Kernel Complexity Reduced Graph Contrastive Learning for Noisy Node Classification

Anonymous authors
Paper under double-blind review

Abstract

Graph Neural Networks (GNNs) have achieved remarkable success in learning node representations and have demonstrated strong performance on node classification. However, their effectiveness can be substantially compromised by noise in real-world graph data. To address this challenge, we propose Kernel Complexity Reduced Graph Contrastive Learning (KCR-GCL), a principled framework for noisy node classification with a provable transductive generalization guarantee. KCR-GCL introduces a novel KCR-GCL encoder, which incorporates a new KCR self-attention layer that adaptively balances different frequency components of the graph inspired by generalized graph convolution and reduces the kernel complexity for provably improved generalization for transductive learning. The KCR-GCL encoder is optimized with a low-rank regularization term through the truncated nuclear norm (TNN) on the gram matrix of the learned features. The learned low-rank representations are then used to train a linear classifier for transductive node classification in noisy graph data. The design of KCR-GCL is inspired by the Low Frequency Property (LFP) widely studied in general deep learning and node-level graph learning, and is further supported by a sharp generalization bound for transductive learning. To the best of our knowledge, KCR-GCL is among the first to theoretically reveal the benefits of low-rank regularization in transductive settings for noisy graph data. Experiments on standard benchmarks highlight the effectiveness and robustness of KCR-GCL in learning node representations under noisy conditions. The code of KCR-GCL is available at https://anonymous.4open.science/r/KCR-GCL/.

1 Introduction

Graph Neural Networks (GNNs) are widely recognized as effective tools for node representation learning (Kipf & Welling, 2017; Bruna et al., 2014; Xu et al., 2019). However, the majority of existing GNN methods do not adequately address the presence of noise in the graph data (Zhu et al., 2024; Zhong et al., 2019). Such noise can arise either in the attributes or labels of nodes, introducing attribute noise and label noise, respectively. Prior studies (Patrini et al., 2017) have demonstrated that noise in the input data can significantly impair the generalization ability of neural networks. This issue is further amplified since noise associated with a few nodes can spread through the graph structure and affect other nodes (Dai et al., 2021; Wang et al., 2023; 2024c) in graph-structured data. As a result, corrupted nodes not only degrade their own representations but also influence those of their neighbors. Such a challenge highlights the need for GNN models that can learn effectively even in the presence of 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 & 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., 2024b; Ju et al., 2025). Detailed settings for the label noise and attribute noise studied in this paper are presented in Section A of the appendix.

Limitations of Existing Robust Graph Learning Methods on Noisy Graph Data. Traditional strategies for robust learning either modify the loss function to accommodate corrupted data (Patrini et al., 2017; Goldberger & Ben-Reuven, 2017) or remove samples suspected to be noisy (Malach & Shalev-Shwartz, 2017; Jiang et al., 2018; Yu et al., 2019; Li et al., 2020; Han et al., 2018). Although such ideas have been adapted for graph settings (Dai et al., 2021; Qian et al., 2023; Li et al., 2024c; Yuan et al., 2023), they often depend on heuristics and lack theoretical backing in the transductive setting. To this end, we introduce Kernel Complexity Reduced Graph Contrastive Learning, or KCR-GCL, which introduces a novel KCR-GCL encoder improving both robustness and generalization for node representation learning. The node representations learned by the KCR-GCL encoder are used by a linear transductive classifier for transductive node classification. In contrast to existing robust graph learning methods, our KCR-GCL encoder is empirically inspired by the low-frequency nature of graph signals, and theoretically supported by a new and sharp generalization bound developed for transductive learning. Our generalization bound features three components capturing the training loss of the classifier when using clean labels, the impact of label noise on the classification loss, and the kernel complexity of the gram matrix of the learned features. To the best of our knowledge, this paper is among the first to provide a principled theoretical justification for the advantage of low-rank representation learning in GNNs on noisy graph data. Experimental evaluations conducted on widely used benchmarks demonstrate that KCR-GCL consistently achieves strong robustness and outperforms the current state-of-the-art.

Although GNNs are known to function as low-pass filters, they do not explicitly target low-frequency signals. Consequently, their ability to exploit the Low-Frequency Property (LFP) 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; 2025) in the presence of noisy graph data remains limited. As illustrated in Figure 1, deferred to Section 4.1, the LFP reveals that clean label information tends to concentrate within the low-rank part of the feature gram matrix. Unlike conventional GNNs, the KCR-GCL encoder explicitly captures LFP by learning low-rank node representations. Prior works such as (Cheng et al., 2021) have illustrated the benefit of such low-rank learning in mitigating attribute noise by introducing learnable filtering mechanisms. In comparison, the KCR-GCL encoder explicitly promotes low-frequency information through the low-rank regularization by the Truncated Nuclear Norm (TNN), which aligns with the LFP. Existing works on GCL (Li et al., 2024c; 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.7, where KCR-GCL significantly outperforms KCR-GCN, which is the Kernel Complexity Reduced GCN, thereby justifying the use of the GCL encoder in this paper. The KCR-GCN is also detailed in Section B.7.

The node representations by our KCR-GCL encoder are generated through a novel KCR self-attention layer that explicitly balances low-frequency and high-frequency components of the graph. Inspired by polynomial graph filters commonly used in graph signal processing (Choi et al., 2024; Zhang et al., 2024a; Marques et al., 2020), the KCR self-attention layer learns to combine multiple powers of the attention weight matrix, enabling the model to adaptively emphasize structural patterns across different spectral ranges while improving the generalization capability of the KCR-GCL encoder. Recent studies involving graph attention and transformer-based architectures emphasize the need to balance both low- and high-frequency components for improved node representations (Choi et al., 2024; Zhang et al., 2024a). Although GFSA (Choi et al., 2024) and HONGAT (Zhang et al., 2024a) also learn to combine different powers of the attention matrix based on the generalized graph convolution (Marques et al., 2020), their learning objective is solely the cross-entropy loss on the training labels alone. In contrast, the KCR self-attention layer in our KCR-GCL encoder roots in our novel theoretical result about the generalization bound for transductive learning, and it offers a principled balance between low-frequency and high-frequency by reducing the principled and well-defined kernel complexity, aiming for the provable performance improvement for transductive learning through the TNN regularizer. Reduction of the TNN leads to the reduction of the kernel complexity, leading to lower generalization error bound for transductive learning thus better performance of node classification. Performance comparisons in Table 2 in Section 5.2 and Table 8-9 in Section B.1 of the appendix demonstrate that the KCR-GCL encoder outperforms the current state-of-the-art based on attention and transformers, GFSA (Choi et al., 2024) and HONGAT (Zhang et al., 2024a), when evaluated under label and attribute

noise. As shown in Table 4-5 in Section 5.4, the KCR-GCL encoder achieves a lower kernel complexity, resulting in a lower upper bound for the test loss of transductive node classification, compared to competing graph contrastive and attention-based methods.

1.1 Contributions

Our contributions are as follows.

First, we introduce Kernel Complexity Reduced Graph Contrastive Learning, or KCR-GCL, that learns robust node representations by a novel KCR-GCL encoder. The learned node representations are subsequently used by a linear classifier for transductive node classification. The KCR-GCL encoder features a novel KCR self-attention layer, which explicitly learns to balance low-frequency and high-frequency components in the graph, inspired by the generalized graph convolution (Marques et al., 2020). The optimization of the KCR-GCL encoder incorporates the TNN on the gram matrix of the learned features as a low-rank regularization term into the standard prototypical GCL objective. The design of the KCR-GCL encoder is motivated by the LFP, 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.

Second, we provide a rigorous theoretical analysis that establishes generalization guarantee for the linear transductive classifier trained on the low-rank node representations produced by the KCR-GCL encoder. In particular, we derive a novel and sharp upper bound on the test loss of unlabeled nodes. To the best of our knowledge, this is among the first results to theoretically demonstrate the advantage of learning low-rank node representations for robust transductive classification under noisy conditions. Moreover, our theoretical result establishes the connection between the generalization guarantee and the TNN regularizer, that is, reduced TNN indicates reduced kernel complexity and lower generalization error bound for transductive learning. Furthermore, the KCR self-attention balances different frequency components of the graph and sharpens the derived upper bound through even lower kernel complexity than that without KCR self-attention. Extensive evaluation results on public graph benchmarks for semi-supervised node classification with asymmetric label noise, symmetric label noise, and attribute noise in Table 2 in Section 5.2 and Table 8-9 in Section B.1 of the appendix demonstrate the effectiveness of learning low-rank representation with KCR-GCL for improving robustness. As shown in Table 4-5 in Section 5.4, the KCR-GCL encoder achieves a lower kernel complexity, resulting in a lower upper bound for the test loss of transductive node classification, compared to competing graph contrastive and attention-based methods.

2 Related Works

2.1 Graph Neural Network and its Training on Noisy Data

The increasing adoption of contrastive learning has significantly advanced unsupervised representation learning on graphs (Suresh et al., 2021; Thakoor et al., 2021; Li et al., 2024a; Lee et al., 2022; Feng et al., 2022a; Zhang et al., 2023; Lin et al., 2023). In the graph domain, a large number of graph contrastive learning (GCL) methods (Velickovic et al., 2019; Sun et al., 2020; Hu et al., 2020b; Jiao et al., 2020; Peng et al., 2020; You et al., 2021; Jin et al., 2021; Mo et al., 2022) work by maximizing agreement between corresponding node embeddings across augmented views. Several recent works have incorporated semantic prototypes (Snell et al., 2017; Arik & Pfister, 2020; Allen et al., 2019; Xu et al., 2020) into the contrastive objective (Xu et al., 2021; Guo et al., 2022; Li et al., 2021). Meanwhile, GNNs remain central to node representation learning (Bruna et al., 2014; Kipf & Welling, 2017; Veličković et al., 2018; Xu et al., 2019). However, it has been well established that GNNs are inherently vulnerable to noisy inputs, such as corrupted labels or features (Zhang et al., 2021). To improve robustness, prior work has explored loss correction, which modifies the training objective to account for label noise (Patrini et al., 2017), and sample selection, which focuses training on selected clean samples (Yu et al., 2019; Li et al., 2020; Han et al., 2018). Within graph-based learning, robustness has also been addressed through label denoising, structural regularization, and auxiliary self-supervised tasks (Dai et al., 2021; Qian et al., 2023; Zhuang & Al Hasan, 2022; Li et al., 2024c; Yuan et al., 2023). While these methods build on external objectives or correction heuristics, our approach

introduces a new perspective of enhancing GNN robustness by directly integrating low-rank regularization into the encoder training process of GCL.

2.2 Balancing the Frequency Components with Graph Attention

The low-frequency bias of GNNs emphasizes the importance of leveraging smooth, low-frequency components embedded in both graph topology and node features (NT & Maehara, 2019; Wu et al., 2019; Yu & Qin, 2020). However, relying solely on these low-frequency signals can lead to over-smoothing, where node representations become indistinguishable (Bo et al., 2021; Zhang et al., 2024b; Dong et al., 2025; Sun et al., 2022). To mitigate over-smoothing, recent efforts have proposed to dynamically balance low- and high-frequency components of the graph (Dong et al., 2021; Tang et al., 2025; Bo et al., 2021; Ju et al., 2022; Chang et al., 2021; Sun et al., 2024; Wang et al., 2024a). In parallel, approaches that emphasize low-rank modeling of graph signals and structures have demonstrated greater resilience under noisy conditions (Tang et al., 2024; Yang et al., 2023). Moreover, the attention-based GNN, GFSA (Choi et al., 2024), learns to balance the original attention matrix and its high-order approximation, thereby enriching frequency information and alleviating over-smoothing. HONGAT (Zhang et al., 2024a) explicitly addresses over-smoothing by integrating high-order dependencies and introducing sparsity within the attention mechanism. Whereas GFSA (Choi et al., 2024) and HONGAT (Zhang et al., 2024a) learn the combination weights for different powers of the attention matrix solely by fitting to the training labels, our approach explicitly optimizes the combination weights to minimize the kernel complexity, leading to tighter theoretical generalization bounds and improved robustness.

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 $\mathbf{A} \in \{0,1\}^{N \times N}$. $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$, and the corresponding diagonal degree matrix is defined as $\tilde{\mathbf{D}}$. $[N]$ refers to the set of integers from 1 to N inclusive. A subset $\mathcal{L} \subseteq [N]$ contains m labeled nodes, and its complement $\mathcal{U} = [N] \backslash \mathcal{L}$ has cardinality u. 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 datasets, noise commonly arises either in the node attributes or in the labels. This noise can severely undermine the quality of node embeddings learned by standard GCL encoders, thereby degrading the performance of classifiers built on top of them. Our goal is to train a GCL encoder whose node representations remain robust under two transductive node classification settings, one where the labels of $\mathcal{V}_\mathcal{L}$ are corrupted, and another where the input attributes $\mathbf{X}$ contain noise. The node representations learned by a GCL encoder are given by $\mathbf{H}(\boldsymbol{\theta}) = g_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{A})$, where $g_{\boldsymbol{\theta}}(\cdot)$ denotes the GCL encoder parameterized by $\boldsymbol{\theta}$. In this study, $g_{\boldsymbol{\theta}}$ is instantiated as a two-layer GCN (Kipf & Welling, 2017). The resulting representations, $\mathbf{H}(\boldsymbol{\theta}) = \{\mathbf{H}_1(\boldsymbol{\theta}), \mathbf{H}_2(\boldsymbol{\theta}), \ldots, \mathbf{H}_N(\boldsymbol{\theta})\} \in \mathbb{R}^{N \times d}$, serve as input for the transductive node classification task, for which a linear classifier is first trained on $\mathcal{V}_\mathcal{L}$, and then evaluated for predicting the labels of $\mathcal{V}_\mathcal{U}$. We abbreviate $\mathbf{H}(\boldsymbol{\theta})$ as $\mathbf{H}$ for simplicity of the notations in this paper.

Preliminary: Prototypical GCL (PGCL). We adopt a contrastive learning framework to optimize the GCL encoder $g(\cdot)$, a two-layer GCN (Kipf & 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 $\mathcal{L}_\text{node}(\boldsymbol{\theta}) = -\frac{1}{N} \sum_{i=1}^N \log \frac{s(\mathbf{H}_i^1, \mathbf{H}_i^2)}{s(\mathbf{H}_i^1, \mathbf{H}_i^2) + \sum_{j=1}^N s(\mathbf{H}_i^1, \mathbf{H}_j^2)}$ and $\mathcal{L}_\text{proto}(\boldsymbol{\theta}) = -\frac{1}{N} \sum_{i=1}^N \log \frac{\exp(\mathbf{H}_i \cdot \mathbf{c}_k / \tau)}{\sum_{k=1}^K \exp(\mathbf{H}_i \cdot \mathbf{c}_k / \tau)}$, 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 $\mathcal{L}_{\text{GCL}}(\boldsymbol{\theta}) = \mathcal{L}_{\text{node}}(\boldsymbol{\theta}) + \mathcal{L}_{\text{proto}}(\boldsymbol{\theta})$.

