

A Novel Contrastive Signal Generative Framework for Accurate Graph Learning

Anonymous Authors

Abstract—Graph contrastive learning (GCL) has garnered significant attention recently since it learns complex structural information from graphs through self-supervised learning manner. However, prevalent GCL models may suffer from performance degradation due to inappropriate contrastive signals. Concretely, they commonly generate augmented views based on random perturbation, which leads to biased essential structures due to the introduction of noise. In addition, they assign equal weight to both hard and easy sample pairs, thereby ignoring the difference in importance of the sample pairs. To address these issues, this study proposes a novel Contrastive Signal Generative Framework for Accurate Graph Learning (CSG²L) with the following two-fold ideas: a) building a singular value decomposition (SVD)-directed augmented module (SVD-aug) to obtain the global interactions as well as avoiding the random noise perturbation; b) designing a local-global dependency learning module (LGDL) with an adaptive reweighting strategy which can differentiate the effects of hard and easy sample pairs. Extensive experiments on benchmark datasets demonstrate that the proposed CSG²L outperforms the state-of-art baselines. Moreover, CSG²L is compatible with a variety of GNNs.

Index Terms—Graph contrastive learning, Singular value decomposition, Local-global dependency, Adaptive reweighting

I. INTRODUCTION

Graph neural networks (GNNs) have shown great performance in recent years due to their ability to effectively utilize graph structure and node feature information. They have been widely applied in tasks such as node classification [26]–[31], link prediction [1], and graph classification [2]. The core idea of GNN is to aggregate information within the neighborhood of a graph through a message passing mechanism, thereby learning a representation that can capture local graph structure and node features [5]–[7]. This mechanism allows the GNN model to simultaneously utilize node feature information and the topological structure of the graph when performing tasks. However, existing GNN methods face several limitations. For example, in sparse graphs, neighborhood information is often insufficient to provide meaningful signals for representation learning, and traditional GNNs only focus primarily on the local neighborhoods, this local dependency characteristic limits the model’s learning capabilities and therefore makes it difficult to learn robust node representations. In addition, since graph data usually comes from real-world application scenarios, it often contains noisy edges or abnormal nodes. In this case, the simple local feature aggregation mechanism cannot effectively distinguish useful information from noise, resulting in inaccurate graph learning by the model, affecting the final task effect.

To address these issues, GCL has received increasing attention as an emerging self-supervised learning method [14]–[18]. The core idea of GCL is to construct positive and negative sample pairs and learn the representation of nodes or graphs by pulling positive pairs closer and pushing negative pairs apart, thereby improving the quality of representation learning [15]. Unlike traditional supervised learning methods, GCL does not rely on manually labeled data and can mine useful information from unlabeled data [3]. In addition, GCL can alleviate the problems of data sparsity and missing labels to a certain extent by comparing positive and negative sample pairs [4].

Some studies have shown that contrastive learning can benefit from hard sample pairs that are indistinguishable from anchor samples [19]. Distinguishing between hard and easy sample pairs is crucial for enhancing the effectiveness of contrastive learning. Hard sample pairs are those that are misclassified during training and contain valuable information that can help the model capture complex relationships within the graph. Learning from these pairs allows the model to improve its ability to discriminate between subtle differences. For easy sample pairs that the model can easily and correctly classify, their learning value is relatively low because they have limited contribution to the discrimination task.

Based on some existing research, we found that they usually face two main problems:

1. When generating contrastive views, they often rely on random perturbations or simple graph structure transformations. Although this strategy can generate diverse contrastive views, it may lose some important graph structure information and introduce noise, which hinder the model’s ability to learn accurate contrastive signals.

2. During the training process, all hard and easy sample pairs are usually assigned the same weight, assuming that all sample pairs contribute equally to the learning of the model, ignoring the difference between hard sample pairs and easy sample pairs. Hard sample pairs contain more useful information, and simply treating all sample pairs equally often leads to the model failing to fully learn the complex structure and potential information of hard samples, limiting the effect of graph learning.

To overcome these challenges, we propose CSG²L, a novel graph contrastive signal generative framework which consists of two essential components: a SVD-directed augmented module (SVD-aug) and a local-global dependency learning module (LGDL). The SVD-aug module captures global collaborative signals, while the LGDL module integrates local and global information. Specifically, in the SVD-aug module, we first

perform SVD on the original graph structure to obtain a low-rank approximation of the graph, thereby generating an augmented graph containing global information. On this basis, the LGDL module constructs contrastive sample pairs by using the augmented graph and the original graph. This approach ensures that the model learns both local and global graph interactions. To further optimize the learning process, we introduce an adaptive reweighting strategy into the LGDL module that assigns greater importance to hard sample pairs during training. Through these methods, the model can generate more accurate graph contrastive signals and improve the robustness of learned representations.

