Multi-Semantic Aware Self-Supervised Learning for Multi-Label Node Classification

Jiayu Zhang[1] , **Jitao Zhao**[2] , **Dongxiao He**[1,2,*] , **Cuiying Huo**[2] and **Zhiyong Feng**[2]
[1]Tianjin International Engineering Institute, Tianjin University, Tianjin, China
[2]College of Intelligence and Computing, Tianjin University, Tianjin, China
jiayuzhang0206@gmail.com, {zjtao, hedongxiao, huocuiying, zyfeng}@tju.edu.cn

Abstract

Graph self-supervised learning aims to mine intrinsic signals from graph data itself to train models. It enables the acquisition of high-quality representations without manual annotations, making it suitable for various label-scarce scenarios and thus garnering substantial interest. Existing graph self-supervised methods typically assume that each node possesses a single semantic meaning. However, many real-world entities exhibit multiple semantics, where a single node may simultaneously belong to multiple categories. Since existing approaches are mainly designed for single-semantic settings, they often struggle to accurately capture the multiple semantics within nodes. Meanwhile, existing multi-label graph learning methods mainly depend on extensive manual annotations, which are costly and of limited applicability. To address this, we make a bold attempt to extend graph self-supervised learning to multi-label graphs. It is particularly challenging, as lacking labels makes it difficult to identify multiple semantics, let alone determine the number of underlying categories for nodes. To this end, we propose a **M**ulti-semantic **A**ware **S**elf-**S**upervised pretraining method (MASS) for multi-label graphs. Specifically, we propose a multi-pseudo-label decomposition mechanism, enabling adaptive learning of multiple semantics within nodes without annotations. Extensive results validates the effectiveness of MASS, and it achieved competitive performances with supervised baselines.

1 Introduction

Graph Self-Supervised Learning (GSSL) aims to pre-train graph neural network models without human-annotated labels [Liu *et al.*, 2023; Xie *et al.*, 2023]. Unlike supervised learning, GSSL constructs proxy tasks by mining self-supervised signals from the graph data itself. This enables the model to effectively leverage large amounts of unlabeled graph data to learn meaningful representations and generalize to various downstream scenarios [Hu *et al.*, 2020c; Qiu *et al.*, 2020; Hu *et al.*, 2020b]. Therefore, GSSL has garnered significant attention and has found widespread applications such as molecular prediction [Wang *et al.*, 2023; Zhang *et al.*, 2021], anomaly detection [Zheng *et al.*, 2023; Liu *et al.*, 2022], and e-commerce recommendations [Wu *et al.*, 2019; Ren *et al.*, 2026].

Despite achieving significant success, existing GSSL methods are mainly designed for single-label scenarios, i.e., each node belongs to only one class [Zhu *et al.*, 2020; Kipf and Welling, 2016]. While, in the real world, nodes often possess multiple semantics, i.e., one node may belong to multiple labels [Sun *et al.*, 2018; Hu *et al.*, 2020a]. For example, in social networks [Sun *et al.*, 2018], a user may have multiple interests, such as being a "music fan," "movie enthusiast," and "amateur photographer". This multi-label nature is pervasive. Meanwhile, compared to single-label graphs, multi-label graphs is more complex, making manual annotation more difficult and expensive. Therefore, it is very important to explore self-supervised representation learning on multi-label graphs. Unfortunately, most existing multi-label graph learning methods [Gao *et al.*, 2019; Xiao *et al.*, 2022; Bei *et al.*, 2025] are limited to supervised settings. These methods typically heavily rely on manual annotation to learn the dependencies among node and multiple labels.

Therefore, we make a bold attempt to extend graph self-supervised learning to multi-label scenarios. It is full of challenges, specifically manifested in the following two aspects: (i) **Multi-semantic entanglement (Challenge 1).** In multi-label settings, a node can belong to multiple categories simultaneously. Thus, two nodes may overlap in some semantics while differing in others. A pair of ideal representation should capture this partially aligned, partially separated semantics, thus being able to fully express the complex semantics of various categories hidden in nodes. However, existing GSSL paradigms typically assume that each node possess semantic of only one category. Thus, they [Zhu *et al.*, 2020] often treat a pair of node in either positive or negative pair. This binary alignment cannot adapt to complex semantics among multi-label nodes. The generative GSSLs [Kipf and Welling, 2016] also neglect the entangled semantics brought by multiple labels. (ii) **Hard self-supervised signal mining (Challenge 2).** While existing works [Bei *et al.*, 2025] have

*Corresponding author.

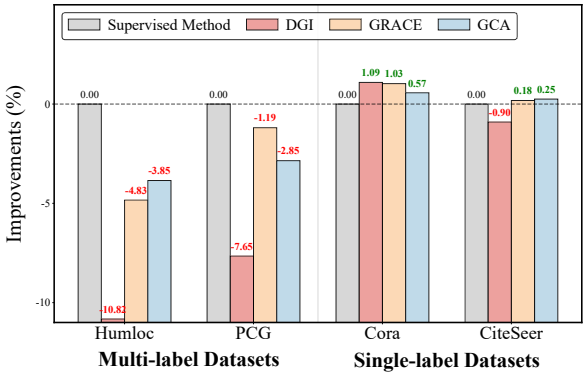

Figure 1: Motivation. We compare performances of existing GSSL methods and supervised methods on multi-label and single-label datasets, respectively. For the multi-label dataset, the supervised baseline is CorGCN, while for the single-label dataset, the supervised GCN is the baseline. We calculate the performance improvement of GSSL methods compared to the supervised baseline, with positive values highlighted in green and negative values in red.

explored multi-label graph learning, their learning of the associations among representations and multi-labels heavily relies on manual annotations. However, acquiring high-quality multi-label annotations is costly, imposing limitations on supervised approaches. In the self-supervised scenario, it is difficult to extract appropriate signals to model multi-label dependencies. We even don't know how many categories the nodes possess. Therefore, as shown in Figure 1, although existing GSSL methods can achieve comparable performance to supervised methods in the single-label domain, their performance declines when transferred to the multi-label scenario.

