Jianan Zhao, Yanfang Ye, and Liang Zhao. Adaptive kernel graph neural network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 7051–7058, 2022.

Wei Ju, Siyu Yi, Yifan Wang, Zhiping Xiao, Zhengyang Mao, Hourun Li, Yiyang Gu, Yifang Qin, Nan Yin, Senzhang Wang, et al. A survey of graph neural networks in real world: Imbalance, noise, privacy and ood challenges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*, 2017.

Vignesh Kothapalli, Tom Tirer, and Joan Bruna. A neural collapse perspective on feature evolution in graph neural networks. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine, editors, *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023.

Junnan Li, Richard Socher, and Steven C. H. Hoi. Dividemix: Learning with noisy labels as semi-supervised learning. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020.

Junnan Li, Pan Zhou, Caiming Xiong, and Steven C. H. Hoi. Prototypical contrastive learning of unsupervised representations. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*, 2021.

Shikun Li, Xiaobo Xia, Shiming Ge, and Tongliang Liu. Selective-supervised contrastive learning with noisy labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 316–325, 2022.

Shiyuan Li, Yixin Liu, Qingfeng Chen, Geoffrey I Webb, and Shirui Pan. Noise-resilient unsupervised graph representation learning via multi-hop feature quality estimation. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pages 1255–1265, 2024a.

Xianxian Li, Qiyu Li, Haodong Qian, Jinyan Wang, et al. Contrastive learning of graphs under label noise. *Neural Networks*, 172:106113, 2024b.

Lu Lin, Jinghui Chen, and Hongning Wang. Spectral augmentation for self-supervised learning on graphs. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*, 2023.

Eran Malach and Shai Shalev-Shwartz. Decoupling "when to update" from "how to update". In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 960–970, 2017.

Péter Mernyei and Cătălina Cangea. Wiki-cs: A wikipedia-based benchmark for graph neural networks. *arXiv preprint arXiv:2007.02901*, 2020.

Yujie Mo, Liang Peng, Jie Xu, Xiaoshuang Shi, and Xiaofeng Zhu. Simple unsupervised graph representation learning. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*, pages 7797–7805, 2022.

Raj Rao Nadakuditi and Mark E. J. Newman. Graph spectra and the detectability of community structure in networks. *Physical Review Letters*, 108(18):188701, 2012.

Hoang NT and Takanori Maehara. Revisiting graph neural networks: All we have is low-pass filters. *CoRR*, abs/1905.09550, 2019.

Giorgio Patrini, Alessandro Rozza, Aditya Krishna Menon, Richard Nock, and Lizhen Qu. Making deep neural networks robust to label noise: A loss correction approach. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 2233–2241, 2017.

Hongbin Pei, Bingzhe Wei, Kevin Chen-Chuan Chang, Yu Lei, and Bo Yang. Geom-gcn: Geometric graph convolutional networks. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*, 2020.

Zhen Peng, Wenbing Huang, Minnan Luo, Qinghua Zheng, Yu Rong, Tingyang Xu, and Junzhou Huang. Graph representation learning via graphical mutual information maximization. In *Proceedings of The Web Conference 2020*, pages 259–270, 2020.

Oleg Platonov, Denis Kuznedelev, Michael Diskin, Artem Babenko, and Liudmila Prokhorenkova. A critical look at the evaluation of gnns under heterophily: Are we really making progress? In *International Conference on Learning Representations*, 2023.

Siyi Qian, Haochao Ying, Renjun Hu, Jingbo Zhou, Jintai Chen, Danny Z. Chen, and Jian Wu. Robust training of graph neural networks via noise governance. In Tat-Seng Chua, Hady W. Lauw, Luo Si, Evimaria Terzi, and Panayiotis Tsaparas, editors, *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining, WSDM 2023, Singapore, 27 February 2023 - 3 March 2023*, pages 607–615, 2023.

Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 5301–5310, 09–15 Jun 2019.

Benedek Rozemberczki and Rik Sarkar. Twitch gamers: a dataset for evaluating proximity preserving and structural role-based node embeddings, 2021.

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

Oleksandr Shchur, Maximilian Mumme, Aleksandar Bojchevski, and Stephan Günnemann. Pitfalls of graph neural network evaluation. *arXiv preprint arXiv:1811.05868*, 2018.

Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. *Advances in neural information processing systems*, 30, 2017.

Hwanjun Song, Minseok Kim, Dongmin Park, Yooju Shin, and Jae-Gil Lee. Learning from noisy labels with deep neural networks: A survey. *IEEE transactions on neural networks and learning systems*, 34(11):8135–8153, 2022.

Jiaqi Sun, Lin Zhang, Shenglin Zhao, and Yujiu Yang. Improving your graph neural networks: a high-frequency booster. In *2022 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 748–756. IEEE, 2022.

Yukuan Sun, Yutai Duan, Haoran Ma, Yuelong Li, and Jianming Wang. High-frequency and low-frequency dual-channel graph attention network. *Pattern Recognition*, 156:110795, 2024.

Tingting Tang, Yue Niu, Salman Avestimehr, and Murali Annavaram. Edge private graph neural networks with singular value perturbation. *Proc. Priv. Enhancing Technol.*, 2024(3):391–406, 2024.

Yu Tang, Lilan Peng, Zhendong Wu, Jie Hu, Pengfei Zhang, and Hongchun Lu. Fahc: frequency adaptive hypergraph constraint for collaborative filtering. *Applied Intelligence*, 55(3):242, 2025.

Ilya O. Tolstikhin, Gilles Blanchard, and Marius Kloft. Localized complexities for transductive learning. In Maria-Florina Balcan, Vitaly Feldman, and Csaba Szepesvári, editors, *Conference on Learning Theory*, volume 35 of *JMLR Workshop and Conference Proceedings*, pages 857–884, 2014.

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

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018.

Botao Wang, Jia Li, Yang Liu, Jiashun Cheng, Yu Rong, Wenjia Wang, and Fugee Tsung. Deep insights into noisy pseudo labeling on graph data. *Advances in Neural Information Processing Systems*, 36:76214–76228, 2023.

Tianfeng Wang, Zhisong Pan, Guyu Hu, and Yahao Hu. Attention-enabled adaptive markov graph convolution. *Neural Computing and Applications*, 36(9):4979–4993, 2024a.

Yancheng Wang, Rajeev Goel, Utkarsh Nath, Alvin C. Silva, Teresa Wu, and Yingzhen Yang. Learning low-rank feature for thorax disease classification. In Amir Globersons,

Lester Mackey, Danielle Belgrave, Angela Fan, Ulrich Paquet, Jakub M. Tomczak, and Cheng Zhang, editors, *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, 2024b.

Zhonghao Wang, Danyu Sun, Sheng Zhou, Haobo Wang, Jiapei Fan, Longtao Huang, and Jiajun Bu. Noisygl: A comprehensive benchmark for graph neural networks under label noise. In Amir Globersons, Lester Mackey, Danielle Belgrave, Angela Fan, Ulrich Paquet, Jakub M. Tomczak, and Cheng Zhang, editors, *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, 2024c.

Felix Wu, Amauri H. Souza Jr., Tianyi Zhang, Christopher Fifty, Tao Yu, and Kilian Q. Weinberger. Simplifying graph convolutional networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pages 6861–6871, 2019.

Bingbing Xu, Huawei Shen, Qi Cao, Keting Cen, and Xueqi Cheng. Graph convolutional networks using heat kernel for semi-supervised learning. In Sarit Kraus, editor, *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pages 1928–1934, 2019a.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*, 2019b.

Minghao Xu, Hang Wang, Bingbing Ni, Hongyu Guo, and Jian Tang. Self-supervised graph-level representation learning with local and global structure. In *International Conference on Machine Learning*, pages 11548–11558. PMLR, 2021.

Wenjia Xu, Yongqin Xian, Jiuniu Wang, Bernt Schiele, and Zeynep Akata. Attribute prototype network for zero-shot learning. *Advances in Neural Information Processing Systems*, 33:21969–21980, 2020.

Yuchen Yan, Yuzhong Chen, Huiyuan Chen, Minghua Xu, Mahashweta Das, Hao Yang, and Hanghang Tong. From trainable negative depth to edge heterophily in graphs. *Advances in Neural Information Processing Systems*, 36: 70162–70178, 2023.

Liang Yang, Qiuliang Zhang, Runjie Shi, Wenmiao Zhou, Bingxin Niu, Chuan Wang, Xiaochun Cao, Dongxiao He, Zhen Wang, and Yuanfang Guo. Graph neural networks without propagation. In *Proceedings of the ACM Web Conference 2023*, pages 469–477, 2023.

Yingzhen Yang. A new concentration inequality for sampling without replacement and its application for transductive learning. In *Forty-second International Conference on Machine Learning*, 2025a.

Yingzhen Yang. Improved generalization bounds for transductive learning by transductive local complexity and its applications. *arXiv preprint arXiv:2309.16858*, 2025b. URL `https://arxiv.org/abs/2309.16858`.

Yazhou Yao, Zeren Sun, Chuanyi Zhang, Fumin Shen, Qi Wu, Jian Zhang, and Zhenmin Tang. Jo-src: A contrastive approach for combating noisy labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5192–5201, 2021.

Yuning You, Tianlong Chen, Yang Shen, and Zhangyang Wang. Graph contrastive learning automated. In *International Conference on Machine Learning*, pages 12121–12132. PMLR, 2021.

Wenhui Yu and Zheng Qin. Graph convolutional network for recommendation with low-pass collaborative filters. In *International Conference on Machine Learning*, pages 10936–10945. PMLR, 2020.

Xingrui Yu, Bo Han, Jiangchao Yao, Gang Niu, Ivor W. Tsang, and Masashi Sugiyama. How does disagreement help generalization against label corruption? In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pages 7164–7173, 2019.

Jingyang Yuan, Xiao Luo, Yifang Qin, Yusheng Zhao, Wei Ju, and Ming Zhang. Learning on graphs under label noise. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2023.

Hanqing Zeng, Hongkuan Zhou, Ajitesh Srivastava, Rajgopal Kannan, and Viktor Prasanna. Graphsaint: Graph sampling based inductive learning method. In *International Conference on Learning Representations*, 2020.

Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, 64(3):107–115, 2021.

Heng-Kai Zhang, Yi-Ge Zhang, Zhi Zhou, and Yu-Feng Li. Hongat: Graph attention networks in the presence of high-order neighbors. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 16750–16758, 2024a.

Qi Zhang, Jinghua Li, Yanfeng Sun, Shaofan Wang, Junbin Gao, and Baocai Yin. Beyond low-pass filtering on large-scale graphs via adaptive filtering graph neural networks. *Neural Networks*, 169:1–10, 2024b.

Yifei Zhang, Hao Zhu, Zixing Song, Piotr Koniusz, and Irwin King. Spectral feature augmentation for graph contrastive learning and beyond. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 11289–11297, 2023.

Zhilu Zhang and Mert Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels. *Advances in neural information processing systems*, 31, 2018.

Jia-Xing Zhong, Nannan Li, Weijie Kong, Shan Liu, Thomas H Li, and Ge Li. Graph convolutional label noise cleaner: Train a plug-and-play action classifier for anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1237–1246, 2019.

Hao Zhu and Piotr Koniusz. Simple spectral graph convolution. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*, 2021.

Yonghua Zhu, Lei Feng, Zhenyun Deng, Yang Chen, Robert Amor, and Michael Witbrock. Robust node classification on graph data with graph and label noise. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pages 17220–17227, 2024.

Jun Zhuang and Mohammad Al Hasan. Defending graph convolutional networks against dynamic graph perturbations via bayesian self-supervision. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 4405–4413, 2022.

Zhuang et al. Refine then classify: Robust graph neural networks with reliable neighborhood contrastive refinement. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025.

Graph Contrastive Learning with Low-Rank Regularization and Low-Rank Attention for Noisy Node Classification (Supplementary Material)

Yancheng Wang[1] **Ping Li**[2] **Yingzhen Yang**[1]

[1]School of Computing and Augmented Intelligence, Arizona State University, Tempe, AZ 85281, USA, yancheng.wang@asu.edu, yingzhen.yang@asu.edu
[2]VecML Inc., Bellevue, WA 98004, USA, pingli98@gmail.com

A DETAILED EXPERIMENT SETTINGS

Datasets. We evaluate our method on eight public benchmarks that are widely used for node representation learning, namely Cora, Citeseer, PubMed [Sen et al., 2008], Coauthor CS, ogbn-arxiv [Hu et al., 2020], Wiki-CS [Mernyei and Cangea, 2020], Amazon-Computers, and Amazon-Photos [Shchur et al., 2018]. Cora, Citeseer, and PubMed are the three most widely used citation networks. Coauthor CS is a co-authorship graph. The ogbn-arxiv is a directed citation graph. Wiki-CS is a hyperlink network of computer science articles. Amazon-Computers and Amazon-Photos are co-purchase networks of products selling on Amazon.com. We summarize the statistics of all the datasets in Table 3.

Table 3: Statistics of the datasets.