Our contributions are as follows:

- We propose a novel graph contrastive signal generative framework CSG^2L which can be combined with existing GNN models. It includes two modules, i.e., SVD-aug and LGDL, to effectively generate accurate graph contrastive signals. 一个整合现有模型的框架
- We build a SVD-directed augmented module (SVD-aug). By utilizing SVD-directed global augmentation, CSG^2L alleviates the impact of noisy contrastive signals and captures both local and global interactions. 全局模块SVD 降噪模块 CSG^2L
- We design a local-global dependency learning module (LGDL) by introducing an adaptive reweighting strategy, which can effectively improve the performance of CSG^2L in the contrastive learning process and generate more accurate contrastive signals. 局部自适应模块LGDL
- We conduct extensive experiments on six real datasets, and the results show that CSG^2L significantly improves the performance of GNN models, fully proving the superiority of our framework.

II. RELATED WORK

A. Graph Neural Networks

Graph neural networks have made significant progress in processing non-Euclidean structured data, with wide applications in node classification, link prediction, and graph classification. Classic models include GCN [5], GraphSAGE [6] and GAT [7] etc., which capture local structural through efficient neighbor aggregation mechanisms. GCN uses the spectral method to perform weighted averaging of neighbor node features, aiming to improve the expressiveness of node representation through Laplace smoothing. GraphSAGE introduces a neighbor sampling method on this basis, thereby improving the scalability to large-scale graph. GAT introduces the attention mechanism, allowing the model to assign different weights to different neighbor nodes, and then dynamically adjusting the effect of message passing. Although these methods perform well on many tasks, they primarily rely on local information aggregation, which can lead to the **neglect of global structural patterns in the graph**, particularly obvious in scenarios **with sparse data or high noise**. To address this limitation, recent studies have explored enhancing GNNs with global graph signals. For example, The SIGN [8] model injects global information into the node representation through a global feature

expansion mechanism, effectively capturing low-frequency signals; GloGNN [9] uses **spectral methods** to extract low-frequency structural information of the graph, enhancing the model's ability to capture global collaborative relationship. In addition, ChebNet [10] approximates the Laplace matrix through Chebyshev polynomials, thereby capturing global structural information while retaining local information; SGC [11] further simplifies GCN and directly aggregates global neighbor node information through linearized convolution operations, effectively reducing computational complexity. In addition to **spectral domain methods**, some researchers have also explored the effective use of global information in the spatial domain. For example, Graph Heat Kernel [12] and PPR (Personalized PageRank) [13] define a global propagation matrix based on the propagation mechanism between nodes to capture the collaborative relationship between long-distance nodes. These works have made progress in global structure modeling, **but how to efficiently combine global information with local information remains an issue worthy of further exploration.**

B. Graph Contrastive Learning

As a self-supervised learning method, graph contrastive learning can effectively mine the latent features of graphs and reduce dependence on manual labels. Existing research mainly focuses on the generation of contrastive views and the optimization of contrastive loss functions, and proposes a variety of innovative solutions to improve model performance. In terms of view generation, GraphCL [14] generates different contrasting views **by randomly perturbing nodes, edges, and features**, enhancing the diversity of the model. GCA [15] further introduces an **enhancement strategy** based on importance weights to avoid the risk of invalid perturbations. However, these methods generally rely on randomness and not preserve the global interactions of the graph effectively. In addition, these augmented methods are sensitive to noise, and their performance will decrease significantly especially with noisy or low-quality data, therefore generate inaccurate contrastive signals. For contrastive loss, InfoNCE loss is one of the most widely used loss function. It effectively improves the distinguishability of node representation by maximizing the similarity of positive sample pairs and minimizing the similarity of negative sample pairs [14], [15], but this method usually gives the same weight to all sample pairs, **ignoring the importance of hard sample pairs for model learning**. In order to improve the learning ability of hard sample pairs, SUGAR [16] effectively improves the quality of contrastive learning signals by comparing the potential distribution information of the graph. BGRL [17] simplifies the optimization process by removing negative sample pairs. ProGCL [18] introduce a refined approach to build a more suitable measure for the hard negative sample pairs together with similarity. Although these methods have achieved certain results, the **weight allocation of hard and easy sample pairs still lacks flexibility**, and it is **difficult to effectively capture the importance of hard sample pairs, leading to inaccurate graph learning.**

