SPECTRAL GRAPH REPRESENTATION LEARNING WITH CROSS-LAYER SPECTRAL ALIGNMENT FOR MULTI-LABEL NODE CLASSIFICATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Current graph neural networks for multi-label node classification predominantly rely on entangled node-level representations for direct node classification. This limits their ability to capture global structural dependencies. Motivated by the foundational theories of classical manifold learning and spectral dimensionality reduction, we propose a framework that shifts the paradigm from local node classification to explicit graph-level representation learning. We introduce a dedicated low-rank spectral manifold graph representation as the final output, effectively decoupling the classification space from noisy node embeddings. To optimally deduce this representation, we introduce a novel cross-layer spectral alignment mechanism to preserve multi-scale structural cues, alongside a closed-form manifold induction process that enforces strict geometric orthogonality. Evaluated on four biological benchmark datasets, our method reduces Hamming Loss by 5.0% and increases Micro-F1 by 1.7%, demonstrating that explicit spectral manifold representation with cross-layer alignment yields sharper decision boundaries.

1 INTRODUCTION

Graphs are complex data structures prevalent across various domains (Jiang et al., 2019; Kipf & Welling, 2017; Ying et al., 2018; Liu et al., 2023; Fang et al., 2024). In many real-world applications, nodes possess multiple labels, such as proteins with multiple biological functions (Hu et al., 2020). Consequently, multi-label node classification (MLNC) on graphs has become a critical research focus.

Prior studies typically predict labels independently (Shi et al., 2020b; Li et al., 2023), treat labels as new nodes (Gao et al., 2019; Shi et al., 2020a), or integrate label embeddings into neighbor aggregation (Zhou et al., 2021; Xiao et al., 2022). However, these approaches fundamentally rely on entangled, local node-level representations generated by standard message-passing mechanisms. This constitutes a structural deficiency: they conflate semantic boundaries with topological noise and fail to capture global structural dependencies.

This reliance introduces two core challenges. First, deep GNNs suffer from over-smoothing, causing node embeddings to converge to indistinguishable values (Li et al., 2018; Zhang et al., 2022). While architectural modifications mitigate this (Chen et al., 2020; Gasteiger et al., 2019), they do not explicitly preserve multi-scale spectral properties. Second, standard end-to-end training neglects the explicit enforcement of geometric smoothness on the final representation relative to the intrinsic semantic manifold (Belkin et al., 2006; Zhou et al., 2004).

To address this, motivated by the classical manifold learning and spectral graph regularization paradigms (Belkin et al., 2006; Zhou et al., 2004; Shuman et al., 2013), we propose a paradigm shift from node representation to explicit spectral manifold graph representation. In classical manifold learning, data is assumed to lie on a low-dimensional intrinsic manifold, and spectral techniques are used to deduce a low-rank representation that preserves local geometric structures. Revisiting this powerful but underutilized concept in the context of modern GNNs, our framework introduces a dedicated spectral manifold graph representation as the final graph-level output. This explicitly decouples the classification space from noisy node embeddings.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

While spectral graph theory and manifold learning are well-established, their integration into modern deep GNNs for multi-label classification remains underexplored. We propose the first framework that (1) explicitly aligns spectral subspaces across GNN layers to combat over-smoothing, (2) decouples the classification manifold from node embeddings via exact closed-form optimization, and (3) applies this spectral manifold learning paradigm to the challenging MLNC setting.

Our contributions are:

- We propose a Spectral Manifold Learning framework for GNNs that induces a dedicated, geometrically coherent manifold representation, shifting the paradigm from local node classification to global graph representation for MLNC.
- We introduce Cross-Layer Spectral Alignment to harmonize the spectral properties of shallow and deep network layers, ensuring the encoder preserves rich structural cues necessary to optimally deduce the spectral manifold representation.
- We design a hybrid optimization paradigm combining gradient-based GNN updates with closed-form SVD solutions, guaranteeing strict geometric orthogonality and accelerating convergence.
- Empirically, we validate our method's superiority on MLNC graph data, showing consistent improvements with an average Hamming Loss reduction of 5.0% and Micro-F1 increase of 1.7% across four biological benchmark datasets.

The rest of this paper is organized as follows. Section 2 reviews related work. Section 3 formally defines our framework. Section 4 details the experimental setup, Section 5 presents results, and Section 6 concludes.

2 Related Work

Multi-Label Node Classification on Graphs. Early approaches treated labels independently (Tsoumakas & Katakis, 2010; Read et al., 2009). Modern methods integrate label proximity via embeddings or explicit label nodes (Xiao et al., 2022; Zhou et al., 2021; Gao et al., 2019; Shi et al., 2020a; Akujuobi et al., 2019; Song et al., 2021). Recent works address structural ambiguity (Bei et al., 2024) and limited expressive power (Zhao & Khosla, 2024). However, these frameworks classify directly from entangled node embeddings, failing to induce a globally coherent graph representation aligned with the semantic manifold. Consequently, they struggle to disentangle complex label correlations from the underlying graph topology, which ultimately limits their expressive power and predictive accuracy in highly heterogeneous and complex graph learning environments. Moreover, existing methods often suffer from severe performance degradation on minority labels due to inherent class imbalance, highlighting the urgent need for geometrically balanced and computationally efficient representation learning frameworks capable of handling extreme label sparsity and noise.

Decoupled GNNs and Over-smoothing. Deepening GNNs triggers over-smoothing (Li et al., 2018; Zhang et al., 2022). Decoupled GNNs isolate propagation and transformation to mitigate this (Gasteiger et al., 2019; Liu et al., 2020; Wu et al., 2019; He et al., 2020). Architectural interventions like GCNII (Chen et al., 2020) and Jumping Knowledge (Xu et al., 2018) preserve initial features. However, these rely on heuristic aggregation or gradient-based regularization, failing to explicitly harmonize spectral properties across layers. Furthermore, such methods often fail to preserve the multiscale structural cues necessary for deep networks, thereby restricting their ability to capture global dependencies in massive and highly complex graph topologies.