To address these challenges, we propose a **M**ulti-semantic **A**ware **S**elf-**S**upervised graph pretraining method (MASS), which aims to learn discriminative and semantically disentangled node representations without labels. For **Challenge 1**, we design a **multi-semantic feature disentanglement module**. This module does not treat node representations as a whole. Instead, it projects them into multiple distinct semantic subspaces, governed by learnable prototypes. These subspaces enable the model to separately model different semantic aspects of each node, capturing nuanced relationships where nodes may align in certain semantic dimensions while diverging in others. For **Challenge 2**, we introduce a **contextual structure encoding module**, aiming to extract richer structural signals from the graph topology to complement and enhance the semantic representations. Specifically, this module captures high-order neighborhood context information of nodes through random walk to generate structural embeddings. The structural embeddings can provide topological self-supervised signals independent of node features, offering additional self-supervised constraints in the absence of label guidance. This design enables the model to understand nodes more comprehensively from both structural and semantic perspectives, effectively alleviating the signal sparsity problem.

Our contributions can be summarized as follows:

- We pioneer the exploration of self-supervised representation learning on multi-label graphs, effectively freeing it from dependence on manual annotations and offering a fresh perspective in this field.
- We propose MASS, a multi-semantic aware self-supervised graph pretraining method, enabling model to capture complementary self-supervised signals that integrate both semantic disentanglement from node features and contextual information from graph topology.
- Comprehensive experiments on four datasets show that our MASS achieves competitive performance against supervised methods on seven key metrics, convincingly demonstrating its effectiveness and superiority.

2 Related Work

2.1 Graph Self-Supervised Learning

Graph Self-Supervised Learning (GSSL) has emerged as a powerful paradigm for learning representations from unlabeled graph data [Xie *et al.*, 2023; Liu *et al.*, 2023; Wu *et al.*, 2023]. Existing methods can be primarily categorized into generative and contrastive methods [Liu *et al.*, 2023]. **Generative GSSL methods** derive supervision signals by reconstructing node attributes or graph topology. Representative works, GAE and VGAE [Kipf and Welling, 2016], adopt an encoder-decoder architecture to reconstruct the graph's adjacency matrix. GraphMAE [Hou *et al.*, 2022] introduces feature masking and learns representations by reconstructing the masked node features. **Contrastive GSSL methods** learn node representations by comparing similar and dissimilar instances across different granularities. Node-level approaches like GRACE [Zhu *et al.*, 2020] and GCA [Zhu *et al.*, 2021] create positive pairs via augmentations and contrast them against negative instances, thereby capturing local semantic invariance. Cross-level approaches such as DGI [Velickovic *et al.*, 2019] maximize mutual information between local node embeddings and global graph summaries to learn discriminative representations.

Despite notable success, they typically treat node features as an indivisible whole, overlooking the inherent multi-semantic nature of nodes in multi-label graphs. Consequently, directly applying existing self-supervised frameworks to multi-label data often results in entangled representations and limited semantic discriminability.

2.2 Graph Multi-Label Learning

Traditional graph multi-label learning is usually based on supervised or semi-supervised frameworks, focusing primarily on modeling label correlations and incorporating label semantics into graph neural networks [Shi *et al.*, 2020; Akujuobi *et al.*, 2019]. Early studies, such as ML-GCN [Gao *et al.*, 2019], propose projecting labels and nodes into a shared embedding space and constructing label-label and node-label correlation matrices to enhance node representations. Subsequently, some methods like LANC [Zhou *et al.*, 2021a] and LARN [Xiao *et al.*, 2022] incorporate label-aware attention mechanisms to capture fine-grained label dependencies. Then, recent work CorGCN [Bei *et al.*, 2025] further addresses challenges such as label ambiguity and structural noise through its correlation-aware graph convolution mechanism. LIP [Sun *et al.*, 2025] dynamically guides the model

learning process by computing higher-order label correlations and propagating influence between labels.

Despite their effectiveness, these methods rely heavily on extensive high-quality labeled data, which is often costly or difficult to obtain in real-world multi-label scenarios.

2.3 Self-Supervised Learning on Multi-Label Data

Researches on self-supervised multi-label learning remains scarce, with only a few works focusing on the field of computer vision [Zhu *et al.*, 2023; Chen and Yeh, 2024; Lin *et al.*, 2021]. For instance, OASS [Xu *et al.*, 2022] leverages object detection to generate region-level pseudo-labels. By designing region-specific contrastive learning tasks, it guides the model to learn semantically distinct representations for different visual concepts within a single image, thereby achieving semantic disentanglement at the object level.

However, in the graph domain, self-supervised learning for multi-label graphs remains unexplored. To the best of our knowledge, no existing work can address the core challenges of multi-label graph learning under a label-free setting. It's the gap that our work aims to bridge.

3 Preliminaries

Notations. Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, consisting of N vertices $\mathcal{V} = \{v_0, v_1, \cdots, v_N\}$ and M edges $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$. The graph structure is represented by an adjacency matrix $\mathbf{A} \in \{0,1\}^{N \times N}$ with $\mathbf{A}_{ij} = 1$ if $(v_i, v_j) \in \mathcal{E}$. The node features are represented by a feature matrix $\mathbf{X} = [\mathbf{x}_1^\top, \mathbf{x}_2^\top, \cdots, \mathbf{x}_N^\top]^\top \in \mathbb{R}^{N \times F}$, where each row $\mathbf{x}_i \in \mathbb{R}^F$ denotes the F-dimensional feature vector of node v_i. In multi-label graph learning, each node is associated with a label set $\mathcal{Y}_i \subseteq \mathcal{C}$, where $\mathcal{C} = \{c_1, c_2, \cdots, c_C\}$ denotes the label space of C semantic categories. This association can be represented by a binary vector $\mathbf{y}_i = [y_{i1}, y_{i2}, \cdots, y_{iC}] \in \{0,1\}^C$, where $y_{ij} = 1$ if node v_i belongs to category c_j, and $y_{ij} = 0$ otherwise. The collection of all node labels forms the label matrix $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N]^\top \in \{0,1\}^{N \times C}$.

Problem Definition. In this work, we address the self-supervised multi-label graph representation learning problem, where **the label matrix Y is unavailable during training.** Our objective is to learn an encoder $f_\theta(\mathbf{X}, \mathbf{A})$ that generates low-dimensional node embeddings $\mathbf{Z} = f_\theta(\mathbf{X}, \mathbf{A}) \in \mathbb{R}^{N \times d}$ where $d \ll F$, with $\mathbf{z}_i$ denoting the embedding of node v_i. The learned representations should capture the multiple semantic facets within nodes without any label, such that they serve as effective representations for downstream multi-label node classification tasks.

4 Methodology

In this section, we present our proposed MASS framework in detail, comprising multi-semantic feature disentanglement (Sec. 4.1) and contextual structure encoding (Sec. 4.2). Subsequently, we elaborate on the loss function (Sec. 4.3). The overall framework of MASS is illustrated in Figure 2.