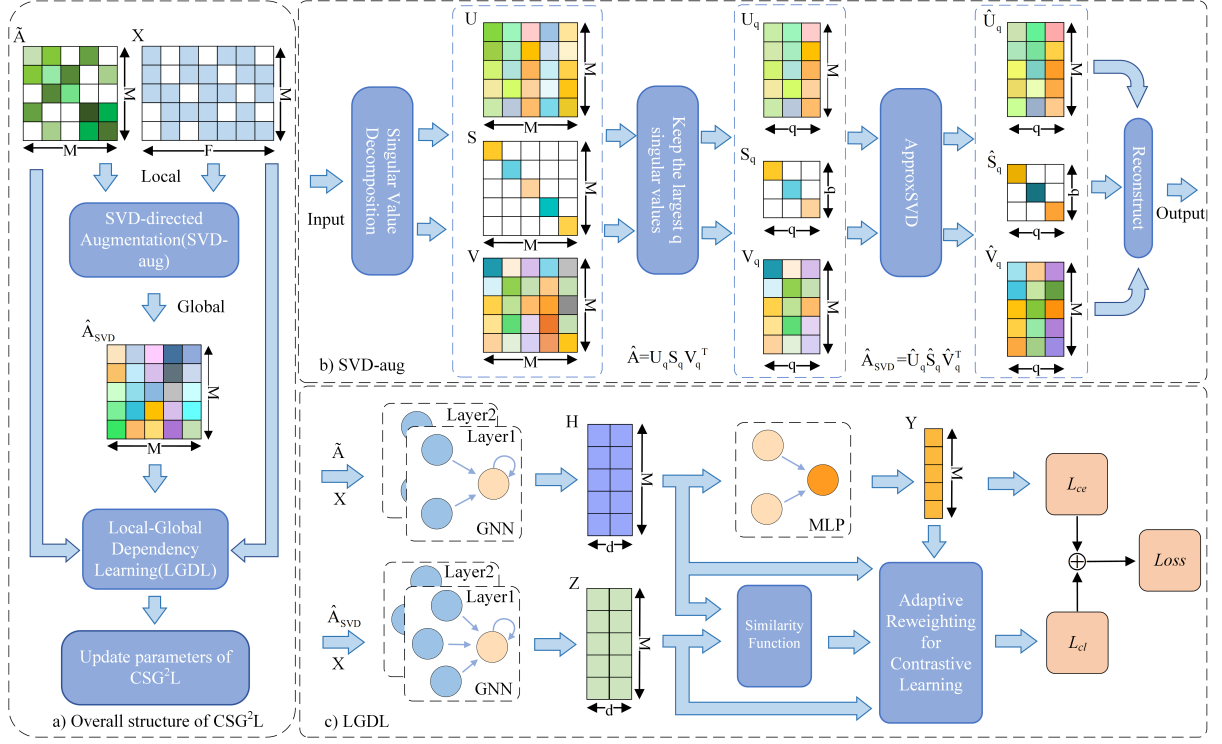


Fig. 1. a) The overall structure of CSG²L. It includes two components: b) SVD-aug module performs SVD-directed reconstruction of the original graph to obtain augmented graph. c) LGDL module inputs both graphs into a shared GNN, and further input the embedding of original graph into an MLP classifier to compute the classification loss. In addition, an adaptive reweighting strategy is introduced for contrastive learning. Finally, the overall objective function can be obtained.

III. METHODOLOGY

We now introduce our CSG²L model, which captures global interactions and introduces an adaptive reweighting strategy into contrastive learning, aiming to address the main challenges associated with contrastive learning for accurate graph learning. As shown in Fig. 1, CSG²L consists of two components: a SVD-directed augmented module (SVD-aug) and a local-global dependency learning module (LGDL). In this section, we first define some basic concepts and then discuss the details of the two components.

A. Problem Definition

We consider a graph $G = (N, E, X, A)$, where $N = \{n_1, n_2, \dots, n_M\}$ and $E \subseteq N \times N$ represent the set of nodes and the set of edges, M is the number of nodes. The node feature matrix is denoted as $X \in \mathbb{R}^{M \times F}$, where F is the number of node feature dimensions. The adjacency matrix is denoted as $A = a_{ij} \in \{0, 1\}^{M \times M}$, $a_{ij} = 1$ when $(n_i, n_j) \in E$, the diagonal matrix is $D = \text{diag}(d_1, d_2, \dots, d_M) \in \mathbb{R}^{M \times M}$. And we define the normalized adjacency matrix as $\tilde{A} \in \mathbb{R}^{M \times M}$ and $\tilde{A} = D^{-1/2}(A + I)D^{-1/2}$, where $I \in \mathbb{R}^{M \times M}$ is an identity matrix.

B. SVD-directed Augmented Module

Global interactions plays a crucial role in capturing long-range dependencies and enhancing the expressiveness of node representations in complex graph structures. By effectively

utilizing global interactions, models can better understand **long-distance relationships** between nodes, leading to improved performance in graph learning tasks. However, existing graph contrastive learning methods usually rely on random augmentations with local neighborhood information, ignoring **the global structure of the graph**. To address these limitations, we build a SVD-directed augmented module (SVD-aug) to extract global information from the original graph and generate an augmented view that incorporate **richer global interactions**. Specifically, SVD-aug consists of two steps: graph reconstruction and approximation of SVD.