Dataset	Nodes	Edges	Features	Classes
Cora	2,708	5,429	1,433	7
CiteSeer	3,327	4,732	3,703	6
PubMed	19,717	44,338	500	3
Coauthor CS	18,333	81,894	6,805	15
ogbn-arxiv	169,343	1,166,243	128	40
Wiki-CS	11,701	215,863	300	10
Amazon-Computers	13,752	245,861	767	10
Amazon-Photos	7,650	119,081	745	8

Implementation Details of the Label Noise. To simulate label corruption, we follow established protocols from prior work [Dai et al., 2022, Qian et al., 2023] and generate two types of noisy labels: (1) Symmetric noise, where each label is randomly flipped to any other class with equal probability; and (2) Asymmetric noise, where label flips occur preferentially between semantically similar classes. Specifically, we adopt the formal label noise model introduced in [Song et al., 2022]. Let $\mathbf{T} \in [0,1]^{C\times C}$ denote the noise transition matrix, where $\mathbf{T}_{ij} := \mathbb{P}(\tilde{y} = j \mid y = i)$ defines the probability that a clean label $y = i$ is flipped to a noisy label $\tilde{y} = j$. Under symmetric noise with rate $\tau \in [0,1]$, the transition matrix is defined by $\mathbf{T}_{ii} = 1 - \tau$ and $\mathbf{T}_{ij} = \frac{\tau}{C-1}$ for all $j \neq i$. For asymmetric noise, the matrix satisfies $\mathbf{T}_{ii} = 1 - \tau$ with $\mathbf{T}_{ij} > \mathbf{T}_{ik}$ for some $j \neq i$, $k \neq i$, capturing the structured mislabeling observed in real-world scenarios. To assess performance under attribute noise, we follow the approach of [Ding et al., 2022], wherein a fixed proportion of each node's input features is randomly shuffled. The ratio of perturbed features defines the attribute noise level used in our experiments.

Implementation Details of the Attribute Noise. To evaluate the performance of our method with attribute noise, we randomly shuffle a certain percent of input attributes for each node following [Ding et al., 2022, Li et al., 2024a, Ju et al., 2025]. The percentage of shuffled attributes is defined as the attribute noise level in our experiments.

Traininig Settings. For all experiments, we adopt the standard dataset splits for training, validation, and testing as established in prior works [Shchur et al., 2018, Mernyei and Cangea, 2020, Hu et al., 2020]. Noise is introduced exclusively into the training and validation sets, while the test set remains unaltered to ensure unbiased evaluation. Hyperparameter tuning is conducted via five-fold cross-validation on the training data of each benchmark. Specifically, we explore the learning rate over

the range $1 \times 10^{-4}, 5 \times 10^{-4}, 1 \times 10^{-3}, 5 \times 10^{-3}, 1 \times 10^{-2}, 3 \times 10^{-2}, 6 \times 10^{-2}, 1 \times 10^{-1}, 5 \times 10^{-1}$ and weight decay from $1 \times 10^{-5}, 5 \times 10^{-5}, 1 \times 10^{-4}, 5 \times 10^{-4}, 1 \times 10^{-3}, 5 \times 10^{-3}$. Dropout rates are chosen from $0.3, 0.4, 0.5, 0.6, 0.7$. The optimal hyperparameters for each dataset are those that yield the lowest validation loss. All models are trained using the Adam optimizer for up to 500 epochs, with early stopping triggered if the validation loss fails to improve over 20 successive epochs. To ensure robustness against initialization variance, each experiment is repeated 10 times with different random seeds.

Cross-Validation for Tuning r_0 and τ. We determine the rank parameter r_0 and the weight τ of the TNN loss via cross-validation on each dataset. The rank is set as $r_0 = \lceil \gamma \min\{N, d\} \rceil$, where γ denotes the rank ratio. We select γ from $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ and τ from $\{0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5\}$ using five-fold cross-validation on 20% of the training set. The chosen values for each dataset are listed in Table 4.

Table 4: Selected rank ratio γ and TNN weight λ for each dataset.

Parameters	Cora	Citeseer	PubMed	Coauthor CS	ogbn-arxiv	Wiki-CS	Amazon-Computers	Amazon-Photos
τ	0.10	0.10	0.10	0.20	0.10	0.25	0.20	0.20
γ	0.2	0.2	0.3	0.3	0.4	0.2	0.2	0.3

B ADDITIONAL EXPERIMENT RESULTS

B.1 EVALUATION ON HETEROPHILIC GRAPHS

In this section, we examine the effectiveness of GCL-LRR for semi-supervised node classification on two commonly used heterophilic graph datasets: Texas and Chameleon [Pei et al., 2020]. We begin by analyzing the Low Frequency Property (LFP) on these datasets using eigen-projection visualizations and concentration ratios, as depicted in Figure 2. The results indicate that the LFP is also present in heterophilic graphs, akin to its presence in homophilic datasets. This analysis is conducted under asymmetric label noise with a corruption level of 60%. Using a rank parameter of $0.2 \min\{N, d\}$, the resulting concentration entropy values are 0.762 for Chameleon and 0.725 for Texas.

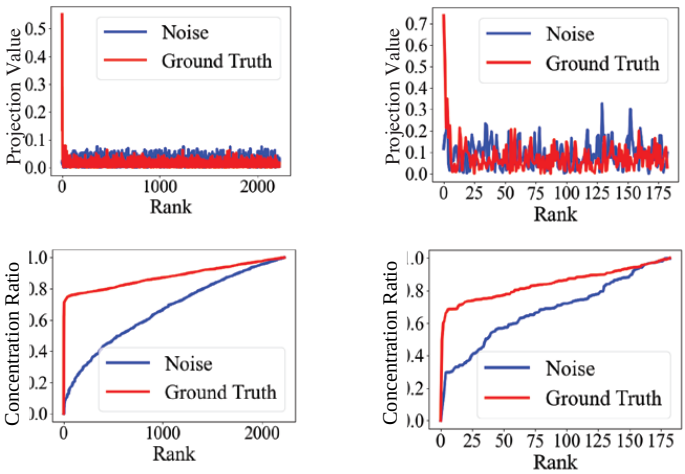
Subsequently, we conduct semi-supervised node classification experiments on Texas and Chameleon following the experimental protocol described in Section 5.2. We employ TEDGCN [Yan et al., 2023], a widely adopted GNN architecture for heterophilic graphs, as the encoder backbone for both GCL-LRR and GCL-LR-Attention. The classification results are reported in Table 5. As shown, both GCL-LRR and GCL-LR-Attention achieve substantial performance improvements over the base heterophilic GNN, demonstrating robustness to various types of noise in these challenging graph settings.

Table 5: Performance comparison for node classification on Texas and Chameleon with asymmetric label noise, symmetric label noise, and attribute noise.

Dataset	Methods	Noise Type									
		0	40			60			80		
		-	Asymmetric	Symmetric	Attribute	Asymmetric	Symmetric	Attribute	Asymmetric	Symmetric	Attribute
Texas	TEDGCN	0.771±0.025	0.525±0.023	0.528±0.018	0.541±0.022	0.402±0.016	0.418±0.019	0.445±0.021	0.312±0.015	0.328±0.017	0.341±0.020
	GCL-LRR	0.780±0.013	0.547±0.019	0.557±0.016	0.568±0.017	0.438±0.015	0.444±0.017	0.463±0.018	0.336±0.012	0.353±0.014	0.365±0.016
	GCL-LR-Attention	**0.785±0.018**	**0.556±0.016**	**0.563±0.013**	**0.576±0.015**	**0.451±0.012**	**0.452±0.014**	**0.472±0.016**	**0.338±0.010**	**0.367±0.012**	**0.372±0.014**
Chameleon	TEDGCN	0.569±0.009	0.382±0.021	0.401±0.018	0.425±0.020	0.298±0.017	0.315±0.019	0.328±0.022	0.225±0.016	0.241±0.018	0.254±0.021
	GCL-LRR	0.584±0.011	0.407±0.019	0.436±0.015	0.447±0.018	0.332±0.015	0.342±0.016	0.356±0.018	0.251±0.013	0.269±0.015	0.283±0.017
	GCL-LR-Attention	**0.585±0.008**	**0.412±0.016**	**0.444±0.013**	**0.452±0.014**	**0.341±0.011**	**0.352±0.013**	**0.361±0.015**	**0.262±0.010**	**0.282±0.012**	**0.290±0.014**

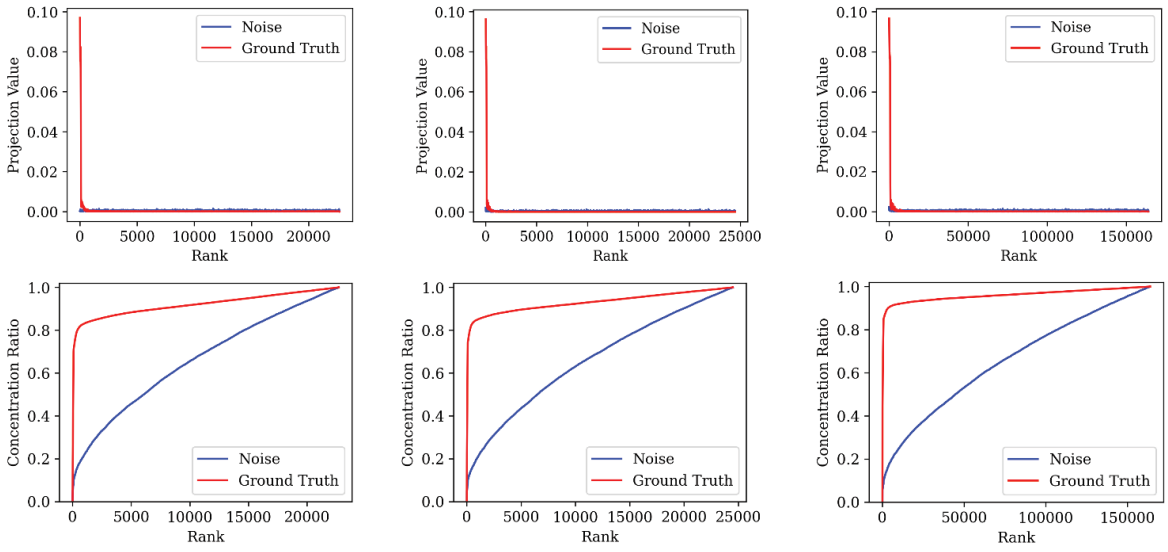
B.2 EVALUATION ON ADDITIONAL HETEROPHILIC GRAPHS

We further evaluate GCL-LRR and GCL-LR-Attention on three larger heterophilic graph datasets: Twitch-Gamers [Rozemberczki and Sarkar, 2021], Roman-Empire, and Amazon-Ratings [Platonov et al., 2023]. Twitch-Gamers contains $168,114$ nodes and $6,797,557$ edges, Roman-Empire contains $22,662$ nodes and $32,927$ undirected edges, and Amazon-Ratings contains $24,492$ nodes and $93,050$ undirected edges. Figure 3 illustrates the corresponding eigen-projection analysis, further confirming that the clean-label signal is concentrated in the leading eigenspace while label noise is more broadly distributed. Following the protocol in Section B.1, we use TEDGCN as the heterophilic GNN backbone and evaluate all methods under asymmetric label noise, symmetric label noise, and attribute noise. As shown in Table 6, GCL-LRR consistently improves over TEDGCN, and GCL-LR-Attention further improves over GCL-LRR across all three datasets and all noise settings.



(a) Chameleon (b) Texas

Figure 2: Eigen-projection (first row) and concentration ratios (second row) on Chameleon and Texas. In the second row, the red curves denote the clean-label concentration ratios, whereas the blue curves denote the noise concentration ratios. The study in this figure is performed for asymmetric label noise with a noise level of 60%. By the rank of $0.2 \min\{N, d\}$, the concentration entropy on Chameleon and Texas are 0.762 and 0.725.



(a) Roman-Empire (b) Amazon-Ratings (c) Twitch-Gamers

Figure 3: Eigen-projection (first row) and concentration ratios (second row) on Roman-Empire, Amazon-Ratings, and Twitch-Gamers. In the second row, the red curves denote the clean-label concentration ratios, whereas the blue curves denote the noise concentration ratios. The study in this figure is performed for asymmetric label noise with a noise level of 60%.

B.3 EVALUATION ON ADDITIONAL GRAPH TYPES

To further evaluate generalizability beyond citation and co-purchase networks, we conduct node classification experiments on a business-review graph, Yelp-All, a social network graph, Reddit, and a protein-protein interaction graph, PPI [Hamilton et al., 2017, Zeng et al., 2020]. The results in Table 7 show that GCL-LRR outperforms the strongest competing baseline across these graph types, and GCL-LR-Attention further improves the performance of GCL-LRR.

Table 6: Performance comparison for node classification on larger heterophilic graph datasets with asymmetric label noise, symmetric label noise, and attribute noise.

Dataset	Methods	Noise Type									
		0	40			60			80		
		-	Asymmetric	Symmetric	Attribute	Asymmetric	Symmetric	Attribute	Asymmetric	Symmetric	Attribute
Twitch-Gamers	TEDGCN	0.612	0.503	0.486	0.521	0.451	0.426	0.478	0.382	0.357	0.421
	GCL-LRR	0.624	0.526	0.512	0.544	0.478	0.459	0.502	0.407	0.388	0.448
	GCL-LR-Attention	0.631	0.538	0.524	0.556	0.491	0.472	0.514	0.421	0.401	0.461
Roman-Empire	TEDGCN	0.758	0.612	0.586	0.638	0.528	0.501	0.572	0.421	0.386	0.486
	GCL-LRR	0.771	0.641	0.619	0.662	0.562	0.536	0.598	0.453	0.421	0.515
	GCL-LR-Attention	0.779	0.655	0.633	0.674	0.576	0.551	0.611	0.469	0.438	0.529
Amazon-Ratings	TEDGCN	0.487	0.362	0.351	0.384	0.308	0.292	0.341	0.246	0.231	0.287
	GCL-LRR	0.501	0.385	0.377	0.407	0.335	0.321	0.366	0.271	0.256	0.313
	GCL-LR-Attention	0.508	0.397	0.389	0.419	0.348	0.334	0.379	0.284	0.269	0.326

Table 7: Performance comparison for node classification on Yelp-All, Reddit, and PPI with asymmetric label noise, symmetric label noise, and attribute noise.