4 Methods

4.1 Kernel Complexity Reduced GCL (KCR-GCL) Encoder

In order to perform node classification with provable generalization guarantee, we propose a new Kernel Complexity Reduced GCL (KCR-GCL) encoder, which applies a novel KCR self-attention to the node representations $\mathbf{H}$ generated by the PGCL encoder, $g_{\boldsymbol{\theta}}$. The output of the KCR self-attention layer on top of the node features $\mathbf{H}$ is $\mathbf{F} = \mathbf{B}\mathbf{H}$, where $\mathbf{F}$ denotes the attention-transformed features, and $\mathbf{B} \in \mathbb{R}^{N \times N}$ is the attention weight matrix. The gram matrix of the node features is given by $\mathbf{K} = \mathbf{H}\mathbf{H}^\top$. Let $\mathbf{B}_0 = \mathbf{K}/\lambda_1$, where λ_1 is the largest eigenvalue of $\mathbf{K}$. In our KCR self-attention, the attention matrix is defined as $\mathbf{B} \coloneqq \sum_{m=1}^{M} \kappa_m \mathbf{B}_0^m$, where $M \geq 1$ is the maximum degree. The coefficients $\{\kappa_m\}_{m=1}^{M}$ are computed by $\kappa_m = \frac{\exp(\boldsymbol{\alpha}_m)}{\sum_{j=1}^{M} \exp(\boldsymbol{\alpha}_j)}$, where $\boldsymbol{\alpha} \in \mathbb{R}^M$ are learnable parameters and $\boldsymbol{\alpha}_m$ is the m-th element of $\boldsymbol{\alpha}$. The design of the attention matrix $\mathbf{B} = \sum_{m=1}^{M} \kappa_m \mathbf{B}_0^m$ in KCR-GCL is inspired by polynomial graph filters widely used in graph signal processing (GSP) for both undirected and directed graphs (Choi et al., 2024; Zhang et al., 2024a; Marques et al., 2020). Each term $\mathbf{B}_0^m$ captures m-hop feature propagation over the graph defined by the kernel matrix $\mathbf{B}_0$, and the learnable coefficients κ_m determine the relative influence of different neighborhood scales, similar to spectral mixing in generalized graph convolutions (Marques et al., 2020). In addition, the design of the attention matrix $\mathbf{B}$ can also reduce the eigenvalues of the gram matrix of the attention-transformed features, thus leading to lower kernel complexity. In the KCR-GCL encoder, $\mathbf{F}(\boldsymbol{\alpha}, \boldsymbol{\theta}) = \mathbf{B}(\boldsymbol{\alpha})\mathbf{H}(\boldsymbol{\theta})$. We abbreviate $\mathbf{F}(\boldsymbol{\alpha}, \boldsymbol{\theta})$ and $\mathbf{B}(\boldsymbol{\alpha})$ as $\mathbf{F}$ and $\mathbf{B}$ for simplicity of the notations. The resulting gram matrix of the transformed features $\mathbf{F}$ is $\mathbf{K}_{\mathbf{F}}(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \mathbf{F}\mathbf{F}^\top = \mathbf{B}\mathbf{K}\mathbf{B}$, and $\mathbf{K}_{\mathbf{F}}(\boldsymbol{\theta}, \boldsymbol{\alpha})$ is the learned feature kernel matrix of the KCR-GCL encoder. We also abbreviate $\mathbf{K}_{\mathbf{F}}(\boldsymbol{\theta}, \boldsymbol{\alpha})$ as $\mathbf{K}_{\mathbf{F}}$ for simplicity of the notations in the sequel.

We propose to reduce the TNN of the gram matrix $\mathbf{K}_{\mathbf{F}}$. 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 $\mathbf{K}_{\mathbf{F}}$. In order to encourage the features $\mathbf{F}$ or the gram matrix $\mathbf{K}_{\mathbf{F}}$ to be low-rank, we explicitly add the TNN $\|\mathbf{K}_{\mathbf{F}}\|_{r_0} \coloneqq \sum_{i=r_0+1}^{N} \widehat{\lambda}_i$ to the loss function of the KCR-GCL encoder. The starting rank $r_0 < \min(N, d)$ is the rank of the gram matrix of the features we aim to obtain with the KCR-GCL encoder, that is, if $\|\mathbf{K}_{\mathbf{F}}\|_{r_0} = 0$, then $\operatorname{rank}(\mathbf{K}_{\mathbf{F}}) = r_0$. The training of the KCR-GCL encoder performs the following optimization,

$$\begin{aligned} &\min_{\boldsymbol{\theta}, \boldsymbol{\alpha}} \mathcal{L}_{\text{KCR-GCL}}(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \mathcal{L}_{\text{node}}(\boldsymbol{\theta}) + \mathcal{L}_{\text{proto}}(\boldsymbol{\theta}) + \tau \|\mathbf{K}_{\mathbf{F}}(\boldsymbol{\theta}, \boldsymbol{\alpha})\|_{r_0}, \\ &\text{s.t. } \mathbf{B}_{ij} = 0 \text{ if } \mathbf{A}_{ij} = 0, \forall i, j \in [N]. \end{aligned} \tag{1}$$

where τ is a weighting parameter chosen by cross-validation described in Section 5.1. We note that the constraint in (1) ensures that the attention weight matrix $\mathbf{B}$ has the same sparsity as the adjacency matrix $\mathbf{A}$ for efficiency of the KCR-GCL encoder, and the TNN of $\mathbf{K}_{\mathbf{F}}$ is still reduced by optimization of the objective $\mathcal{L}_{\text{KCR-GCL}}$. In our experiments, we select the TNN rank r_0 via standard cross-validation across all graph datasets. As reported in Table 7 in Section 5.1, the optimal rank r_0 consistently falls within the range of $0.1 \min\{N, d\}$ to $0.3 \min\{N, d\}$. The reduction of the TNN is also inspired by the reduction of the kernel complexity, which is to be defined later in Section 4.2, leading to provable and sharp generalization error of the linear transductive node classifier using the node representations of the KCR-GCL encoder. Our KCR-GCL encoder also outperforms an ablation study model, Low-Rank GCL without the self-attention matrix $\mathbf{B}$, to be detailed in Section 5.3. Algorithm 1 details the full training procedure for KCR-GCL. Starting from the GCL encoder pretrained by GCL, the algorithm iteratively updates the encoder parameters $\boldsymbol{\theta}$, the classifier weights $\mathbf{W}$, and the KCR self-attention weights $\boldsymbol{\alpha}$. At each training step, the encoder is first updated by optimizing the contrastive loss. Then, based on the updated encoder, node representations are recomputed and used to optimize the KCR-GCL objective $L(\mathbf{W}, \boldsymbol{\theta}, \boldsymbol{\alpha})$, which includes both classification loss and low-rank regularization. This two-stage update ensures that the final representations achieve both semantic alignment and low kernel complexity.

Algorithm 1: Training Algorithm for the KCR-GCL Encoder

1: **Input:** Attribute matrix $\mathbf{X}$, adjacency matrix $\mathbf{A}$, training epochs $t_{\max}$, parameters of the pre-trained GCL encoder $\boldsymbol{\theta}^{(0)}$, learning rate η
2: **Output:** The GCL encoder $g_{\boldsymbol{\theta}}$, the weights of the classifier $\mathbf{W}$, and the weights of the KCR self-attention $\boldsymbol{\alpha}$
3: Randomly initialize $\mathbf{W}^{(0)}$ and $\boldsymbol{\alpha}^{(0)}$
for $t = 1$ *to* t_{max} **do**
Compute node representations $\mathbf{H} = g_{\boldsymbol{\theta}^{(t-1)}}(\mathbf{X}, \mathbf{A})$
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 $\mathbf{c}_k = \frac{1}{|S_k|}\sum_{\mathbf{H}_i \in S_k} \mathbf{H}_i$
Generate augmented views $G^1 = (\mathbf{X}^1, \mathbf{A}^1)$ and $G^2 = (\mathbf{X}^2, \mathbf{A}^2)$
Compute $\mathbf{H}^1 = g_{\boldsymbol{\theta}^{(t-1)}}(\mathbf{X}^1, \mathbf{A}^1)$, $\mathbf{H}^2 = g_{\boldsymbol{\theta}^{(t-1)}}(\mathbf{X}^2, \mathbf{A}^2)$
Perform gradient descent on $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ by $\boldsymbol{\theta}^{(t)} \leftarrow \boldsymbol{\theta}^{(t-1)} - \eta \nabla_{\boldsymbol{\theta}} \mathcal{L}_{\text{KCR-GCL}}(\boldsymbol{\theta}^{(t-1)}, \boldsymbol{\alpha}^{(t-1)})$ and $\boldsymbol{\alpha}^{(t)} \leftarrow \boldsymbol{\alpha}^{(t-1)} - \eta \nabla_{\boldsymbol{\alpha}} \mathcal{L}_{\text{KCR-GCL}}(\boldsymbol{\theta}^{(t-1)}, \boldsymbol{\alpha}^{(t-1)})$. $\mathbf{B}_{ij}^{(t)} = 0$ if $\mathbf{A}_{ij} = 0$ for all $i, j \in [N]$, where $\mathbf{B}_{ij}^{(t)}$ is the attention weight matrix computed by $\boldsymbol{\alpha}^{(t)}$.
4: Compute node representations $\mathbf{F} = \mathbf{B}\mathbf{H} = \mathbf{B}g_{\boldsymbol{\theta}^{(t)}}(\mathbf{X}, \mathbf{A})$.
5: **Return** the KCR-GCL encoder, and the node representations $\mathbf{F}$.

Motivation of Learning Low-Rank Features by the KCR-GCL Encoder. We investigate how the information from the ground-truth clean labels and the label noise is distributed across different eigenvectors of the feature gram matrix $\mathbf{K_F}$ 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 $\mathbf{K_F}$. 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 $\tilde{\mathbf{Y}} \in \{0,1\}^{N \times C}$ consists 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 $\mathbf{K_F}$, while the signal 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, the signal concentration ratio at rank r of the noise is defined as $\frac{1}{C}\sum_{c=1}^{C} \left\| \mathbf{U}^{(r)\top} \mathbf{N}^{(c)} \right\|_2^2 / \left\| \mathbf{N}^{(c)} \right\|_2^2 \in \mathbb{R}^N$. Empirical results, shown as red curves in Figure 1, indicate that the clean label signals are primarily concentrated on the leading eigenvectors of $\mathbf{K_F}$. 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 using low-rank features $\mathbf{F}$, or equivalently the low-rank gram matrix $\mathbf{K_F}$, for node classification with label noise. This is because the low-rank part of the feature matrix $\mathbf{F}$ or the gram matrix $\mathbf{K_F}$ 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

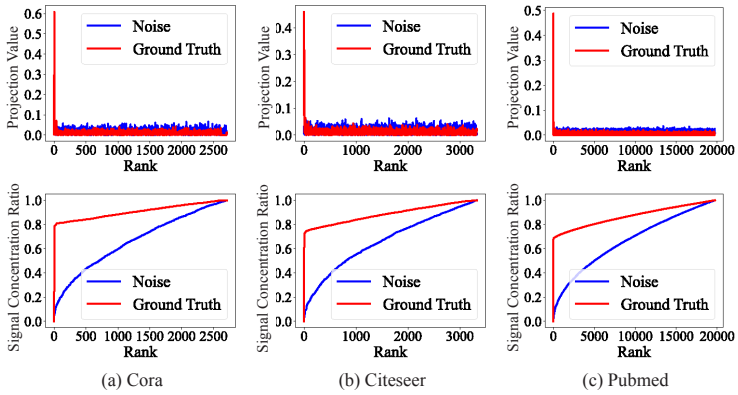


Figure 1: Eigen-projection (first row) and signal concentration ratio (second row) on Cora, Citeseer, and Pubmed, as the illustration of the Low Frequency Property (LFP). The study in this figure is performed for asymmetric label noise with a noise level of 60%. By the rank $r = 0.2 \min\{N, d\}$, the signal concentration ratio of $\tilde{\mathbf{Y}}$ for Cora, Citeseer, and Pubmed are 0.844, 0.809, and 0.784 respectively. Figure 2 in Section B.4 further illustrates the eigen-projection and signal concentration ratio on more datasets.

Low Frequency Property (LFP), which has been widely studied in deep learning and node-level graph learning (Rahaman et al., 2019; Arora et al., 2019; Cao et al., 2021; Choraria et al., 2022; Wang et al., 2024b; 2025). Moreover, we remark that the regularization term $\|\mathbf{K_F}\|_{r_0}$ in the loss function (1) of KCR-GCL is also theoretically motivated by our sharp upper bound for the test loss using a linear transductive classifier, to be detailed in the next subsection.

4.2 Transductive Node Classification With Provable Generalization Guarantee

In this section, we present a linear transductive node classification method based on the node representations $\mathbf{F}$ obtained from the KCR-GCL encoder. 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. The classifier's linear prediction is computed by $\mathbf{FW}$, where $\mathbf{W} \in \mathbb{R}^{d\times C}$ is the learnable weight matrix. Final predictions are made using the softmax transformation $\text{softmax}(\mathbf{FW}) \in \mathbb{R}^{N\times C}$ to estimate class probabilities for the test nodes. We then train the transductive linear classifier on top of $\mathbf{F}$ by minimizing the loss function,

$$\min_{\mathbf{W}} L(\mathbf{W}, \boldsymbol{\theta}, \boldsymbol{\alpha}) = \frac{1}{m} \sum_{v_i \in \mathcal{V}_\mathcal{L}} \text{KL}\left(\mathbf{y}_i, \left[\text{softmax}\left(\mathbf{FW}\right)\right]_i\right). \tag{2}$$

where $\text{KL}(\cdot,\cdot)$ measures the KL divergence. 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 $\mathbf{F}(\mathbf{W}, t) \coloneqq \mathbf{FW}^{(t)}$ 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 $\mathbf{F}$ with $r_0 \in [N]$, which is the generalization error bound for the linear transductive classifier using $\mathbf{F}$ to predict the labels of the unlabeled nodes. Similar to existing works such as (Kothapalli et al., 2023) that use the Mean Squared Error (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, 2023; 2025).

Theorem 4.1. Let $m \geq cN$ for a constant $c \in (0, 1)$, and $r_0 \in [N]$. Assume that a set $\mathcal{L}$ with $|\mathcal{L}| = m$ is sampled uniformly without replacement from $[N]$, and the remaining nodes $\mathcal{V}_U = \mathcal{V} \setminus \mathcal{V}_L$ are the test nodes. Then for every $x > 0$, with probability at least $1 - \exp(-x)$, after the t-th iteration of gradient descent for all $t \geq 1$, we have

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

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

This theorem is proved in Section C, and the sharpness of the upper bound for the generalization error, as the RHS of (3), is proved in (Yang, 2025; 2023). 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: $L_1(\mathbf{K_F}, \tilde{\mathbf{Y}}, t)$, $L_2(\mathbf{K_F}, \mathbf{N}, t)$, and $\text{KC}(\mathbf{K_F})$, each serving a distinct role. $L_1(\mathbf{K_F}, \tilde{\mathbf{Y}}, t)$ reflects the training loss of the classifier when using clean labels. $L_2(\mathbf{K_F}, \mathbf{N}, t)$ captures the impact of label noise on the classification loss. $\text{KC}(\mathbf{K_F})$ denotes the kernel complexity (KC) of the gram matrix $\mathbf{K_F}$.