4.1 Multi-Semantic Feature Disentanglement

To address **Challenge 1**, we introduce a multi-semantic feature disentanglement module that decomposes holistic node representations into multiple independent semantic subspaces, each corresponding to a potential semantic category. By this, the module can capture the differentiated characteristics of nodes across various semantic dimensions.

Semantic Subspaces

We first adopt two-layer GCN [Kipf and Welling, 2017] as the base encoder to obtain initial nodes feature representations $\mathbf{Z}_f$. Nevertheless, $\mathbf{Z}_f$ remains a fused representation where different semantics are entangled. To explicitly disentangle these semantics, we propose to decompose the unified representation $\mathbf{Z}_f$ into C independent semantic subspaces, each corresponding to a potential semantic category $c \in \{1, 2, \ldots, C\}$. Specifically, we design C independent projection headers, formulated as:

$$\mathbf{R}_c = \mathbf{Z}_f \mathbf{W}_c + \mathbf{b}_c, \tag{1}$$

where $\mathbf{W}_c \in \mathbb{R}^{d_h \times d_c}$ and $\mathbf{b}_c \in \mathbb{R}^{d_c}$ are learnable parameters, and $\mathbf{R}_c \in \mathbb{R}^{N \times d_c}$ denotes the representation of all nodes projected into the c-th semantic subspace.

This design allows each semantic to have an independent space, avoiding interference between different semantics. Thus, we obtain a set of independent semantic representations $\{\mathbf{R}_c\}_{c=1}^C$, each capturing a distinct semantic facet of the nodes without mutual interference.

Codebook-Based Pseudo-Labels

Due to lacking guidance from ground-truth labels in self-supervised learning setting, the model should seek supervisory signals from data itself. To address this, we try to learn a set of pseudo-labels as alternatives to guide model by leveraging a codebook-based prototype learning mechanism.

Specifically, we maintain a codebook of K prototypes for each category c. It can be denoted as $\mathbf{p}_c = \{\mathbf{p}_c^1, \mathbf{p}_c^2, \ldots, \mathbf{p}_c^K\}$, where each prototype $\mathbf{p}_c^k$ represents potential semantic within category c. These prototypes are dynamically aggregated through an attention mechanism to form the category's central representation as pseudo-labels:

$$\alpha_c^k = \frac{\exp(\boldsymbol{w}_a^\top \boldsymbol{p}_c^k)}{\sum_{j=1}^K \exp(\boldsymbol{w}_a^\top \boldsymbol{p}_c^j)}, \bar{\boldsymbol{p}}_c = \sum_{k=1}^K \alpha_c^k \boldsymbol{p}_c^k, \tag{2}$$

where $\boldsymbol{w}_a^\top \in \mathbb{R}^{d_c}$ is a learnable attention vector.

Label Alignment Loss

To leverage the learned pseudo-labels to guide representation learning, we design a label alignment loss that encourages node embeddings to align with their corresponding semantic central representation.

Specifically, given the pseudo-label $\bar{\boldsymbol{p}}_c$ and the node representation $\mathbf{r}_i^c$ in semantic subspace c, we first compute the confidence score that node v_i belongs to category c:

$$s_i^c = \sigma(\mathbf{r}_i^c \bar{\boldsymbol{p}}_c), \tag{3}$$

where $\sigma(\cdot)$ is the sigmoid function, mapping it to $(0, 1)$.

Based on this confidence score, we define positive and negative sample sets for each category c with a threshold $\tau \in (0, 1)$:

$$\mathcal{P}_c = \{v_i | s_i^c > \tau\}, \mathcal{N}_c = \{v_i | s_i^c \leq \tau\}. \tag{4}$$

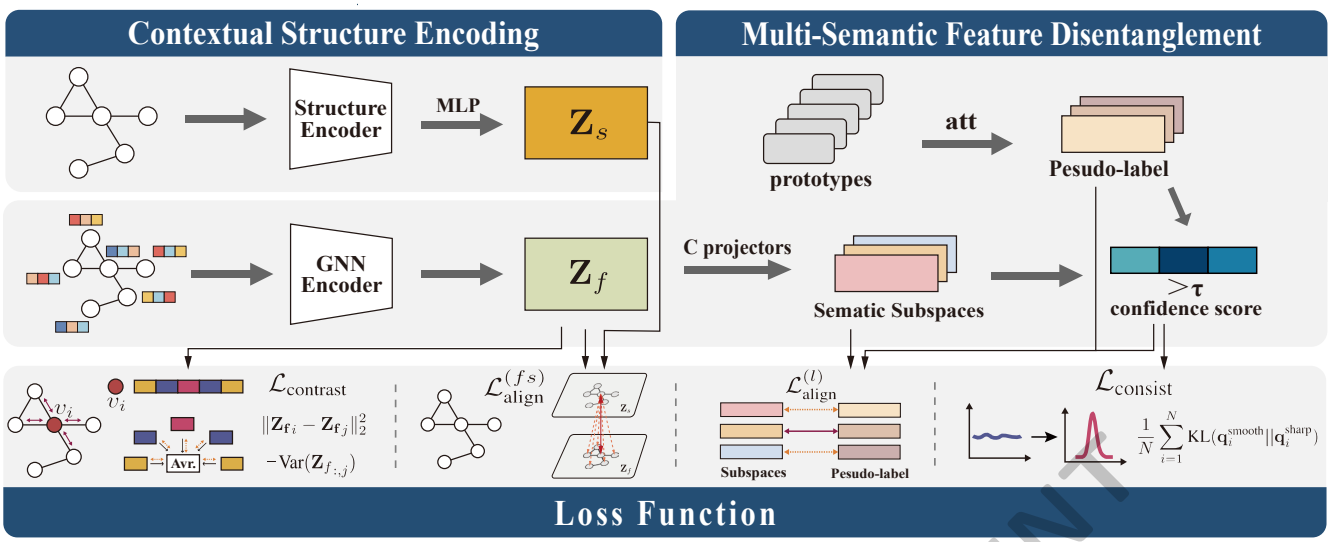

Figure 2: The overall framework of our proposed MASS. (a) Multi-Semantic Feature Disentanglement: it decomposes holistic node representations into multiple independent semantic subspaces by the guidance of codebook-based pesudo-labels. (Sec. 4.1) (b) Contextual Structure Encoding: it generates structural representations to provide complementary self-supervision signals from graph topology. (Sec. 4.2)

Then, to encourage positive samples to align with their corresponding pseudo-label while pushing negative samples away, the label alignment loss is formulated as:

$$\mathcal{L}_{\text{align}}^{(l)} = \sum_{c=1}^{C}(\mathbb{E}_{v_i\in\mathcal{P}_c}[1-\cos(\mathbf{r}_i^c, \bar{\boldsymbol{p}}_c)] + \mathbb{E}_{v_i\in\mathcal{N}_c}[\cos(\mathbf{r}_i^c, \bar{\boldsymbol{p}}_c)]), \tag{5}$$

where $\cos(\cdot,\cdot)$ denotes cosine similarity, $\mathbb{E}$ denotes the mathematical expectation.