Dataset	Methods	Noise Type									
		0	40			60			80		
		-	Asymmetric	Symmetric	Attribute	Asymmetric	Symmetric	Attribute	Asymmetric	Symmetric	Attribute
Yelp-All	GCN	0.612	0.425	0.448	0.470	0.352	0.361	0.391	0.288	0.271	0.302
	S^2GC	0.620	0.437	0.462	0.485	0.363	0.376	0.405	0.296	0.282	0.311
	GCE	0.616	0.433	0.456	0.477	0.359	0.367	0.395	0.293	0.277	0.306
	NRGNN	0.621	0.451	0.482	0.492	0.381	0.405	0.414	0.316	0.302	0.319
	RTGNN	0.628	0.456	0.490	0.500	0.389	0.415	0.421	0.323	0.309	0.326
	SUGRL	0.625	0.450	0.486	0.497	0.383	0.410	0.418	0.318	0.305	0.324
	MERIT	0.626	0.447	0.481	0.493	0.381	0.407	0.414	0.316	0.300	0.320
	ARIEL	0.630	0.455	0.491	0.500	0.389	0.416	0.422	0.323	0.310	0.327
	SFA	0.627	0.452	0.487	0.498	0.385	0.411	0.420	0.319	0.306	0.325
	Sel-Cl	0.624	0.453	0.489	0.499	0.386	0.413	0.421	0.321	0.308	0.326
	Jo-SRC	0.625	0.454	0.490	0.501	0.388	0.415	0.423	0.322	0.309	0.328
	GRAND+	0.636	0.456	0.492	0.502	0.389	0.416	0.424	0.323	0.310	0.329
	GFSA	0.626	0.449	0.483	0.495	0.382	0.408	0.416	0.317	0.303	0.322
	HONGAT	0.623	0.445	0.478	0.490	0.377	0.402	0.411	0.313	0.297	0.318
	CRGNN	0.631	0.455	0.491	0.501	0.388	0.415	0.423	0.322	0.309	0.328
	CGNN	0.624	0.446	0.480	0.492	0.379	0.404	0.413	0.314	0.299	0.319
	GCL-LRR	0.645	0.478	0.518	0.525	0.414	0.446	0.451	0.348	0.337	0.355
	GCL-LR-Attention	0.650	0.493	0.535	0.541	0.433	0.465	0.468	0.371	0.358	0.374
Reddit	GCN	0.921	0.653	0.671	0.714	0.526	0.522	0.552	0.438	0.405	0.427
	S^2GC	0.924	0.665	0.682	0.727	0.536	0.535	0.570	0.447	0.414	0.436
	GCE	0.923	0.668	0.676	0.719	0.534	0.525	0.557	0.452	0.408	0.425
	NRGNN	0.925	0.686	0.706	0.725	0.565	0.579	0.571	0.472	0.430	0.436
	RTGNN	0.927	0.689	0.710	0.732	0.575	0.589	0.585	0.468	0.436	0.433
	SUGRL	0.927	0.684	0.711	0.733	0.568	0.583	0.584	0.461	0.431	0.445
	MERIT	0.928	0.688	0.706	0.728	0.570	0.585	0.584	0.464	0.423	0.441
	ARIEL	0.929	0.692	0.715	0.733	0.574	0.589	0.580	0.466	0.435	0.443
	SFA	0.929	0.691	0.709	0.735	0.573	0.588	0.587	0.470	0.422	0.444
	Sel-Cl	0.927	0.690	0.714	0.734	0.571	0.590	0.585	0.469	0.434	0.444
	Jo-SRC	0.926	0.691	0.715	0.736	0.575	0.591	0.588	0.471	0.433	0.446
	GRAND+	0.934	0.690	0.714	0.735	0.574	0.590	0.587	0.468	0.433	0.445
	GFSA	0.929	0.687	0.708	0.731	0.570	0.586	0.583	0.465	0.429	0.441
	HONGAT	0.925	0.682	0.704	0.727	0.565	0.580	0.579	0.460	0.423	0.436
	CRGNN	0.930	0.691	0.713	0.735	0.574	0.590	0.587	0.469	0.433	0.445
	CGNN	0.927	0.684	0.705	0.729	0.566	0.582	0.580	0.462	0.425	0.438
	GCL-LRR	0.941	0.713	0.735	0.752	0.607	0.622	0.624	0.495	0.461	0.474
	GCL-LR-Attention	0.944	0.728	0.752	0.766	0.626	0.641	0.639	0.517	0.483	0.492
PPI	GCN	0.775	0.548	0.566	0.605	0.431	0.438	0.483	0.336	0.315	0.372
	S^2GC	0.783	0.561	0.581	0.621	0.445	0.455	0.497	0.347	0.326	0.383
	GCE	0.778	0.557	0.573	0.612	0.439	0.446	0.488	0.342	0.321	0.377
	NRGNN	0.785	0.578	0.605	0.628	0.468	0.489	0.510	0.372	0.351	0.396
	RTGNN	0.790	0.584	0.612	0.637	0.478	0.501	0.519	0.381	0.361	0.405
	SUGRL	0.789	0.579	0.608	0.634	0.472	0.495	0.516	0.376	0.357	0.402
	MERIT	0.791	0.576	0.603	0.630	0.470	0.492	0.513	0.374	0.352	0.398
	ARIEL	0.793	0.584	0.613	0.638	0.478	0.502	0.520	0.382	0.362	0.406
	SFA	0.790	0.581	0.609	0.635	0.474	0.497	0.518	0.377	0.358	0.403
	Sel-Cl	0.788	0.582	0.611	0.636	0.476	0.500	0.519	0.380	0.360	0.405
	Jo-SRC	0.789	0.583	0.612	0.638	0.477	0.501	0.521	0.381	0.361	0.407
	GRAND+	0.801	0.584	0.614	0.639	0.479	0.502	0.522	0.382	0.362	0.408
	GFSA	0.792	0.578	0.605	0.632	0.471	0.493	0.515	0.375	0.354	0.400
	HONGAT	0.787	0.573	0.599	0.626	0.465	0.487	0.509	0.369	0.348	0.394
	CRGNN	0.794	0.584	0.613	0.638	0.478	0.501	0.521	0.381	0.361	0.407
	CGNN	0.788	0.574	0.601	0.628	0.467	0.489	0.511	0.371	0.350	0.396
	GCL-LRR	0.812	0.610	0.642	0.661	0.513	0.541	0.556	0.418	0.402	0.445
	GCL-LR-Attention	0.817	0.626	0.659	0.677	0.533	0.561	0.573	0.441	0.424	0.464

B.4 ADDITIONAL STUDY IN THE KERNEL COMPLEXITY

In this section, we evaluate the KC of the gram matrix derived from node representations learned by GCL-LRR and other competing GCL methods across various datasets, under symmetric label noise at a corruption level of 40, as discussed in Theorem 4.1. The corresponding results are reported in Table 8. As observed, the Gram matrices produced by GCL-LRR exhibit substantially lower complexity compared to those of baseline methods. This indicates that transductive classifiers trained on GCL-LRR representations are likely to achieve lower generalization errors on the unlabeled nodes.

Table 8: Comparisons in the kernel complexity defined in Theorem 4.1. The evaluation is performed on the semi-supervised node classification task with 40% of symmetric label noise.

Datasets		MERIT	SFA	Jo-SRC	GCN	GFSA	HONGAT	GCL-LRR	GCL-LR-Attention
Cora	KC	0.37	0.42	0.48	0.44	0.35	0.40	0.20	0.18
	r_0	1420	1478	1665	1511	1262	1450	440	395
Citeseer	KC	0.47	0.45	0.55	0.64	0.47	0.50	0.24	0.21
	r_0	1214	1180	1405	1590	1224	1285	405	369
PubMed	KC	0.54	0.50	0.62	0.71	0.52	0.66	0.30	0.28
	r_0	1644	1562	1785	1993	1588	1874	1197	1090
Wiki-CS	KC	0.42	0.44	0.40	0.49	0.43	0.45	0.19	0.17
	r_0	1805	1993	1746	2130	1842	2048	970	904
Amazon-Computers	KC	0.39	0.37	0.40	0.45	0.35	0.37	0.12	0.11
	r_0	1450	1428	1489	1632	1370	1415	874	820
Amazon-Photos	KC	0.38	0.38	0.43	0.47	0.39	0.41	0.14	0.12
	r_0	1872	1884	1990	2145	1895	1921	750	722
Coauthor-CS	KC	0.29	0.28	0.32	0.34	0.31	0.32	0.12	0.11
	r_0	1774	1725	1896	1903	1872	1890	1120	1039
ogbn-arxiv	KC	0.12	0.13	0.12	0.14	0.12	0.13	0.05	0.05
	r_0	1860	1936	1852	1996	1845	1920	1354	1328

B.5 COMPLETE NODE CLASSIFICATION RESULTS

Table 9 shows the average classification accuracy and standard deviation across 10 runs on the Cora, Citeseer, PubMed, and ogbn-arxiv datasets, comparing GCL-LR-Attention with the strongest baselines. Moreover, Table 10 presents additional results on Coauthor-CS, Wiki-CS, Amazon-Computers, and Amazon-Photos under both types of label noise and varying degrees of attribute noise. It is observed that GCL-LR-Attention consistently achieves the best performance across all datasets and noise levels.

B.6 COMPARISONS WITH RECENT ROBUST GCL METHODS

We further compare GCL-LRR and GCL-LR-Attention with recent robust GCL baselines, including DKGCCL [Chen et al., 2026], GRANCE [Zhuang et al., 2025], and ArnoldiGCL [Coskun et al., 2025]. The results in Table 11 show that GCL-LRR and GCL-LR-Attention consistently outperform these recent baselines across all eight benchmark datasets and all three noise types. For example, on Coauthor-CS under 60% asymmetric label noise, GCL-LRR achieves 57.5% accuracy, outperforming the strongest recent baseline, GRANCE, by 1.3%, and GCL-LR-Attention further improves the accuracy to 59.4%.

B.7 EIGEN-PROJECTION AND CONCENTRATION ENTROPY ANALYSIS ON ADDITIONAL DATASETS

Figure 4 presents the eigen-projection visualizations and corresponding concentration ratios for Coauthor-CS, Amazon-Computers, Amazon-Photos, and ogbn-arxiv.

B.8 NODE CLASSIFICATION RESULTS FOR GCL METHODS WITH DIFFERENT TYPES OF CLASSIFIERS

Existing GCL approaches, including MERIT [Jin et al., 2021], SUGRL [Mo et al., 2022], and SFA [Zhang et al., 2023], typically follow a two-stage procedure: they first train a graph encoder using contrastive objectives such as InfoNCE [Jin et al., 2021], and subsequently train a linear classifier in a supervised manner on the resulting node representations. In

Table 9: Performance comparison for node classification on Cora, Citeseer, PubMed, and ogbn-arxiv with asymmetric label noise, symmetric label noise, and attribute noise. The highest values for each dataset under each setting in the table are bold. The results represent the mean values computed over 10 independent runs, with the standard deviation reported after ±.