Graph Reconstruction. To effectively capture the global interactions and enhance the ability to model global signals, we adopts **SVD** [20] to extract important graph structure information from the **global perspective**. Specifically, we perform **SVD on the normalized adjacency matrix \tilde{A}** , decomposing it into two orthogonal matrices and a singular value matrix, denoted as $\tilde{A} = USV^T$. Here, U and V are both $M \times M$ standard orthogonal matrices, with their columns representing the eigenvectors of the row-row and column-column correlation matrix of \tilde{A} . S is an $M \times M$ diagonal matrix storing the singular values of \tilde{A} , which reflect the important information of the matrix in different directions. The singular vectors corresponding to the top-ranked singular values usually contain the main information. The largest singular values are associated with the principal components of the matrix. Retaining the largest singular values and their corresponding

singular vectors preserve the key information of the graph while reducing noise and ensuring effective reconstruction. Thus, we truncate the singular value list to keep the largest q singular values. Using these truncated singular values and their corresponding singular vectors, we reconstruct the normalized adjacency matrix as $\hat{A} = U_q S_q V_q^T$, where $U_q \in \mathbb{R}^{M \times q}$ and $V_q \in \mathbb{R}^{M \times q}$ contain the first q columns of U and V respectively, $S_q \in \mathbb{R}^{q \times q}$ is the diagonal matrix of the q largest singular values. The reconstructed matrix \hat{A} is a low-rank approximation of the normalized adjacency matrix \tilde{A} .

Approximation of SVD. However, performing exact SVD on the normalized adjacency matrix of large graphs is highly expensive and may result in significant loss of efficiency. Therefore, we use the randomized algorithm which is usually used in the ApproxSVD algorithm [21]. This method first transform the original matrix into a low-rank orthogonal matrix, and continuously approximate the range of the original matrix, and then SVD is applied to this smaller-scale matrix:

$$\hat{U}_q, \hat{S}_q, \hat{V}_q^T = \text{ApproxSVD}(\tilde{A}, q), \quad \hat{A}_{SVD} = \hat{U}_q \hat{S}_q \hat{V}_q^T. \quad (1)$$

The graph structure learning based on SVD can effectively reduce the influence of noise in the network, demonstrating strong robustness in most scenarios, and can also maintain the global information while considering the characteristics of each sample pair. q值是如何确定的？

C. Local-Global Dependency Learning

In graph contrastive learning, accurate contrastive signals are crucial for model learning. Effective contrastive signals can not only improve model performance but also help discriminate subtle differences between sample pairs more effectively. The combination of local and global information plays a critical role in generating high-quality contrastive signals. Local information captures neighborhood relationship, while global information provides an overall perspective of the graph. The combination of the two can improve the model's discriminative ability, leading to more effective graph representation learning. However, most existing graph contrastive learning methods usually focus on generating randomly augmented views and then treat all sample pairs with equal importance during training. This approach does not consider local and global information which may result in inaccurate contrastive signals and hinder model performance. To address this issue, we design a local-global dependency learning module (LGDL) to integrate local and global information, improve the discrimination between hard and easy sample pairs, and then generate more accurate contrastive signals. Specifically, LGDL contains two steps: local-global encoding and adaptive reweighting contrastive learning. 混合全局和局部信息生成高质量信号

Local-Global Encoding. GNNs capture the complex structural features in graphs by propagating information between neighbors and integrating local structure and features into high-dimensional node embeddings. In this module, we calculate node embeddings based on the original graph and the reconstructed graph respectively to better combine global information and local relationship between nodes, providing

stronger representation capabilities for contrastive learning. Specifically, we input the original graph node features X and normalized adjacency matrix \tilde{A} into a GNN model $f(\cdot, \cdot)$, and calculate the node embeddings through two-layer message propagation:

$$H = f(\tilde{A}, X), \quad (2)$$

where $H \in \mathbb{R}^{M \times m}$ represents node embeddings in the original graph, and m is the dimension of the embedding. In our framework, a variety of GNN models can be used, such as GCN [5], GIN [25], and GPRGNN [26]. Then, we rewrite the message propagation rule in equation (2) with the approximated matrices in equation (1):

$$Z = f(\hat{A}_{SVD}, X) = f(\hat{U}_q \hat{S}_q \hat{V}_q^T, X), \quad (3)$$

where $Z \in \mathbb{R}^{M \times m}$ is node embeddings of the reconstructed graph. In this way, the nodes high-order features can further integrate local and global information during the propagation process of the GNN.