Thus, $\mathcal{L}_{align}^{(l)}$ provides self-supervised alignment signals to the model and enables each semantic subspace to autonomously learn category prototypes, achieving alignment of node representations in shared semantic subspaces and separation in non-shared ones.

Contrastive Regularization

Following the smoothing assumption of graphs, we introduce an aggregation constraint to preserve local consistency. Meanwhile, to prevent the representation space from collapsing into trivial solutions, we add a separation constraint to promote variance across semantic dimensions. Therefore, the overall contrastive regularization loss can be defined as:

$$\mathcal{L}_{\text{contrast}} = \mathcal{L}_{\text{agg}} + \gamma_{\text{sep}}\mathcal{L}_{\text{sep}}, \tag{6}$$

where γ_{sep} is the weight coefficient, and $\mathcal{L}_{\text{agg}}$ and $\mathcal{L}_{\text{sep}}$ can be formulated as:

$$\begin{aligned}\mathcal{L}_{\text{agg}} &= \frac{1}{|\mathcal{E}'|}\sum_{(v_i,v_j)\in\mathcal{E}'}\|\mathbf{Z}_{\mathbf{f}i} - \mathbf{Z}_{\mathbf{f}j}\|_2^2,\\ \mathcal{L}_{\text{sep}} &= -\frac{1}{d_h}\sum_{j=1}^{d_h}\text{Var}(\hat{\mathbf{Z}}_{f_{:,j}}),\end{aligned} \tag{7}$$

where $|\mathcal{E}'| = \{(v_i, v_j) \in \mathcal{E} | i \neq j\}$ denotes the set of edges excluding self-loops. $\mathbf{Z}_{\mathbf{f}i}$ represents the feature representation of node v_i. $\hat{\mathbf{Z}}_{:,j}$ denotes the normalized representation of all nodes in the j-th semantic dimension.

By enhancing local consistency and dimensionality diversity, $\mathcal{L}_{\text{contrast}}$ preserves local semantic similarity between neighboring nodes and effectively prevents representation space from collapsing into trivial solutions during self-supervised learning.

Consistency Regularization

Notably, in self-supervised learning, models are prone to premature convergence to suboptimal solutions or overconfident predictions, where predictions may collapse into a few dominant categories while ignoring other plausible semantic associations. This issue is particularly severe in multi-label learning, as the model must identify and balance multiple coexisting semantics for each node without guidance from ground-truth labels.

Thus, we design a temperature-scaled consistency regularization loss to guide the learning of semantic subspaces and pesudo-labels. This loss controls the entropy of the predicted distribution through a temperature parameter, dynamically balancing exploration and discrimination of model.

Specifically, for the confidence score $\mathbf{s}_i = [s_i^1, s_i^2, \cdots, s_i^C]$ of node v_i, we introduce two temperature parameters T and T', where $T > T' > 0$, to generate smoothed and sharpened distributions respectively:

$$\mathbf{q}_i^{\text{smooth}} = \text{softmax}(\mathbf{s}_i/T), \mathbf{q}_i^{\text{sharp}} = \text{softmax}(\mathbf{s}_i/T') \tag{8}$$

The smoothed distribution $\mathbf{q}_i^{\text{smooth}}$ has higher entropy with more uniform probability distribution across labels, encouraging the model to explore potential relationships between different labels. The sharpened distribution $\mathbf{q}_i^{\text{sharp}}$ has lower entropy with more prominent probabilities for high-confidence labels, prompting the model to make clearer distinctions. The consistency loss minimizes the KL divergence

between them:

$$\mathcal{L}_{\text{consist}} = \frac{1}{N}\sum_{i=1}^{N} \text{KL}(\mathbf{q}_i^{\text{smooth}} || \mathbf{q}_i^{\text{sharp}}) \tag{9}$$

This design enables the model to focus more on exploring inter-label relationships in early training stages and gradually enhance discriminative capability as training progresses. Meanwhile, the signals provided by the soft distribution are richer than hard assignments, transmitting more detailed information through the KL divergence constraint between distributions, effectively mitigating pseudo-label noise issues.

4.2 Contextual Structure Encoding

To address **Challenge 2**, we aim to mine richer information from graph data in a self-supervised setting. Besides node attributes, we seek to extract more comprehensive signals from the topological structure of the graph. Therefore, we design a contextual structure encoding module that learns node embeddings from higher-order neighborhood patterns, complementing the multi-semantic feature disentanglement module.

We adopt a DeepWalk-based [Perozzi *et al.*, 2014] structural encoding approach to obtain structural embeddings of the graph. Specially, we employ random walks to generate sequences that capture L-hop neighborhood structures. Starting from each node v_i, we generate M random walk sequences of length L : $\mathcal{W}_i = \{\text{walk}_i^1, \text{walk}_i^2, \cdots, \text{walk}_i^M\}$, where $\text{walk}_i^m = (v_0, v_1, \cdots, v_{L-1})$ with $v_0 = v_i$.

Then, we follow the Skip-gram [Mikolov *et al.*, 2013] model to learn structural embeddings of nodes by maximizing the log probability of co-occurrence between nodes and their context nodes in the walk sequences:

$$\max \frac{1}{N}\sum_{i=1}^{N}\sum_{m=1}^{M}\sum_{t=1}^{L}\sum_{-w\le j\le w, j\neq 0} \log P(v_{i,t+j}|v_{i,t}), \tag{10}$$

$$P(v_j|v_i) = \exp(\mathbf{u}_j^\top \mathbf{u}_i) / \sum_{v_k\in\mathcal{V}} \exp(\mathbf{u}_k^\top \mathbf{u}_i) \tag{11}$$

where w is the window size, $\mathbf{u}_i$ is the structural embeddings vector of node v_i. By optimizing the above objective, we obtain structural embeddings U. We apply a multi-layer perceptron (MLP) to the structural embeddings U to obtain the final structural representations $Z_s = \text{MLP}(U) \in \mathbb{R}^{N\times d_s}$.