Dataset	Methods	Noise Type									
		0	40			60			80		
		-	Asymmetric	Symmetric	Attribute	Asymmetric	Symmetric	Attribute	Asymmetric	Symmetric	Attribute
Cora	GCN	0.815±0.005	0.547±0.015	0.636±0.007	0.639±0.008	0.405±0.014	0.517±0.010	0.439±0.012	0.265±0.012	0.354±0.014	0.317±0.013
	S^2GC	0.835±0.002	0.569±0.007	0.664±0.007	0.661±0.007	0.422±0.010	0.535±0.010	0.454±0.011	0.279±0.014	0.366±0.014	0.320±0.013
	GCE	0.819±0.004	0.573±0.011	0.652±0.008	0.650±0.014	0.449±0.011	0.509±0.011	0.445±0.015	0.280±0.013	0.353±0.013	0.325±0.015
	NRGNN	0.822±0.006	0.571±0.019	0.676±0.007	0.645±0.012	0.472±0.014	0.548±0.014	0.451±0.011	0.282±0.022	0.373±0.012	0.326±0.010
	RTGNN	0.828±0.003	0.570±0.010	0.682±0.008	0.678±0.011	0.474±0.011	0.555±0.010	0.457±0.009	0.280±0.011	0.386±0.014	0.342±0.016
	SUGRL	0.834±0.005	0.564±0.011	0.674±0.012	0.675±0.009	0.468±0.011	0.552±0.011	0.452±0.012	0.280±0.012	0.381±0.012	0.338±0.014
	MERIT	0.831±0.005	0.560±0.008	0.670±0.008	0.671±0.009	0.467±0.013	0.547±0.013	0.450±0.014	0.277±0.013	0.385±0.013	0.335±0.009
	ARIEL	0.843±0.004	0.573±0.013	0.681±0.010	0.675±0.009	0.471±0.012	0.553±0.012	0.455±0.014	0.284±0.014	0.389±0.013	0.343±0.013
	SFA	0.839±0.010	0.564±0.011	0.677±0.013	0.676±0.015	0.473±0.014	0.549±0.014	0.457±0.014	0.282±0.016	0.389±0.013	0.344±0.017
	Sel-Cl	0.828±0.002	0.570±0.010	0.685±0.012	0.676±0.009	0.472±0.013	0.554±0.014	0.455±0.011	0.282±0.017	0.389±0.013	0.341±0.015
	Jo-SRC	0.825±0.005	0.571±0.006	0.684±0.013	0.679±0.007	0.473±0.011	0.556±0.008	0.458±0.012	0.285±0.013	0.387±0.018	0.345±0.018
	GRAND+	0.858±0.006	0.570±0.009	0.682±0.007	0.678±0.011	0.472±0.010	0.554±0.008	0.456±0.012	0.284±0.015	0.387±0.015	0.345±0.013
	GFSA	0.837±0.006	0.568±0.012	0.676±0.010	0.672±0.009	0.466±0.012	0.545±0.013	0.451±0.012	0.279±0.012	0.384±0.015	0.336±0.013
	HONGAT	0.833±0.004	0.566±0.011	0.673±0.011	0.667±0.010	0.464±0.010	0.543±0.011	0.449±0.010	0.278±0.013	0.381±0.014	0.334±0.014
	CRGNN	0.842±0.005	0.572±0.010	0.678±0.010	0.674±0.010	0.472±0.012	0.551±0.013	0.454±0.013	0.283±0.014	0.386±0.014	0.341±0.015
	CGNN	0.835±0.006	0.567±0.009	0.670±0.012	0.669±0.011	0.462±0.013	0.544±0.011	0.450±0.013	0.281±0.012	0.380±0.013	0.337±0.014
	GCL-LRR	0.858±0.006	0.589±0.011	0.713±0.007	0.695±0.011	0.492±0.011	0.587±0.013	0.477±0.012	0.306±0.012	0.419±0.012	0.363±0.011
	GCL-LR-Attention	**0.861±0.006**	**0.602±0.011**	**0.724±0.007**	**0.708±0.011**	**0.510±0.011**	**0.605±0.013**	**0.492±0.012**	**0.329±0.012**	**0.436±0.012**	**0.382±0.011**
Citeseer	GCN	0.703±0.005	0.475±0.023	0.501±0.013	0.529±0.009	0.351±0.014	0.341±0.014	0.372±0.011	0.291±0.022	0.281±0.019	0.290±0.014
	S^2GC	0.736±0.005	0.488±0.013	0.528±0.013	0.553±0.008	0.363±0.012	0.367±0.014	0.390±0.013	0.304±0.024	0.284±0.019	0.288±0.011
	GCE	0.705±0.004	0.490±0.016	0.512±0.014	0.540±0.014	0.362±0.015	0.352±0.010	0.381±0.009	0.309±0.012	0.285±0.014	0.285±0.011
	NRGNN	0.710±0.006	0.498±0.015	0.546±0.015	0.538±0.011	0.382±0.016	0.412±0.016	0.377±0.012	0.336±0.021	0.309±0.018	0.284±0.009
	RTGNN	0.746±0.008	0.498±0.007	0.556±0.007	0.550±0.012	0.392±0.010	0.424±0.013	0.390±0.014	0.348±0.017	0.308±0.016	0.302±0.011
	SUGRL	0.730±0.005	0.493±0.011	0.541±0.011	0.544±0.010	0.376±0.009	0.421±0.009	0.388±0.009	0.339±0.010	0.305±0.010	0.300±0.009
	MERIT	0.740±0.007	0.496±0.012	0.536±0.012	0.542±0.010	0.383±0.011	0.425±0.011	0.387±0.008	0.344±0.014	0.301±0.014	0.295±0.009
	SFA	0.740±0.011	0.502±0.014	0.532±0.015	0.547±0.013	0.390±0.014	0.433±0.014	0.389±0.012	0.347±0.016	0.312±0.015	0.299±0.013
	ARIEL	0.727±0.007	0.500±0.008	0.550±0.013	0.548±0.008	0.391±0.009	0.427±0.012	0.389±0.014	0.349±0.014	0.307±0.013	0.299±0.013
	Sel-Cl	0.725±0.008	0.499±0.012	0.551±0.010	0.549±0.008	0.389±0.011	0.426±0.008	0.391±0.020	0.350±0.018	0.310±0.015	0.300±0.017
	Jo-SRC	0.730±0.005	0.500±0.013	0.555±0.011	0.551±0.011	0.394±0.013	0.425±0.013	0.393±0.013	0.351±0.013	0.305±0.018	0.303±0.013
	GRAND+	0.756±0.004	0.497±0.010	0.553±0.010	0.552±0.011	0.390±0.013	0.422±0.013	0.387±0.013	0.348±0.013	0.309±0.014	0.302±0.012
	GFSA	0.743±0.006	0.495±0.012	0.546±0.012	0.546±0.011	0.386±0.011	0.418±0.011	0.386±0.012	0.342±0.013	0.308±0.015	0.298±0.012
	HONGAT	0.738±0.007	0.492±0.014	0.540±0.011	0.545±0.009	0.380±0.012	0.413±0.010	0.384±0.013	0.340±0.014	0.306±0.016	0.296±0.011
	CRGNN	0.751±0.006	0.497±0.011	0.552±0.010	0.549±0.012	0.389±0.014	0.423±0.013	0.388±0.012	0.347±0.015	0.310±0.014	0.301±0.012
	CGNN	0.741±0.007	0.493±0.013	0.544±0.012	0.546±0.010	0.385±0.013	0.419±0.012	0.385±0.011	0.343±0.013	0.307±0.013	0.297±0.012
	GCL-LRR	0.757±0.010	0.520±0.013	0.581±0.013	0.570±0.007	0.410±0.014	0.455±0.014	0.406±0.012	0.369±0.012	0.335±0.014	0.318±0.010
	GCL-LR-Attention	**0.762±0.010**	**0.533±0.013**	**0.597±0.013**	**0.588±0.007**	**0.430±0.014**	**0.472±0.014**	**0.423±0.012**	**0.392±0.012**	**0.352±0.014**	**0.335±0.010**
PubMed	GCN	0.790±0.007	0.584±0.022	0.574±0.012	0.595±0.012	0.405±0.025	0.386±0.011	0.488±0.013	0.305±0.022	0.295±0.013	0.423±0.013
	S^2GC	0.802±0.005	0.585±0.023	0.589±0.013	0.610±0.009	0.421±0.030	0.401±0.014	0.497±0.012	0.310±0.039	0.290±0.019	0.431±0.010
	GCE	0.792±0.009	0.589±0.018	0.581±0.011	0.590±0.014	0.430±0.012	0.399±0.012	0.491±0.010	0.311±0.021	0.301±0.011	0.424±0.012
	NRGNN	0.797±0.008	0.602±0.022	0.618±0.013	0.603±0.008	0.443±0.012	0.434±0.012	0.499±0.009	0.330±0.023	0.325±0.013	0.433±0.011
	RTGNN	0.797±0.004	0.610±0.008	0.622±0.010	0.614±0.012	0.455±0.010	0.455±0.011	0.501±0.011	0.335±0.013	0.338±0.017	0.452±0.013
	SUGRL	0.819±0.005	0.603±0.013	0.615±0.013	0.615±0.010	0.445±0.011	0.441±0.011	0.501±0.007	0.321±0.009	0.321±0.009	0.446±0.010
	MERIT	0.801±0.004	0.593±0.011	0.612±0.011	0.613±0.011	0.447±0.012	0.443±0.012	0.497±0.009	0.328±0.011	0.323±0.011	0.445±0.009
	ARIEL	0.800±0.003	0.610±0.013	0.622±0.010	0.615±0.011	0.453±0.012	0.453±0.012	0.502±0.014	0.331±0.014	0.336±0.018	0.457±0.013
	SFA	0.804±0.010	0.596±0.011	0.615±0.011	0.609±0.011	0.447±0.014	0.446±0.017	0.499±0.014	0.330±0.011	0.327±0.011	0.447±0.014
	Sel-Cl	0.799±0.005	0.605±0.014	0.625±0.012	0.614±0.012	0.455±0.014	0.449±0.010	0.502±0.008	0.334±0.021	0.332±0.014	0.456±0.014
	Jo-SRC	0.801±0.005	0.613±0.010	0.624±0.013	0.617±0.013	0.453±0.008	0.455±0.013	0.504±0.013	0.330±0.015	0.334±0.018	0.459±0.018
	GRAND+	0.845±0.006	0.610±0.011	0.624±0.013	0.617±0.013	0.453±0.008	0.453±0.011	0.503±0.010	0.331±0.014	0.337±0.013	0.458±0.014
	GFSA	0.823±0.005	0.608±0.012	0.621±0.011	0.616±0.009	0.450±0.013	0.452±0.012	0.500±0.010	0.333±0.013	0.334±0.011	0.455±0.012
	HONGAT	0.818±0.006	0.606±0.011	0.619±0.012	0.613±0.010	0.448±0.014	0.447±0.012	0.498±0.012	0.328±0.012	0.326±0.013	0.450±0.011
	CRGNN	0.829±0.005	0.612±0.010	0.623±0.009	0.618±0.011	0.452±0.011	0.455±0.013	0.503±0.009	0.335±0.013	0.333±0.014	0.457±0.012
	CGNN	0.822±0.006	0.607±0.013	0.620±0.011	0.615±0.010	0.449±0.012	0.451±0.014	0.499±0.010	0.332±0.014	0.330±0.012	0.454±0.013
	GCL-LRR	0.845±0.009	0.637±0.014	0.645±0.015	0.637±0.011	0.479±0.011	0.484±0.013	0.526±0.011	0.356±0.011	0.360±0.012	0.482±0.014
	GCL-LR-Attention	**0.846±0.009**	**0.652±0.014**	**0.662±0.015**	**0.655±0.011**	**0.498±0.011**	**0.503±0.013**	**0.544±0.011**	**0.379±0.011**	**0.379±0.012**	**0.498±0.014**
ogbn-arxiv	GCN	0.717±0.003	0.401±0.014	0.421±0.014	0.478±0.010	0.336±0.011	0.346±0.021	0.339±0.012	0.286±0.022	0.256±0.010	0.294±0.013
	S^2GC	0.712±0.003	0.417±0.017	0.429±0.014	0.492±0.010	0.344±0.016	0.353±0.031	0.343±0.009	0.297±0.023	0.266±0.013	0.284±0.012
	GCE	0.720±0.004	0.410±0.018	0.428±0.008	0.480±0.014	0.348±0.019	0.344±0.019	0.342±0.015	0.310±0.014	0.260±0.011	0.275±0.015
	NRGNN	0.721±0.006	0.449±0.014	0.466±0.009	0.485±0.012	0.371±0.020	0.379±0.008	0.342±0.011	0.330±0.018	0.271±0.018	0.300±0.010
	RTGNN	0.718±0.004	0.443±0.012	0.464±0.012	0.484±0.014	0.380±0.011	0.384±0.013	0.340±0.017	0.335±0.011	0.285±0.015	0.301±0.006
	SUGRL	0.693±0.002	0.439±0.010	0.467±0.010	0.480±0.012	0.365±0.013	0.385±0.011	0.341±0.009	0.327±0.011	0.275±0.011	0.295±0.011
	MERIT	0.717±0.004	0.442±0.009	0.463±0.009	0.483±0.010	0.368±0.011	0.381±0.011	0.341±0.012	0.324±0.012	0.272±0.010	0.304±0.009
	ARIEL	0.717±0.004	0.448±0.013	0.471±0.013	0.482±0.011	0.379±0.014	0.384±0.015	0.342±0.015	0.334±0.014	0.280±0.013	0.300±0.010
	SFA	0.718±0.009	0.445±0.012	0.463±0.013	0.486±0.012	0.368±0.011	0.378±0.014	0.338±0.015	0.325±0.014	0.273±0.012	0.302±0.013
	Sel-Cl	0.719±0.002	0.447±0.007	0.469±0.007	0.486±0.010	0.375±0.008	0.389±0.025	0.344±0.013	0.331±0.008	0.284±0.019	0.304±0.012
	Jo-SRC	0.715±0.005	0.445±0.011	0.466±0.009	0.481±0.010	0.377±0.013	0.387±0.013	0.340±0.013	0.333±0.013	0.282±0.018	0.297±0.009
	GRAND+	0.725±0.004	0.445±0.008	0.466±0.011	0.481±0.011	0.378±0.010	0.385±0.012	0.344±0.010	0.332±0.010	0.282±0.016	0.303±0.009
	GFSA	0.719±0.004	0.443±0.012	0.460±0.010	0.482±0.011	0.370±0.012	0.379±0.012	0.342±0.011	0.328±0.012	0.278±0.013	0.299±0.011
	HONGAT	0.716±0.005	0.440±0.011	0.458±0.012	0.480±0.012	0.366±0.013	0.373±0.013	0.339±0.012	0.324±0.014	0.276±0.014	0.296±0.012
	CRGNN	0.721±0.003	0.446±0.010	0.465±0.010	0.483±0.009	0.372±0.010	0.382±0.011	0.343±0.010	0.330±0.012	0.281±0.012	0.302±0.010
	CGNN	0.717±0.006	0.441±0.013	0.462±0.011	0.481±0.010	0.368±0.014	0.376±0.012	0.340±0.011	0.326±0.015	0.277±0.013	0.298±0.012
	GCL-LRR	0.728±0.006	0.472±0.013	0.492±0.011	0.508±0.014	0.405±0.014	0.411±0.012	0.405±0.012	0.359±0.015	0.307±0.013	0.335±0.013
	GCL-LR-Attention	**0.731±0.006**	**0.487±0.013**	**0.507±0.011**	**0.523±0.014**	**0.423±0.014**	**0.430±0.012**	**0.423±0.012**	**0.374±0.015**	**0.332±0.013**	**0.350±0.013**

contrast, GCL-LRR integrates a transductive classifier directly atop the contrastively learned representations, enabling label propagation during training. To ensure a fair comparison, we retrain all baseline GCL methods using the same transductive classifier employed in GCL-LRR, as well as an additional two-layer transductive GCN classifier. The results with different types of classifiers are shown in Table 12. It is observed that GCL-LR-Attention still outperforms SOTA GCL methods even

Table 10: Performance comparison for node classification on Coauthor-CS, Wiki-CS, Amazon-Computers, and Amazon-Photos with asymmetric label noise, symmetric label noise, and attribute noise.