The design of the self-attention matrix $\mathbf{B}$ is to reduce the eigenvalues of the gram matrix $\mathbf{K}$ of the PGCL. Let $\{\lambda_i\}_{i=1}^N$ denote the eigenvalues of $\mathbf{K}$, ordered as $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N \geq 0$. Since $\mathbf{B}$ has the same eigenvectors as $\mathbf{K}$, and the maximum eigenvalue of $\mathbf{B}$ falls in $(0, 1]$, it can be verified that $\widehat{\lambda}_i \leq \lambda_i$. As a result, the KCR

self-attention layer reduces the kernel complexity of the original gram matrix $\mathbf{K}$, rendering a sharper upper bound for transductive node classification. This is reflected in Table 4, where the ablation study model, LR-GCL without the KCR self-attention layer, exhibits larger kernel complexity and generalization upper bounds than those of our KCR-GCL. Importantly, the TNN $\|\mathbf{K_F}\|_{r_0}$ appears in the upper bound in (3), thereby providing theoretical justification for incorporating the TNN regularizer $\|\mathbf{K_F}\|_{r_0}$ to promote low-rank feature learning in our KCR-GCL encoder. Furthermore, under the LFP, which is consistently supported by the empirical evidence shown in Figure 1, $L_1(\mathbf{K_F}, \tilde{\mathbf{Y}}, t)$ diminishes as the number of training iterations t increases. Simultaneously, $L_2(\mathbf{K_F}, \mathbf{N}, t)$ remains small due to the approximately uniform eigen-projection of label noise, while $\mathbf{K_F}$ remains close to a rank-r_0 matrix, as the TNN is effectively minimized through the KCR-GCL training objective (1). We note that while the theoretical guarantee in Theorem 4.1 is for label noise, the LFP also holds for attribute noise, to be shown in Section B.4, which motivates KCR-GCL for node classification under either noisy labels or noisy attributes in the next section.

5 Experiments

In this section, we present a thorough evaluation of KCR-GCL across multiple standard graph benchmarks. The experiment settings are detailed in Section 5.1. Performance under semi-supervised node classification settings with different types of label noise and attribute noise is discussed in Section 5.2. We perform an ablation study to verify the effectiveness of the KCR self-attention layer in KCR-GCL in Section 5.3. The kernel complexity (KC) and the theoretical upper bound on test loss for both models are analyzed in Section 5.4. Section 5.5 compares the computational efficiency of KCR-GCL with other baselines. Additional experiment results are presented in Section B of the supplementary. Section B.1 presents additional results for semi-supervised node classification with different types of label noise and attribute noise. Section B.2 compares our method against existing graph contrastive learning approaches with diverse classifier designs. The sensitivity analysis on hyperparameters τ, M, r_0 are conducted in Section B.3. Additional results of eigen-projection visualizations and signal concentration ratios are provided in Section B.4. To evaluate the reliability of the observed gains in Section 5.2, Section B.6, and Section B.1, we perform Student's t-test in Section B.5. Section B.6 explores the applicability of our models on heterophilic graphs. The ablation study in Section B.7 demonstrates that the KCR self-attention layer with low-rank regularization as that in the KCR-GCL objective (1) is also applicable to GCN so as to enhance the robustness of standard GCN (Kipf & Welling, 2017), and the resultant GCN-based model is termed KCR-GCN detailed in Section B.7. However, KCR-GCL substantially outperforms KCR-GCN, justifying the use of the GCL encoder in this paper.

Table 1: 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
Texas	183	309	1,703	5
Chameleon	2,277	36,101	2,325	5

5.1 Experimental Settings

We evaluate our proposed approaches on ten widely recognized graph benchmark datasets: Cora, Citeseer, PubMed (Sen et al., 2008), Coauthor CS, ogbn-arxiv (Hu et al., 2020a), Wiki-CS (Mernyei & Cangea, 2020), the Amazon-Computers and Amazon-Photos datasets (Shchur et al., 2018), and the two heterophilic graph benchmarks Texas and Chameleon (Pei et al., 2020). A summary of the key statistics for all datasets is provided in Table 1 in Section A of the appendix. All experiments utilize the standard train/validation/test splits defined in prior studies (Shchur et al., 2018). Noise is only introduced into the training and validation sets to preserve the integrity of the test data for fair performance evaluation. As the public graph datasets do not inherently contain label or feature noise, we synthetically introduce the symmetric and asymmetric label

noise following (Han et al., 2020; Dai et al., 2022; Qian et al., 2023). Detailed settings for the generation of the symmetric label noise, asymmetric label noise, and attribute noise, and the cross-validation for tuning r_0, τ, and M are presented in Section A of the appendix. The statistics of the datasets are presented in Table 1 in Section A of the appendix.

Table 2: Comparison for node classification accuracy (%) on Cora, Citeseer, 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 KCR-GCL and all baseline methods for symmetric label noise is deferred to Table 8 in Section B.1 of the appendix. Table 9 in Section B.1 of the appendix 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	81.5±0.5	54.7±1.5	63.9±0.8	40.5±1.4	43.9±1.2	26.5±1.2	31.7±1.3
	NRGNN	82.2±0.6	57.1±1.9	64.5±1.2	47.0±1.4	45.1±1.1	28.2±2.2	32.6±1.0
	RTGNN	82.8±0.3	57.0±1.0	67.8±1.1	47.4±1.1	45.7±0.9	28.0±1.1	34.2±1.6
	SUGRL	83.4±0.5	56.4±1.1	67.5±0.9	46.8±1.1	45.2±1.2	28.0±1.2	33.8±1.4
	MERIT	83.1±0.5	56.0±0.8	67.1±0.9	46.7±1.3	45.0±1.4	27.7±1.3	33.5±0.9
	SFA	83.9±1.0	56.4±1.1	67.6±1.5	47.3±1.4	45.7±1.4	28.2±1.6	34.4±1.7
	Sel-Cl	82.8±0.2	57.0±1.0	67.6±0.9	47.2±1.3	45.5±1.1	28.2±1.7	34.1±1.5
	Jo-SRC	82.5±0.5	57.1±0.6	67.9±0.7	47.3±1.1	45.8±1.2	28.5±1.3	34.5±1.8
	GRAND+	85.8±0.6	57.0±0.9	67.8±1.1	47.2±1.0	45.6±1.2	28.4±1.5	34.5±1.3
	GFSA	83.7±0.6	56.8±1.2	67.2±0.9	46.6±1.2	45.1±1.2	27.9±1.2	33.6±1.3
	HONGAT	83.3±0.4	56.6±1.1	66.7±1.0	46.4±1.0	44.9±1.0	27.8±1.3	33.4±1.4
	CRGNN	84.2±0.5	57.2±1.0	67.4±1.0	47.0±1.2	45.4±1.3	28.3±1.4	34.1±1.5
	CGNN	83.5±0.6	56.7±0.9	66.9±1.1	46.2±1.3	45.0±1.3	28.1±1.2	33.7±1.4
	ProCon	84.8±0.6	57.2±1.1	67.8±0.7	47.3±1.1	45.8±1.2	28.4±1.3	34.4±1.8
	KCR-GCL	**86.1±0.6**	**60.2±1.1**	**70.8±1.1**	**51.0±1.1**	**49.2±1.2**	**32.9±1.2**	**38.2±1.1**
Citeseer	GCN	70.3±0.5	47.5±2.3	52.9±0.9	35.1±1.4	37.2±1.1	29.1±2.2	29.0±1.4
	NRGNN	71.0±0.6	49.8±1.5	53.8±1.1	38.2±1.6	37.7±1.2	33.6±2.1	28.4±0.9
	RTGNN	74.6±0.8	49.8±0.7	55.0±1.2	39.2±1.0	39.0±1.4	34.8±1.7	30.2±1.1
	SUGRL	73.0±0.5	49.3±1.1	54.4±1.0	37.6±0.9	38.8±0.9	33.9±1.0	30.0±0.9
	MERIT	74.0±0.7	49.6±1.2	54.2±1.0	38.3±1.1	38.7±0.8	34.4±1.4	29.5±0.9
	SFA	74.0±1.1	50.2±1.4	54.7±1.3	39.0±1.4	38.9±1.2	34.7±1.6	29.9±1.3
	Jo-SRC	73.0±0.5	50.0±1.3	55.1±1.1	39.4±1.3	39.3±1.3	35.1±1.3	30.3±1.3
	GRAND+	75.6±0.4	49.7±1.0	55.2±1.1	39.0±1.3	38.7±1.3	34.8±1.3	30.2±1.2
	GFSA	74.3±0.6	49.5±1.2	54.6±1.1	38.6±1.1	38.6±1.2	34.2±1.3	29.8±1.2
	HONGAT	73.8±0.7	49.2±1.4	54.5±0.9	38.0±1.2	38.4±1.3	34.0±1.4	29.6±1.1
	CRGNN	75.1±0.6	49.7±1.1	54.9±1.2	38.9±1.4	38.8±1.2	34.7±1.5	30.1±1.2
	CGNN	74.1±0.7	49.3±1.3	54.6±1.0	38.5±1.3	38.5±1.1	34.3±1.3	29.7±1.2
	ProCon	75.3±0.4	50.4±1.4	55.2±0.8	39.6±1.3	39.2±1.3	35.3±1.3	30.2±1.3
	KCR-GCL	**76.2±1.0**	**53.3±1.3**	**58.8±0.7**	**43.0±1.4**	**42.3±1.2**	**39.2±1.2**	**33.5±1.0**
PubMed	GCN	79.0±0.7	58.4±2.2	59.5±1.2	40.5±2.5	48.8±1.3	30.5±2.2	42.3±1.3
	NRGNN	79.7±0.8	60.2±2.2	60.3±0.8	44.3±1.2	49.9±0.9	33.0±2.3	43.3±1.1
	RTGNN	79.7±0.4	61.0±0.8	61.4±1.2	45.5±1.0	50.1±1.1	33.5±1.3	45.2±1.3
	SUGRL	81.9±0.5	60.3±1.3	61.5±1.0	44.5±1.1	50.1±0.7	32.1±0.9	44.6±1.0
	MERIT	80.1±0.4	59.3±1.1	61.3±1.1	44.7±1.2	49.7±0.9	32.8±1.1	44.5±0.9
	SFA	80.4±1.0	59.6±1.1	60.9±1.1	44.7±1.4	49.9±1.4	33.0±1.1	44.7±1.4
	Jo-SRC	80.1±0.5	61.3±1.0	61.7±1.3	45.3±0.8	50.4±1.3	33.0±1.5	45.9±1.8
	GRAND+	84.5±0.6	61.0±1.1	61.7±1.3	45.3±0.8	50.3±1.0	33.1±1.4	45.8±1.4
	GFSA	82.3±0.5	60.8±1.2	61.6±0.9	45.0±1.3	50.0±1.0	33.3±1.3	45.5±1.2
	HONGAT	81.8±0.6	60.6±1.1	61.3±1.0	44.8±1.4	49.8±1.2	32.8±1.2	45.0±1.1
	CRGNN	82.9±0.5	61.2±1.0	61.8±1.1	45.2±1.1	50.3±0.9	33.5±1.3	45.7±1.2
	CGNN	82.2±0.6	60.7±1.3	61.5±1.0	44.9±1.2	49.9±1.0	33.2±1.4	45.4±1.3
	ProCon	83.4±0.6	61.2±1.0	62.0±1.1	45.4±1.0	50.6±1.3	33.4±1.3	46.1±1.8
	KCR-GCL	**84.6±0.9**	**65.2±1.4**	**65.5±1.1**	**49.8±1.1**	**54.4±1.1**	**37.9±1.1**	**49.8±1.4**
ogbn-arxiv	GCN	71.7±0.3	40.1±1.4	47.8±1.0	33.6±1.1	33.9±1.2	28.6±2.2	29.4±1.3
	NRGNN	72.1±0.6	44.9±1.4	48.5±1.2	37.1±2.0	34.2±1.1	33.0±1.8	30.0±1.0
	RTGNN	71.8±0.4	44.3±1.2	48.4±1.4	38.0±1.1	34.0±1.7	33.5±1.1	30.1±0.6
	SUGRL	69.3±0.2	43.9±1.0	48.0±1.2	36.5±1.3	34.1±0.9	32.7±1.1	29.5±1.1
	MERIT	71.7±0.4	44.2±0.9	48.3±1.0	36.8±1.1	34.1±1.2	32.4±1.2	30.4±0.9
	SFA	71.8±0.9	44.5±1.2	48.6±1.2	36.8±1.1	33.8±1.5	32.5±1.4	30.2±1.3
	Jo-SRC	71.5±0.5	44.5±1.1	48.1±1.0	37.7±1.3	34.0±1.3	33.3±1.3	29.7±0.9
	GRAND+	72.5±0.4	44.5±0.8	48.1±1.1	37.8±1.0	34.4±1.0	33.2±1.0	30.3±0.9
	GFSA	71.9±0.4	44.3±1.2	48.2±1.1	37.0±1.2	34.2±1.1	32.8±1.2	29.9±1.1
	HONGAT	71.6±0.5	44.0±1.1	48.0±1.2	36.6±1.3	33.9±1.2	32.4±1.4	29.6±1.2
	CRGNN	72.1±0.3	44.6±1.0	48.3±0.9	37.2±1.0	34.3±1.0	33.0±1.2	30.2±1.0
	CGNN	71.7±0.6	44.1±1.3	48.1±1.0	36.8±1.4	34.0±1.1	32.6±1.5	29.8±1.2
	ProCon	72.3±0.4	44.8±1.4	48.9±1.0	38.0±1.1	34.3±1.3	33.5±1.1	30.3±0.9
	KCR-GCL	**73.1±0.6**	**48.7±1.3**	**52.3±1.4**	**42.3±1.4**	**42.3±1.2**	**37.4±1.5**	**35.0±1.3**

5.2 Node Classification

To rigorously assess the robustness of KCR-GCL, we conduct extensive experiments on graphs affected by both symmetric and asymmetric label noise, with corruption rates ranging from 40% to 80% in increments of 20%. In parallel, we examine the impact of attribute perturbations under the same levels of noise. We perform an extensive comparison of KCR-GCL against a broad range of semi-supervised node representation learning baselines. This includes classical methods such as GCN (Kipf & Welling, 2017), GCE (Zhang & Sabuncu, 2018), S^2GC (Zhu & Koniusz, 2021), and GRAND+(Feng et al., 2022b). To assess contrastive learning capabilities, we compare KCR-GCL with leading GCL-based methods, including MERIT (Jin et al., 2021), SUGRL (Mo et al., 2022), and SFA (Zhang et al., 2023). We further evaluate KCR-GCL against CRGNN (Li et al., 2024c) and CGNN (Yuan et al., 2023), which incorporate contrastive paradigms tailored to noisy label environments. Additionally, we compare KCR-GCL with attention-based GNN architectures, GFSA (Choi et al., 2024) and HONGAT (Zhang et al., 2024a), both of which are designed to integrate information across different frequency components of the graph spectrum. To evaluate performance under label corruption, we also include three specialized baselines designed to address noisy labels on graphs, NRGNN (Dai et al., 2021), RTGNN (Qian et al., 2023), and ProCon (Li et al., 2025). We run all experiments 10 times and report both the mean classification accuracy and the standard deviation. Table 2 compares KCR-GCL with the most competitive baseline methods on Cora, Citeseer, PubMed, and ogbn-arxiv with asymmetric label noise and attribute label noise. Additional results on Cora, Citeseer, PubMed, and ogbn-arxiv are deferred to Table 8 in Section B.1 of the appendix. Moreover, Table 9 in Section B.1 of the appendix 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 KCR-GCL consistently achieves the best performance across different datasets with different types of noise. For example, under 80% asymmetric label noise on PubMed, KCR-GCL outperforms CRGNN, the strongest baseline, by 4.4% in node classification accuracy, demonstrating the effectiveness of low-rank representation learning.