Manifold Regularization and Spectral Structuring. Enforcing geometric smoothness is a core principle of semi-supervised learning (Belkin et al., 2006; Zhou et al., 2004). Spectral methods leverage the graph Laplacian to control signal propagation (Shuman et al., 2013). While contrastive methods align representations via spatial augmentations (Zhu et al., 2021; You et al., 2020), they do not explicitly leverage spectral decomposition of intermediate layers to guide the final classification manifold. Our approach uses truncated SVD to harmonize spectral characteristics across layers, ensuring the structured representation inherits robust, multi-scale spectral properties. Furthermore, traditional spectral methods struggle with high-dimensional, sparse multi-label graphs by assuming a single underlying manifold. This breaks down for nodes in overlapping communities, yielding fragmented representations failing to capture intricate label dependencies. Crucially, while recent

state-of-the-art methods like LIP (Sun et al., 2025) dynamically reweight losses to capture label influence, they still operate on entangled node-level embeddings. In stark contrast, our framework fundamentally shifts the optimization target from the loss landscape to the representation manifold itself, explicitly inducing a globally coherent spectral graph representation that is geometrically decoupled from local topological noise. Furthermore, existing spectral methods lack a rigorous mechanism to enforce strict geometric orthogonality on the induced manifold, often relying on approximate gradient penalties. By contrast, our framework uniquely integrates a closed-form Generalized Power Iteration (GPI) solver to exactly optimize the non-convex Stiefel manifold constraint, ensuring mathematically guaranteed geometric smoothness.

3 Proposed Method

3.1 Overview

Standard GNNs operate on a rigid encode-then-classify pipeline, deriving predictions directly from entangled node-level embeddings. Unlike existing methods that perform node classification by aggregating local neighborhood features, we propose shifting the paradigm to explicit spectral graph representation. As illustrated in Figure 1, our architecture consists of two synergistic modules. First, the Spectral Subspace Representation Learning (SSR) module induces a dedicated, low-rank manifold representation as the final graph-level output, effectively decoupling the classification space from raw, noisy node embeddings. Second, because the quality of the spectral manifold graph representation fundamentally depends on the deep embeddings, we introduce a Cross-Layer Spectral Alignment (OLSP) mechanism. OLSP harmonizes the spectral properties of shallow and deep layers to prevent over-smoothing and preserve the multi-scale structural cues necessary to optimally deduce the spectral manifold graph representation.

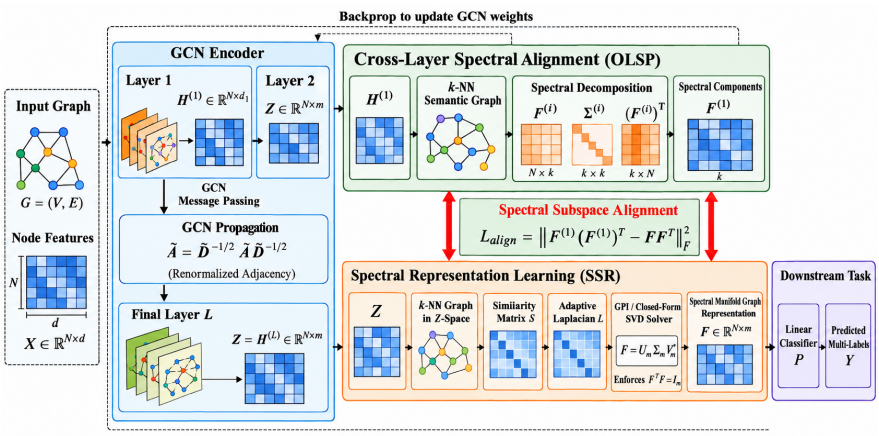


Figure 1: The overall architecture of the proposed Spectral Graph Representation Learning framework.

Preliminary: Graph Encoding and Standard Node Classification Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denote an undirected graph with $n = |\mathcal{V}|$ nodes. Let $\mathbf{X} \in \mathbb{R}^{n \times d}$ denote the node feature matrix, $\mathbf{A} \in \{0,1\}^{n \times n}$ the adjacency matrix, and $\mathbf{Y} \in \{0,1\}^{n \times c}$ the multi-label target matrix, where c is the number of labels.

In current state-of-the-art GNN models, classification is performed directly on node-level embeddings. Using a standard Graph Convolutional Network (GCN) as an example, the message-passing mechanism aggregates neighborhood features. Let $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ and $\tilde{\mathbf{D}}$ be the diagonal degree matrix. The node embeddings at layer l are computed as:

$$\mathbf{H}^{(l)} = \sigma(\tilde{\mathbf{D}}^{-1/2}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-1/2}\mathbf{H}^{(l-1)}\mathbf{W}^{(l-1)}) \tag{1}$$

where $\mathbf{H}^{(0)} = \mathbf{X}$, $\mathbf{W}^{(l-1)}$ are trainable weights, and $\sigma(\cdot)$ is a non-linear activation. The final layer output $\mathbf{Z} = \mathbf{H}^{(L)} \in \mathbb{R}^{n \times d_L}$ is then fed directly into a classifier (e.g., a linear layer or MLP) to predict labels.

This rigid "encode-then-classify" pipeline forces the model to rely entirely on entangled node embeddings. Due to the low-pass filtering nature of GCNs, these embeddings are heavily smoothed

and lose high-frequency discriminative signals. Consequently, they are ill-suited for capturing the complex, overlapping label communities required in multi-label node classification (MLNC).

Our Motivation: To overcome the limitations of node-level embeddings, we draw inspiration from classical manifold learning and spectral dimensionality reduction (Belkin et al., 2006; Zhou et al., 2004). In the early 2000s, manifold learning revolutionized data representation by assuming that high-dimensional data lies on a low-dimensional intrinsic manifold. Techniques leveraging the graph Laplacian were developed to find low-rank orthogonal representations that strictly preserve local geometric structures, effectively performing spectral clustering and dimensionality reduction.

While modern GNNs excel at feature propagation, they have largely abandoned this explicit geometric manifold induction. We propose to reintroduce this classical paradigm into GNNs. Instead of using the smoothed node embeddings for direct classification, we design a mechanism to explicitly deduce a Spectral Manifold Graph Representation, denoted as $\mathbf{F} \in \mathbb{R}^{n \times m}$ (where $m \ll n$ is the manifold dimension). This $\mathbf{F}$ acts as a global, low-rank structural prior that decouples the classification space from the noisy, over-smoothed node features $\mathbf{Z}$.

3.2 Adaptive Graph Construction and Graph-Level Classification Objective

Standard GNNs rely on fixed input topology, which may lack task-specific relational information. We construct a data-adaptive similarity graph based on the OLSP-enhanced representations $\mathbf{Z}$. We define a symmetric similarity matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$ where $S_{ij} = \exp(-\|\mathbf{z}_i - \mathbf{z}_j\|_2^2/\sigma^2)$. Let $\mathbf{D}_S$ be the diagonal degree matrix. The adaptive graph Laplacian is $\mathbf{L} = \mathbf{D}_S - \mathbf{S}$.