Adaptive Reweighting Contrastive Learning. Graph contrastive learning is a self-supervised learning method that can improve the model's ability to discriminate features by measuring the similarity of sample pairs. In traditional graph contrastive learning, InfoNCE loss is a commonly used optimization objective. Its key idea is to learn effective representations by minimizing the distance between positive sample pairs and maximizing the distance between negative ones. The InfoNCE loss is formulated as:

$$l_{info}(h_i, z_i) = -\log \frac{e^{\theta(h_i, z_i)/\tau}}{e^{\theta(h_i, z_i)/\tau} + \sum_{k \neq i} (e^{\theta(h_i, z_k)/\tau} + e^{\theta(h_i, h_k)/\tau})}, \quad (4)$$

where h_i, z_i denote the embeddings of the i -th sample on the original graph and the reconstructed graph, and τ is a temperature parameter. Besides, we define $\theta(h_i, z_i) = s(g(h), g(z))$, $s(\cdot, \cdot)$ is the cosine similarity and $g(\cdot)$ is a non-linear projection to enhance the expression power of the critic function. The disadvantage of InfoNCE loss is that it treats both hard sample pairs and easy sample pairs equally. This strategy may lead to insufficient learning for hard samples which can weaken the model's representation learning. At the same time, over-optimization of easy samples wastes computing resources and limits the model's discriminative capability. To address these issues, we introduce a reweighting function R to adaptively adjust the weights of sample pairs. Concretely, after encoding, the node embeddings H learned from the original graph are input into an MLP classifier $c(\cdot)$ to obtain the prediction $P = c(H) \in \mathbb{R}^{M \times C}$, where C denotes the number of classes. Since the prediction may be noisy, we only consider the nodes with the most confident prediction for each class, that is, for the high-confidence nodes whose prediction values for a specific class exceed a given threshold, we calculate their

pseudo-labels $Y \in \mathbb{R}^M$ based on P , and then use Y to obtain the sample pair pseudo-label matrix $Q \in \mathbb{R}^{M \times M}$ as follows:

$$Q_{ik} = \begin{cases} 1, & Y_i = Y_k, \text{ 两两标签相同为1} \\ 0, & Y_i \neq Y_k, \end{cases} \quad (5)$$

where Q_{ik} represents the pseudo-label relationship between the i -th sample and the k -th sample, that is, when the pseudo-labels are the same ($Q_{ik} = 1$), they are more likely to be the positive sample pair, and when the pseudo-labels are not the same ($Q_{ik} = 0$), it means that they are more likely to be the negative sample pair.

Based on the similarity function s and the generated pseudo-label matrix Q , we define the **reweighting function** R as:

$$R(h_i, z_k) = \begin{cases} 1, & \text{others,} \\ |Q_{ik} - \text{Norm}(s(h_i, z_k))|, & i, k \in O, \end{cases} \quad (6)$$

where $\text{Norm}(\cdot)$ denotes the min-max normalization and **O denotes the high-confidence nodes**. When the sample pair is not with high confidence, we keep the original InfoNCE setting for it. For high-confidence sample pairs, i.e., $i, k \in O$, their weights are adaptively reweighted with the pseudo-labels and the samples similarity. Concretely, when the i -th and k -th sample nodes are identified as a positive sample pair ($Q_{ik} = 1$), if the calculated similarity of their embeddings after passing through the GNN model is low, it means that the positive sample pair is likely to be a hard sample pair. Therefore, a larger weight is given to the sample pair, and the positive sample pairs with high similarity (the easy sample pairs) are down-weighted. For negative sample pair with different pseudo-labels ($Q_{ik} = 0$), **if their embedding similarity is low, it means that the negative sample pair is likely to be an easy sample pair, so a smaller weight is given to the sample pair**, while negative sample pairs with higher similarity (the hard sample pairs) are up-weighted. This adaptive reweighting strategy focuses more on hard sample pairs, less on easy sample pairs, thereby further improving the discriminative capability of sample pairs and generating more accurate contrastive signals.

Based on R and θ , the reweighted InfoNCE can be defined as follows:

$$l(h_i, z_i) = -\log\left(\frac{e^{(R(h_i, z_i) \cdot \theta(h_i, z_i)) / \tau}}{e^{(R(h_i, z_i) \cdot \theta(h_i, z_i)) / \tau} + \sum_{k \neq i} e^{(R(h_i, z_k) \cdot \theta(h_i, z_k)) / \tau} + \sum_{k \neq i} e^{(R(h_i, h_k) \cdot \theta(h_i, h_k)) / \tau}}\right), \quad (7)$$

The overall contrastive learning objective is the average over all positive pairs, formally given by:

$$L_{cl} = \frac{1}{2M} \sum_{i=1}^M [l(h_i, z_i) + l(z_i, h_i)]. \quad (8)$$

Finally, we jointly optimize the main objective function of the node classification model with the contrastive loss:

$$L = L_{ce} + \lambda L_{cl}, \quad (9)$$

TABLE I
DATASETS STATISTICS

Datasets	#Nodes	#Edges	#Features	#Classes
Wisconsin	251	499	1703	5
Texas	183	309	1703	5
Cornell	183	295	1703	5
Chameleon	2277	36101	2325	5
Cora	2708	5429	1433	7
Citeseer	3327	4732	3703	6

where L_{ce} denotes the cross entropy loss and λ is the weight of the contrastive loss L_{cl} .