4.3 Loss Function

Considering that representations obtained from the feature space and structural space for the same node should exhibit consistency at the semantic level, we design a feature-structure alignment loss function:

$$\mathcal{L}_{\text{align}}^{(fs)} = \underbrace{\frac{1}{N}\sum_{i=1}^{N}(1-\mathbf{S}_{ii})}_{\text{Self-alignment}} + \underbrace{\frac{1}{N(N-1)}\sum_{i\neq j}\max(0,\mathbf{S}_{ij})}_{\text{Mutual Exclusion}} \tag{12}$$

where $\mathbf{S} = \hat{\mathbf{Z}}_f \hat{\mathbf{H}}_s^\top$ is the similarity matrix between feature representations and projected structural representations. Here, $\hat{\mathbf{Z}}_f$ and $\hat{\mathbf{H}}_s$ represent the normalized feature and structural representations, respectively, with $\mathbf{H}_s = \text{proj}(\mathbf{Z}_s)$ obtained via a projection network $\text{proj}(\cdot)$ that maps $\mathbf{Z}_s$ into the same semantic space as $\mathbf{Z}_f$.

The self-alignment component of this loss function minimizes the distance between the feature and structural representations of the same node. Concurrently, the exclusion component penalizes positive similarities between different nodes, thereby preventing their representations from becoming excessively close. From an information-theoretic perspective, this formulation maximizes mutual information between the two views of each individual node while minimizing mutual information between different nodes.

With the losses mentioned above, the overall loss function of MASS is as follows:

$$\mathcal{L} = \mathcal{L}_{\text{align}}^{(fs)} + \alpha\mathcal{L}_{\text{align}}^{(l)} + \beta\mathcal{L}_{\text{contrast}} + \gamma\mathcal{L}_{\text{consist}} \tag{13}$$

where α, β and γ are the controllable hyper-parameters.

After optimization, the model produces two complementary representations: the feature representation $\mathbf{Z}_f$ and the structure representation $\mathbf{Z}_s$. Then, we obtain the final representation by concatenating them, i.e., $\mathbf{Z} = \text{Concat}(\mathbf{Z}_f, \mathbf{Z}_s)$. The combined representation $\mathbf{Z} \in \mathbb{R}^{N\times(d_h+d_s)}$ captures both feature semantics and graph topology, providing a comprehensive embedding for downstream tasks.

5 Experiments

In this section, we evaluate our proposed MASS framework through comprehensive experiments, aiming to answer three fundamental questions: (i) **RQ1**: How does MASS compare against state-of-the-art methods across diverse multi-label graph datasets? (ii) **RQ2**: What is the contribution of each component to the overall performance? (iii) **RQ3**: How sensitive is MASS to key hyperparameter settings?

5.1 Experimental Setup

Datasets. We conduct comprehensive experiments on four widely-used multi-label graph datasets: **Humloc** [Zhao *et al.*, 2023], **PCG** [Zhao *et al.*, 2023], **Blogcatalog** [Zhou *et al.*, 2021b] and **PPI** [Zeng *et al.*, 2020] for evaluation, covering various domains including social networks, academic networks, and biological networks.

Baselines. We compare MASS with the following seven state-of-the-art models: (i) Semi-supervised multi-label learning methods: LARN [Xiao *et al.*, 2022], LIP [Sun *et al.*, 2025], and CorGCN [Bei *et al.*, 2025]; (ii) Self-supervised learning methods: GAE [Kipf and Welling, 2016], DGI [Velickovic *et al.*, 2019], GRACE [Zhu *et al.*, 2020], and FUG [Zhao *et al.*, 2024]. All baseline methods are implemented using official code or reproduced according to original papers. We equip all models with the GCN backbone architecture to ensure fair comparison.

Evaluation Metrics. To comprehensively evaluate multi-label node classification performance, we employ seven standard evaluation metrics. **Ranking Loss** and **Hamming Loss**