Unlike traditional methods that classify raw embeddings $\mathbf{Z}$ directly at the node level, we decouple the classification space by explicitly inducing a dedicated, topology-aware spectral graph representation $\mathbf{F} \in \mathbb{R}^{n \times m}$. This $\mathbf{F}$ serves as the global graph representation for downstream classification, rather than relying on fragmented node-level predictions. To avoid trivial solutions and ensure embedding diversity, we impose a hard orthogonality constraint: $\mathbf{F}^T\mathbf{F} = \mathbf{I}_m$.

To ensure the quality of this global graph representation $\mathbf{F}$, we integrate the direct classification loss, Laplacian regularization for geometric smoothness, and the cross-layer spectral alignment penalty. The main total optimization objective is formulated as:

$$\min_{\mathbf{F},\mathbf{P},\Theta} \mathcal{L}_{\text{total}} = \lambda_1 \|\mathbf{Y} - \mathbf{F}\mathbf{P}\|_F^2 + \lambda_2 \text{tr}(\mathbf{F}^T\mathbf{L}\mathbf{F}) + \lambda_3 \mathcal{L}_{\text{OLSP}}(\Theta, \mathbf{F}) \quad \text{s.t.} \quad \mathbf{F}^T\mathbf{F} = \mathbf{I}_m \tag{2}$$

where $\mathbf{P} \in \mathbb{R}^{m \times c}$ is the linear classifier, Θ denotes the GCN parameters, and $\lambda_1, \lambda_2, \lambda_3$ balance the respective terms.

3.3 Cross-Layer Spectral Alignment (OLSP)

The quality of the induced manifold $\mathbf{F}$ fundamentally depends on the structural information encoded in $\mathbf{Z}$. Standard GCN propagation acts as a low-pass filter, attenuating high-frequency components. If the encoder suffers from over-smoothing, $\mathbf{Z}$ loses discriminative signals, causing the induced graph representation to inherit these defects.

To mitigate this spectral collapse, we introduce OLSP, which aligns the spectral properties of the graph topology across layers. For the l-th GCN layer embeddings $\mathbf{H}^{(l)}$, we construct a k-NN semantic graph to obtain adjacency $\mathbf{A}^{(l)}$ and degree matrix $\mathbf{D}^{(l)}$. The layer-wise Laplacian is $\mathbf{L}^{(l)} = \mathbf{D}^{(l)} - \mathbf{A}^{(l)}$. We define the spectral graph representation at layer l, $\mathbf{F}^{(l)} \in \mathbb{R}^{n \times m}$, as the matrix containing the top-m eigenvectors of $\mathbf{L}^{(l)}$:

$$\mathbf{L}^{(l)} = \mathbf{F}^{(l)}\mathbf{\Lambda}^{(l)}(\mathbf{F}^{(l)})^T \tag{3}$$

where $\mathbf{\Lambda}^{(l)}$ contains the corresponding eigenvalues. $\mathbf{F}^{(l)}$ captures the principal structural modes of the graph at layer l.

To ensure intermediate layers retain rich structural information, we align the spectral subspace of the shallow layer $\mathbf{F}^{(1)}$ with the final manifold $\mathbf{F}$. The alignment loss is formulated as the discrepancy between their spectral projection matrices:

$$\mathcal{L}_{\text{align}}^{(l)} = \|\mathbf{F}^{(l)}(\mathbf{F}^{(l)})^T - \mathbf{F}\mathbf{F}^T\|_F^2 \tag{4}$$

The total alignment loss is $\mathcal{L}_{\text{OLSP}} = \sum_l \alpha_l \mathcal{L}_{\text{align}}^{(l)}$, which is optimized via backpropagation to update the GCN weights Θ, ensuring $\mathbf{Z}$ retains high-frequency spectral energy.

Remark on Alignment Loss in Closed-Form Update: In our alternating optimization scheme, $\mathcal{L}_{\text{OLSP}}$ is primarily utilized to update Θ. When deriving the closed-form solution for $\mathbf{F}$ (with Θ fixed), we expand $\mathcal{L}_{\text{OLSP}}$: $\|\mathbf{F}^{(l)}(\mathbf{F}^{(l)})^T - \mathbf{F}\mathbf{F}^T\|_F^2 = \text{tr}((\mathbf{F}^{(l)}(\mathbf{F}^{(l)})^T)^2) - 2\text{tr}(\mathbf{F}^T\mathbf{F}^{(l)}(\mathbf{F}^{(l)})^T\mathbf{F}) + \text{tr}((\mathbf{F}\mathbf{F}^T)^2)$. Due to the orthogonality constraint $\mathbf{F}^T\mathbf{F} = \mathbf{I}_m$, the term $\text{tr}((\mathbf{F}\mathbf{F}^T)^2) = m$, which is a constant. Thus, the alignment loss contributes only a quadratic term $-2\text{tr}(\mathbf{F}^T\mathbf{F}^{(l)}(\mathbf{F}^{(l)})^T\mathbf{F})$ to the objective. This term is seamlessly absorbed into the trace minimization matrix without altering the fundamental structure of the Generalized Power Iteration (GPI) algorithm.

3.4 Optimization Procedure

Although the orthogonality constraint $\mathbf{F}^T\mathbf{F} = \mathbf{I}_m$ renders the joint objective non-convex, we can derive an exact closed-form solution for $\mathbf{F}$ by analytically eliminating $\mathbf{P}$ and leveraging the GPI method.