5.3 Ablation Study on the KCR Graph-Attention Layer

To study the effectiveness of the KCR self-attention layer, we design an ablation model of KCR-GCL, referred to as Low Rank-GCL (LR-GCL), without the KCR self-attention layer. The TNN, $\|\mathbf{K}(\boldsymbol{\theta})\|_{r_0}$, is incorporated as the low-rank regularization term on the kernel gram matrix $\mathbf{K}(\boldsymbol{\theta}) = \mathbf{H}(\boldsymbol{\theta})\mathbf{H}(\boldsymbol{\theta})^\top$ to train the GCL encoder $g_{\boldsymbol{\theta}}$, instead of the KCR-GCL encoder. Compared to KCR-GCL, LR-GCL lacks the polynomial graph filtering or graph convolutions implemented by the KCR self-attention layer. We train the LR-GCL encoder by minimizing $\mathcal{L}_{\text{LR-GCL}}(\boldsymbol{\theta}) = \mathcal{L}_{\text{node}}(\boldsymbol{\theta}) + \mathcal{L}_{\text{proto}}(\boldsymbol{\theta}) + \tau_0\|\mathbf{K}(\boldsymbol{\theta})\|_{r_0}$, where $\tau_0 > 0$ is the weighting parameter for the TNN, $\|\mathbf{K}(\boldsymbol{\theta})\|_{r_0}$, which is decided by the same cross-validation process as described in Section 5.1. The study is performed on four benchmark datasets, Cora, Citeseer, PubMed, and ogbn-arxiv, under various types and levels of noise, following Section 5.2. Table 3 demonstrates that incorporating the KCR self-attention layer consistently improves node classification performance across all datasets and noise settings. For example, on the Cora dataset under 60% symmetric label noise, KCR-GCL outperforms LR-GCL by 1.8%, highlighting the effectiveness of the KCR self-attention layer $\mathbf{B}$ with graph convolution, which learns rich frequency information and alleviates over-smoothing, similar to GFSA (Choi et al., 2024) and HONGAT (Zhang et al., 2024a).

Table 3: Ablation study on the KCR self-attention layer in KCR-GCL for node classification accuracy (%) on Cora, Citeseer, PubMed, and ogbn-arxiv.

Dataset	Methods	Noise Type									
		0	40			60			80		
		-	Asymmetric	Symmetric	Attribute	Asymmetric	Symmetric	Attribute	Asymmetric	Symmetric	Attribute
Cora	LR-GCL	85.8±0.6	58.9±1.1	71.3±0.7	69.5±1.1	49.2±1.1	58.7±1.3	47.7±1.2	30.6±1.2	41.9±1.2	36.3±1.1
	KCR-GCL	**86.1±0.6**	**60.2±1.1**	**72.4±0.7**	**70.8±1.1**	**51.0±1.1**	**60.5±1.3**	**49.2±1.2**	**32.9±1.2**	**43.6±1.2**	**38.2±1.1**
Citeseer	LR-GCL	75.7±1.0	52.0±1.3	58.1±1.3	57.0±0.7	41.0±1.4	45.5±1.4	40.6±1.2	36.9±1.2	33.5±1.4	31.8±1.0
	KCR-GCL	**76.2±1.0**	**53.3±1.3**	**59.7±1.3**	**58.8±0.7**	**43.0±1.4**	**47.2±1.4**	**42.3±1.2**	**39.2±1.2**	**35.2±1.4**	**33.5±1.0**
PubMed	LR-GCL	84.5±0.9	63.7±1.4	64.5±1.5	63.7±1.1	47.9±1.1	48.4±1.3	52.6±1.1	35.6±1.1	36.0±1.2	48.2±1.4
	KCR-GCL	**84.6±0.9**	**65.2±1.4**	**66.2±1.5**	**65.5±1.1**	**49.8±1.1**	**50.3±1.3**	**54.4±1.1**	**37.9±1.1**	**37.9±1.2**	**49.8±1.4**
ogbn-arxiv	LR-GCL	72.8±0.6	47.2±1.3	49.2±1.1	50.8±1.4	40.5±1.4	41.1±1.2	40.5±1.2	35.9±1.5	30.7±1.3	33.5±1.3
	KCR-GCL	**73.1±0.6**	**48.7±1.3**	**50.7±1.1**	**52.3±1.4**	**42.3±1.4**	**43.0±1.2**	**42.3±1.2**	**37.4±1.5**	**33.2±1.3**	**35.0±1.3**

5.4 Study on the Kernel Complexity and the Upper Bound of the Test Loss

We present a detailed comparison of the individual components that constitute the upper bound of the test loss defined in Equation (3), namely $L_1(\mathbf{K_F}, \tilde{\mathbf{Y}}, t)$, $L_2(\mathbf{K_F}, \mathbf{N}, t)$, and the kernel complexity term $\mathrm{KC}(\mathbf{K_F})$, based on node representations obtained from various methods. The corresponding evaluation results are summarized in Table 4. All experiments are conducted on the Cora, Citeseer, and PubMed datasets under symmetric label noise with a corruption rate of 40%. The results demonstrate that KCR-GCL consistently achieves significantly lower values across all three terms when compared to baseline models. These reductions indicate a stronger capacity for generalization in semi-supervised node classification tasks, even under the presence of label noise.

Table 4: Comparisons on $L_1(\mathbf{K_F}, \tilde{\mathbf{Y}}, t)$, $L_2(\mathbf{K_F}, \mathbf{N}, t)$, $\mathrm{KC}(\mathbf{K_F})$ and the value of the upper bound of the test loss from Theorem 4.1. The lowest values for each dataset in the table are bold.

Datasets		MERIT	SFA	Jo-SRC	GCN	GFSA	HONGAT	ProCon	LR-GCL	KCR-GCL
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	4.84 ± 0.35	3.72 ± 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	3.40 ± 0.35	2.97 ± 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	0.33 ± 0.20	0.20 ± 0.02	**0.14** ± 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	10.28 ± 0.35	7.05 ± 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	4.14 ± 0.35	4.02 ± 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	3.85 ± 0.35	3.75 ± 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	0.43 ± 0.20	0.24 ± 0.18	**0.18** ± 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	9.00 ± 0.35	8.20 ± 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	3.55 ± 0.35	3.38 ± 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	2.42 ± 0.35	2.32 ± 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	0.48 ± 0.20	0.30 ± 0.29	**0.25** ± 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	6.88 ± 0.35	6.25 ± 0.30	**6.16** ± 0.40

In addition, we further compare the KC of the gram matrix computed from node representations generated by KCR-GCL, its ablation model LR-GCL, and several competing baseline methods across more benchmark datasets. This evaluation is conducted under symmetric label noise with a corruption rate of 40%. As shown in Table 5, the node representations learned by KCR-GCL exhibit consistently lower kernel complexity, suggesting that transductive classifiers trained on such representations are likely to achieve smaller generalization errors on previously unseen nodes.

Table 5: Comparisons in the kernel complexity defined in Theorem 4.1 of the main paper. 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	ProCon	LR-GCL	KCR-GCL
Cora	KC	0.37	0.42	0.48	0.44	0.35	0.40	0.33	0.20	0.14
	r_0	1420	1478	1665	1511	1262	1450	1232	440	395
Citeseer	KC	0.47	0.45	0.55	0.64	0.47	0.50	0.43	0.24	0.18
	r_0	1214	1180	1405	1590	1224	1285	1150	405	369
PubMed	KC	0.54	0.50	0.62	0.71	0.52	0.66	0.48	0.30	0.25
	r_0	1644	1562	1785	1993	1588	1874	1532	1197	1090
Wiki-CS	KC	0.42	0.44	0.40	0.49	0.43	0.45	0.38	0.19	0.14
	r_0	1805	1993	1746	2130	1842	2048	1716	970	904
Amazon-Computers	KC	0.39	0.37	0.40	0.45	0.35	0.37	0.33	0.12	0.10
	r_0	1450	1428	1489	1632	1370	1415	1340	874	820
Amazon-Photos	KC	0.38	0.38	0.43	0.47	0.39	0.41	0.36	0.14	0.12
	r_0	1872	1884	1990	2145	1895	1921	1842	750	722
Coauthor-CS	KC	0.29	0.28	0.32	0.34	0.31	0.32	0.26	0.12	0.10
	r_0	1774	1725	1896	1903	1872	1890	1695	1120	1039
ogbn-arxiv	KC	0.12	0.13	0.12	0.14	0.12	0.13	0.10	0.05	0.04
	r_0	1860	1936	1852	1996	1845	1920	1815	1354	1328

5.5 Training Time Comparison

In this section, we report a comparative analysis of the training time for KCR-GCL and other baseline methods across all benchmark datasets. The total training time for KCR-GCL encompasses three components: the time required for robust graph contrastive learning, the computation time for the singular value decomposition (SVD) of the kernel matrix, and the training time of the transductive classifier.

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 GPU. The detailed results are provided in Table 6. As shown, the overall training time of KCR-GCL is comparable to that of state-of-the-art GCL methods such as SFA and MERIT.

Table 6: Training time (seconds) comparisons for node classification.

Methods	Cora	Citeseer	PubMed	Coauthor CS	Wiki-CS	Amazon Computer	Amazon Photo	ogbn-arxiv
GCN	11.5	13.7	38.6	43.2	22.3	30.2	19.0	215.1
S^2GC	20.7	22.5	47.2	57.2	27.6	38.5	22.2	243.7
GCE	32.6	36.9	67.3	80.8	37.6	50.1	32.2	346.1
NRGNN	72.4	80.5	142.7	189.4	74.3	97.2	62.4	650.2
RTGNN	143.3	169.5	299.5	353.5	153.7	201.5	124.2	1322.2
ProCon	128.5	151.2	268.4	318.7	138.9	181.3	111.7	1187.5
SUGRL	100.3	122.1	207.4	227.1	107.7	142.8	87.7	946.8
MERIT	167.2	179.2	336.7	375.3	172.3	226.5	140.6	1495.1
ARIEL	156.9	164.3	284.3	332.6	145.1	190.4	118.3	1261.4
SFA	237.5	269.4	457.1	492.3	233.5	304.5	187.2	2013.1
Sel-Cl	177.3	189.9	313.5	352.5	161.7	211.1	130.9	1401.1
Jo-SRC	148.2	157.1	281.0	306.1	144.5	188.0	118.5	1256.0
GRAND+	57.4	68.4	101.7	124.2	54.8	73.8	44.5	479.2
KCR-GCL	166.2	185.4	372.7	399.5	195.4	253.6	159.2	1674.8

6 Conclusions

This paper introduces Kernel Complexity Reduced Graph Contrastive Learning, or KCR-GCL, which consists of a KCR-GCL encoder that learns robust node representation, which will be used for transductive node classification. The KCR-GCL encoder integrates a novel self-attention mechanism that adaptively combines multiple powers of the feature kernel matrix to balance spectral components and reduce kernel complexity. The encoder is trained within a prototypical graph contrastive learning (GCL) framework, with a truncated nuclear norm (TNN) on the gram matrix of the learned features incorporated as a regularizer. The TNN regularizer encourages the KCR-GCL encoder to learn low-rank representations, motivated by the prevalence of low-frequency components in real-world graphs and the theoretical tightness of generalization bounds in transductive settings. Extensive empirical results across diverse graph benchmarks demonstrate that KCR-GCL exhibits strong robustness and consistently outperforms state-of-the-art in learning effective node representations for transductive node classification under noisy conditions, where graphs are subjected to either label corruption or attribute perturbations.

References

Kelsey Allen, Evan Shelhamer, Hanul Shin, and Joshua Tenenbaum. Infinite mixture prototypes for few-shot learning. In *International Conference on Machine Learning*, pp. 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*, pp. 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, pp. 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 *International Joint Conference on Artificial Intelligence*, pp. 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*, pp. 2905–2909, 2021.

Xiuyuan Cheng, Zichen Miao, and Qiang Qiu. Graph convolution with low-rank learnable local filters. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=9OHFhefeB86.

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, 2024.

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.

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*, pp. 181–191, 2022.

Kaize Ding, Zhe Xu, Hanghang Tong, and Huan Liu. Data augmentation for deep graph learning: A survey. *ACM SIGKDD Explorations Newsletter*, 24(2):61–77, 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*, pp. 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*, pp. 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*, pp. 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*, pp. 9706–9715, 2022.

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 *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pp. 8536–8546, 2018.

Bo Han, Quanming Yao, Tongliang Liu, Gang Niu, Ivor W Tsang, James T Kwok, and Masashi Sugiyama. A survey of label-noise representation learning: Past, present and future. *arXiv preprint arXiv:2011.04406*, 2020.

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 *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020a.

Weihua Hu, Bowen Liu, Joseph Gomes, Marinka Zitnik, Percy Liang, Vijay S. Pande, and Jure Leskovec. Strategies for pre-training graph neural networks. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*, 2020b.

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*, pp. 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)*, pp. 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, pp. 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 *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.

Namkyeong Lee, Junseok Lee, and Chanyoung Park. Augmentation-free self-supervised learning on graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 7372–7380, 2022.

Haifeng Li, Jun Cao, Jiawei Zhu, Qinyao Luo, Silu He, and Xuying Wang. Augmentation-free graph contrastive learning of invariant-discriminative representations. *IEEE Trans. Neural Networks Learn. Syst.*, 35(8):11157–11167, 2024a.

Junnan Li, Richard Socher, and Steven CH Hoi. Dividemix: Learning with noisy labels as semi-supervised learning. In *International Conference on Learning Representations*, 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.

Kailai Li, Jiawei Sun, Jiong Lou, Zhanbo Feng, Hefeng Zhou, Chentao Wu, Guangtao Xue, Wei Zhao, and Jie Li. Leveraging peer-informed label consistency for robust graph neural networks with noisy labels. In James Kwok (ed.), *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, pp. 5598–5606. International Joint Conferences on Artificial Intelligence Organization, 8 2025. doi: 10.24963/ijcai.2025/623. Main Track.

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*, pp. 1255–1265, 2024b.

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

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 *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pp. 960–970, 2017.

Antonio G. Marques, Santiago Segarra, and Gonzalo Mateos. Signal processing on directed graphs: The role of edge directionality when processing and learning from network data. *IEEE Signal Process. Mag.*, 37(6): 99–116, 2020.

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*, pp. 7797–7805, 2022.

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*, pp. 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*, pp. 259–270, 2020.

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 *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining, WSDM 2023, Singapore, 27 February 2023 - 3 March 2023*, pp. 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 *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 5301–5310, 09–15 Jun 2019.

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.

Fan-Yun Sun, Jordan Hoffmann, Vikas Verma, and Jian Tang. Infograph: Unsupervised and semi-supervised graph-level representation learning via mutual information maximization. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*, 2020.

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)*, pp. 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.

Susheel Suresh, Pan Li, Cong Hao, and Jennifer Neville. Adversarial graph augmentation to improve graph contrastive learning. *Advances in Neural Information Processing Systems*, 34:15920–15933, 2021.

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.

Shantanu Thakoor, Corentin Tallec, Mohammad Gheshlaghi Azar, Rémi Munos, Petar Veličković, and Michal Valko. Bootstrapped representation learning on graphs. In *ICLR 2021 Workshop on Geometrical and Topological Representation Learning*, 2021.

Ilya O. Tolstikhin, Gilles Blanchard, and Marius Kloft. Localized complexities for transductive learning. In *Conference on Learning Theory*, volume 35 of *JMLR Workshop and Conference Proceedings*, pp. 857–884, 2014.

Petar Velickovic, William Fedus, William L. Hamilton, Pietro Liò, Yoshua Bengio, and R. Devon Hjelm. Deep graph infomax. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*, 2019.

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 *Advances in Neural Information Processing Systems*, 2024b.

Yancheng Wang, Changyu Liu, and Yingzhen Yang. Diffusion on graph: Augmentation of graph structure for node classification. *Trans. Mach. Learn. Res.*, 2025, 2025.

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 *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 *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*, pp. 6861–6871, 2019.

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*, 2019.

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*, pp. 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*, pp. 469–477, 2023.