		Semi-supervised			Self-supervised				
	Metrics	LARN	LIP	CorGCN	GAE	DGI	GRACE	FUG	MASS
Humloc	Ranking (↓)	18.01±0.69	14.55±0.49	14.93±0.75	17.97±0.48	19.76±0.77	16.28±0.09	17.16±0.59	**14.53±0.40**
	Hamming (↓)	9.01±0.25	8.05±0.15	8.04±0.23	8.62±0.15	8.65±0.02	8.30±0.21	8.28±0.22	**7.80±0.03**
	Ma-AUC (↑)	65.69±1.27	**69.45±0.46**	67.67±3.40	60.04±0.80	61.72±0.85	64.57±1.08	65.51±0.96	68.85±0.62
	Mi-AUC (↑)	81.95±0.54	85.11±0.41	85.52±0.56	79.82±0.37	76.27±0.90	81.39±0.38	82.11±1.07	**85.67±0.27**
	Ma-AP (↑)	15.40±0.10	19.09±0.26	21.67±1.41	15.19±0.09	17.00±0.25	19.92±0.76	18.21±0.55	**22.22±0.45**
	Mi-AP (↑)	32.32±1.85	40.29±0.57	42.17±1.18	30.88±0.25	31.60±0.81	37.42±1.06	37.88±1.63	**42.82±0.65**
	LRAP (↑)	55.49±0.74	**62.66±0.79**	61.58±1.71	55.14±0.55	56.37±0.46	60.54±0.53	59.20±0.84	62.52±0.42
PCG	Ranking (↓)	31.96±0.59	28.27±0.45	28.65±1.37	32.14±0.19	34.04±0.30	29.55±0.30	28.89±0.62	**27.57±0.34**
	Hamming (↓)	13.50±0.04	12.93±0.09	13.01±0.28	14.07±0.02	14.34±0.19	13.18±0.22	12.90±0.15	**12.86±0.13**
	Ma-AUC (↑)	53.28±0.63	57.58±0.34	58.33±0.79	58.20±0.85	57.13±0.24	58.44±0.17	59.46±0.49	**60.49±1.25**
	Mi-AUC (↑)	65.78±0.58	70.18±0.36	70.45±0.76	66.33±0.42	65.06±0.25	69.12±0.42	70.29±0.47	**71.63±0.42**
	Ma-AP (↑)	14.56±0.15	16.75±0.08	**19.32±1.36**	18.87±0.89	17.80±0.31	18.39±0.55	18.67±0.89	19.19±0.80
	Mi-AP (↑)	22.24±0.40	26.07±0.57	26.38±1.28	24.87±0.70	23.45±0.44	26.34±0.36	27.24±1.00	**28.14±0.42**
	LRAP (↑)	42.47±0.64	45.69±0.67	45.56±1.59	43.49±0.96	41.76±0.54	44.97±0.31	45.94±0.61	**47.18±0.10**
Blogcatalog	Ranking (↓)	25.81±0.21	25.54±0.07	25.69±0.20	25.87±0.05	25.52±0.08	25.74±0.27	25.65±0.11	**21.83±0.21**
	Hamming (↓)	3.60±0.01	3.59±0.01	3.60±0.00	3.61±0.00	3.61±0.00	3.60±0.01	3.60±0.00	**3.54±0.02**
	Ma-AUC (↑)	51.51±0.28	51.92±0.28	51.87±0.79	52.71±0.62	52.03±0.76	51.02±0.30	52.18±0.59	**63.35±0.51**
	Mi-AUC (↑)	73.55±0.25	73.30±0.15	73.72±0.20	72.99±0.12	73.55±0.16	73.63±0.10	73.78±0.27	**77.72±0.27**
	Ma-AP (↑)	3.83±0.01	4.04±0.03	4.00±0.04	4.45±0.12	3.94±0.09	3.92±0.06	4.04±0.02	**10.15±0.29**
	Mi-AP (↑)	9.37±0.14	9.28±0.04	9.43±0.10	9.46±0.15	9.32±0.02	9.48±0.07	9.50±0.03	**19.86±1.55**
	LRAP (↑)	27.87±0.14	27.68±0.12	27.75±0.15	28.01±0.12	27.73±0.01	27.71±0.17	27.83±0.09	**37.49±1.26**
PPI	Ranking (↓)	24.48±0.14	22.99±0.11	20.23±0.52	21.20±0.39	22.53±0.11	22.71±0.57	21.37±0.08	**19.91±0.05**
	Hamming (↓)	26.40±0.29	25.18±0.15	**23.76±0.41**	24.24±0.19	24.75±0.09	24.86±0.16	24.60±0.03	23.91±0.07
	Ma-AUC (↑)	57.94±0.70	62.43±0.26	69.52±0.85	67.86±0.53	65.28±0.17	65.37±1.12	65.79±0.19	**69.65±0.13**
	Mi-AUC (↑)	71.75±0.08	74.29±0.18	77.82±0.48	77.01±0.35	75.48±0.08	75.54±0.54	75.92±0.04	**78.06±0.13**
	Ma-AP (↑)	36.98±0.29	42.18±0.36	**50.03±1.27**	48.07±0.63	45.00±0.31	44.55±1.18	45.09±0.27	49.34±0.06
	Mi-AP (↑)	36.98±0.29	60.13±0.31	64.36±0.72	63.83±0.38	61.98±0.06	61.91±0.61	61.90±0.14	**64.56±0.11**
	LRAP (↑)	62.94±0.17	64.07±0.14	**66.98±0.60**	65.90±0.45	64.65±0.23	64.24±0.61	65.58±0.09	66.88±0.07
Average Rank		6.89	4.66	2.98	5.27	6.21	5.02	3.64	**1.32**

Table 1: The performances of various methods on multi-label node classification. We report mean ± std over three runs for each method on four datasets. The top-1 and top-2 results per metric are highlighted in **bold** and underlined. (↓: the lower, the better, ↑: the higher, the better)

measure the proportion of misclassified labels and label ranking quality, respectively; **Macro-AUC** and **Micro-AUC** evaluate classifier discrimination capability at different granularities; **Macro-AP** and **Micro-AP** assess prediction accuracy based on precision-recall curves; **LRAP** specifically evaluates average precision for label ranking. We adopt a standard experimental protocol: randomly split 10% of nodes as the training set, 10% as the validation set, and the remaining 80% as the test set. To ensure reliability, all experiments are repeated 20 times with different random seeds, and we report average results with standard deviation, providing stable performance evaluation. The source code is available at https://github.com/hedongxiao-tju/MASS.

5.2 Experimental Analysis

For **RQ1**, we evaluate the performance of our proposed MASS framework against seven state-of-the-art baseline methods on four multi-label graph datasets. The results of seven metrics are reported in Table 1. We adopt *Friedman test* to assess the statistical significance. The results of the *Friedman test* for all seven evaluation metrics are shown in Table 2. With all p-values far less than the significance level $\alpha = 0.05$, the null hypothesis that all methods perform equally is rejected, thereby confirming statistically significant differences across the compared methods.

From Table 1 and Table 2, we can observe that: (i) **Our proposed MASS achieves significant improvements over other baselines across four experimental datasets.** MASS achieves the best or highly competitive results on seven evaluation metrics, with an average rank of 1.32, significantly outperforming all other methods. Specifically, in terms of Ranking Loss and Hamming Loss, MASS generally outperforms the baselines, demonstrating clear advantages in label ranking quality and classification error rate. Regarding AUC and AP metrics, MASS achieves the best or second-best performances in macro and micro measures, highlighting its strong discriminative capability even under class imbalance and label scarcity. Especially on Blogcatalog, MASS achieves a Micro-AUC of 77.72%, outperforming the second-best model by 5.34%. (ii) **MASS performs better than existing self-supervised learning methods.** Compared with self-supervised baselines, MASS shows

Evaluation Metric	F_F	p-value
Ranking Loss	44.33	1.84e-07
Hamming Loss	37.12	4.44e-06
Ma-AUC	41.38	6.82e-07
Mi-AUC	50.31	1.26e-08
Ma-AP	52.78	4.10e-09
Mi-AP	49.00	2.27e-08
LARP	40.69	9.27e-07

Table 2: Friedman statistics F_F for each evaluation metric ($\alpha = 0.05$)

Method	Humloc	PCG	Blogcatalog	PPI
w/o MFD	83.07	68.23	77.09	74.98
w/o CSE	84.26	71.07	73.45	73.28
w/o Align	82.32	69.69	77.39	77.44
w/o Contr	85.22	71.09	75.87	77.20
w/o Consist	85.20	70.99	75.53	77.28
MASS	**85.67**	**71.36**	**77.72**	**78.06**

Table 3: Ablation study with its five variants (Micro-AUC)

clear gains across nearly all metrics. This verifies that our model MASS can effectively capture the multi-semantics of nodes in multi-label graphs, which are overlooked by conventional self-supervised approaches designed for single-label scenarios. (iii) **MASS achieves competitive performance even compared to semi-supervised multi-label learning methods.** Notably, without manual annotation during training, MASS matches or surpasses semi-supervised baselines on multiple metrics. This indicates that the proposed pretext tasks in our model successfully extract rich supervisory signals from the graph data itself, thereby guiding model to learn more discriminative and semantically disentangled node representations. Overall, the strong performance validates the effectiveness of our proposed MASS in self-supervised multi-label graph learning.