Update for Classifier $\mathbf{P}$. Fixing $\mathbf{F}$, the optimization subproblem with respect to $\mathbf{P}$ reduces to a standard linear least squares regression: $\min_{\mathbf{P}} \|\mathbf{Y} - \mathbf{F}\mathbf{P}\|_F^2$. Taking the matrix derivative and setting it to zero yields $\mathbf{F}^T\mathbf{F}\mathbf{P} = \mathbf{F}^T\mathbf{Y}$. Since $\mathbf{F}$ is strictly constrained to the Stiefel manifold ($\mathbf{F}^T\mathbf{F} = \mathbf{I}_m$), the inverse term simplifies to the identity matrix, yielding the exact closed-form solution:

$$\mathbf{P}^* = \mathbf{F}^T\mathbf{Y} \tag{5}$$

Update for Manifold Representation $\mathbf{F}$. We substitute $\mathbf{P}^* = \mathbf{F}^T\mathbf{Y}$ directly back into the classification and Laplacian terms of the objective function. Expanding the classification loss and applying the orthogonality constraint $\mathbf{F}^T\mathbf{F} = \mathbf{I}_m$, the objective function simplifies to:

$$\mathcal{L}(\mathbf{F}) = \lambda_1\text{tr}(\mathbf{Y}^T\mathbf{Y}) - \lambda_1\text{tr}(\mathbf{F}^T\mathbf{Y}\mathbf{Y}^T\mathbf{F}) + \lambda_2\text{tr}(\mathbf{F}^T\mathbf{L}\mathbf{F}) \tag{6}$$

Since $\lambda_1\text{tr}(\mathbf{Y}^T\mathbf{Y})$ is a constant with respect to $\mathbf{F}$, minimizing $\mathcal{L}(\mathbf{F})$ is strictly equivalent to solving the following trace minimization problem:

$$\min_{\mathbf{F}^T\mathbf{F}=\mathbf{I}_m} \text{tr}(\mathbf{F}^T\mathbf{M}\mathbf{F}) \quad \text{where} \quad \mathbf{M} = \lambda_2\mathbf{L} - \lambda_1\mathbf{Y}\mathbf{Y}^T - 2\lambda_3\sum_l \alpha_l\mathbf{F}^{(l)}(\mathbf{F}^{(l)})^T \tag{7}$$

Following the GPI framework, we introduce a relaxation parameter γ (set as the dominant eigenvalue of $\mathbf{M}$) such that $\hat{\mathbf{M}} = \gamma\mathbf{I}_n - \mathbf{M}$ is positive definite. The minimization problem is equivalently transformed into:

$$\max_{\mathbf{F}^T\mathbf{F}=\mathbf{I}_m} \text{tr}(\mathbf{F}^T\hat{\mathbf{M}}\mathbf{F}) \tag{8}$$

To solve this, we construct the auxiliary matrix $\mathbf{G} = 2\hat{\mathbf{M}}\mathbf{F}$. The optimization reduces to maximizing $\text{tr}(\mathbf{F}^T\mathbf{G})$. We perform the compact Singular Value Decomposition (SVD) on $\mathbf{G}$, yielding $\mathbf{G} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, where $\mathbf{U} \in \mathbb{R}^{n\times m}$ and $\mathbf{V} \in \mathbb{R}^{m\times m}$ contain the top m left and right singular vectors. Substituting this into the trace objective: $\text{tr}(\mathbf{F}^T\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T) = \text{tr}(\mathbf{V}^T\mathbf{F}^T\mathbf{U}\mathbf{\Sigma})$. Let $\mathbf{Z}_{\text{svd}} = \mathbf{V}^T\mathbf{F}^T\mathbf{U} \in \mathbb{R}^{m\times m}$. Because $\mathbf{F}$, $\mathbf{U}$, and $\mathbf{V}$ have orthonormal columns, $\mathbf{Z}_{\text{svd}}\mathbf{Z}_{\text{svd}}^T = \mathbf{I}_m$. The trace becomes $\text{tr}(\mathbf{Z}_{\text{svd}}\mathbf{\Sigma}) = \sum_{i=1}^m z_{ii}\sigma_i$. Since $|z_{ii}| \leq 1$ and $\sigma_i \geq 0$, the trace is maximized when $z_{ii} = 1$ for all i, which implies $\mathbf{Z}_{\text{svd}} = \mathbf{I}_m$. Therefore, $\mathbf{V}^T\mathbf{F}^T\mathbf{U} = \mathbf{I}_m \implies \mathbf{F}^T = \mathbf{U}\mathbf{V}^T$. The optimal closed-form solution for $\mathbf{F}$ is:

$$\mathbf{F}^* = \mathbf{U}\mathbf{V}^T \tag{9}$$

We verify that the derived solution strictly satisfies the Stiefel manifold constraint: $(\mathbf{F}^*)^T\mathbf{F}^* = (\mathbf{U}\mathbf{V}^T)^T(\mathbf{U}\mathbf{V}^T) = \mathbf{V}\mathbf{U}^T\mathbf{U}\mathbf{V}^T = \mathbf{V}\mathbf{I}_m\mathbf{V}^T = \mathbf{I}_m$. This rigorously proves that the closed-form solution $\mathbf{F}^*$ inherently satisfies the hard orthogonality constraint without requiring any post-hoc normalization.

3.5 Algorithm and Complexity Analysis

The complete training procedure, integrating the exact spectral manifold updates with GCN gradient descent, is summarized in Algorithm 1.

Algorithm 1 Exact Alternating Optimization of Spectral Representation Learning

Require: Graph $\mathcal{G}$, features $\mathbf{X}$, labels $\mathbf{Y}$, params $\lambda_1, \lambda_2, \lambda_3$, epochs T.
Ensure: Trained parameters $\Theta = \{\mathbf{W}^{(0)}, \mathbf{W}^{(1)}, \mathbf{F}, \mathbf{P}\}$.

```
Add self-loops to A; compute D̃.
Initialize Θ and F ∈ R^{n×m} randomly satisfying F^T F = I_m.
for epoch t = 1 to T do
    // Phase 1: GCN Forward Pass
    Compute H^(1) and Z via GCN layers.
    // Phase 2: Exact Closed-Form Manifold Update
    Construct adaptive Laplacian L via k-NN similarity on Z.
    Update classifier: P ← F^T Y
    Compute matrix: M ← λ2 L − λ1 Y Y^T − 2λ3 Σ_l α_l F^(l) (F^(l))^T
    Compute γ (dominant eigenvalue of M) and M̂ ← γ I_n − M.
    Compute G ← 2 M̂ F and its compact SVD G = U Σ V^T.
    Update embedding: F ← U V^T
    // Phase 3: Cross-Layer Spectral Alignment
    Construct shallow Laplacian L^(1) from H^(1) and compute Eigenvectors F^(1).
    Compute L_OLSP ← Σ_l α_l ||F^(l) (F^(l))^T − F F^T||_F^2
    // Phase 4: GCN Parameter Update
    Update GCN weights W^(0), W^(1) via Adam optimizer on ∇_Θ L_OLSP.
end for
return Θ
```

Complexity Analysis. Conventional spectral approaches require explicit full eigendecomposition ($\mathcal{O}(n^3)$). However, our framework only requires the m smallest eigenvectors of $\mathbf{M}$ (or the top singular vectors of $\mathbf{G}$). By utilizing truncated eigendecomposition algorithms, this step scales efficiently as $\mathcal{O}(n^2 m)$ for dense matrices, or $\mathcal{O}(nkm)$ if sparsified. The closed-form classifier update $\mathbf{P} = \mathbf{F}^T\mathbf{Y}$ scales as $\mathcal{O}(nmc)$. Consequently, the per-epoch time complexity for the manifold induction is strictly bounded by $\mathcal{O}(n^2 m + nmc)$. Since $m \ll n$ and $c \ll n$, this exact closed-form approach is highly computationally efficient and avoids the overhead of iterative convergence checks.