Yingzhen Yang. Sharp generalization of transductive learning: A transductive local rademacher complexity approach. *CoRR*, abs/2309.16858, 2023. doi: 10.48550/ARXIV.2309.16858. URL https://doi.org/10.48550/arXiv.2309.16858.

Yingzhen Yang. A new concentration inequality for sampling without replacement and its application for transductive learning. In *International Conference on Machine Learning (ICML)*, 2025.

Yuning You, Tianlong Chen, Yang Shen, and Zhangyang Wang. Graph contrastive learning automated. In *International Conference on Machine Learning*, pp. 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*, pp. 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 *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*, pp. 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)*, pp. 1–5. IEEE, 2023.

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, pp. 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, pp. 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*, pp. 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, pp. 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, pp. 4405–4413, 2022.

A Detailed Experiment Settings

Implementation Details of the Label Noise. To introduce label noise, we follow established methodologies from the literature (Han et al., 2020; Dai et al., 2022; Qian et al., 2023), adopting (1) Symmetric noise, where each label is replaced by a randomly chosen label from the remaining classes with uniform probability; and (2) Asymmetric noise, in which labels are more likely to be flipped to semantically related classes. We implement the formal noise model described in (Song et al., 2022), where a noise transition matrix $\mathbf{T} \in [0,1]^{C\times C}$ is used, with entries $\mathbf{T}_{ij} := \mathbb{P}(\tilde{y} = j \mid y = i)$ denoting the probability of a clean label i being corrupted into a noisy label j. For symmetric corruption at noise rate τ, we define $\mathbf{T}_{ii} = 1 - \tau$ and $\mathbf{T}_{ij} = \frac{\tau}{C-1}$ for all $j \neq i$. In the asymmetric case, $\mathbf{T}_{ii} = 1 - \tau$ while the off-diagonal entries are structured such that $\mathbf{T}_{ij} > \mathbf{T}_{ik}$ for certain pairs $j, k \neq i$, reflecting realistic label confusion patterns.

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., 2024b; Ju et al., 2025). The percentage of shuffled attributes is defined as the attribute noise level in our experiments.

The hyperparameters are selected via five-fold cross-validation on the training set. We sweep the learning rate over $\{1\times10^{-4}, 5\times10^{-4}, 1\times10^{-3}, 5\times10^{-3}, 1\times10^{-2}, 5\times10^{-2}, 1\times10^{-1}, 5\times10^{-1}\}$ and the weight decay from $\{1\times10^{-5}, 5\times10^{-5}, 1\times10^{-4}, 5\times10^{-4}, 1\times10^{-3}, 5\times10^{-3}\}$. The best hyperparameters are identified as those minimizing the validation loss. We train all models using the Adam optimizer for a maximum of 500 epochs, employing early stopping if the validation loss does not improve for 20 consecutive epochs.

Table 7: Selected rank ratio γ and TNN weight λ for each dataset.

Parameters	Cora	Citeseer	PubMed	Coauthor CS	ogbn-arxiv	Wiki-CS	Computers	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
M	3	2	4	3	3	3	4	3

Cross-Validation for Tuning r_0, τ, and M. The rank parameter r_0, the regularization weight τ associated with the TNN loss, and the value of the maximum power, M, in KCR attention are selected through cross-validation tailored to each dataset. We define the rank as $r_0 = \lceil \gamma \min\{N, d\} \rceil$, where γ represents the rank ratio. The hyperparameter γ is searched over the set $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$, while the TNN weight τ is chosen from $\{0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5\}$. The value of the maximum power, M, in KCR attention is selected from $\{1, 2, 3, 4, 5\}$. All the above parameters are tuned using five-fold cross-validation, conducted on a randomly sampled 20% subset of the training data. The final selected values for each dataset are reported in Table 7.

Table 8: Comparison for node classification accuracy (%) 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	81.5±0.5	54.7±1.5	63.6±0.7	63.9±0.8	40.5±1.4	51.7±1.0	43.9±1.2	26.5±1.2	35.4±1.4	31.7±1.3
	S^2GC	83.5±0.2	56.9±0.7	66.4±0.7	66.1±0.7	42.2±1.0	53.5±1.0	45.4±1.1	27.9±1.4	36.6±1.4	32.0±1.3
	GCE	81.9±0.4	57.3±1.1	65.2±0.8	65.0±1.4	44.9±1.1	50.9±1.1	44.5±1.5	28.0±1.3	35.3±1.3	32.5±1.5
	NRGNN	82.2±0.6	57.1±1.9	67.6±0.7	64.5±1.2	47.0±1.4	54.8±1.4	45.1±1.1	28.2±2.2	37.3±1.2	32.6±1.0
	RTGNN	82.8±0.3	57.0±1.0	68.2±0.8	67.8±1.1	47.4±1.1	55.5±1.0	45.7±0.9	28.0±1.1	38.6±1.4	34.2±1.6
	SUGRL	83.4±0.5	56.4±1.1	67.4±1.2	67.5±0.9	46.8±1.1	55.2±1.1	45.2±1.2	28.0±1.2	38.1±1.2	33.8±1.4
	MERIT	83.1±0.5	56.0±0.8	67.0±0.8	67.1±0.9	46.7±1.3	54.7±1.3	45.0±1.4	27.7±1.3	38.5±1.3	33.5±0.9
	ARIEL	84.3±0.4	57.3±1.3	68.1±1.0	67.5±0.9	47.1±1.2	55.3±1.2	45.5±1.4	28.4±1.4	38.9±1.3	34.3±1.3
	SFA	83.9±1.0	56.4±1.1	67.7±1.3	67.6±1.5	47.3±1.4	54.9±1.4	45.7±1.4	28.2±1.6	38.9±1.3	34.4±1.7
	Sel-Cl	82.8±0.2	57.0±1.0	68.5±1.2	67.6±0.9	47.2±1.3	55.4±1.4	45.5±1.1	28.2±1.7	38.9±1.3	34.1±1.5
	Jo-SRC	82.5±0.5	57.1±0.6	68.4±1.3	67.9±0.7	47.3±1.1	55.6±0.8	45.8±1.2	28.5±1.3	38.7±1.8	34.5±1.8
	GRAND+	85.8±0.6	57.0±0.9	68.2±0.7	67.8±1.1	47.2±1.0	55.4±0.8	45.6±1.2	28.4±1.5	38.7±1.5	34.5±1.3
	GFSA	83.7±0.6	56.8±1.2	67.6±1.0	67.2±0.9	46.6±1.2	54.5±1.3	45.1±1.2	27.9±1.2	38.4±1.5	33.6±1.3
	HONGAT	83.3±0.4	56.6±1.1	67.3±1.1	66.7±1.0	46.4±1.0	54.3±1.1	44.9±1.0	27.8±1.3	38.1±1.4	33.4±1.4
	CRGNN	84.2±0.5	57.2±1.0	67.8±1.0	67.4±1.0	47.0±1.2	55.1±1.3	45.4±1.3	28.3±1.4	38.6±1.4	34.1±1.5
	CGNN	83.5±0.6	56.7±0.9	67.0±1.2	66.9±1.1	46.2±1.3	54.4±1.1	45.0±1.3	28.1±1.2	38.0±1.3	33.7±1.4
	ProCon	84.8±0.6	57.2±1.1	68.7±1.2	67.8±0.7	47.3±1.1	55.8±0.8	45.8±1.2	28.4±1.3	39.1±1.3	34.4±1.8
	KCR-GCL	**86.1±0.6**	**60.2±1.1**	**72.4±0.7**	**70.8±1.1**	**51.0±1.1**	**60.5±1.3**	**49.2±1.2**	**32.9±1.2**	**43.6±1.2**	**38.2±1.1**
Citeseer	GCN	70.3±0.5	47.5±2.3	50.1±1.3	52.9±0.9	35.1±1.4	34.1±1.4	37.2±1.1	29.1±2.2	28.1±1.9	29.0±1.4
	S^2GC	73.6±0.5	48.8±1.3	52.8±1.3	55.3±0.8	36.3±1.2	36.7±1.4	39.0±1.3	30.4±2.4	28.4±1.9	28.8±1.1
	GCE	70.5±0.4	49.0±1.6	51.2±1.4	54.0±1.4	36.2±1.5	35.2±1.0	38.1±0.9	30.9±1.2	28.5±1.4	28.5±1.1
	NRGNN	71.0±0.6	49.8±1.5	54.6±1.5	53.8±1.1	38.2±1.6	41.2±1.6	37.7±1.2	33.6±2.1	30.9±1.8	28.4±0.9
	RTGNN	74.6±0.8	49.8±0.7	55.6±0.7	55.0±1.2	39.2±1.0	42.4±1.3	39.0±1.4	34.8±1.7	30.8±1.6	30.2±1.1
	SUGRL	73.0±0.5	49.3±1.1	54.1±1.1	54.4±1.0	37.6±0.9	42.1±0.9	38.8±0.9	33.9±1.0	30.5±1.0	30.0±0.9
	MERIT	74.0±0.7	49.6±1.2	53.6±1.2	54.2±1.0	38.3±1.1	42.5±1.1	38.7±0.8	34.4±1.4	30.1±1.4	29.5±0.9
	SFA	74.0±1.1	50.2±1.4	53.2±1.5	54.7±1.3	39.0±1.4	43.3±1.4	38.9±1.2	34.7±1.6	31.2±1.5	29.9±1.3
	ARIEL	72.7±0.7	50.0±0.8	55.0±1.3	54.8±0.8	39.1±0.9	42.7±1.2	38.9±1.4	34.9±1.4	30.7±1.3	29.9±1.3
	Sel-Cl	72.5±0.8	49.9±1.2	55.1±1.0	54.9±0.8	38.9±1.1	42.6±0.8	39.1±2.0	35.0±1.8	31.0±1.5	30.0±1.7
	Jo-SRC	73.0±0.5	50.0±1.3	55.5±1.1	55.1±1.1	39.4±1.3	42.5±1.3	39.3±1.3	35.1±1.3	30.5±1.8	30.3±1.3
	GRAND+	75.6±0.4	49.7±1.0	55.3±1.0	55.2±1.1	39.0±1.3	42.2±1.3	38.7±1.3	34.8±1.3	30.9±1.4	30.2±1.2
	GFSA	74.3±0.6	49.5±1.2	54.6±1.2	54.6±1.1	38.6±1.1	41.8±1.1	38.6±1.2	34.2±1.3	30.8±1.5	29.8±1.2
	HONGAT	73.8±0.7	49.2±1.4	54.0±1.1	54.5±0.9	38.0±1.2	41.3±1.0	38.4±1.3	34.0±1.4	30.6±1.6	29.6±1.1
	CRGNN	75.1±0.6	49.7±1.1	55.2±1.0	54.9±1.2	38.9±1.4	42.3±1.3	38.8±1.2	34.7±1.5	31.0±1.4	30.1±1.2
	CGNN	74.1±0.7	49.3±1.3	54.4±1.2	54.6±1.0	38.5±1.3	41.9±1.2	38.5±1.1	34.3±1.3	30.7±1.3	29.7±1.2
	ProCon	75.3±0.4	50.4±1.4	55.5±0.7	55.2±0.8	39.6±1.3	43.0±1.4	39.2±1.3	35.3±1.3	31.1±1.5	30.2±1.3
	KCR-GCL	**76.2±1.0**	**53.3±1.3**	**59.7±1.3**	**58.8±0.7**	**43.0±1.4**	**47.2±1.4**	**42.3±1.2**	**39.2±1.2**	**35.2±1.4**	**33.5±1.0**
PubMed	GCN	79.0±0.7	58.4±2.2	57.4±1.2	59.5±1.2	40.5±2.5	38.6±1.1	48.8±1.3	30.5±2.2	29.5±1.3	42.3±1.3
	S^2GC	80.2±0.5	58.5±2.3	58.9±1.3	61.0±0.9	42.1±3.0	40.1±1.4	49.7±1.2	31.0±3.9	29.0±1.9	43.1±1.0
	GCE	79.2±0.9	58.9±1.8	58.1±1.1	59.0±1.4	43.0±1.2	39.9±1.2	49.1±1.0	31.1±2.1	30.1±1.1	42.4±1.2
	NRGNN	79.7±0.8	60.2±2.2	61.8±1.3	60.3±0.8	44.3±1.2	43.4±1.2	49.9±0.9	33.0±2.3	32.5±1.3	43.3±1.1
	RTGNN	79.7±0.4	61.0±0.8	62.2±1.0	61.4±1.2	45.5±1.0	45.5±1.1	50.1±1.1	33.5±1.3	33.8±1.7	45.2±1.3
	SUGRL	81.9±0.5	60.3±1.3	61.5±1.3	61.5±1.0	44.5±1.1	44.1±1.1	50.1±0.7	32.1±0.9	32.1±0.9	44.6±1.0
	MERIT	80.1±0.4	59.3±1.1	61.2±1.1	61.3±1.1	44.7±1.2	44.3±1.2	49.7±0.9	32.8±1.1	32.3±1.1	44.5±0.9
	ARIEL	80.0±0.3	61.0±1.3	62.2±1.0	61.5±1.1	45.3±1.2	45.3±1.2	50.2±1.4	33.1±1.4	33.6±1.8	45.7±1.3
	SFA	80.4±1.0	59.6±1.1	61.5±1.1	60.9±1.1	44.7±1.4	44.6±1.7	49.9±1.4	33.0±1.1	32.7±1.1	44.7±1.4
	Sel-Cl	79.9±0.5	60.5±1.4	62.5±1.2	61.4±1.2	45.5±1.4	44.9±1.0	50.2±0.8	33.4±2.1	33.2±1.4	45.6±1.4
	Jo-SRC	80.1±0.5	61.3±1.0	62.4±1.3	61.7±1.3	45.3±0.8	45.5±1.3	50.4±1.3	33.0±1.5	33.4±1.8	45.9±1.8
	GRAND+	84.5±0.6	61.0±1.1	62.4±1.3	61.7±1.3	45.3±0.8	45.3±1.1	50.3±1.0	33.1±1.4	33.7±1.3	45.8±1.4
	GFSA	82.3±0.5	60.8±1.2	62.1±1.1	61.6±0.9	45.0±1.3	45.2±1.2	50.0±1.0	33.3±1.3	33.4±1.1	45.5±1.2
	HONGAT	81.8±0.6	60.6±1.1	61.9±1.2	61.3±1.0	44.8±1.4	44.7±1.2	49.8±1.2	32.8±1.2	32.6±1.3	45.0±1.1
	CRGNN	82.9±0.5	61.2±1.0	62.3±0.9	61.8±1.1	45.2±1.1	45.5±1.3	50.3±0.9	33.5±1.3	33.3±1.4	45.7±1.2
	CGNN	82.2±0.6	60.7±1.3	62.0±1.1	61.5±1.0	44.9±1.2	45.1±1.4	49.9±1.0	33.2±1.4	33.0±1.2	45.4±1.3
	ProCon	83.4±0.6	61.2±1.0	62.5±1.2	62.0±1.1	45.4±1.0	45.4±1.1	50.6±1.3	33.4±1.3	33.8±1.7	46.1±1.8
	KCR-GCL	**84.6±0.9**	**65.2±1.4**	**66.2±1.5**	**65.5±1.1**	**49.8±1.1**	**50.3±1.3**	**54.4±1.1**	**37.9±1.1**	**37.9±1.2**	**49.8±1.4**
ogbn-arxiv	GCN	71.7±0.3	40.1±1.4	42.1±1.4	47.8±1.0	33.6±1.1	34.6±2.1	33.9±1.2	28.6±2.2	25.6±1.0	29.4±1.3
	S^2GC	71.2±0.3	41.7±1.7	42.9±1.4	49.2±1.0	34.4±1.6	35.3±3.1	34.3±0.9	29.7±2.3	26.6±1.3	28.4±1.2
	GCE	72.0±0.4	41.0±1.8	42.8±0.8	48.0±1.4	34.8±1.9	34.4±1.9	34.2±1.5	31.0±1.4	26.0±1.1	27.5±1.5
	NRGNN	72.1±0.6	44.9±1.4	46.6±0.9	48.5±1.2	37.1±2.0	37.9±0.8	34.2±1.1	33.0±1.8	27.1±1.8	30.0±1.0
	RTGNN	71.8±0.4	44.3±1.2	46.4±1.2	48.4±1.4	38.0±1.1	38.4±1.3	34.0±1.7	33.5±1.1	28.5±1.5	30.1±0.6
	SUGRL	69.3±0.2	43.9±1.0	46.7±1.0	48.0±1.2	36.5±1.3	38.5±1.1	34.1±0.9	32.7±1.1	27.5±1.1	29.5±1.1
	MERIT	71.7±0.4	44.2±0.9	46.3±0.9	48.3±1.0	36.8±1.1	38.1±1.1	34.1±1.2	32.4±1.2	27.2±1.0	30.4±0.9
	ARIEL	71.7±0.4	44.8±1.3	47.1±1.3	48.2±1.1	37.9±1.4	38.4±1.5	34.2±1.5	33.4±1.4	28.0±1.3	30.0±1.0
	SFA	71.8±0.9	44.5±1.2	46.3±1.3	48.6±1.2	36.8±1.1	37.8±1.4	33.8±1.5	32.5±1.4	27.3±1.2	30.2±1.3
	Sel-Cl	71.9±0.2	44.7±0.7	46.9±0.7	48.6±1.0	37.5±0.8	38.9±2.5	34.4±1.3	33.1±0.8	28.4±1.9	30.4±1.2
	Jo-SRC	71.5±0.5	44.5±1.1	46.6±0.9	48.1±1.0	37.7±1.3	38.7±1.3	34.0±1.3	33.3±1.3	28.2±1.8	29.7±0.9
	GRAND+	72.5±0.4	44.5±0.8	46.6±1.1	48.1±1.1	37.8±1.0	38.5±1.2	34.4±1.0	33.2±1.0	28.2±1.6	30.3±0.9
	GFSA	71.9±0.4	44.3±1.2	46.0±1.0	48.2±1.1	37.0±1.2	37.9±1.2	34.2±1.1	32.8±1.2	27.8±1.3	29.9±1.1
	HONGAT	71.6±0.5	44.0±1.1	45.8±1.2	48.0±1.2	36.6±1.3	37.3±1.3	33.9±1.2	32.4±1.4	27.6±1.4	29.6±1.2
	CRGNN	72.1±0.3	44.6±1.0	46.5±1.0	48.3±0.9	37.2±1.0	38.2±1.1	34.3±1.0	33.0±1.2	28.1±1.2	30.2±1.0
	CGNN	71.7±0.6	44.1±1.3	46.2±1.1	48.1±1.0	36.8±1.4	37.6±1.2	34.0±1.1	32.6±1.5	27.7±1.3	29.8±1.2
	ProCon	72.3±0.4	44.8±1.4	47.3±1.3	48.9±1.0	38.0±1.1	39.1±2.5	34.3±1.3	33.5±1.1	28.7±1.5	30.3±0.9
	KCR-GCL	**73.1±0.6**	**48.7±1.3**	**50.7±1.1**	**52.3±1.4**	**42.3±1.4**	**43.0±1.2**	**42.3±1.2**	**37.4±1.5**	**33.2±1.3**	**35.0±1.3**