Dataset	Methods	Noise Type									
		0	40			60			80		
		-	Asymmetric	Symmetric	Attribute	Asymmetric	Symmetric	Attribute	Asymmetric	Symmetric	Attribute
Coauthor-CS	GCN	0.918±0.001	0.645±0.009	0.656±0.006	0.702±0.010	0.511±0.013	0.501±0.009	0.531±0.010	0.429±0.022	0.389±0.011	0.415±0.013
	S^2GC	0.918±0.001	0.657±0.012	0.663±0.006	0.713±0.010	0.516±0.013	0.514±0.009	0.556±0.009	0.437±0.020	0.396±0.010	0.422±0.012
	GCE	0.922±0.003	0.662±0.017	0.659±0.007	0.705±0.014	0.515±0.016	0.502±0.007	0.539±0.009	0.443±0.017	0.389±0.012	0.412±0.011
	NRGNN	0.919±0.002	0.678±0.014	0.689±0.009	0.705±0.012	0.545±0.021	0.556±0.011	0.546±0.011	0.461±0.012	0.410±0.012	0.417±0.007
	RTGNN	0.920±0.005	0.678±0.012	0.691±0.009	0.712±0.008	0.559±0.010	0.569±0.011	0.560±0.008	0.455±0.015	0.415±0.015	0.412±0.014
	SUGRL	0.922±0.005	0.675±0.010	0.695±0.010	0.714±0.006	0.550±0.011	0.560±0.011	0.561±0.007	0.449±0.011	0.411±0.011	0.429±0.008
	MERIT	0.924±0.004	0.679±0.011	0.689±0.008	0.709±0.005	0.552±0.014	0.562±0.014	0.562±0.011	0.452±0.013	0.403±0.013	0.426±0.005
	ARIEL	0.925±0.004	0.682±0.011	0.699±0.009	0.712±0.005	0.555±0.011	0.566±0.011	0.556±0.011	0.454±0.014	0.415±0.019	0.427±0.013
	SFA	0.925±0.009	0.682±0.011	0.690±0.012	0.715±0.012	0.555±0.015	0.567±0.014	0.565±0.013	0.458±0.013	0.402±0.013	0.429±0.015
	Sel-Cl	0.922±0.008	0.684±0.009	0.694±0.012	0.714±0.010	0.557±0.013	0.568±0.013	0.566±0.010	0.457±0.013	0.412±0.017	0.425±0.009
	Jo-SRC	0.921±0.005	0.684±0.011	0.695±0.004	0.709±0.007	0.560±0.011	0.566±0.011	0.561±0.009	0.456±0.013	0.410±0.018	0.428±0.010
	GRAND+	0.927±0.004	0.682±0.011	0.693±0.006	0.715±0.008	0.554±0.008	0.568±0.013	0.557±0.011	0.455±0.012	0.416±0.013	0.428±0.011
	GFSA	0.923±0.004	0.679±0.010	0.687±0.009	0.711±0.009	0.550±0.012	0.559±0.011	0.558±0.010	0.453±0.014	0.410±0.012	0.426±0.011
	HONGAT	0.924±0.003	0.681±0.012	0.692±0.010	0.713±0.008	0.553±0.013	0.563±0.013	0.560±0.012	0.456±0.013	0.411±0.015	0.427±0.010
	CRGNN	0.926±0.005	0.683±0.011	0.690±0.011	0.712±0.007	0.551±0.015	0.561±0.012	0.559±0.011	0.454±0.012	0.412±0.014	0.426±0.012
	CGNN	0.925±0.006	0.680±0.012	0.689±0.012	0.710±0.010	0.549±0.014	0.560±0.012	0.557±0.012	0.452±0.013	0.409±0.015	0.425±0.012
	GCL-LRR	0.933±0.006	0.699±0.015	0.721±0.011	0.742±0.015	0.575±0.014	0.595±0.018	0.588±0.015	0.469±0.015	0.438±0.015	0.453±0.017
	GCL-LR-Attention	**0.934±0.006**	**0.714±0.015**	**0.736±0.011**	**0.758±0.015**	**0.594±0.014**	**0.612±0.018**	**0.606±0.015**	**0.489±0.015**	**0.453±0.015**	**0.472±0.017**
Wiki-CS	GCN	0.801±0.004	0.612±0.008	0.625±0.010	0.647±0.009	0.497±0.013	0.483±0.012	0.502±0.011	0.401±0.016	0.365±0.017	0.382±0.016
	S^2GC	0.806±0.003	0.621±0.009	0.630±0.011	0.659±0.010	0.503±0.014	0.492±0.012	0.516±0.012	0.411±0.018	0.373±0.016	0.397±0.015
	GCE	0.808±0.004	0.618±0.007	0.629±0.008	0.651±0.009	0.495±0.012	0.481±0.010	0.510±0.010	0.404±0.015	0.361±0.013	0.383±0.014
	NRGNN	0.809±0.003	0.635±0.008	0.642±0.009	0.665±0.008	0.514±0.012	0.518±0.011	0.526±0.010	0.426±0.014	0.386±0.016	0.403±0.012
	RTGNN	0.811±0.004	0.638±0.010	0.645±0.010	0.667±0.008	0.517±0.011	0.522±0.011	0.528±0.009	0.428±0.013	0.391±0.015	0.406±0.012
	SUGRL	0.812±0.004	0.640±0.010	0.646±0.010	0.669±0.009	0.518±0.012	0.523±0.011	0.529±0.010	0.427±0.013	0.389±0.014	0.406±0.012
	MERIT	0.813±0.004	0.641±0.009	0.648±0.010	0.670±0.009	0.519±0.012	0.525±0.012	0.532±0.010	0.432±0.014	0.392±0.013	0.410±0.013
	ARIEL	0.814±0.003	0.645±0.010	0.652±0.009	0.674±0.008	0.523±0.013	0.528±0.011	0.535±0.011	0.434±0.012	0.394±0.012	0.412±0.012
	SFA	0.815±0.005	0.643±0.011	0.650±0.010	0.673±0.009	0.520±0.013	0.527±0.012	0.533±0.010	0.430±0.015	0.391±0.014	0.408±0.012
	Sel-Cl	0.813±0.004	0.644±0.009	0.651±0.010	0.672±0.009	0.521±0.012	0.526±0.011	0.531±0.009	0.429±0.013	0.390±0.014	0.407±0.013
	Jo-SRC	0.812±0.004	0.646±0.010	0.652±0.008	0.671±0.009	0.522±0.012	0.528±0.011	0.534±0.011	0.431±0.014	0.393±0.015	0.409±0.012
	GRAND+	0.816±0.003	0.647±0.011	0.653±0.009	0.676±0.008	0.524±0.010	0.529±0.011	0.536±0.010	0.432±0.014	0.395±0.013	0.411±0.011
	GFSA	0.810±0.004	0.636±0.010	0.642±0.010	0.666±0.009	0.515±0.012	0.520±0.011	0.526±0.010	0.424±0.013	0.386±0.014	0.403±0.012
	HONGAT	0.814±0.004	0.642±0.011	0.648±0.010	0.670±0.009	0.518±0.012	0.523±0.011	0.530±0.010	0.427±0.013	0.388±0.013	0.405±0.012
	CGNN	0.813±0.004	0.643±0.010	0.649±0.009	0.669±0.009	0.519±0.011	0.524±0.011	0.531±0.010	0.428±0.014	0.389±0.015	0.406±0.012
	CRGNN	0.815±0.005	0.645±0.010	0.652±0.010	0.671±0.009	0.522±0.013	0.528±0.012	0.533±0.011	0.431±0.013	0.392±0.013	0.410±0.011
	GCL-LRR	0.821±0.006	0.664±0.015	0.679±0.011	0.690±0.015	0.540±0.014	0.558±0.018	0.554±0.015	0.442±0.015	0.415±0.015	0.423±0.017
	GCL-LR-Attention	**0.826±0.004**	**0.678±0.013**	**0.699±0.010**	**0.707±0.012**	**0.553±0.014**	**0.572±0.013**	**0.569±0.011**	**0.459±0.014**	**0.426±0.014**	**0.443±0.012**
Amazon-Computers	GCN	0.872±0.005	0.619±0.012	0.638±0.011	0.658±0.013	0.471±0.014	0.484±0.012	0.501±0.010	0.377±0.017	0.354±0.016	0.368±0.015
	S^2GC	0.876±0.004	0.625±0.010	0.642±0.012	0.664±0.011	0.479±0.013	0.491±0.013	0.509±0.012	0.382±0.016	0.359±0.015	0.375±0.014
	GCE	0.879±0.006	0.623±0.011	0.641±0.010	0.661±0.012	0.475±0.014	0.486±0.012	0.505±0.012	0.380±0.015	0.356±0.016	0.370±0.014
	NRGNN	0.878±0.004	0.639±0.010	0.656±0.010	0.672±0.011	0.491±0.013	0.503±0.011	0.518±0.011	0.391±0.014	0.364±0.015	0.380±0.014
	RTGNN	0.880±0.005	0.641±0.010	0.658±0.010	0.674±0.009	0.494±0.012	0.507±0.012	0.521±0.010	0.392±0.014	0.366±0.013	0.383±0.012
	SUGRL	0.882±0.005	0.643±0.011	0.659±0.010	0.676±0.010	0.495±0.012	0.507±0.011	0.522±0.011	0.392±0.014	0.366±0.013	0.384±0.013
	MERIT	0.883±0.004	0.644±0.011	0.660±0.010	0.676±0.009	0.496±0.012	0.508±0.012	0.523±0.011	0.394±0.015	0.368±0.013	0.386±0.012
	ARIEL	0.884±0.004	0.645±0.010	0.662±0.009	0.679±0.010	0.498±0.011	0.510±0.011	0.526±0.011	0.396±0.013	0.369±0.014	0.388±0.012
	SFA	0.885±0.005	0.643±0.011	0.661±0.010	0.677±0.010	0.497±0.012	0.509±0.011	0.525±0.012	0.395±0.013	0.368±0.012	0.387±0.013
	Sel-Cl	0.882±0.006	0.646±0.009	0.663±0.011	0.678±0.010	0.499±0.011	0.511±0.011	0.527±0.012	0.396±0.013	0.369±0.013	0.389±0.012
	Jo-SRC	0.881±0.004	0.644±0.010	0.661±0.009	0.675±0.009	0.495±0.011	0.508±0.010	0.523±0.011	0.393±0.014	0.367±0.013	0.385±0.013
	GRAND+	0.886±0.004	0.647±0.009	0.665±0.009	0.680±0.010	0.501±0.010	0.513±0.010	0.529±0.010	0.398±0.013	0.370±0.013	0.390±0.011
	GFSA	0.879±0.005	0.640±0.011	0.656±0.010	0.673±0.010	0.492±0.012	0.504±0.011	0.519±0.011	0.389±0.014	0.364±0.013	0.381±0.013
	HONGAT	0.883±0.005	0.640±0.010	0.658±0.010	0.674±0.010	0.492±0.011	0.505±0.011	0.520±0.010	0.390±0.014	0.365±0.014	0.382±0.012
	CGNN	0.884±0.005	0.642±0.010	0.659±0.010	0.676±0.010	0.494±0.011	0.507±0.011	0.522±0.011	0.392±0.014	0.366±0.013	0.384±0.013
	CRGNN	0.885±0.004	0.644±0.009	0.662±0.009	0.678±0.009	0.496±0.011	0.509±0.010	0.524±0.011	0.395±0.013	0.368±0.012	0.387±0.011
	GCL-LRR	0.889±0.006	0.659±0.011	0.682±0.007	0.690±0.011	0.514±0.011	0.523±0.013	0.535±0.012	0.414±0.012	0.390±0.012	0.403±0.011
	GCL-LR-Attention	**0.896±0.005**	**0.676±0.014**	**0.694±0.011**	**0.701±0.013**	**0.534±0.014**	**0.548±0.013**	**0.545±0.013**	**0.432±0.014**	**0.401±0.015**	**0.418±0.014**
Amazon-Photos	GCN	0.899±0.004	0.638±0.011	0.649±0.009	0.665±0.010	0.487±0.012	0.498±0.011	0.509±0.011	0.395±0.014	0.361±0.013	0.374±0.012
	S^2GC	0.903±0.005	0.645±0.010	0.655±0.010	0.672±0.010	0.495±0.011	0.506±0.010	0.517±0.011	0.399±0.014	0.366±0.013	0.379±0.013
	GCE	0.905±0.004	0.642±0.011	0.654±0.009	0.670±0.010	0.492±0.012	0.503±0.010	0.513±0.011	0.397±0.013	0.364±0.012	0.377±0.012
	NRGNN	0.906±0.003	0.653±0.010	0.663±0.010	0.680±0.010	0.503±0.011	0.514±0.010	0.526±0.010	0.408±0.013	0.371±0.012	0.386±0.012
	RTGNN	0.908±0.004	0.656±0.009	0.665±0.010	0.682±0.009	0.506±0.010	0.517±0.010	0.529±0.011	0.411±0.012	0.373±0.012	0.388±0.011
	SUGRL	0.909±0.004	0.657±0.009	0.667±0.010	0.683±0.010	0.507±0.010	0.518±0.011	0.530±0.010	0.412±0.012	0.373±0.012	0.389±0.011
	MERIT	0.910±0.005	0.659±0.010	0.667±0.009	0.684±0.010	0.508±0.011	0.519±0.010	0.531±0.010	0.413±0.012	0.374±0.012	0.390±0.012
	ARIEL	0.911±0.005	0.660±0.010	0.669±0.010	0.686±0.010	0.511±0.010	0.521±0.011	0.533±0.010	0.415±0.013	0.376±0.013	0.392±0.011
	SFA	0.912±0.004	0.658±0.009	0.668±0.010	0.685±0.009	0.509±0.010	0.520±0.010	0.532±0.010	0.414±0.012	0.375±0.011	0.391±0.012
	Sel-Cl	0.909±0.005	0.661±0.009	0.670±0.009	0.687±0.009	0.512±0.010	0.522±0.010	0.534±0.010	0.416±0.012	0.377±0.011	0.393±0.011
	Jo-SRC	0.908±0.004	0.659±0.009	0.668±0.009	0.684±0.010	0.510±0.010	0.520±0.010	0.532±0.011	0.413±0.012	0.374±0.012	0.390±0.012
	GRAND+	0.913±0.005	0.662±0.010	0.671±0.009	0.688±0.010	0.513±0.010	0.523±0.010	0.535±0.010	0.417±0.012	0.378±0.012	0.394±0.011
	GFSA	0.907±0.004	0.654±0.009	0.664±0.010	0.680±0.010	0.504±0.010	0.515±0.011	0.527±0.010	0.409±0.012	0.371±0.012	0.387±0.011
	HONGAT	0.910±0.004	0.655±0.010	0.666±0.009	0.682±0.010	0.505±0.010	0.516±0.010	0.528±0.010	0.410±0.012	0.372±0.012	0.387±0.012
	CGNN	0.911±0.004	0.657±0.009	0.667±0.010	0.683±0.010	0.507±0.010	0.518±0.011	0.530±0.010	0.412±0.012	0.373±0.012	0.389±0.011
	CRGNN	0.912±0.004	0.659±0.010	0.669±0.010	0.685±0.010	0.509±0.011	0.521±0.011	0.533±0.011	0.414±0.012	0.375±0.012	0.391±0.012
	GCL-LRR	0.919±0.010	0.680±0.013	0.692±0.013	0.702±0.007	0.530±0.014	0.541±0.014	0.544±0.012	0.435±0.012	0.390±0.014	0.410±0.010
	GCL-LR-Attention	**0.922±0.005**	**0.692±0.013**	**0.708±0.011**	**0.717±0.011**	**0.545±0.014**	**0.558±0.013**	**0.553±0.013**	**0.442±0.014**	**0.408±0.014**	**0.421±0.013**

Table 11: Performance comparison with recent robust GCL methods on all eight benchmark datasets under asymmetric label noise, symmetric label noise, and attribute noise.