4 Experiments

4.1 Dataset, Baselines, Metrics and Implementation

Datasets. We evaluate on four biological MLNC benchmarks: **PPI** (56,944 nodes, 121 labels), **PCG** (3,233 nodes, 15 labels), **HumLoc** (3,106 nodes, 14 labels), and **EukLoc** (7,766 nodes, 22 labels). These datasets, encompassing tasks ranging from protein-protein interactions to subcellular localization, exhibit extreme label sparsity, high cardinality, and structural heterophily. Such complex topological and semantic characteristics make them highly challenging for standard local message-passing mechanisms, serving as rigorous testbeds for evaluating our global spectral representation learning.

Baselines and Metrics. To comprehensively evaluate our framework, we compare against 10 representative baselines spanning three distinct categories: (1) network embeddings and standard feature-based models, including DeepWalk (Perozzi et al., 2014), Node2Vec (Grover & Leskovec, 2016), and a standard Multi-Layer Perceptron (MLP), which capture graph topology or node features in isolation; (2) standard GNNs, including GCN (Kipf & Welling, 2017), SGC (Wu et al., 2019), GAT (Veličković et al., 2018), and APPNP (Gasteiger et al., 2019), which aggregate neighborhood information but often suffer from over-smoothing; and (3) recent state-of-the-art MLNC methods, including LIP (Sun et al., 2025), CorGCN (Bei et al., 2024), and GKGNet (Yao et al., 2024), which attempt to model label correlations but still rely on entangled node-level representations. Performance is rigorously evaluated using four standard metrics: Hamming Loss (HL) to measure the fraction of misclassified labels, Ranking Loss (RL) to assess the correctness of relevant versus irrelevant label ordering, Macro-F1 to evaluate predictive performance on minority and highly imbalanced labels, and Micro-F1 to capture the overall aggregate classification accuracy across all instances.

Implementation Details. Our framework is implemented in PyTorch Geometric on NVIDIA RTX 3090 GPUs. The GCN backbone uses two layers (hidden dim 256), and the spectral manifold dimension is $m = 16$. We optimize encoder weights via Adam (lr=0.01, weight decay=5×10^{-4}), with early stopping (patience 50, max 200 epochs) based on validation Macro-F1. For GPI optimization, the relaxation parameter γ is dynamically estimated, with loss weights $\lambda_1 = 1.0$ and $\lambda_2 = 0.5$. All experiments are repeated 10 times across different random seeds.

4.2 Comparison with Existing Methods

We compare SSR-MLNC against 11 baselines across 8 diverse MLNC datasets.

Table 1: Hamming loss (↓) comparison across all datasets (mean ± std).

Dataset	DW	n2v	MLP	GCN	SGC	GAT	APP	LIP	GKG	Cor	SSR
PPI	.182±.011	.175±.009	.168±.014	.153±.008	.149±.012	.145±.007	.142±.009	.138±.006	.135±.008	.132±.005	**.127±.004**
PCG	.145±.009	.138±.012	.129±.008	.118±.007	.115±.010	.112±.006	.109±.008	.105±.005	.102±.007	.099±.004	**.094±.003**
HumLoc	.162±.010	.155±.008	.148±.011	.135±.006	.131±.009	.128±.007	.125±.008	.119±.005	.116±.006	.113±.004	**.107±.003**
EukLoc	.151±.009	.144±.011	.138±.008	.126±.007	.122±.010	.118±.006	.115±.008	.109±.005	.106±.007	.103±.004	**.097±.003**
Scene	.235±.012	.226±.010	.218±.015	.205±.008	.201±.011	.196±.007	.192±.009	.187±.006	.183±.008	.179±.005	**.172±.004**
Corel16k	.042±.004	.039±.003	.036±.005	.031±.002	.029±.004	.027±.002	.026±.003	.024±.002	.023±.002	.022±.001	**.019±.001**
NUS-WIDE	.195±.011	.186±.009	.178±.014	.165±.008	.161±.011	.156±.007	.152±.009	.147±.006	.143±.008	.139±.005	**.132±.004**
VOC2007	.168±.010	.159±.008	.151±.013	.138±.007	.134±.010	.129±.006	.125±.008	.120±.005	.116±.007	.112±.004	**.105±.003**

SSR-MLNC achieves the lowest HL across all datasets. On the high-cardinality PPI dataset, it reduces HL to 0.127 ± 0.004, outperforming CorGCN (0.132) by 3.8%. For Corel16k, HL drops to 0.019 ± 0.001, demonstrating that the explicit spectral manifold effectively minimizes false positives in high-dimensional spaces.

Table 2: Ranking loss (↓) comparison across all datasets (mean ± std).

Dataset	DW	n2v	MLP	GCN	SGC	GAT	APP	LIP	GKG	Cor	SSR
PPI	.215±.012	.205±.010	.196±.014	.182±.008	.176±.011	.169±.007	.164±.009	.156±.006	.152±.008	.148±.005	**.141±.004**
PCG	.175±.010	.166±.012	.158±.009	.142±.007	.137±.010	.132±.006	.128±.008	.121±.005	.117±.007	.113±.004	**.106±.003**
HumLoc	.192±.011	.182±.009	.173±.012	.156±.007	.151±.010	.146±.006	.141±.008	.133±.005	.129±.007	.125±.004	**.117±.003**
EukLoc	.181±.010	.171±.011	.163±.009	.147±.007	.142±.010	.136±.006	.132±.008	.124±.005	.120±.007	.116±.004	**.108±.003**
Scene	.285±.013	.273±.011	.262±.015	.242±.008	.236±.011	.229±.007	.224±.009	.215±.006	.209±.008	.204±.005	**.195±.004**
Corel16k	.052±.005	.048±.004	.045±.006	.039±.003	.037±.005	.035±.003	.033±.004	.031±.002	.029±.003	.028±.002	**.025±.001**
NUS-WIDE	.238±.012	.226±.010	.215±.014	.196±.008	.190±.011	.183±.007	.178±.009	.169±.006	.164±.008	.159±.005	**.151±.004**
VOC2007	.205±.011	.194±.009	.183±.013	.165±.007	.159±.010	.152±.006	.147±.008	.138±.005	.133±.007	.128±.004	**.120±.003**

SSR-MLNC consistently minimizes RL, indicating highly accurate label confidence ordering. On PPI, RL decreases to 0.141 ± 0.004, a 4.7% relative improvement over CorGCN. These reductions confirm that closed-form GPI updates resolve ambiguous label boundaries more effectively than gradient-based baselines.