Table 9: Comparison for node classification accuracy (%) 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	91.8±0.1	64.5±0.9	65.6±0.6	70.2±1.0	51.1±1.3	50.1±0.9	53.1±1.0	42.9±2.2	38.9±1.1	41.5±1.3
	S^2GC	91.8±0.1	65.7±1.2	66.3±0.6	71.3±1.0	51.6±1.3	51.4±0.9	55.6±0.9	43.7±2.0	39.6±1.0	42.2±1.2
	GCE	92.2±0.3	66.2±1.7	65.9±0.7	70.5±1.4	51.5±1.6	50.2±0.7	53.9±0.9	44.3±1.7	38.9±1.2	41.2±1.1
	NRGNN	91.9±0.2	67.8±1.4	68.9±0.9	70.5±1.2	54.5±2.1	55.6±1.1	54.6±1.1	46.1±1.2	41.0±1.2	41.7±0.7
	RTGNN	92.0±0.5	67.8±1.2	69.1±0.9	71.2±0.8	55.9±1.0	56.9±1.1	56.0±0.8	45.5±1.5	41.5±1.5	41.2±1.4
	SUGRL	92.2±0.5	67.5±1.0	69.5±1.0	71.4±0.6	55.0±1.1	56.0±1.1	56.1±0.7	44.9±1.1	41.1±1.1	42.9±0.8
	MERIT	92.4±0.4	67.9±1.1	68.9±0.8	70.9±0.5	55.2±1.4	56.2±1.4	56.2±1.1	45.2±1.3	40.3±1.3	42.6±0.5
	ARIEL	92.5±0.4	68.2±1.1	69.9±0.9	71.2±0.5	55.5±1.1	56.6±1.1	55.6±1.1	45.4±1.4	41.5±1.9	42.7±1.3
	SFA	92.5±0.9	68.2±1.1	69.0±1.2	71.5±1.2	55.5±1.5	56.7±1.4	56.5±1.3	45.8±1.3	40.2±1.3	42.9±1.5
	Sel-Cl	92.2±0.8	68.4±0.9	69.4±1.2	71.4±1.0	55.7±1.3	56.8±1.3	56.6±1.0	45.7±1.3	41.2±1.7	42.5±0.9
	Jo-SRC	92.1±0.5	68.4±1.1	69.5±0.4	70.9±0.7	56.0±1.1	56.6±1.1	56.1±0.9	45.6±1.3	41.0±1.8	42.8±1.0
	GRAND+	92.7±0.4	68.2±1.1	69.3±0.6	71.5±0.8	55.4±0.8	56.8±1.3	55.7±1.1	45.5±1.2	41.6±1.3	42.8±1.1
	GFSA	92.3±0.4	67.9±1.0	68.7±0.9	71.1±0.9	55.0±1.2	55.9±1.1	55.8±1.0	45.3±1.4	41.0±1.2	42.6±1.1
	HONGAT	92.4±0.3	68.1±1.2	69.2±1.0	71.3±0.8	55.3±1.3	56.3±1.3	56.0±1.2	45.6±1.3	41.1±1.5	42.7±1.0
	CRGNN	92.6±0.5	68.3±1.1	69.0±1.1	71.2±0.7	55.1±1.5	56.1±1.2	55.9±1.1	45.4±1.2	41.2±1.4	42.6±1.2
	CGNN	92.5±0.6	68.0±1.2	68.9±1.2	71.0±1.0	54.9±1.4	56.0±1.2	55.7±1.2	45.2±1.3	40.9±1.5	42.5±1.2
	ProCon	92.7±0.4	68.6±0.9	69.7±0.9	71.4±1.2	56.2±1.1	56.8±1.1	56.5±1.0	46.3±1.2	41.5±1.3	42.8±0.8
	KCR-GCL	**93.4±0.6**	**71.4±1.5**	**73.6±1.1**	**75.8±1.5**	**59.4±1.4**	**61.2±1.8**	**60.6±1.5**	**48.9±1.5**	**45.3±1.5**	**47.0±1.7**
Wiki-CS	GCN	80.1±0.4	61.2±0.8	62.5±1.0	64.7±0.9	49.7±1.3	48.3±1.2	50.2±1.1	40.1±1.6	36.5±1.7	38.2±1.6
	S^2GC	80.6±0.3	62.1±0.9	63.0±1.1	65.9±1.0	50.3±1.4	49.2±1.2	51.6±1.2	41.1±1.8	37.3±1.6	39.7±1.5
	GCE	80.8±0.4	61.8±0.7	62.9±0.8	65.1±0.9	49.5±1.2	48.1±1.0	51.0±1.0	40.4±1.5	36.1±1.3	38.3±1.4
	NRGNN	80.9±0.3	63.5±0.8	64.2±0.9	66.5±0.8	51.4±1.2	51.8±1.1	52.6±1.0	42.6±1.4	38.6±1.6	40.3±1.2
	RTGNN	81.1±0.4	63.8±1.0	64.5±1.0	66.7±0.8	51.7±1.1	52.2±1.1	52.8±0.9	42.8±1.3	39.1±1.5	40.6±1.2
	SUGRL	81.2±0.4	64.0±1.0	64.6±1.0	66.9±0.9	51.8±1.2	52.3±1.1	52.9±1.0	42.7±1.3	38.9±1.4	40.6±1.2
	MERIT	81.3±0.4	64.1±0.9	64.8±1.0	67.0±0.9	51.9±1.2	52.5±1.2	53.2±1.0	43.2±1.4	39.2±1.3	41.0±1.3
	ARIEL	81.4±0.3	64.5±1.0	65.2±0.9	67.4±0.8	52.3±1.3	52.8±1.1	53.5±1.1	43.4±1.2	39.4±1.2	41.2±1.2
	SFA	81.5±0.5	64.3±1.1	65.0±1.0	67.3±0.9	52.0±1.3	52.7±1.2	53.3±1.0	43.0±1.5	39.1±1.4	40.8±1.2
	Sel-Cl	81.3±0.4	64.4±0.9	65.1±1.0	67.2±0.9	52.1±1.2	52.6±1.1	53.1±0.9	42.9±1.3	39.0±1.4	40.7±1.3
	Jo-SRC	81.2±0.4	64.6±1.0	65.2±0.8	67.1±0.9	52.2±1.2	52.8±1.1	53.4±1.1	43.1±1.4	39.3±1.5	40.9±1.2
	GRAND+	81.6±0.3	64.7±1.1	65.3±0.9	67.6±0.8	52.4±1.0	52.9±1.1	53.6±1.0	43.2±1.4	39.5±1.3	41.1±1.1
	GFSA	81.0±0.4	63.6±1.0	64.2±1.0	66.6±0.9	51.5±1.2	52.0±1.1	52.6±1.0	42.4±1.3	38.6±1.4	40.3±1.2
	HONGAT	81.4±0.4	64.2±1.1	64.8±1.0	67.0±0.9	51.8±1.2	52.3±1.1	53.0±1.0	42.7±1.3	38.8±1.3	40.5±1.2
	CGNN	81.3±0.4	64.3±1.0	64.9±0.9	66.9±0.9	51.9±1.1	52.4±1.1	53.1±1.0	42.8±1.4	38.9±1.5	40.6±1.2
	CRGNN	81.5±0.5	64.5±1.0	65.2±1.0	67.1±0.9	52.2±1.3	52.8±1.2	53.3±1.1	43.1±1.3	39.2±1.3	41.0±1.1
	ProCon	81.8±0.3	64.7±1.1	65.2±0.9	67.8±0.8	52.3±1.0	52.8±1.1	53.8±1.0	43.3±1.2	39.5±1.3	41.4±1.2
	KCR-GCL	**82.6±0.4**	**67.8±1.3**	**69.9±1.0**	**70.7±1.2**	**55.3±1.4**	**57.2±1.3**	**56.9±1.1**	**45.9±1.4**	**42.6±1.4**	**44.3±1.2**
Amazon-Computers	GCN	87.2±0.5	61.9±1.2	63.8±1.1	65.8±1.3	47.1±1.4	48.4±1.2	50.1±1.0	37.7±1.7	35.4±1.6	36.8±1.5
	S^2GC	87.6±0.4	62.5±1.0	64.2±1.2	66.4±1.1	47.9±1.3	49.1±1.3	50.9±1.2	38.2±1.6	35.9±1.5	37.5±1.4
	GCE	87.9±0.6	62.3±1.1	64.1±1.0	66.1±1.2	47.5±1.4	48.6±1.2	50.5±1.2	38.0±1.5	35.6±1.6	37.0±1.4
	NRGNN	87.8±0.4	63.9±1.0	65.6±1.0	67.2±1.1	49.1±1.3	50.3±1.1	51.8±1.1	39.1±1.4	36.4±1.5	38.0±1.4
	RTGNN	88.0±0.5	64.1±1.0	65.8±1.0	67.4±0.9	49.4±1.2	50.7±1.2	52.1±1.0	39.2±1.4	36.6±1.3	38.3±1.2
	SUGRL	88.2±0.5	64.3±1.1	65.9±1.0	67.6±1.0	49.5±1.2	50.7±1.1	52.2±1.1	39.2±1.4	36.6±1.3	38.4±1.3
	MERIT	88.3±0.4	64.4±1.1	66.0±1.0	67.6±0.9	49.6±1.2	50.8±1.2	52.3±1.1	39.4±1.5	36.8±1.3	38.6±1.2
	ARIEL	88.4±0.4	64.5±1.0	66.2±0.9	67.9±1.0	49.8±1.1	51.0±1.1	52.6±1.1	39.6±1.3	36.9±1.4	38.8±1.2
	SFA	88.5±0.5	64.3±1.1	66.1±1.0	67.7±1.0	49.7±1.2	50.9±1.1	52.5±1.2	39.5±1.3	36.8±1.2	38.7±1.3
	Sel-Cl	88.2±0.6	64.6±0.9	66.3±1.1	67.8±1.0	49.9±1.1	51.1±1.1	52.7±1.2	39.6±1.3	36.9±1.3	38.9±1.2
	Jo-SRC	88.1±0.4	64.4±1.0	66.1±0.9	67.5±0.9	49.5±1.1	50.8±1.0	52.3±1.1	39.3±1.4	36.7±1.3	38.5±1.3
	GRAND+	88.6±0.4	64.7±0.9	66.5±0.9	68.0±1.0	50.1±1.0	51.3±1.0	52.9±1.0	39.8±1.3	37.0±1.3	39.0±1.1
	GFSA	87.9±0.5	64.0±1.1	65.6±1.0	67.3±1.0	49.2±1.2	50.4±1.1	51.9±1.1	38.9±1.4	36.4±1.3	38.1±1.3
	HONGAT	88.3±0.5	64.0±1.0	65.8±1.0	67.4±1.0	49.2±1.1	50.5±1.1	52.0±1.0	39.0±1.4	36.5±1.4	38.2±1.2
	CGNN	88.4±0.5	64.2±1.0	65.9±1.0	67.6±1.0	49.4±1.1	50.7±1.1	52.2±1.1	39.2±1.4	36.6±1.3	38.4±1.3
	CRGNN	88.5±0.4	64.4±0.9	66.2±0.9	67.8±0.9	49.6±1.1	50.9±1.0	52.4±1.1	39.5±1.3	36.8±1.2	38.7±1.1
	ProCon	88.5±0.4	64.7±0.9	66.7±0.9	68.0±1.0	50.0±1.0	51.5±1.0	52.8±1.0	39.7±1.3	37.2±1.3	39.0±1.1
	KCR-GCL	**89.6±0.5**	**67.6±1.4**	**69.4±1.1**	**70.1±1.3**	**53.4±1.4**	**54.8±1.3**	**54.5±1.3**	**43.2±1.4**	**40.1±1.5**	**41.8±1.4**
Amazon-Photos	GCN	89.9±0.4	63.8±1.1	64.9±0.9	66.5±1.0	48.7±1.2	49.8±1.1	50.9±1.1	39.5±1.4	36.1±1.3	37.4±1.2
	S^2GC	90.3±0.5	64.5±1.0	65.5±1.0	67.2±1.0	49.5±1.1	50.6±1.0	51.7±1.1	39.9±1.4	36.6±1.3	37.9±1.3
	GCE	90.5±0.4	64.2±1.1	65.4±0.9	67.0±1.0	49.2±1.2	50.3±1.0	51.3±1.1	39.7±1.3	36.4±1.2	37.7±1.2
	NRGNN	90.6±0.3	65.3±1.0	66.3±1.0	68.0±1.0	50.3±1.1	51.4±1.0	52.6±1.0	40.8±1.3	37.1±1.2	38.6±1.2
	RTGNN	90.8±0.4	65.6±0.9	66.5±1.0	68.2±0.9	50.6±1.0	51.7±1.0	52.9±1.1	41.1±1.2	37.3±1.2	38.8±1.1
	SUGRL	90.9±0.4	65.7±0.9	66.7±1.0	68.3±1.0	50.7±1.0	51.8±1.1	53.0±1.0	41.2±1.2	37.3±1.2	38.9±1.1
	MERIT	91.0±0.5	65.9±1.0	66.7±0.9	68.4±1.0	50.8±1.1	51.9±1.0	53.1±1.0	41.3±1.2	37.4±1.2	39.0±1.2
	ARIEL	91.1±0.5	66.0±1.0	66.9±1.0	68.6±1.0	51.1±1.0	52.1±1.1	53.3±1.0	41.5±1.3	37.6±1.3	39.2±1.1
	SFA	91.2±0.4	65.8±0.9	66.8±1.0	68.5±0.9	50.9±1.0	52.0±1.0	53.2±1.0	41.4±1.2	37.5±1.1	39.1±1.2
	Sel-Cl	90.9±0.5	66.1±0.9	67.0±0.9	68.7±0.9	51.2±1.0	52.2±1.0	53.4±1.0	41.6±1.2	37.7±1.1	39.3±1.1
	Jo-SRC	90.8±0.4	65.9±0.9	66.8±0.9	68.4±1.0	51.0±1.0	52.0±1.0	53.2±1.1	41.3±1.2	37.4±1.2	39.0±1.2
	GRAND+	91.3±0.5	66.2±1.0	67.1±0.9	68.8±1.0	51.3±1.0	52.3±1.0	53.5±1.0	41.7±1.2	37.8±1.2	39.4±1.1
	GFSA	90.7±0.4	65.4±0.9	66.4±1.0	68.0±1.0	50.4±1.0	51.5±1.1	52.7±1.0	40.9±1.2	37.1±1.2	38.7±1.1
	HONGAT	91.0±0.4	65.5±1.0	66.6±0.9	68.2±1.0	50.5±1.0	51.6±1.0	52.8±1.0	41.0±1.2	37.2±1.2	38.7±1.2
	CGNN	91.1±0.4	65.7±0.9	66.7±1.0	68.3±1.0	50.7±1.0	51.8±1.1	53.0±1.0	41.2±1.2	37.3±1.2	38.9±1.1
	CRGNN	91.2±0.4	65.9±1.0	66.9±1.0	68.5±1.0	50.9±1.1	52.1±1.1	53.3±1.1	41.4±1.2	37.5±1.2	39.1±1.2
	ProCon	91.2±0.5	66.4±1.0	67.0±0.9	68.8±1.0	51.5±1.0	52.2±1.0	53.5±1.0	41.9±1.2	37.8±1.2	39.3±1.1
	KCR-GCL	**92.2±0.5**	**69.2±1.3**	**70.8±1.1**	**71.7±1.1**	**54.5±1.4**	**55.8±1.3**	**55.3±1.3**	**44.2±1.4**	**40.8±1.4**	**42.1±1.3**