IV. EXPERIMENTS

In this section, we evaluate the performance of the CSG²L framework on node classification tasks on six real-world benchmark datasets. We first introduce the datasets, baseline models, and hyper-parameter settings. Then, we compare CSG²L with the baseline models and analyze the results. Finally, hyper-parameter analysis and comprehensive ablation study are performed to thoroughly demonstrate the effectiveness of our method.

A. Datasets

We evaluate the proposed CSG²L on six widely used real-world datasets, including four heterophilic graph datasets Wisconsin, Texas, Cornell [22], and Chameleon [23], and **two homophilic graph** datasets Cora and Citeseer [24]. Table I summarizes the characteristics of all datasets in detail. For all benchmarks, we use the same feature vectors, graph structures, and class labels provided in [22], and perform standard 10 fixed random split of the datasets (60%/20%/20% of nodes per class for training/validation/testing). When the validation accuracy of the trained model reaches a maximum value of each run, we start the testing procedure and calculate the average test accuracy of these ten runs.

B. Baselines

We compare the proposed model with ten baseline models, including four GNN models: GCN [5], GAT [7], GIN [25], GraphSAGE [6], and six recent state-of-the-art special structure GNN models tackling heterophily: MixHop [27], H2GCN [28], GPRGNN [26], GOAL [29] OrderedGNN [30], ES-GNN [31]. In order to demonstrate the effectiveness and generalizability of our framework, we select three models, GCN, GIN, and GPRGNN, as the enhanced models.

C. Hyper-parameter Setting

We uniformly set the hyper-parameters of all models as follows: the number of GNN layers is 2, the hidden layer dimension is 64, the temperature parameter is 0.5, the rank of SVD is 5, the dropout rate is set to 0.5, the learning rate is set to 0.05, the weight decay is selected from the set $\{1e-3, 5e-3, 5e-4, 5e-5, 5e-6\}$, and the Adam optimizer is adopted. For other hyper-parameters, we use their best default parameters in the original paper. For fair comparison, in the

TABLE II
AVERAGE ACCURACY(%) AND STANDARD DEVIATION FOR TEN DIFFERENT DATA SPLITS ON SIX REAL-WORLD GRAPH DATASETS. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD. THE IMPROVEMENT OF THE CSG²L MODEL RELATIVE TO ITS COUNTERPART IS CALCULATED, WITH THE IMPROVEMENT INDICATED BY AN UPWARD ARROW \uparrow .

Method	Wisconsin	Texas	Cornell	Chameleon	Cora	Citeseer	Average
GAT	56.31 \pm 0.55	58.27 \pm 1.28	56.59 \pm 2.41	54.69 \pm 1.95	86.13 \pm 0.84	75.46 \pm 1.37	64.58
GraphSAGE	79.09 \pm 3.23	78.02 \pm 4.71	67.12 \pm 4.27	58.73 \pm 1.68	86.30 \pm 1.02	76.14 \pm 0.98	74.23
MixHop	75.88 \pm 4.90	77.84 \pm 7.73	73.51 \pm 6.34	60.50 \pm 2.53	83.10 \pm 2.03	73.05 \pm 2.65	73.98
H2GCN	82.08 \pm 3.22	79.70 \pm 5.16	78.16 \pm 4.05	56.85 \pm 1.68	86.26 \pm 1.08	75.51 \pm 1.33	76.43
GOAL	73.55 \pm 3.04	74.77 \pm 4.11	73.88 \pm 3.29	67.32 \pm 0.66	85.92 \pm 0.80	74.77 \pm 0.46	75.04
OrderedGNN	79.46 \pm 3.79	78.31 \pm 4.64	79.47 \pm 1.18	65.28 \pm 1.15	84.48 \pm 0.91	73.38 \pm 0.67	76.73
ES-GNN	75.50 \pm 4.21	75.47 \pm 2.15	74.06 \pm 4.57	67.58 \pm 0.63	84.17 \pm 1.12	73.47 \pm 0.56	75.04
GPRGNN	83.25 \pm 0.63	81.31 \pm 1.03	79.93 \pm 0.92	67.07 \pm 0.53	85.32 \pm 0.26	76.06 \pm 0.53	78.82
CSG²L_{GPRGNN}	86.08\pm1.53	85.57\pm0.66	82.98\pm0.64	68.12\pm1.02	86.89\pm0.49	76.79\pm0.27	81.07
<i>improvements</i>	\uparrow 2.83	\uparrow 4.26	\uparrow 3.05	\uparrow 1.05	\uparrow 1.57	\uparrow 0.73	\uparrow 2.25

TABLE III
IMPROVEMENT RESULTS OF CSG²L WITH THREE BASELINE GNN MODELS.