Dataset	Methods	Noise Type									
		0	40			60			80		
		-	Asymmetric	Symmetric	Attribute	Asymmetric	Symmetric	Attribute	Asymmetric	Symmetric	Attribute
Cora	DKGCCL	0.850	0.565	0.690	0.672	0.463	0.558	0.449	0.272	0.385	0.330
	GRANCE	0.845	0.577	0.701	0.684	0.479	0.575	0.465	0.292	0.406	0.351
	ArnoldiGCL	0.842	0.571	0.697	0.679	0.472	0.569	0.459	0.284	0.399	0.343
	GCL-LRR	0.858	0.589	0.713	0.695	0.492	0.587	0.477	0.306	0.419	0.363
	GCL-LR-Attention	0.861	0.602	0.724	0.708	0.510	0.605	0.492	0.329	0.436	0.382
Citeseer	DKGCCL	0.739	0.496	0.558	0.547	0.381	0.426	0.378	0.335	0.301	0.285
	GRANCE	0.744	0.508	0.569	0.559	0.397	0.443	0.394	0.355	0.322	0.306
	ArnoldiGCL	0.741	0.502	0.565	0.554	0.390	0.437	0.388	0.347	0.315	0.298
	GCL-LRR	0.757	0.520	0.581	0.570	0.410	0.455	0.406	0.369	0.335	0.318
	GCL-LR-Attention	0.762	0.533	0.597	0.588	0.430	0.472	0.423	0.392	0.352	0.335
PubMed	DKGCCL	0.827	0.613	0.622	0.614	0.450	0.455	0.498	0.322	0.326	0.449
	GRANCE	0.832	0.625	0.633	0.626	0.466	0.472	0.514	0.342	0.347	0.470
	ArnoldiGCL	0.829	0.619	0.629	0.621	0.459	0.466	0.508	0.334	0.340	0.462
	GCL-LRR	0.845	0.637	0.645	0.637	0.479	0.484	0.526	0.356	0.360	0.482
	GCL-LR-Attention	0.846	0.652	0.662	0.655	0.498	0.503	0.544	0.379	0.379	0.498
ogbn-arxiv	DKGCCL	0.710	0.448	0.469	0.485	0.376	0.382	0.377	0.325	0.273	0.302
	GRANCE	0.715	0.460	0.480	0.497	0.392	0.399	0.393	0.345	0.294	0.323
	ArnoldiGCL	0.712	0.454	0.476	0.492	0.385	0.393	0.387	0.337	0.287	0.315
	GCL-LRR	0.728	0.472	0.492	0.508	0.405	0.411	0.405	0.359	0.307	0.335
	GCL-LR-Attention	0.731	0.487	0.507	0.523	0.423	0.430	0.423	0.374	0.332	0.350
Coauthor-CS	DKGCCL	0.915	0.675	0.698	0.719	0.546	0.566	0.560	0.435	0.404	0.420
	GRANCE	0.920	0.687	0.709	0.731	0.562	0.583	0.576	0.455	0.425	0.441
	ArnoldiGCL	0.917	0.681	0.705	0.726	0.555	0.577	0.570	0.447	0.418	0.433
	GCL-LRR	0.933	0.699	0.721	0.742	0.575	0.595	0.588	0.469	0.438	0.453
	GCL-LR-Attention	0.934	0.714	0.736	0.758	0.594	0.612	0.606	0.489	0.453	0.472
Wiki-CS	DKGCCL	0.803	0.640	0.656	0.667	0.511	0.529	0.526	0.408	0.381	0.390
	GRANCE	0.808	0.652	0.667	0.679	0.527	0.546	0.542	0.428	0.402	0.411
	ArnoldiGCL	0.805	0.646	0.663	0.674	0.520	0.540	0.536	0.420	0.395	0.403
	GCL-LRR	0.821	0.664	0.679	0.690	0.540	0.558	0.554	0.442	0.415	0.423
	GCL-LR-Attention	0.826	0.678	0.699	0.707	0.553	0.572	0.569	0.459	0.426	0.443
Amazon-Computers	DKGCCL	0.871	0.635	0.659	0.667	0.485	0.494	0.507	0.380	0.356	0.370
	GRANCE	0.876	0.647	0.670	0.679	0.501	0.511	0.523	0.400	0.377	0.391
	ArnoldiGCL	0.873	0.641	0.666	0.674	0.494	0.505	0.517	0.392	0.370	0.383
	GCL-LRR	0.889	0.659	0.682	0.690	0.514	0.523	0.535	0.414	0.390	0.403
	GCL-LR-Attention	0.896	0.676	0.694	0.701	0.534	0.548	0.545	0.432	0.401	0.418
Amazon-Photos	DKGCCL	0.901	0.656	0.669	0.679	0.501	0.512	0.516	0.401	0.356	0.377
	GRANCE	0.906	0.668	0.680	0.691	0.517	0.529	0.532	0.421	0.377	0.398
	ArnoldiGCL	0.903	0.662	0.676	0.686	0.510	0.523	0.526	0.413	0.370	0.390
	GCL-LRR	0.919	0.680	0.692	0.702	0.530	0.541	0.544	0.435	0.390	0.410
	GCL-LR-Attention	0.922	0.692	0.708	0.717	0.545	0.558	0.553	0.442	0.408	0.421

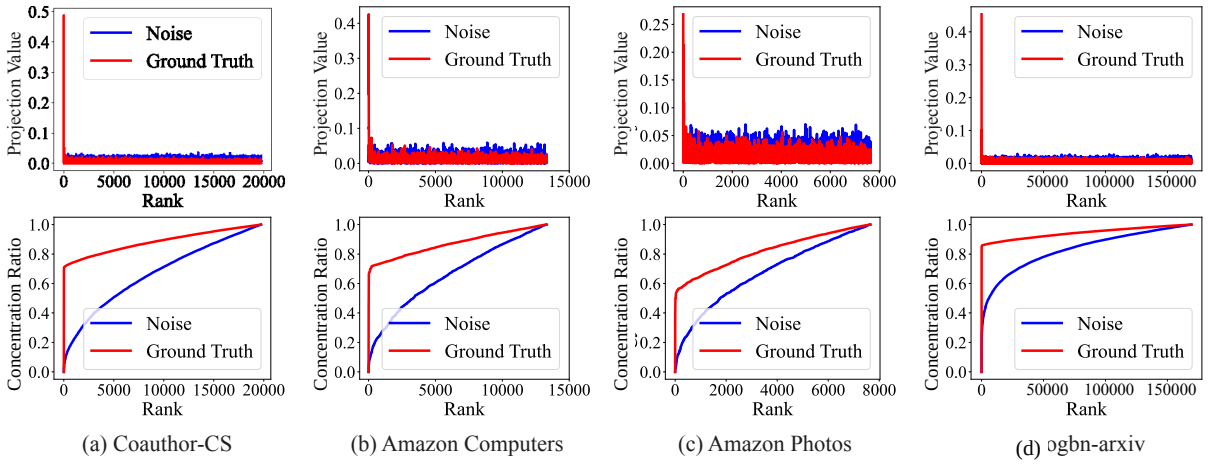


Figure 4: Eigen-projection (first row) and energy concentration (second row) on Coauthor-CS, Amazon-Computers, Amazon-Photos, and ogbn-arxiv. In the second row, the red curves denote the clean-label concentration ratios, whereas the blue curves denote the noise concentration ratios. By the rank of $0.2 \min\{N, d\}$, the concentration entropy on Coauthor-CS, Amazon-Computers, Amazon-Photos, and ogbn-arxiv are 0.779, 0.809, 0.752, and 0.787.

when the transductive classifiers are employed.

Table 12: Performance comparison for node classification by inductive linear classifier, transductive GCN classifier, and transductive classifier used in GCL-LRR. The comparisons are performed on Cora.

Methods	Noise Type									
	0	40			60			80		
	-	Asymmetric	Symmetric	Attribute	Asymmetric	Symmetric	Attribute	Asymmetric	Symmetric	Attribute
SUGRL (original, inductive classifier)	0.834±0.005	0.564±0.011	0.674±0.012	0.675±0.009	0.468±0.011	0.552±0.011	0.452±0.012	0.280±0.012	0.381±0.012	0.338±0.014
SUGRL + transductive GCN	0.833±0.006	0.562±0.013	0.675±0.015	0.673±0.012	0.472±0.011	0.551±0.011	0.454±0.012	0.280±0.012	0.380±0.012	0.340±0.014
SUGRL + linear transductive classifier	0.836±0.007	0.568±0.013	0.677±0.010	0.674±0.011	0.472±0.011	0.555±0.011	0.457±0.012	0.284±0.012	0.383±0.012	0.341±0.014
MERIT (original, inductive classifier)	0.831±0.005	0.560±0.008	0.670±0.008	0.671±0.009	0.467±0.013	0.547±0.013	0.450±0.014	0.277±0.013	0.385±0.013	0.335±0.009
MERIT + transductive GCN	0.831±0.007	0.562±0.011	0.668±0.013	0.672±0.014	0.466±0.013	0.549±0.015	0.451±0.016	0.276±0.012	0.382±0.014	0.337±0.013
MERIT + linear transductive classifier	0.833±0.003	0.562±0.014	0.673±0.012	0.673±0.011	0.466±0.015	0.546±0.016	0.453±0.017	0.280±0.016	0.386±0.011	0.336±0.014
SFA (original, inductive classifier)	0.839±0.010	0.564±0.011	0.677±0.013	0.676±0.015	0.473±0.014	0.549±0.014	0.457±0.014	0.282±0.016	0.389±0.013	0.344±0.017
SFA + transductive GCN	0.837±0.013	0.565±0.011	0.673±0.017	0.673±0.018	0.474±0.016	0.551±0.015	0.453±0.018	0.277±0.016	0.389±0.015	0.343±0.019
SFA + linear transductive classifier	0.841±0.015	0.566±0.013	0.678±0.014	0.679±0.014	0.477±0.015	0.552±0.012	0.456±0.016	0.284±0.017	0.391±0.015	0.348±0.019
GCL-LRR	0.858±0.006	0.589±0.011	0.713±0.007	0.695±0.011	0.492±0.011	0.587±0.013	0.477±0.012	0.306±0.012	0.419±0.012	0.363±0.011
GCL-LR-Attention	**0.861±0.006**	**0.602±0.011**	**0.724±0.007**	**0.708±0.011**	**0.510±0.011**	**0.605±0.013**	**0.492±0.012**	**0.329±0.012**	**0.436±0.012**	**0.382±0.011**

B.9 STATISTICAL SIGNIFICANCE ANALYSIS

In this section, we compute the p-values from paired t-tests comparing the performance of GCL-LRR and GCL-LR-Attention against the strongest baseline methods to assess the statistical significance of the observed improvements. As shown in Table 13, the p-values for both GCL-LRR and GCL-LR-Attention remain consistently below the threshold of 0.05 across all datasets and noise conditions, thereby confirming that the performance gains over the top baselines are statistically significant.

B.10 SENSITIVITY ANALYSIS ON THE HYPERPARAMETERS

We perform a sensitivity analysis on the weighting parameter τ associated with the TNN $\|\mathbf{K}\|_{r_0}$ in Equation 1. This analysis is conducted using GCL-LR-Attention on the Coauthor-CS dataset under the setting of semi-supervised node classification with 60% asymmetric label noise. We vary τ over the set $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$, and the corresponding classification accuracies are reported in Table 14. The highest accuracy is achieved at $\tau = 0.5$, though GCL-LR-Attention maintains stable and competitive performance across the entire range. Even in the least favorable case, with $\tau = 0.1$, the accuracy declines by only 0.6%, highlighting the robustness of the model to variations in τ.

Table 13: P-values of the t-tests for GCL-LRR and GCL-LR-Attention against the top baseline methods under each noise setting on all the benchmark datasets.