B Additional Experiment Results

B.1 Complete Node Classification Results

Table 8 shows the average classification accuracy and standard deviation across 10 runs on the Cora, Citeseer, PubMed, and ogbn-arxiv datasets, comparing KCR-GCL with the strongest baselines. Moreover, Table 9 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 KCR-GCL consistently achieves the best performance across all datasets and noise levels.

B.2 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 contrast, KCR-GCL 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 KCR-GCL, as well as an additional two-layer transductive GCN classifier. The results with different types of classifiers are shown in Table 10. It is observed that KCR-GCL still outperforms SOTA GCL methods even when the transductive classifiers are employed.

Table 10: Comparison for node classification accuracy (%) by inductive linear classifier, transductive two-layer GCN classifier, and transductive classifier used in KCR-GCL. 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)	83.4±0.5	56.4±1.1	67.4±1.2	67.5±0.9	46.8±1.1	55.2±1.1	45.2±1.2	28.0±1.2	38.1±1.2	33.8±1.4
SUGRL + transductive GCN	83.3±0.6	56.2±1.3	67.5±1.5	67.3±1.2	47.0±1.1	55.1±1.1	45.4±1.2	28.0±1.2	38.0±1.2	34.0±1.4
SUGRL + linear transductive classifier	83.6±0.7	56.8±1.3	67.7±1.0	67.4±1.1	47.2±1.1	55.5±1.1	45.7±1.2	28.4±1.2	38.3±1.2	34.1±1.4
MERIT (original, inductive classifier)	83.1±0.5	56.0±0.8	67.0±0.8	67.1±0.9	46.7±1.3	54.7±1.3	45.0±1.4	27.7±1.3	38.5±1.3	33.5±0.9
MERIT + transductive GCN	83.1±0.7	56.2±1.1	66.8±1.3	67.2±1.4	46.6±1.3	54.9±1.5	45.1±1.6	27.6±1.2	38.2±1.4	33.7±1.3
MERIT + linear transductive classifier	83.3±0.3	56.2±1.4	67.3±1.2	67.3±1.1	46.6±1.5	54.6±1.6	45.3±1.7	28.0±1.6	38.6±1.1	33.6±1.4
SFA (original, inductive classifier)	83.9±1.0	56.4±1.1	67.7±1.3	67.6±1.5	47.3±1.4	54.9±1.4	45.7±1.4	28.2±1.6	38.9±1.3	34.4±1.7
SFA + transductive GCN	83.7±1.3	56.5±1.1	67.3±1.7	67.3±1.8	47.4±1.6	55.1±1.5	45.3±1.8	27.7±1.6	38.9±1.5	34.3±1.9
SFA + linear transductive classifier	84.1±1.5	56.6±1.3	67.8±1.4	67.9±1.4	47.7±1.5	55.2±1.2	45.6±1.6	28.4±1.7	39.1±1.5	34.8±1.9
LR-GCL	85.8±0.6	58.9±1.1	71.3±0.7	69.5±1.1	49.2±1.1	58.7±1.3	47.7±1.2	30.6±1.2	41.9±1.2	36.3±1.1
KCR-GCL	**86.1±0.6**	**60.2±1.1**	**72.4±0.7**	**70.8±1.1**	**51.0±1.1**	**60.5±1.3**	**49.2±1.2**	**32.9±1.2**	**43.6±1.2**	**38.2±1.1**

B.3 Sensitivity Analysis on the Hyperparameters

We perform a sensitivity analysis on the weighting parameter τ associated with the TNN. This analysis is conducted using KCR-GCL 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 11. The highest accuracy is achieved at $\tau = 0.5$, though KCR-GCL 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.

In addition, we conduct an ablation study on the hyperparameter M, which is the maximum power in the KCR self-attention. As discussed in Section 4.1, larger values of M allow the attention matrix $\mathbf{B}$ to incorporate higher-order feature propagation through polynomial kernel expansions. We vary M from 1 to 9, and the corresponding classification accuracies are also reported in Table 11. The results indicate that KCR-GCL is robust to the choice of M, with the best performance observed when $M = 3$.

We also perform an ablation study to examine the influence of the rank parameter $r_0 = \lceil \gamma \min \{N, d\} \rceil$ in the regularization term employed in the training loss of the KCR-GCL. Table 12 shows the classification performance of LR-GCL under various choices of r_0. The results indicate that KCR-GCL consistently

Table 11: Sensitivity analysis on the weighting parameter τ for the TNN and the maximum power M in the KCR self-attention. The study is performed using KCR-GCL on the Coauthor-CS dataset for semi-supervised node classification accuracy (%) 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	58.8	59.4	59.3	59.1	59.4	59.0	59.1	59.0	58.9
M	1	2	3	4	5	6	7	8	9
Accuracy	59.0	59.2	59.4	59.3	59.4	59.2	59.3	58.7	58.8

maintains near-optimal accuracy across a wide range of rank settings, particularly when γ is chosen within the interval 0.1 to 0.3.

Table 12: Ablation study on the value of rank r_0 in the optimization problem (1) 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. 'Asy' and 'Sy' in this table denote the asymmetric label noise and symmetric label noise.

γ	Noise Type						
	0	40		60		80	
	-	Asy	Sy	Asy	Sy	Asy	Sy
0.1	-0.2	-0.1	-0.2	-0.2	-0.1	-0.1	-0.0
0.2	-0.0	-0.0	-0.0	-0.0	-0.0	-0.0	-0.0
0.3	-0.0	-0.0	-0.1	-0.2	-0.1	-0.0	-0.1
0.4	-0.1	-0.3	-0.2	-0.1	-0.2	-0.2	-0.2
0.5	-0.1	-0.2	-0.3	-0.3	-0.3	-0.1	-0.2
0.6	-0.3	-0.2	-0.2	-0.3	-0.2	-0.2	-0.3
0.7	-0.3	-0.4	-0.3	-0.4	-0.4	-0.4	-0.5
0.8	-0.2	-0.5	-0.6	-0.6	-0.6	-0.7	-0.7
0.9	-0.4	-0.4	-0.5	-0.7	-0.8	-0.8	-0.6
1.0	-0.4	-0.4	-0.7	-0.7	-0.8	-1.0	-0.8
optimal	85.8	58.9	71.3	49.2	58.7	30.6	41.9

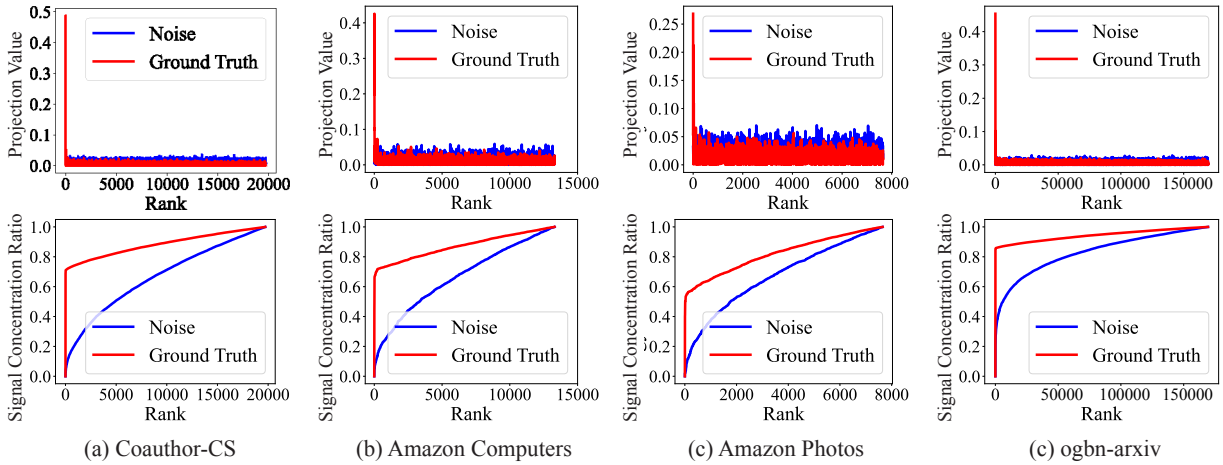
B.4 Eigen-Projection and Concentration Entropy Analysis on Additional Datasets

Figure 2 illustrates the eigen-projection visualizations and corresponding signal concentration ratios for Coauthor-CS, Amazon-Computers, Amazon-Photos, and ogbn-arxiv. We also investigate the presence of the Low Frequency Property (LFP) in heterophilic graph benchmarks, Texas and Chameleon (Pei et al., 2020), through eigen-projection plots and signal concentration ratio analysis, as illustrated in Figure 3. The findings reveal that the LFP persists in heterophilic settings, similar to homophilic graphs. The analysis is performed under asymmetric label noise with a corruption rate of 60%. When setting the rank parameter to $0.2 \min\{N, d\}$, the corresponding concentration entropy scores are observed to be 0.762 for Chameleon and 0.725 for Texas.

Study of the Low Frequency Property (LFP) for Attribute Noise. We study how the information from the ground-truth labels is distributed across different eigenvectors of the feature gram matrix $\mathbf{K_F}$ when the feature $\mathbf{F}$ is learned from the graph with attribute noise. It is noted that the observed label $\mathbf{Y} \in \mathbb{R}^{N \times C}$ is the clean ground-truth label without any noise in this setting. Following Figure 1 in Section 4.1, we compute the eigen-projection score of the label $\mathbf{Y}$ on the eigenvectors of the gram matrix $\mathbf{K_F}$ and the corresponding signal concentration ratios. Figure 4 illustrates that the ground-truth label signals are primarily concentrated on the leading eigenvectors of $\mathbf{K_F}$, even when the feature $\mathbf{F}$ is learned from the graph with attribute noise. The above observation motivates learning low-rank features for node classification with attribute noise.

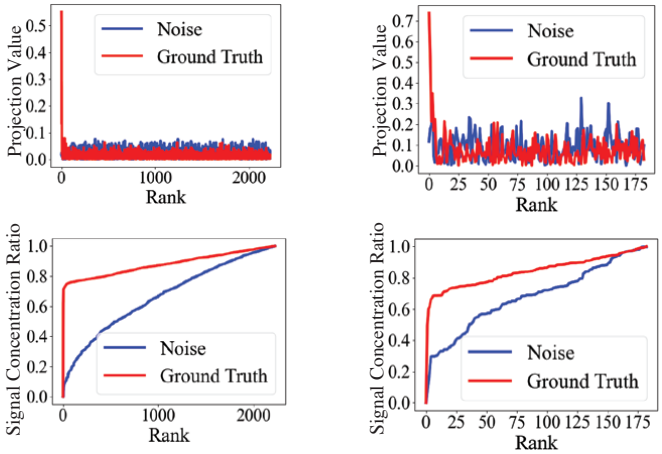
B.5 Statistical Significance Analysis

In this section, we compute the p-values from paired t-tests comparing the performance of KCR-GCL and its ablation model, LR-GCL, against the strongest baseline methods to assess the statistical significance of the observed improvements. As shown in Table 13, the p-values for both KCR-GCL and LR-GCL remain



(a) Coauthor-CS (b) Amazon Computers (c) Amazon Photos (c) ogbn-arxiv

Figure 2: Eigen-projection (first row) and energy concentration (second row) on Coauthor-CS, Amazon-Computers, Amazon-Photos, and ogbn-arxiv. 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.



(a) Chameleon (b) Texas

Figure 3: Eigen-projection (first row) and signal concentration ratio (second row) on Chameleon and Texas. 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.

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.6 Evaluation on Heterophilic Graphs

We evaluate the performance of KCR-GCL on semi-supervised node classification tasks involving two widely used heterophilic graph benchmarks, Texas and Chameleon (Pei et al., 2020). To begin, we illustrate the LFP on Texas and Chameleon in Figure 3 in Section B.4. We adopt TEDGCN (Yan et al., 2023), a GNN tailored for heterophilic graphs, as the encoder backbone for KCR-GCL. As shown in Table 14, KCR-GCL yields substantial improvements over the base TEDGCN model, demonstrating the benefits of reducing kernel complexity under noisy and heterophilic conditions.