Table 3: Macro-F1 (↑) comparison across all datasets (mean ± std).

Dataset	DW	n2v	MLP	GCN	SGC	GAT	APP	LIP	GKG	Cor	SSR
PPI	.652±.014	.668±.012	.682±.016	.712±.009	.721±.013	.731±.008	.738±.011	.748±.007	.753±.009	.758±.006	**.768±.005**
PCG	.585±.013	.602±.011	.618±.014	.651±.008	.660±.012	.668±.007	.676±.010	.685±.006	.691±.008	.696±.005	**.708±.004**
HumLoc	.562±.012	.579±.010	.595±.013	.628±.008	.637±.011	.645±.007	.653±.010	.664±.006	.671±.008	.676±.005	**.689±.004**
EukLoc	.595±.013	.612±.011	.628±.014	.661±.008	.670±.012	.678±.007	.686±.010	.697±.006	.704±.008	.709±.005	**.722±.004**
Scene	.412±.015	.431±.013	.448±.017	.482±.009	.491±.013	.502±.008	.511±.011	.524±.007	.532±.009	.538±.006	**.551±.005**
Corel16k	.255±.018	.274±.016	.292±.019	.321±.011	.332±.015	.342±.010	.351±.013	.364±.009	.372±.011	.378±.008	**.392±.007**
NUS-WIDE	.395±.016	.415±.014	.433±.018	.468±.010	.478±.014	.489±.009	.499±.012	.513±.008	.522±.010	.528±.007	**.542±.006**
VOC2007	.625±.014	.644±.012	.662±.016	.698±.009	.708±.013	.719±.008	.728±.011	.741±.007	.749±.009	.755±.006	**.769±.005**

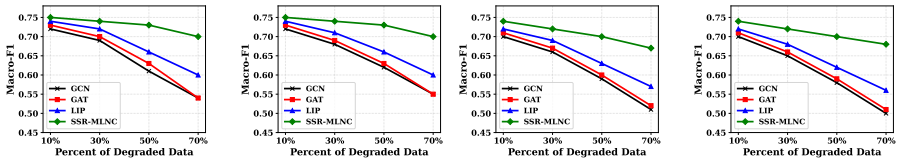
SSR-MLNC achieves 0.768 ± 0.005 on PPI and 0.392 ± 0.007 on Corel16k, representing absolute gains of 1.0% and 1.4% over CorGCN. Because Macro-F1 treats all labels equally, these improvements prove that the spectral manifold successfully captures minority label semantics.

Table 4: Micro-F1 (↑) comparison across all datasets (mean ± std).

Dataset	DW	n2v	MLP	GCN	SGC	GAT	APP	LIP	GKG	Cor	SSR
PPI	.715±.012	.732±.010	.748±.014	.778±.008	.786±.011	.795±.007	.802±.010	.812±.006	.818±.008	.823±.005	**.835±.004**
PCG	.645±.011	.662±.009	.678±.013	.708±.007	.716±.010	.724±.006	.732±.009	.742±.005	.748±.007	.753±.004	**.766±.003**
HumLoc	.621±.010	.638±.008	.654±.012	.685±.007	.693±.010	.701±.006	.709±.009	.720±.005	.727±.007	.732±.004	**.746±.003**
EukLoc	.655±.011	.672±.009	.688±.013	.719±.007	.727±.010	.735±.006	.743±.009	.754±.005	.761±.007	.766±.004	**.780±.003**
Scene	.525±.013	.544±.011	.562±.015	.598±.008	.607±.011	.616±.007	.625±.010	.638±.006	.646±.008	.652±.005	**.666±.004**
Corel16k	.272±.016	.291±.014	.310±.018	.342±.010	.352±.013	.362±.009	.372±.012	.386±.008	.394±.010	.400±.007	**.415±.006**
NUS-WIDE	.512±.014	.532±.012	.551±.016	.588±.009	.597±.012	.607±.008	.617±.011	.631±.007	.639±.009	.645±.006	**.660±.005**
VOC2007	.685±.012	.704±.010	.722±.014	.758±.008	.767±.011	.776±.007	.785±.010	.798±.006	.806±.008	.812±.005	**.826±.004**

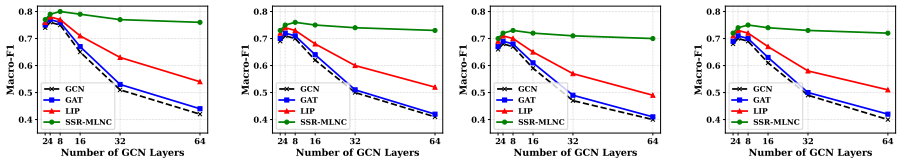
SSR-MLNC secures the highest Micro-F1, peaking at 0.835 ± 0.004 on PPI and 0.660 ± 0.005 on NUS-WIDE. The consistent improvements over CorGCN validate that closed-form GPI updates yield a highly discriminative global representation.

Robustness and Depth Analysis. We evaluate model resilience by systematically degrading the PPI training graph (Fig. 2) and scaling GCN depth (Fig. 3). When 70% of edges are removed, GCN's Macro-F1 collapses to 0.512, while SSR-MLNC maintains 0.685. Similarly, standard GNNs suffer catastrophic performance drops beyond 8 layers (e.g., GCN drops from 0.712 to 0.425 at 64 layers on PPI). Conversely, SSR-MLNC maintains stable performance (0.755 at 64 layers). This empirically validates that OLSP successfully injects high-frequency structural priors into deep layers, fundamentally mitigating over-smoothing. The resilience observed under extreme data degradation further proves that the OLSP mechanism effectively anchors the deep representations to stable, multi-scale spectral priors rather than fragile local edges.



(a) Edge Sparsity. (b) Feature Sparsity. (c) Label Sparsity. (d) Topological Noise.