Datasets	Methods	Noise Type								
		40			60			80		
		Asymmetric	Symmetric	Attribute	Asymmetric	Symmetric	Attribute	Asymmetric	Symmetric	Attribute
Cora	GCL-LRR	0.038	0.024	0.021	0.035	0.021	0.041	0.025	0.028	0.031
	GCL-LR-Attention	0.027	0.022	0.018	0.018	0.038	0.035	0.031	0.023	0.030
Citeseer	GCL-LRR	0.043	0.035	0.022	0.021	0.030	0.041	0.027	0.024	0.022
	GCL-LR-Attention	0.037	0.038	0.019	0.027	0.025	0.040	0.035	0.037	0.030
PubMed	GCL-LRR	0.028	0.043	0.030	0.026	0.027	0.043	0.036	0.040	0.042
	GCL-LR-Attention	0.025	0.030	0.026	0.023	0.024	0.041	0.033	0.035	0.037
Coauthor-CS	GCL-LRR	0.041	0.032	0.036	0.043	0.044	0.040	0.027	0.037	0.042
	GCL-LR-Attention	0.036	0.030	0.034	0.041	0.042	0.039	0.025	0.033	0.036
ogbn-arxiv	GCL-LRR	0.040	0.032	0.036	0.043	0.044	0.040	0.027	0.037	0.042
	GCL-LR-Attention	0.036	0.030	0.034	0.041	0.042	0.039	0.025	0.033	0.036
Wiki-CS	GCL-LRR	0.044	0.035	0.033	0.028	0.034	0.041	0.036	0.040	0.042
	GCL-LR-Attention	0.040	0.036	0.031	0.026	0.029	0.039	0.034	0.038	0.040
Amazon-Computers	GCL-LRR	0.033	0.043	0.034	0.031	0.034	0.041	0.036	0.040	0.042
	GCL-LR-Attention	0.031	0.038	0.030	0.028	0.030	0.038	0.034	0.037	0.040
Amazon-Photos	GCL-LRR	0.040	0.041	0.038	0.043	0.044	0.040	0.027	0.037	0.042
	GCL-LR-Attention	0.037	0.039	0.036	0.041	0.042	0.039	0.025	0.033	0.036
Texas	GCL-LRR	0.038	0.024	0.021	0.035	0.021	0.041	0.025	0.028	0.031
	GCL-LR-Attention	0.027	0.022	0.018	0.018	0.038	0.035	0.031	0.023	0.030
Chameleon	GCL-LRR	0.044	0.035	0.033	0.028	0.034	0.041	0.036	0.040	0.042
	GCL-LR-Attention	0.040	0.036	0.031	0.026	0.029	0.039	0.034	0.038	0.040

Table 14: Sensitivity analysis on the weighting parameter τ for the TNN. The study is performed using GCL-LR-Attention on the Coauthor-CS dataset for semi-supervised node classification under 60% asymmetric label noise.

τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Accuracy	0.588	0.590	0.593	0.591	0.594	0.590	0.591	0.590	0.589

B.11 SENSITIVITY ANALYSIS OF THE THEORETICAL UPPER BOUND

To more directly validate the theoretical role of the KC term in Theorem 4.1, we further compute $L_1(\mathbf{K}, \tilde{\mathbf{Y}}, t)$, $L_2(\mathbf{K}, \mathbf{N}, t)$, $\mathrm{KC}(\mathbf{K})$, and the corresponding upper bound under different values of τ, while keeping the model architecture fixed as GCL-LR-Attention. As shown in Table 15, the best accuracy is achieved at $\tau = 0.5$, which also yields the lowest L_1, L_2, $\mathrm{KC}(\mathbf{K})$, and upper bound. This observation directly supports the theoretical claim that reducing kernel complexity tightens the generalization bound and improves robustness.

Table 15: Sensitivity analysis of the empirical accuracy and theoretical upper-bound terms under different values of τ. The study is performed using GCL-LR-Attention on the Coauthor-CS dataset under 60% asymmetric label noise.

τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Accuracy $\uparrow$	0.588	0.590	0.593	0.591	**0.594**	0.590	0.591	0.590	0.589
$L_1(\mathbf{K}, \tilde{\mathbf{Y}}, t) \downarrow$	4.35	4.26	4.14	4.08	**3.96**	4.06	4.13	4.25	4.32
$L_2(\mathbf{K}, \mathbf{N}, t) \downarrow$	3.86	3.78	3.66	3.59	**3.51**	3.58	3.64	3.76	3.83
$\mathrm{KC}(\mathbf{K}) \downarrow$	0.29	0.26	0.23	0.21	**0.18**	0.20	0.22	0.25	0.27
Upper Bound $\downarrow$	8.77	8.57	8.34	8.18	**7.96**	8.13	8.29	8.53	8.69

B.12 ABLATION STUDY ON THE RANK IN THE TRUNCATED NUCLEAR NORM

We conduct an ablation study to investigate the effect of the rank parameter r_0 in the TNN $\|\mathbf{K}\|_{r_0}$ used in the loss function (1) of GCL-LRR. As shown in Table 16, the performance of GCL-LRR remains consistently close to the optimal across different settings of r_0, particularly when the rank lies within the range $0.1 \min\{N, d\}$ to $0.3 \min\{N, d\}$.

Table 16: Ablation study on the value of rank r_0 in the optimization problem (2) on Cora with different levels of asymmetric and symmetric label noise. The accuracy with the optimal rank is shown in the last row. The accuracy difference against the optimal rank is shown for other ranks.

Rank	Noise Type						
	0	40		60		80	
	-	Asymmetric	Symmetric	Asymmetric	Symmetric	Asymmetric	Symmetric
$0.1\min\{N,d\}$	-0.002	-0.001	-0.002	-0.002	-0.001	-0.001	-0.000
$0.2\min\{N,d\}$	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
$0.3\min\{N,d\}$	-0.000	-0.000	-0.001	-0.002	-0.001	-0.000	-0.001
$0.4\min\{N,d\}$	-0.001	-0.003	-0.002	-0.001	-0.002	-0.002	-0.002
$0.5\min\{N,d\}$	-0.001	-0.002	-0.003	-0.003	-0.003	-0.001	-0.002
$0.6\min\{N,d\}$	-0.003	-0.002	-0.002	-0.003	-0.002	-0.002	-0.003
$0.7\min\{N,d\}$	-0.003	-0.004	-0.003	-0.004	-0.004	-0.004	-0.005
$0.8\min\{N,d\}$	-0.002	-0.005	-0.006	-0.006	-0.006	-0.007	-0.007
$0.9\min\{N,d\}$	-0.004	-0.004	-0.005	-0.007	-0.008	-0.008	-0.006
$\min\{N,d\}$	-0.004	-0.004	-0.007	-0.007	-0.008	-0.010	-0.008
optimal	0.858	0.589	0.713	0.492	0.587	0.306	0.419

B.13 RANK ABLATION ON ADDITIONAL DATASETS

We extend the ablation study on the rank parameter r_0 to Coauthor-CS and ogbn-arxiv. As shown in Table 17, GCL-LRR remains stable across a broad range of rank choices, and the optimal performance is achieved at moderate ranks. The full-rank setting consistently degrades accuracy, indicating that the low-rank regularizer improves robustness while preserving sufficient representation capacity.

Table 17: Ablation study on the value of rank r_0 on Coauthor-CS and ogbn-arxiv with different levels of asymmetric and symmetric label noise. The accuracy with the optimal rank is shown in the last row of each dataset block. The accuracy difference against the optimal rank is shown for other ranks.

Dataset	Rank	Noise Type						
		0	40		60		80	
		-	Asymmetric	Symmetric	Asymmetric	Symmetric	Asymmetric	Symmetric
Coauthor-CS	$0.1\min\{N,d\}$	-0.003	-0.004	-0.002	-0.005	-0.003	-0.004	-0.002
	$0.2\min\{N,d\}$	-0.001	-0.002	-0.001	-0.001	-0.002	-0.001	-0.001
	$0.3\min\{N,d\}$	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	$0.4\min\{N,d\}$	-0.001	-0.001	-0.002	-0.002	-0.001	-0.002	-0.001
	$0.5\min\{N,d\}$	-0.002	-0.003	-0.002	-0.003	-0.002	-0.003	-0.002
	$0.6\min\{N,d\}$	-0.004	-0.003	-0.004	-0.005	-0.004	-0.004	-0.003
	$0.7\min\{N,d\}$	-0.005	-0.006	-0.005	-0.006	-0.005	-0.007	-0.005
	$0.8\min\{N,d\}$	-0.007	-0.008	-0.007	-0.008	-0.007	-0.009	-0.008
	$0.9\min\{N,d\}$	-0.009	-0.010	-0.011	-0.010	-0.011	-0.012	-0.010
	$\min\{N,d\}$	-0.011	-0.013	-0.012	-0.014	-0.013	-0.015	-0.012
	optimal	0.933	0.699	0.721	0.575	0.595	0.469	0.438
ogbn-arxiv	$0.1\min\{N,d\}$	-0.006	-0.005	-0.006	-0.005	-0.006	-0.004	-0.005
	$0.2\min\{N,d\}$	-0.003	-0.004	-0.003	-0.003	-0.004	-0.003	-0.002
	$0.3\min\{N,d\}$	-0.001	-0.002	-0.001	-0.001	-0.002	-0.001	-0.001
	$0.4\min\{N,d\}$	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	$0.5\min\{N,d\}$	-0.001	-0.001	-0.002	-0.001	-0.001	-0.002	-0.001
	$0.6\min\{N,d\}$	-0.003	-0.002	-0.003	-0.003	-0.004	-0.003	-0.003
	$0.7\min\{N,d\}$	-0.005	-0.004	-0.005	-0.006	-0.005	-0.005	-0.006
	$0.8\min\{N,d\}$	-0.007	-0.006	-0.008	-0.007	-0.008	-0.007	-0.008
	$0.9\min\{N,d\}$	-0.009	-0.008	-0.010	-0.009	-0.011	-0.010	-0.009
	$\min\{N,d\}$	-0.012	-0.010	-0.013	-0.012	-0.014	-0.013	-0.012
	optimal	0.728	0.472	0.492	0.405	0.411	0.359	0.307

B.14 STUDY ON THE EFFECTIVENESS OF THE LOW-RANK REGULARIZATION ON GCN

In this section, we conduct an ablation study by comparing GCL-LR-Attention to an ablation model, GCN-LR-Attention, which applies the LR-Attention layer with low-rank regularization to a standard GCN [Kipf and Welling, 2017] for semi-

supervised node classification. Let $\mathbf{H}_{\text{GCN}} \in \mathbb{R}^{N\times d}$ be the output node features of a standard GCN with the same number of layers and hidden dimensions as the GCL encoder in GCL-LR-Attention. Let $\mathbf{K}_{\mathbf{H}_{\text{GCN}}} = \mathbf{H}_{\text{GCN}}\mathbf{H}_{\text{GCN}}^{\top}/N \in \mathbb{R}^{N\times N}$ be the gram matrix of the node features $\mathbf{H}_{\text{GCN}}$. Let $\mathbf{B}_{\text{GCN}} \in \mathbb{R}^{N\times N}$ be the attention matrix of the LR-Attention layer computed from $\mathbf{K}_{\mathbf{H}_{\text{GCN}}}$ in the same way as defined in Section 4.3. Let $\boldsymbol{\theta}_{\text{GCN}}$ denote the network parameters of the GCN. The parameters of the GCN-LR-Attention are trained jointly with a linear classifier applied to $\mathbf{B}_{\text{GCN}}\mathbf{H}_{\text{GCN}}$ by minimizing the following loss function,

$$\min_{\boldsymbol{\theta}_{\text{GCN}},\mathbf{W}_{\text{GCN}}} \mathcal{L}_{\text{GCN-LR-Attention}}(\boldsymbol{\theta}_{\text{GCN}},\mathbf{W}_{\text{GCN}}) = \mathcal{L}_{\text{CE}}(\boldsymbol{\theta}_{\text{GCN}},\mathbf{W}_{\text{GCN}}) + \tau\|\mathbf{K}_{\mathbf{H}_{\text{GCN}}}(\boldsymbol{\theta}_{\text{GCN}})\|_{r_0}, \tag{5}$$

$$\tag{6}$$

where

$$\mathcal{L}_{\text{CE}}(\boldsymbol{\theta}_{\text{GCN}},\mathbf{W}_{\text{GCN}}) = \frac{1}{m}\sum_{v_i\in\mathcal{V}_{\mathcal{L}}} \text{KL}\left(\mathbf{y}_i, [\text{softmax}\left(\mathbf{B}_{\text{GCN}}\mathbf{H}_{\text{GCN}}\mathbf{W}_{\text{GCN}}\right)]_i\right)$$

is the standard cross-entropy loss, and $\mathbf{W}_{\text{GCN}} \in \mathbb{R}^{d\times C}$ is the weight matrix of the linear classifier. All hyperparameters, including τ and r_0, are selected by cross-validation following the same settings as in Section A of the supplementary.

The comparisons between GCL-LR-Attention and GCN-LR-Attention are conducted on the Cora and Citeseer under three different types of noise, including asymmetric label noise, symmetric label noise, and attribute noise, with noise level fixed at 60%. For comparison, we also report the results of the standard GCN and GCL, both of which are trained without the LR-Attention layer and low-rank regularization. As shown in Table 18, GCN-LR-Attention consistently outperforms the GCN baseline across all noise settings, demonstrating that the LR-Attention layer with low-rank regularization alone is effective in improving the robustness of transductive node classification under noise for GCN. Importantly, GCL improves over the GCN baseline, indicating that GCL exhibits better robustness than GCN. Finally, GCL-LR-Attention, which combines the LR-Attention layer with low-rank regularization and GCL, achieves the best performance in all cases, highlighting the complementary benefits of the two components and explaining why we employ GCL encoders in this paper. For instance, GCL-LR-Attention outperforms GCN-LR-Attention by 4.5% and the GCL baseline by 4.0% in node classification accuracy on Citeseer with asymmetric label noise at a noise level of 60%.

Table 18: Ablation results for comparison between GCN-LR-Attention and GCL-LR-Attention under different noise types. The noise level is fixed at 60% for all noise settings.