5.3 Ablation Study

For **RQ2**, to validate the effectiveness of each component in our proposed MASS framework, we conduct systematic ablation studies by comparing MASS with the following variants: (1) ***w/o MFD*** removes the multi-semantic feature disentanglement module. (2) ***w/o CSE*** removes the contextual structure encoding module. (3) ***w/o Align*** removes the label alignment loss $\mathcal{L}_{\text{align}}^{(l)}$. (4) ***w/o Contr*** removes the contrastive regularization loss $\mathcal{L}_{\text{contrast}}$. (5) ***w/o Consist*** removes the consistency regularization loss $\mathcal{L}_{\text{consist}}$.

Table 3 summarizes the ablation results. We observe that: (i) **Effectiveness of multi-semantic feature disentanglement module and contextual structure encoding module.** Removing either the feature disentanglement module (***w/o MFD***) or the structure encoding module (***w/o CSE***) leads to significant decline in performance across all datasets. This demonstrates that modeling multi-semantics of nodes and capturing signals from topological structures are essential to self-supervised multi-label learning. (ii) **Effectiveness of regularization losses.** The removal of any regularization term impairs model performance, which demonstrates that these regularization terms are not merely auxiliary. These losses respectively prevent representation collapse and mitigate over-confidence in early training, thereby enhancing the model's ability to identify and separate multiple semantics.

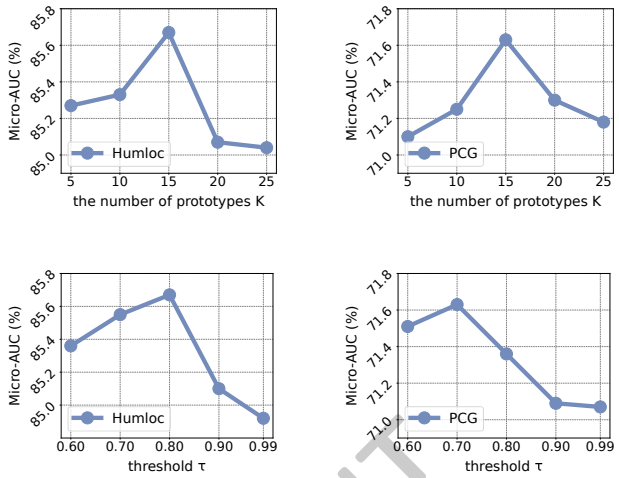

Figure 3: Sensitivity analysis of the hyperparameters K and τ

5.4 Parameter Study

For **RQ3**, we investigate the sensitivity of MASS to key hyperparameters including the number of prototypes K and threshold τ.

As shown in Figure 3, both K and τ influence model performance. When K is too small, the model fails to capture the semantic diversity within each category, This results in limited guidance from pseudo-labels, thereby leading to suboptimal performance. With K increasing, the model can capture finer semantics, and performance improves accordingly. When K becomes excessively large, the performance consequently declines. This indicates that redundant noise may be introduced into the pseudo-labels, causing semantic overlap or interference. Similarly, when τ is too low, some low-confidence samples were incorrectly classified as positive samples, causing the representations to deviate from the ideal semantic space. Conversely, when τ is too high, some high-confidence samples that should belong to the positives are erroneously excluded, hindering effective semantic alignment, which consequently degrades performance. Therefore, appropriate values of K and τ can enhance the model's overall performance.

6 Conclusion

In this paper, we explore self-supervised learning on multi-label graph data and propose a Multi-semantic Aware Self-Supervised graph pretraining method (MASS), which can effectively capture multiple semantics within nodes and free it from dependence on manual annotations. Extensive experiments demonstrates that our proposed MASS achieves competitive performances with supervised baselines.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No.62276187, No.62422210 and No.62372323).

Contribution Statement

Jiayu Zhang and Jitao Zhao contributed equally to this work.

References

[Akujuobi *et al.*, 2019] 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 2019, Beijing, China, November 8-11, 2019*, pages 1–10. IEEE, 2019.

[Bei *et al.*, 2025] Yuanchen Bei, Weizhi Chen, Hao Chen, Sheng Zhou, Carl Ji Yang, Jiapei Fan, Longtao Huang, and Jiajun Bu. Correlation-aware graph convolutional networks for multi-label node classification. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining, V.1, KDD 2025, Toronto, ON, Canada, August 3-7, 2025*, pages 37–48. ACM, 2025.

[Chen and Yeh, 2024] Chun-Yen Chen and Mei-Chen Yeh. Self-supervised multi-label classification with global context and local attention. In Cathal Gurrin, Rachada Kongkachandra, Klaus Schoeffmann, Duc-Tien Dang-Nguyen, Luca Rossetto, Shin'ichi Satoh, and Liting Zhou, editors, *Proceedings of the 2024 International Conference on Multimedia Retrieval, ICMR 2024, Phuket, Thailand, June 10-14, 2024*, pages 934–942. ACM, 2024.

[Gao *et al.*, 2019] Kaisheng Gao, Jing Zhang, and Cangqi Zhou. Semi-supervised graph embedding for multi-label graph node classification. In *Web Information Systems Engineering - WISE 2019 - 20th International Conference, Hong Kong, China, November 26-30, 2019, Proceedings*, volume 11881 of *Lecture Notes in Computer Science*, pages 555–567. Springer, 2019.

[Hou *et al.*, 2022] Zhenyu Hou, Xiao Liu, Yukuo Cen, Yuxiao Dong, Hongxia Yang, Chunjie Wang, and Jie Tang. Graphmae: Self-supervised masked graph autoencoders. In *KDD '22: The 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, August 14 - 18, 2022*, pages 594–604. ACM, 2022.

[Hu *et al.*, 2020a] 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*, 2020.

[Hu *et al.*, 2020b] 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*. OpenReview.net, 2020.

[Hu *et al.*, 2020c] Ziniu Hu, Yuxiao Dong, Kuansan Wang, Kai-Wei Chang, and Yizhou Sun. GPT-GNN: generative pre-training of graph neural networks. In Rajesh Gupta, Yan Liu, Jiliang Tang, and B. Aditya Prakash, editors, *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020*, pages 1857–1867. ACM, 2020.