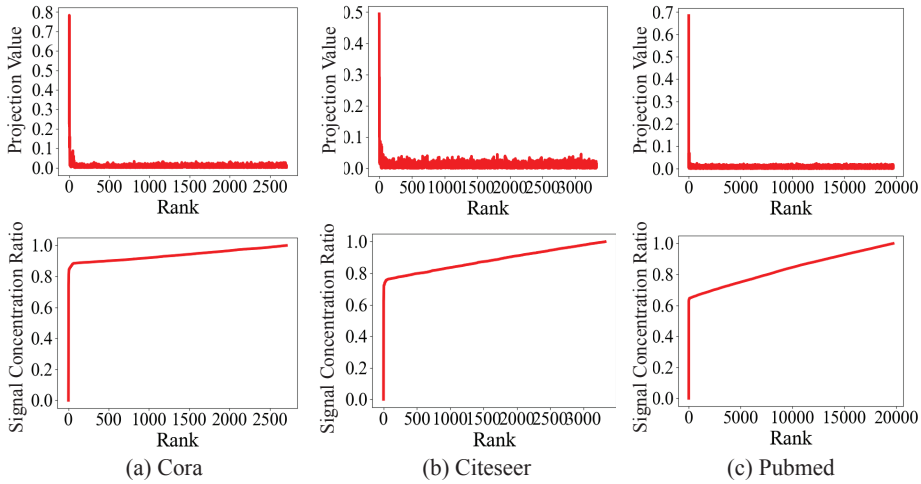


Figure 4: Eigen-projection (first row) and signal concentration ratio (second row) on Cora, Citeseer, and Pubmed, as the illustration of the Low Frequency Property (LFP). The study in this figure is performed on graphs consisting of attribute noise with a noise level of 60%. By the rank $r = 0.2 \min\{N, d\}$, the signal concentration ratio of $\tilde{\mathbf{Y}}$ for Cora, Citeseer, and Pubmed are 0.815, 0.785, and 0.689 respectively.

Table 13: P-values of the t-tests for LR-GCL and KCR-GCL 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	LR-GCL	0.038	0.024	0.021	0.035	0.021	0.041	0.025	0.028	0.031
	KCR-GCL	0.027	0.022	0.018	0.018	0.038	0.035	0.031	0.023	0.030
Citeseer	LR-GCL	0.043	0.035	0.022	0.021	0.030	0.041	0.027	0.024	0.022
	KCR-GCL	0.037	0.038	0.019	0.027	0.025	0.040	0.035	0.037	0.030
PubMed	LR-GCL	0.028	0.043	0.030	0.026	0.027	0.043	0.036	0.040	0.042
	KCR-GCL	0.025	0.030	0.026	0.023	0.024	0.041	0.033	0.035	0.037
Coauthor-CS	LR-GCL	0.041	0.032	0.036	0.043	0.044	0.040	0.027	0.037	0.042
	KCR-GCL	0.036	0.030	0.034	0.041	0.042	0.039	0.025	0.033	0.036
ogbn-arxiv	LR-GCL	0.040	0.032	0.036	0.043	0.044	0.040	0.027	0.037	0.042
	KCR-GCL	0.036	0.030	0.034	0.041	0.042	0.039	0.025	0.033	0.036
Wiki-CS	LR-GCL	0.044	0.035	0.033	0.028	0.034	0.041	0.036	0.040	0.042
	KCR-GCL	0.040	0.036	0.031	0.026	0.029	0.039	0.034	0.038	0.040
Amazon-Computers	LR-GCL	0.033	0.043	0.034	0.031	0.034	0.041	0.036	0.040	0.042
	KCR-GCL	0.031	0.038	0.030	0.028	0.030	0.038	0.034	0.037	0.040
Amazon-Photos	LR-GCL	0.040	0.041	0.038	0.043	0.044	0.040	0.027	0.037	0.042
	KCR-GCL	0.037	0.039	0.036	0.041	0.042	0.039	0.025	0.033	0.036
Texas	LR-GCL	0.038	0.024	0.021	0.035	0.021	0.041	0.025	0.028	0.031
	KCR-GCL	0.027	0.022	0.018	0.018	0.038	0.035	0.031	0.023	0.030
Chameleon	LR-GCL	0.044	0.035	0.033	0.028	0.034	0.041	0.036	0.040	0.042
	KCR-GCL	0.040	0.036	0.031	0.026	0.029	0.039	0.034	0.038	0.040

Table 14: Comparison for node classification accuracy (%) on Texas and Chameleon.

Dataset	Methods	Noise Type									
		0	40			60			80		
		-	Asymmetric	Symmetric	Attribute	Asymmetric	Symmetric	Attribute	Asymmetric	Symmetric	Attribute
Texas	TEDGCN	77.1±2.5	52.5±2.3	52.8±1.8	54.1±2.2	40.2±1.6	41.8±1.9	44.5±2.1	31.2±1.5	32.8±1.7	34.1±2.0
	KCR-GCL	**78.5±1.8**	**55.6±1.6**	**56.3±1.3**	**57.6±1.5**	**45.1±1.2**	**45.2±1.4**	**47.2±1.6**	**33.8±1.0**	**36.7±1.2**	**38.2±1.4**
Chameleon	TEDGCN	56.9±0.9	38.2±2.1	40.1±1.8	42.5±2.0	29.8±1.7	31.5±1.9	32.8±2.2	22.5±1.6	24.1±1.8	25.4±2.1
	KCR-GCL	**58.5±0.8**	**41.2±1.6**	**44.4±1.3**	**45.2±1.4**	**34.1±1.1**	**35.2±1.3**	**36.1±1.5**	**26.2±1.0**	**28.2±1.2**	**29.0±1.4**

B.7 Study on the Effectiveness of the Low-Rank Regularization on GCN

In this section, we conduct an ablation study by comparing KCR-GCL to an ablation model, Kernel Complexity Reduced GCN or KCR-GCN, which applies the KCR self-attention layer with low-rank regularization as that in the KCR-GCL objective (1) to a standard GCN (Kipf & Welling, 2017) for semi-supervised node

classification. Let $\mathbf{F}_{\text{GCN}} \in \mathbb{R}^{N\times d}$ be the output node features of a standard GCN with the same number of layers and hidden dimensions as the GCL encoder in KCR-GCL. Let $\mathbf{K}_{\mathbf{F}_{\text{GCN}}} = \mathbf{F}_{\text{GCN}}\mathbf{F}_{\text{GCN}}^\top \in \mathbb{R}^{N\times N}$ be the gram matrix of the node features $\mathbf{F}_{\text{GCN}}$. Let $\mathbf{B}_{\text{GCN}} \in \mathbb{R}^{N\times N}$ be the attention matrix of the KCR attention layer computed from $\mathbf{K}_{\mathbf{F}_{\text{GCN}}}$ in the same way as defined in Section 4.1 with learnable parameters $\boldsymbol{\alpha}_{\text{GCN}} \in \mathbb{R}^M$. Let $\boldsymbol{\theta}_{\text{GCN}}$ denote the network parameters of the GCN. The parameters of the KCR-GCN, including $\boldsymbol{\theta}_{\text{GCN}}$ and $\boldsymbol{\alpha}_{\text{GCN}}$, are trained jointly with a linear classifier applied to $\mathbf{B}_{\text{GCN}}\mathbf{F}_{\text{GCN}}$ by minimizing the following loss function,

$$\begin{aligned}
&\min_{\boldsymbol{\theta}_{\text{GCN}},\boldsymbol{\alpha}_{\text{GCN}}} \mathcal{L}_{\text{KCR-GCN}}(\boldsymbol{\theta}_{\text{GCN}}, \boldsymbol{\alpha}_{\text{GCN}}, \mathbf{W}_{\text{GCN}}) \\
&\qquad = \mathcal{L}_{\text{CE}}(\boldsymbol{\theta}_{\text{GCN}}, \boldsymbol{\alpha}_{\text{GCN}}) + \tau \|\mathbf{K}_{\mathbf{F}_{\text{GCN}}}(\boldsymbol{\theta}_{\text{GCN}}, \boldsymbol{\alpha}_{\text{GCN}})\|_{r_0}, \\
&\text{s.t. } \mathbf{B}_{ij} = 0 \text{ if } \mathbf{A}_{ij} = 0, \forall i,j \in [N],
\end{aligned} \tag{4}$$

where

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

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, M, and r_0, are selected by cross-validation following the same settings as in Section A of the appendix.

The comparisons between KCR-GCL and KCR-GCN 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 KCR self-attention layer and low-rank regularization. As shown in Table 15, KCR-GCN consistently outperforms the GCN baseline across all noise settings, demonstrating that the KCR self-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, KCR-GCL, which combines the KCR self-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, KCR-GCL outperforms KCR-GCN by 4.2% and the GCL baseline by 4.0% in node classification accuracy on Citeseer with asymmetric label noise at a noise level of 60%.

Table 15: Ablation results for comparison of node classification accuracy (%) between KCR-GCN and KCR-GCL under different noise types. The noise level is fixed at 60% for all noise settings.

Dataset	Noise Type	GCN	KCR-GCN	GCL	KCR-GCL
Cora	Asymmetric	40.5	44.2	47.0	**51.0**
Cora	Symmetric	51.7	54.0	55.6	**60.5**
Cora	Attribute	43.9	46.1	45.8	**49.2**
Citeseer	Asymmetric	35.1	38.8	39.0	**43.0**
Citeseer	Symmetric	34.1	38.7	42.5	**47.0**
Citeseer	Attribute	37.2	39.5	39.0	**42.3**

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

$$\begin{aligned}
\mathbf{W}^{(t)} &= \mathbf{W}^{(t-1)} - \eta\left[\mathbf{F}\right]_{\mathcal{L}}^\top \left[\mathbf{F}\mathbf{W}^{(t-1)} - \mathbf{Y}\right]_{\mathcal{L}} \\
&= \mathbf{W}^{(t-1)} - \eta\left[\mathbf{F}\right]_{\mathcal{L}}^\top \left[\mathbf{F}\mathbf{W}^{(t-1)} - \tilde{\mathbf{Y}}\right]_{\mathcal{L}} + \eta\left[\mathbf{F}\right]_{\mathcal{L}}^\top \left[\mathbf{N}\right]_{\mathcal{L}}.
\end{aligned} \tag{5}$$

It follows by (5) that

$$[\mathbf{F}]_{\mathcal{L}} \mathbf{W}^{(t)} = [\mathbf{F}]_{\mathcal{L}} \mathbf{W}^{(t-1)} - \eta \mathbf{K}_{\mathcal{L},\mathcal{L}} \left[\mathbf{F}\mathbf{W}^{(t-1)} - \tilde{\mathbf{Y}}\right]_{\mathcal{L}} + \eta \left[\mathbf{K}_{\mathbf{F}}\right]_{\mathcal{L},\mathcal{L}} [\mathbf{N}]_{\mathcal{L}}, \tag{6}$$

where $\mathbf{K}_{\mathcal{L},\mathcal{L}} \coloneqq [\mathbf{F}]_{\mathcal{L}} [\mathbf{F}]_{\mathcal{L}}^{\top} \in \mathbb{R}^{m \times m}$. With $\mathbf{F}(\mathbf{W}, t) = \mathbf{F}\mathbf{W}^{(t)}$, it follows by (6) that

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

It follows from the above equality and the recursion that

$$\left[\mathbf{F}(\mathbf{W}, t) - \tilde{\mathbf{Y}}\right]_{\mathcal{L}} = -\left(\mathbf{I}_m - \eta \left[\mathbf{K}_{\mathbf{F}}\right]_{\mathcal{L},\mathcal{L}}\right)^t \left[\tilde{\mathbf{Y}}\right]_{\mathcal{L}} + \eta \left[\mathbf{K}_{\mathbf{F}}\right]_{\mathcal{L},\mathcal{L}} \sum_{t'=0}^{t-1} \left(\mathbf{I}_m - \eta \left[\mathbf{K}_{\mathbf{F}}\right]_{\mathcal{L},\mathcal{L}}\right)^{t'} [\mathbf{N}]_{\mathcal{L}} \tag{7}$$

We apply the sharp tranductive learning bounds in (Yang, 2023; 2025) to obtain the following bound for the test loss $\frac{1}{u}\left\|\left[\mathbf{F}(\mathbf{W}, t) - \tilde{\mathbf{Y}}\right]_{\mathcal{U}}\right\|_{\mathrm{F}}^2$:

$$\frac{1}{u}\left\|\left[\mathbf{F}(\mathbf{W}, t) - \tilde{\mathbf{Y}}\right]_{\mathcal{U}}\right\|_{\mathrm{F}}^2 \leq \frac{c_0}{m}\left\|\left[\mathbf{F}(\mathbf{W}, t) - \tilde{\mathbf{Y}}\right]_{\mathcal{L}}\right\|_{\mathrm{F}}^2 + c_0 \min_{0 \leq Q \leq N} r(u, m, Q) + \frac{c_0 x}{u}, \tag{8}$$

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 $\tau_0^2 = \max_{i \in [N]} [K_{\mathbf{F}}]_{ii}$.

It follows from (7) and (8) that for every $r_0 \in [N]$, we have

$$\begin{aligned}
&\frac{1}{u}\left\|\left[\mathbf{F}(\mathbf{W}, t) - \tilde{\mathbf{Y}}\right]_{\mathcal{U}}\right\|_{\mathrm{F}}^2 \leq \frac{c_0}{m}\left\|\left(\mathbf{I}_m - \eta \left[\mathbf{K}_{\mathbf{F}}\right]_{\mathcal{L},\mathcal{L}}\right)^t \left[\tilde{\mathbf{Y}}\right]_{\mathcal{L}}\right\|_{\mathrm{F}}^2 \\
&+ c_0 r_0 \left(\frac{1}{u} + \frac{1}{m}\right) + c_0 \left(\sqrt{\frac{\sum_{q=r_0+1}^{N} \widehat{\lambda}_q}{u}} + \sqrt{\frac{\sum_{q=r_0+1}^{N} \widehat{\lambda}_q}{m}}\right) + \frac{c_0 x}{u} \\
&\overset{①}{\leq} \frac{2c_0}{m}\left\|\left(\mathbf{I}_m - \eta \left[\mathbf{K}_{\mathbf{F}}\right]_{\mathcal{L},\mathcal{L}}\right)^t \left[\tilde{\mathbf{Y}}\right]_{\mathcal{L}}\right\|_{\mathrm{F}}^2 + \frac{2c_0}{m}\left\|\eta \left[\mathbf{K}_{\mathbf{F}}\right]_{\mathcal{L},\mathcal{L}} \sum_{t'=0}^{t-1} \left(\mathbf{I}_m - \eta \left[\mathbf{K}_{\mathbf{F}}\right]_{\mathcal{L},\mathcal{L}}\right)^{t'} [\mathbf{N}]_{\mathcal{L}}\right\|_{\mathrm{F}}^2 \\
&\quad + c_0 r_0 \left(\frac{1}{u} + \frac{1}{m}\right) + c_0 \sqrt{\|\mathbf{K}_{\mathbf{F}}\|_{r_0}} \left(\sqrt{\frac{1}{u}} + \sqrt{\frac{1}{m}}\right) + \frac{c_0 x}{u},
\end{aligned} \tag{9}$$

where ① follows from the Cauchy-Schwarz inequality, (7), and $\sum_{q=r_0+1}^{N} \widehat{\lambda}_q = \|\mathbf{K}_{\mathbf{F}}\|_{r_0}$. Equation (3) in Theorem 4.1 of the main paper then follows directly from (9). □