Dataset	Noise Type	GCN	GCN-LR-Attention	GCL	GCL-LR-Attention
Cora	Asymmetric	0.405	0.447	0.472	**0.510**
Cora	Symmetric	0.517	0.542	0.556	**0.605**
Cora	Attribute	0.439	0.458	0.458	**0.492**
Citeseer	Asymmetric	0.351	0.385	0.390	**0.430**
Citeseer	Symmetric	0.341	0.389	0.425	**0.472**
Citeseer	Attribute	0.372	0.392	0.390	**0.423**

B.15 EXTENDED COMPONENT ABLATION STUDIES

We extend the component ablation studies to all eight benchmark datasets under 60% asymmetric label noise, symmetric label noise, and attribute noise. Table 19 isolates the effects of the LR-Attention layer, the GCL encoder, the TNN-based low-rank regularization, and standard full-rank attention. The consistent ordering across datasets shows that GCN-LR-Attention improves over GCN, GCL improves over GCN, GCL-LRR improves over vanilla GCL by adding low-rank regularization, and GCL-LR-Attention achieves the strongest performance. The comparison with GCL+Standard Attention [Veličković et al., 2018] further shows that the gains are not merely due to adding an attention module, but come from the proposed low-rank attention design.

B.16 TRAINING TIME COMPARISON

In this section, we report a comparative analysis of the training time of GCL-LR-Attention, GCL-LRR and other baseline methods across all benchmark datasets. For the baseline GCL methods, the reported training time includes both the encoder training phase and the downstream classifier training. All experiments are conducted using a single 80 GB NVIDIA A100

Table 19: Extended component ablation results on all eight benchmark datasets. The noise level is fixed at 60% for all noise settings.

Dataset	Noise Type	GCN	GCN-LR-Attention	GCL	GCL-LRR	GCL+Standard Attention	GCL-LR-Attention
Cora	Asymmetric	0.405	0.447	0.472	0.492	0.486	0.510
	Symmetric	0.517	0.542	0.556	0.587	0.575	0.605
	Attribute	0.439	0.458	0.458	0.477	0.472	0.492
Citeseer	Asymmetric	0.351	0.385	0.390	0.410	0.403	0.430
	Symmetric	0.341	0.389	0.425	0.455	0.442	0.472
	Attribute	0.372	0.392	0.390	0.406	0.404	0.423
PubMed	Asymmetric	0.405	0.444	0.462	0.479	0.476	0.498
	Symmetric	0.386	0.431	0.467	0.484	0.481	0.503
	Attribute	0.488	0.512	0.520	0.526	0.532	0.544
ogbn-arxiv	Asymmetric	0.336	0.368	0.386	0.405	0.400	0.423
	Symmetric	0.346	0.373	0.392	0.411	0.409	0.430
	Attribute	0.339	0.364	0.386	0.405	0.399	0.423
Coauthor-CS	Asymmetric	0.511	0.542	0.558	0.575	0.574	0.594
	Symmetric	0.501	0.548	0.575	0.595	0.592	0.612
	Attribute	0.531	0.552	0.570	0.588	0.588	0.606
Wiki-CS	Asymmetric	0.472	0.501	0.519	0.540	0.536	0.553
	Symmetric	0.486	0.515	0.535	0.558	0.553	0.572
	Attribute	0.493	0.514	0.531	0.554	0.550	0.569
Amazon-Computers	Asymmetric	0.452	0.481	0.501	0.514	0.518	0.534
	Symmetric	0.463	0.493	0.514	0.523	0.531	0.548
	Attribute	0.471	0.497	0.512	0.535	0.529	0.545
Amazon-Photos	Asymmetric	0.461	0.492	0.510	0.530	0.527	0.545
	Symmetric	0.474	0.504	0.523	0.541	0.540	0.558
	Attribute	0.480	0.505	0.520	0.544	0.537	0.553

GPU. As shown in Table 20, the training time of GCL-LRR and GCL-LR-Attention is comparable to that of state-of-the-art GCL and graph attention methods, while achieving substantially better node classification performance under 60% asymmetric label noise. For instance, on the large-scale ogbn-arxiv dataset, GCL-LRR requires 1552.7s of training time, which is 5.2% less than the state-of-the-art robust GCL method, CRGNN, and 14.9% less than the state-of-the-art robust graph attention method, HONGAT, while achieving 3.3% and 3.9% higher node classification accuracy on graphs with 60% asymmetric label noise. GCL-LR-Attention incurs a slightly higher training cost of 1660.3s, corresponding to only a marginal 1.4% increase over CRGNN, while still reducing training time by 9.0% compared with HONGAT, while improving accuracy by 5.1% and 5.7%, respectively, on ogbn-arxiv with 60% asymmetric label noise.

C THEORETICAL RESULTS

We present the proof of Theorem 4.1 in this section.

Proof of Theorem 4.1. Define $\mathbf{N} \coloneqq \mathbf{Y} - \tilde{\mathbf{Y}} \in \mathbb{R}^N$ as the label noise. It can be verified that at the t-th iteration of gradient descent for $t \geq 1$, we have

$$\begin{aligned}\mathbf{W}^{(t)} &= \mathbf{W}^{(t-1)} - \frac{\eta}{m} [\mathbf{H}]_{\mathcal{L}}^{\top} \left[\mathbf{H}\mathbf{W}^{(t-1)} - \mathbf{Y}\right]_{\mathcal{L}} \\ &= \mathbf{W}^{(t-1)} - \frac{\eta}{m} [\mathbf{H}]_{\mathcal{L}}^{\top} \left[\mathbf{H}\mathbf{W}^{(t-1)} - \tilde{\mathbf{Y}}\right]_{\mathcal{L}} + \frac{\eta}{m} [\mathbf{H}]_{\mathcal{L}}^{\top} [\mathbf{N}]_{\mathcal{L}} . \end{aligned} \tag{7}$$

It follows from (7) that

$$[\mathbf{H}]_{\mathcal{L}} \mathbf{W}^{(t)} = [\mathbf{H}]_{\mathcal{L}} \mathbf{W}^{(t-1)} - \eta \mathbf{K}_{\mathcal{L},\mathcal{L}} \left[\mathbf{H}\mathbf{W}^{(t-1)} - \tilde{\mathbf{Y}}\right]_{\mathcal{L}} + \eta \mathbf{K}_{\mathcal{L},\mathcal{L}} [\mathbf{N}]_{\mathcal{L}} , \tag{8}$$

where $\mathbf{K}_{\mathcal{L},\mathcal{L}} = [\mathbf{H}]_{\mathcal{L}} [\mathbf{H}]_{\mathcal{L}}^{\top} / m \in \mathbb{R}^{m \times m}$. With $\mathbf{F}(\mathbf{W}, t) = \mathbf{H}\mathbf{W}^{(t)}$, it follows from (8) that

$$\left[\mathbf{F}(\mathbf{W}, t) - \tilde{\mathbf{Y}}\right]_{\mathcal{L}} = (\mathbf{I}_m - \eta \mathbf{K}_{\mathcal{L},\mathcal{L}}) \left[\mathbf{F}(\mathbf{W}, t-1) - \tilde{\mathbf{Y}}\right]_{\mathcal{L}} + \eta \mathbf{K}_{\mathcal{L},\mathcal{L}} [\mathbf{N}]_{\mathcal{L}} .$$

Table 20: Training time (seconds) comparisons for semi-supervised node classification. Herein we also compare the node classification accuracy of GCL-LRR, GCL-LR-Attention, and other competing methods based on GCL and graph attention under 60% asymmetric label noise. The training time is evaluated for the entire training process.

Methods		Cora	Citeseer	PubMed	Coauthor-CS	Wiki-CS	Computer	Photo	ogbn-arxiv
SUGRL	Training Time	100.3	122.1	207.4	227.1	107.7	142.8	87.7	946.8
	Accuracy	0.468	0.376	0.445	0.550	0.518	0.495	0.507	0.365
MERIT	Training Time	167.2	179.2	336.7	375.3	172.3	226.5	140.6	1495.1
	Accuracy	0.467	0.383	0.447	0.552	0.519	0.496	0.508	0.368
ARIEL	Training Time	156.9	164.3	284.3	332.6	145.1	190.4	118.3	1261.4
	Accuracy	0.471	0.391	0.453	0.555	0.523	0.498	0.511	0.379
SFA	Training Time	237.5	269.4	457.1	492.3	233.5	304.5	187.2	2013.1
	Accuracy	0.473	0.390	0.447	0.555	0.520	0.497	0.509	0.368
GFSA	Training Time	177.8	198.4	398.8	427.5	209.1	271.4	170.3	1792.0
	Accuracy	0.466	0.386	0.450	0.550	0.515	0.492	0.504	0.370
HONGAT	Training Time	181.2	195.1	406.2	435.5	213.0	276.4	173.5	1825.5
	Accuracy	0.464	0.380	0.448	0.553	0.518	0.492	0.505	0.366
CRGNN	Training Time	159.6	178.0	357.8	383.5	187.6	243.5	152.8	1637.8
	Accuracy	0.472	0.389	0.452	0.551	0.522	0.496	0.509	0.372
CGNN	Training Time	157.9	171.1	339.1	379.5	177.6	224.9	141.2	1550.1
	Accuracy	0.462	0.385	0.449	0.549	0.519	0.494	0.507	0.368
GCL-LRR	Training Time	159.9	174.5	350.7	380.9	180.3	235.7	145.5	1552.7
	Accuracy	0.492	0.410	0.479	0.575	0.540	0.514	0.530	0.405
GCL-LR-Attention	Training Time	165.1	182.2	369.7	397.5	193.1	251.2	158.2	1660.3
	Accuracy	**0.510**	**0.430**	**0.498**	**0.594**	**0.553**	**0.534**	**0.545**	**0.423**

It follows from the above equality and the recursion that

$$\left[\mathbf{F}(\mathbf{W},t)-\tilde{\mathbf{Y}}\right]_{\mathcal{L}} = -\left(\mathbf{I}_m-\eta\mathbf{K}_{\mathcal{L},\mathcal{L}}\right)^t\left[\tilde{\mathbf{Y}}\right]_{\mathcal{L}}+\eta\mathbf{K}_{\mathcal{L},\mathcal{L}}\sum_{t'=0}^{t-1}\left(\mathbf{I}_m-\eta\mathbf{K}_{\mathcal{L},\mathcal{L}}\right)^{t'}\left[\mathbf{N}\right]_{\mathcal{L}}. \tag{9}$$

We apply [Yang, 2025b, Eq. (B.70)] to obtain the following bound for the test loss $\frac{1}{u}\left\|\left[\mathbf{F}(\mathbf{W},t)-\tilde{\mathbf{Y}}\right]_{\mathcal{U}}\right\|_{\mathrm{F}}^2$:

$$\frac{1}{u}\left\|\left[\mathbf{F}(\mathbf{W},t)-\tilde{\mathbf{Y}}\right]_{\mathcal{U}}\right\|_{\mathrm{F}}^2 \le \frac{1}{m}\left\|\left[\mathbf{F}(\mathbf{W},t)-\tilde{\mathbf{Y}}\right]_{\mathcal{L}}\right\|_{\mathrm{F}}^2 + c_0\min_{0\le Q\le N} r(u,m,Q)+\frac{c_0x}{N_{u,m,\delta}}, \tag{10}$$

with

$$r(u,m,Q) \coloneqq Q\left(\frac{1}{u}+\frac{1}{m}\right)+\left(\sqrt{\frac{\sum_{q=Q+1}^{N}\widehat{\lambda}_q}{u}}+\sqrt{\frac{\sum_{q=Q+1}^{N}\widehat{\lambda}_q}{m}}\right),$$

where c_0 is a positive constant depending on $\mathbf{U}$, $\left\{\widehat{\lambda}_i\right\}_{i=1}^{r}$, and τ_0 with $\tau_0^2=\max_{i\in[N]}\mathbf{K}_{ii}$.

It follows from (9) and (10) that for every $r_0\in[N]$, we have

$$\begin{aligned}
\frac{1}{u}\left\|\left[\mathbf{F}(\mathbf{W},t)-\tilde{\mathbf{Y}}\right]_{\mathcal{U}}\right\|_{\mathrm{F}}^2 &\le \frac{1}{m}\left\|\left[\mathbf{F}(\mathbf{W},t)-\tilde{\mathbf{Y}}\right]_{\mathcal{L}}\right\|_{\mathrm{F}}^2 \\
&\quad + c_0r_0\left(\frac{1}{u}+\frac{1}{m}\right)+c_0\left(\sqrt{\frac{\sum_{q=r_0+1}^{N}\widehat{\lambda}_q}{u}}+\sqrt{\frac{\sum_{q=r_0+1}^{N}\widehat{\lambda}_q}{m}}\right)+\frac{c_0x}{N_{u,m,\delta}} \\
&\overset{①}{\le} \frac{2}{m}\left\|\left(\mathbf{I}_m-\eta\mathbf{K}_{\mathcal{L},\mathcal{L}}\right)^t\left[\tilde{\mathbf{Y}}\right]_{\mathcal{L}}\right\|_{\mathrm{F}}^2+\frac{2}{m}\left\|\eta\mathbf{K}_{\mathcal{L},\mathcal{L}}\sum_{t'=0}^{t-1}\left(\mathbf{I}_m-\eta\mathbf{K}_{\mathcal{L},\mathcal{L}}\right)^{t'}\left[\mathbf{N}\right]_{\mathcal{L}}\right\|_{\mathrm{F}}^2 \\
&\quad + c_0r_0\left(\frac{1}{u}+\frac{1}{m}\right)+c_0\sqrt{\|\mathbf{K}\|_{r_0}}\left(\sqrt{\frac{1}{u}}+\sqrt{\frac{1}{m}}\right)+\frac{c_0x}{N_{u,m,\delta}},
\end{aligned} \tag{11}$$

where ① follows from the Cauchy-Schwarz inequality, (9), and $\sum_{q=r_0+1}^{N} \widehat{\lambda}_q = \|\mathbf{K}\|_{r_0}$. (3) then follows directly from (11). □