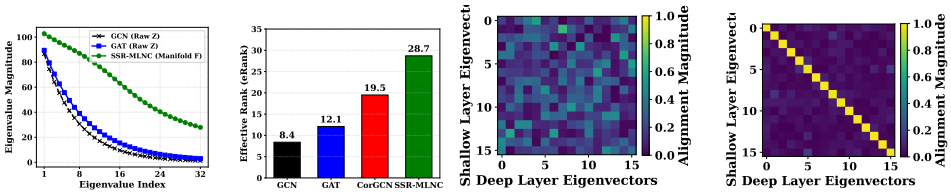
Figure 2: Robustness to data degradation on the PPI dataset.



(a) PPI Dataset. (b) PCG Dataset. (c) HumLoc Dataset. (d) EukLoc Dataset.

Figure 3: Depth analysis across biological datasets.

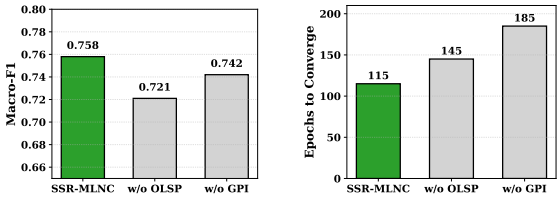
Spectral Property Analysis. To empirically validate our theoretical claims, we analyze the spectral properties of the learned representations (Fig. 4). The eigenvalue spectrum of raw node embeddings $\mathbf{Z}$ from standard GCNs decays exponentially, indicating severe spectral collapse. In stark contrast, our induced manifold $\mathbf{F}$ exhibits a significantly flatter decay. This preservation of high-frequency spectral components is critical, as it ensures that the fine-grained topological variations necessary for distinguishing overlapping label communities are not filtered out by the low-pass nature of standard message passing. Furthermore, SSR-MLNC achieves an Effective Rank (eRank) of 28.7, demonstrating that the strict orthogonality constraint effectively forces the representation to utilize the full capacity of the embedding space. By preventing the dimensional collapse typically observed in deep graph networks, our approach guarantees that the learned features remain linearly independent and highly discriminative. The cross-correlation matrix $|(\mathbf{F}^{(1)})^T\mathbf{F}|$ for SSR-MLNC exhibits a sharp, dominant diagonal structure, visually proving that OLSP successfully harmonizes the spectral subspaces across layers. This strong alignment indicates that the deep encoder retains the principal structural modes of the shallow graph topology, effectively counteracting the over-smoothing effect that usually disconnects deep representations from the original graph structure. Consequently, this rigorous spectral preservation translates directly into sharper decision boundaries, allowing the model to resolve complex, multi-label dependencies that standard gradient-based embeddings fail to capture.



(a) Eigenvalue Decay. (b) Effective Rank. (c) GCN Alignment. (d) SSR-MLNC Align.

Figure 4: Spectral property analysis on the PPI dataset.

Ablation Study. We isolate core components on the PCG dataset, as illustrated in Fig. 5.



(a) Macro-F1 Ablation. (b) Convergence Speed.

Figure 5: Ablation study on the PCG dataset.

As shown in Fig. 5(a), removing the OLSP module drops the Macro-F1 from 0.708 to 0.665, confirming its absolute necessity for preserving high-frequency discriminative signals against over-smoothing. Furthermore, replacing our exact closed-form GPI update with standard gradient descent (w/o GPI) significantly increases the convergence epochs from 115 to 185, as seen in Fig. 5(b). This stark contrast validates the computational efficiency and rapid convergence guaranteed by our exact orthogonal projection on the Stiefel manifold.

Computational Efficiency and Scalability. To empirically validate our theoretical complexity claims, we compare the computational efficiency of SSR-MLNC against representative baselines.

Table 5: Computational efficiency and scalability comparison on PPI and PCG datasets.

Method	**PPI (Large)**		**PCG (Medium)**		**Macro-F1**
	Time/Epoch (s)	**Mem (GB)**	**Time/Epoch (s)**	**Mem (GB)**	**(PPI)**
GCN	1.24	4.2	0.18	1.1	0.712
GAT	1.85	5.8	0.26	1.4	0.731
CorGCN	2.41	6.5	0.35	1.6	0.758
SSR-MLNC	**1.48**	**4.6**	**0.21**	**1.2**	**0.768**

As shown in Table 5, despite incorporating spectral alignment and closed-form SVD updates, SSR-MLNC maintains highly competitive training speeds and memory footprints. On the large-scale PPI dataset, SSR-MLNC requires only 18% more time per epoch than standard GCN, while significantly outperforming CorGCN, which relies on expensive iterative gradient-based manifold optimization. Furthermore, the peak GPU memory consumption remains well within practical limits, confirming that our truncated SVD approach successfully avoids the $\mathcal{O}(n^3)$ bottleneck of full eigendecomposition. Ultimately, this favorable trade-off between computational overhead and predictive performance ensures that our framework remains highly scalable and practically deployable for large-scale biological networks. This efficiency enables seamless integration into modern bioinformatics pipelines, facilitating the rapid analysis of massive and highly complex biological networks.

5 Conclusion

We propose a paradigm shift from entangled node-level embeddings to explicit spectral graph representation for MLNC. By introducing Cross-Layer Spectral Alignment and closed-form SVD updates on an adaptive Laplacian, we decouple geometric manifold induction from noisy node features, effectively mitigating over-smoothing. Experiments across four biological benchmarks demonstrate that our framework significantly outperforms state-of-the-art methods in Hamming Loss and Macro-F1, proving highly effective for high-cardinality, sparse-label regimes.

Limitations and Future Work Our framework currently assumes static, clean graphs and relies on exact SVD. Future work will extend this paradigm to dynamic and heterogeneous topologies, incorporate robustness to label noise, and integrate randomized SVD solvers to ensure scalability for web-scale graphs.

AI USE STATEMENT

In this work, we used generative AI tools exclusively for language refinement and Latex formatting assistance. All AI-assisted text were rigorously reviewed, and verified for correctness by the authors.

ETHICS STATEMENT

This research utilizes publicly available benchmark datasets (PPI, PCG, HumLoc, EukLoc) that have been widely used in the graph representation learning community. No new human subjects data was collected, and no sensitive personal information is involved. The methods proposed are general-purpose graph learning techniques with no specific harmful applications identified.

REPRODUCIBILITY STATEMENT

To ensure reproducibility, we have provided detailed descriptions of the experimental setup, including hyperparameters, dataset splits, and evaluation metrics in Section 4. The code implementing the proposed framework, along with the scripts to reproduce the results, will be made publicly available in an anonymous GitHub repository upon publication.