Method	Wisconsin	Texas	Cornell	Chameleon	Cora	Citeseer	Average
GCN	57.84 \pm 5.69	59.69 \pm 5.18	50.56 \pm 3.09	58.65 \pm 2.73	85.10 \pm 0.62	73.21 \pm 1.26	64.18
CSG²L_{GCN}	61.37\pm3.35	62.16\pm3.84	54.28\pm2.81	66.19\pm1.56	85.59\pm0.53	73.73\pm1.51	67.22
<i>improvements</i>	\uparrow 3.53	\uparrow 2.47	\uparrow 3.72	\uparrow 7.54	\uparrow 0.49	\uparrow 0.52	\uparrow 3.04
GIN	76.35 \pm 1.30	73.87 \pm 2.86	75.40 \pm 5.19	58.53 \pm 2.24	81.17 \pm 0.36	74.35 \pm 0.25	73.28
CSG²L_{GIN}	83.66\pm2.71	78.38\pm1.89	77.66\pm3.83	66.16\pm1.30	82.29\pm0.64	74.84\pm0.36	77.17
<i>improvements</i>	\uparrow 7.31	\uparrow 4.51	\uparrow 2.26	\uparrow 7.63	\uparrow 1.12	\uparrow 0.49	\uparrow 3.89
GPRGNN	83.25 \pm 0.63	81.31 \pm 1.03	79.93 \pm 0.92	67.07 \pm 0.53	85.32 \pm 0.26	76.06 \pm 0.53	78.82
CSG²L_{GPRGNN}	86.08\pm1.53	85.57\pm0.66	82.98\pm0.64	68.12\pm1.02	86.89\pm0.49	76.79\pm0.27	81.07
<i>improvements</i>	\uparrow 2.83	\uparrow 4.26	\uparrow 3.05	\uparrow 1.05	\uparrow 1.57	\uparrow 0.73	\uparrow 2.25

enhanced GNNs(GCN, GIN, GPRGNN), the hyper-parameter settings are the same as those in their counterparts, ensuring a consistent evaluation. We set the weight of contrastive learning loss λ in equation (9) to 0.1 and discuss it later.

D. Results on Node Classification

We use the enhanced model of GPRGNN to perform experimental results on node classification tasks on 6 datasets, and compare with the corresponding baseline and GNN-based SOTA models, as shown in Table II. The average accuracy and standard deviation of ten different data splits are presented, and the best results are highlighted in bold. CSG²L achieves significant performance by combining with the SVD-aug module and the LGDL module. It can be observed that although GPRGNN performs well on some datasets (Wisconsin, Texas, Cornell), it is still inferior to the SOTA model on other datasets. Our CSG²L can help GPRGNN obtain the most competitive results, with the best results on four heterophilic datasets, and even a good improvement on homophilic datasets, with an overall performance improvement of 2.25% on average. This is attributed to the fact that our model can fully combine with local and global information, and further improve the performance by considering the weights of hard and easy sample pairs.

In addition, we also report the improvement results of CSG²L in Table III. The results in the table show that CSG²L consistently improves the performance of three baseline GNN

models on six datasets. Specifically, the proposed framework improves the performance of GCN by 3.04% on average, GIN by 3.89% on average, GPRGNN by 2.25% on average. This result means that CSG²L is an effective general framework that can improve the performance of well-known GNN models that are already recognized to be effective.

E. Hyper-parameter Analysis

We achieve the simultaneous improvement of the node representation learning quality and classification task performance by jointly optimizing the classification loss and contrastive loss. As defined in equation (9), the **hyper-parameter λ is used to adjust the weight of the contrastive loss in the overall loss function**, thereby balancing the contribution of the classification task and contrastive learning. In order to deeply analyze the impact of λ on model performance, we conduct sufficient experiments using CSG²Ls on three representative benchmark datasets: Cora, Texas, and Chameleon.

The experimental results are shown in Fig. 2. It can be observed from the figure that when $\lambda = 0.01$, the performance of our model has not yet reached the optimal level. This is because the weight of the contrastive loss is low at this time, preventing our model from fully utilizing the representation information learned by contrastive learning during the optimization process. As λ gradually increases, the model performance gradually improves. When $\lambda = 0.1$, the model achieves the best average accuracy on the three datasets,