[Kipf and Welling, 2016] Thomas N Kipf and Max Welling. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308*, 2016.

[Kipf and Welling, 2017] 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*. OpenReview.net, 2017.

[Lin *et al.*, 2021] Jinke Lin, Qingling Cai, and Manying Lin. Multi-label classification of fundus images with graph convolutional network and self-supervised learning. *IEEE Signal Process. Lett.*, 28:454–458, 2021.

[Liu *et al.*, 2022] Yixin Liu, Zhao Li, Shirui Pan, Chen Gong, Chuan Zhou, and George Karypis. Anomaly detection on attributed networks via contrastive self-supervised learning. *IEEE Trans. Neural Networks Learn. Syst.*, 33(6):2378–2392, 2022.

[Liu *et al.*, 2023] Yixin Liu, Ming Jin, Shirui Pan, Chuan Zhou, Yu Zheng, Feng Xia, and Philip S. Yu. Graph self-supervised learning: A survey. *IEEE Trans. Knowl. Data Eng.*, 35(6):5879–5900, 2023.

[Mikolov *et al.*, 2013] Tomás Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, pages 3111–3119, 2013.

[Perozzi *et al.*, 2014] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, New York, NY, USA - August 24 - 27, 2014*, pages 701–710. ACM, 2014.

[Qiu *et al.*, 2020] Jiezhong Qiu, Qibin Chen, Yuxiao Dong, Jing Zhang, Hongxia Yang, Ming Ding, Kuansan Wang, and Jie Tang. GCC: graph contrastive coding for graph neural network pre-training. In Rajesh Gupta, Yan Liu, Jiliang Tang, and B. Aditya Prakash, editors, *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020*, pages 1150–1160. ACM, 2020.

[Ren *et al.*, 2026] Xubin Ren, Wei Wei, Lianghao Xia, and Chao Huang. A comprehensive survey on self-supervised learning for recommendation. *ACM Comput. Surv.*, 58(1):22:1–22:38, 2026.

[Shi *et al.*, 2020] Min Shi, Yufei Tang, and Xingquan Zhu. MLNE: multi-label network embedding. *IEEE Trans. Neural Networks Learn. Syst.*, 31(9):3682–3695, 2020.

[Sun *et al.*, 2018] Peijie Sun, Le Wu, and Meng Wang. Attentive recurrent social recommendation. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018, Ann Arbor, MI, USA, July 08-12, 2018*, pages 185–194. ACM, 2018.

[Sun *et al.*, 2025] Yifei Sun, Zemin Liu, Bryan Hooi, Yang Yang, Rizal Fathony, Jia Chen, and Bingsheng He. Multi-label node classification with label influence propagation. In *The Thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24-28, 2025*. OpenReview.net, 2025.

[Velickovic *et al.*, 2019] 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*. OpenReview.net, 2019.

[Wang *et al.*, 2023] Hanchen Wang, Jean Kaddour, Shengchao Liu, Jian Tang, Joan Lasenby, and Qi Liu. Evaluating self-supervised learning for molecular graph embeddings. 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.

[Wu *et al.*, 2019] Shu Wu, Yuyuan Tang, Yanqiao Zhu, Liang Wang, Xing Xie, and Tieniu Tan. Session-based recommendation with graph neural networks. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 346–353. AAAI Press, 2019.

[Wu *et al.*, 2023] Lirong Wu, Haitao Lin, Cheng Tan, Zhangyang Gao, and Stan Z. Li. Self-supervised learning on graphs: Contrastive, generative, or predictive. *IEEE Trans. Knowl. Data Eng.*, 35(4):4216–4235, 2023.

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

[Xie *et al.*, 2023] Yaochen Xie, Zhao Xu, Jingtun Zhang, Zhengyang Wang, and Shuiwang Ji. Self-supervised learning of graph neural networks: A unified review. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(2):2412–2429, 2023.

[Xu *et al.*, 2022] Kaixin Xu, Liyang Liu, Ziyuan Zhao, Zeng Zeng, and Bharadwaj Veeravalli. Object-aware self-supervised multi-label learning. In *2022 IEEE International Conference on Image Processing, ICIP 2022, Bordeaux, France, 16-19 October 2022*, pages 361–365. IEEE, 2022.

[Zeng *et al.*, 2020] Hanqing Zeng, Hongkuan Zhou, Ajitesh Srivastava, Rajgopal Kannan, and Viktor K. Prasanna. Graphsaint: Graph sampling based inductive learning method. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020.

[Zhang *et al.*, 2021] Zaixi Zhang, Qi Liu, Hao Wang, Chengqiang Lu, and Chee-Kong Lee. Motif-based graph self-supervised learning for molecular property prediction. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 15870–15882, 2021.

[Zhao *et al.*, 2023] Tianqi Zhao, Ngan Thi Dong, Alan Hanjalic, and Megha Khosla. Multi-label node classification on graph-structured data. *Trans. Mach. Learn. Res.*, 2023, 2023.

[Zhao *et al.*, 2024] Jitao Zhao, Di Jin, Meng Ge, Lianze Shan, Xin Wang, Dongxiao He, and Zhiyong Feng. FUG: feature-universal graph contrastive pre-training for graphs with diverse node features. 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*, 2024.

[Zheng *et al.*, 2023] Yu Zheng, Ming Jin, Yixin Liu, Lianhua Chi, Khoa Tran Phan, and Yi-Ping Phoebe Chen. Generative and contrastive self-supervised learning for graph anomaly detection. *IEEE Trans. Knowl. Data Eng.*, 35(12):12220–12233, 2023.

[Zhou *et al.*, 2021a] 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 Syst. Appl.*, 180:115063, 2021.

[Zhou *et al.*, 2021b] 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 Syst. Appl.*, 180:115063, 2021.

[Zhu *et al.*, 2020] Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. Deep graph contrastive representation learning. *CoRR*, abs/2006.04131, 2020.

[Zhu *et al.*, 2021] Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. Graph contrastive learning with adaptive augmentation. In *WWW '21: The Web Conference 2021, Virtual Event / Ljubljana, Slovenia, April 19-23, 2021*, pages 2069–2080. ACM / IW3C2, 2021.

[Zhu *et al.*, 2023] Ke Zhu, Minghao Fu, and Jianxin Wu. Multi-label self-supervised learning with scene images. In *IEEE/CVF International Conference on Computer Vision, ICCV 2023, Paris, France, October 1-6, 2023*, pages 6671–6680. IEEE, 2023.