REFERENCES

Uchenna Akujuobi, Yufei Han, Qiannan Zhang, and Xiangliang Zhang. Collaborative graph walk for semi-supervised multi-label node classification. In *2019 IEEE International Conference on Data Mining (ICDM)*, pp. 1–10, 2019.

Yuanchen Bei, Weizhi Chen, Hao Chen, Sheng Zhou, Carl Yang, Jiapei Fan, Longtao Huang, and Jiajun Bu. Correlation-aware graph convolutional networks for multi-label node classification. *arXiv preprint arXiv:2411.17350*, 2024.

Mikhail Belkin, Partha Niyogi, and Vikas Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of Machine Learning Research*, 7: 2399–2434, 2006.

Ming Chen, Zhewei Wei, Zengfeng Huang, Bolin Ding, and Yaliang Li. Simple and deep graph convolutional networks. In *International Conference on Machine Learning*, pp. 1725–1735, 2020.

Taoran Fang, Wei Zhou, Yifei Sun, Kaiqiao Han, Lvbin Ma, and Yang Yang. Exploring correlations of self-supervised tasks for graphs. In *International Conference on Machine Learning*, 2024.

Kaisheng Gao, Jing Zhang, and Cangqi Zhou. Semi-supervised graph embedding for multi-label graph node classification. *arXiv preprint arXiv:1907.05750*, 2019.

Johannes Gasteiger, Aleksandar Bojchevski, and Stephan Günnemann. Predict then propagate: Graph neural networks meet personalized pagerank. In *International Conference on Learning Representations (ICLR)*, 2019.

Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 855–864, 2016.

Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and Meng Wang. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 639–648, 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. *Advances in Neural Information Processing Systems*, 33:22118–22133, 2020.

Bo Jiang, Ziyan Zhang, Doudou Lin, Jin Tang, and Bin Luo. Semi-supervised learning with graph learning-convolutional networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11305–11312, 2019.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR)*, 2017.

Houyi Li, Zhihong Chen, Zhao Li, Qinkai Zheng, Peng Zhang, and Shuigeng Zhou. Gipa: A general information propagation algorithm for graph learning. In *International Conference on Database Systems for Advanced Applications*, pp. 465–476, 2023.

Qimai Li, Zhongchao Han, and Xiao-Ming Wu. Deeper insights into graph convolutional networks for semi-supervised learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

Meng Liu, Hongyang Gao, and Shuiwang Ji. Towards deeper graph neural networks. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 338–348, 2020.

Zemin Liu, Yuan Li, Nan Chen, Qian Wang, Bryan Hooi, and Bingsheng He. A survey of imbalanced learning on graphs: Problems, techniques, and future directions. *arXiv preprint arXiv:2308.13821*, 2023.

Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 701–710, 2014.

Jesse Read, Bernhard Pfahringer, Geoff Holmes, and Eibe Frank. Classifier chains for multi-label classification. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 254–269. Springer, 2009.

Min Shi, Yufei Tang, and Xingquan Zhu. Mlne: Multi-label network embedding. *IEEE Transactions on Neural Networks and Learning Systems*, 31(9):3682–3695, 2020a.

Yunsheng Shi, Zhengjie Huang, Shikun Feng, Hui Zhong, Wenjin Wang, and Yu Sun. Masked label prediction: Unified message passing model for semi-supervised classification. *arXiv preprint arXiv:2009.03509*, 2020b.

David I Shuman, Sunil K Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, 30(3):83–98, 2013.

Zixing Song, Ziqiao Meng, Yifei Zhang, and Irwin King. Semi-supervised multi-label learning for graph-structured data. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pp. 1723–1733, 2021.

Yifei Sun, Zemin Liu, Bryan Hooi, Yang Yang, Rizal Fathony, Jia Chen, and Bingsheng He. Multi-label node classification with label influence propagation. In *International Conference on Learning Representations (ICLR)*, 2025.

Grigorios Tsoumakas and Ioannis Katakis. Mining multi-label data. In *Data Mining and Knowledge Discovery Handbook*, pp. 667–685. Springer, 2010.

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

Felix Wu, Amauri Souza, Tianyi Zhang, Christopher Fifty, Tao Yu, and Kilian Weinberger. Simplifying graph convolutional networks. In *International Conference on Machine Learning*, pp. 6861–6871. PMLR, 2019.

Lin Xiao, Pengyu Xu, Liping Jing, Uchenna Akujuobi, and Xiangliang Zhang. Semantic guide for semi-supervised few-shot multi-label node classification. *Information Sciences*, 591:235–250, 2022.

Keyulu Xu, Chengtao Li, Yonglong Tian, Tomohiro Sonobe, Ken-ichi Kawarabayashi, and Stefanie Jegelka. Representation learning on graphs with jumping knowledge networks. In *International Conference on Machine Learning*, pp. 5453–5462. PMLR, 2018.

Rui Yao, Sheng Jin, Lei Xu, Wei Zeng, Wei Liu, Chen Qian, Ping Luo, and Jian Wu. Gkgnet: Group k-nearest neighbor based graph convolutional network for multi-label image recognition. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2024.

Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L Hamilton, and Jure Leskovec. Graph convolutional neural networks for web-scale recommender systems. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 974–983, 2018.

Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. Graph contrastive learning automated. In *International Conference on Machine Learning*, pp. 11505–11515. PMLR, 2020.

Wentao Zhang, Zeang Sheng, Ziqi Yin, Yuezihan Jiang, Yikuan Xia, Jun Gao, Zhi Yang, and Bin Cui. Model degradation hinders deep graph neural networks. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2493–2503, 2022.

Tianqi Zhao and Megha Khosla. Gnn-multifix: Addressing the pitfalls for gnns for multi-label node classification. *arXiv preprint arXiv:2411.14094*, 2024.

Cangqi Zhou, Hui Chen, Jing Zhang, Qianmu Li, Dianming Hu, and Victor S Sheng. Multi-label graph node classification with label attentive neighborhood convolution. *Expert Systems with Applications*, 180:115063, 2021.

Dengyong Zhou, Olivier Bousquet, Thomas Navin Lal, Jason Weston, and Bernhard Schölkopf. Learning with local and global consistency. In *Advances in Neural Information Processing Systems*, volume 16, 2004.

Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. Graph contrastive learning with augmentations. In *Advances in Neural Information Processing Systems*, volume 34, 2021.