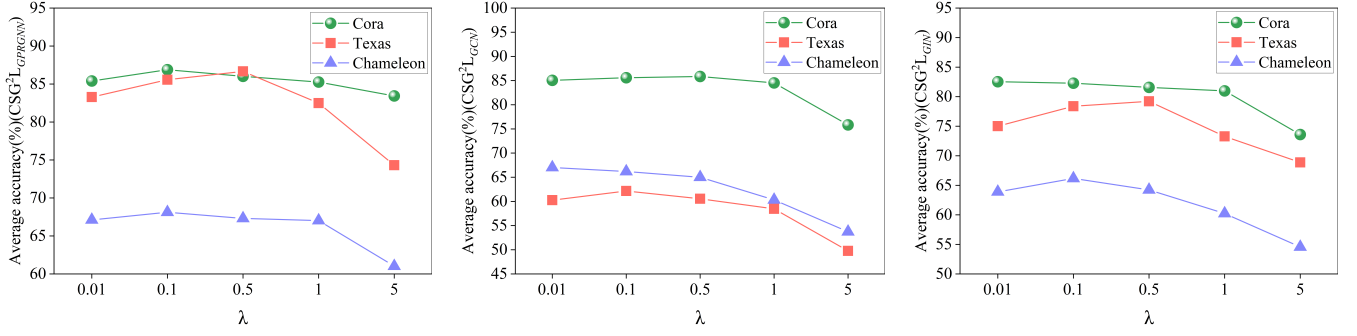


Fig. 2. Hyper-parameter analysis results of CSG²Ls on three benchmark datasets.

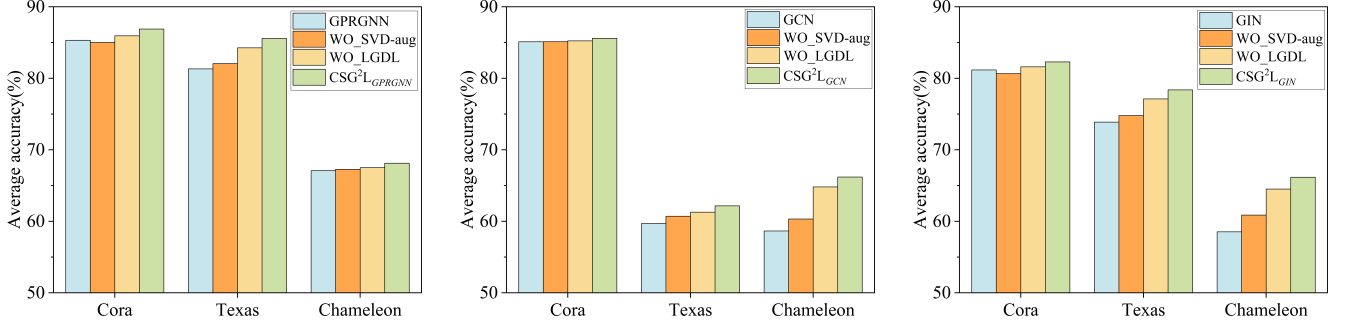


Fig. 3. Ablation study results on three benchmark datasets.

indicating that the contrastive loss and classification loss have reached a good balance at this point. In this case, contrastive learning can effectively explore the structural information and global interactions between nodes, prompting the model to learn a more robust node representation without disrupting the optimization of the classification task. However, when λ continues to increase, the performance begins to decline significantly, especially when $\lambda = 5$, the performance drops to the lowest level. This is because the contrastive loss dominates the overall loss function, resulting in a significant weakening of the contribution of the classification loss. In this case, the model training focuses excessively on bringing the representation of positive sample pairs closer and pulling the representation of negative sample pairs farther away, but ignores the classification objective, which ultimately affects the overall performance. Based on the above analysis, we choose the default value of $\lambda = 0.1$.

F. Ablation Study

In order to evaluate the effectiveness of our proposed CSG²L framework in improving the performance of GNN models, we conduct ablation studies on three benchmark datasets. The goal is to analysis the trend of model performance by gradually removing the two different components, and further explore the contribution and importance of each module in the overall framework. The experiment results are shown in Fig. 3, which shows the ablation study results of the three basic models of GPRGNN, GCN and GIN on three datasets. From the figure, our analysis is as follows:

1) *Original model performance*: In the baseline model, without both SVD-aug and LGDL modules, the model only uses the node feature matrix X and the normalized adjacency matrix \tilde{A} as inputs for node classification, and the performance is relatively low. This is because the original model does not fully capture global interactions and is easily affected by data sparsity and noise, resulting in poor distinguishability of node representation.

2) *W/O SVD-aug*: In this step, SVD-aug is removed. We introduce random augmentation such as edge perturbations and attribute masking to generate augmented graph, and InfoNCE loss is used to contrast embeddings of the original and augmented graphs, the average performance of the three models is improved. This shows that through simple random perturbations, the model can capture more local structural change information and enhance the robustness of node representation. However, since the transformation introduced by random augmentations lacks pertinence and fails to effectively capture the global collaborative signals, the performance improvement is relatively limited.

3) *W/O LGDL*: In this step, LGDL module is removed. When only introducing the SVD-aug module, the model uses both the original normalized adjacency matrix \tilde{A} and the reconstructed normalized adjacency matrix \hat{A}_{SVD} , along with the node feature matrix X as inputs, and then InfoNCE loss is used to contrast embeddings of the original and augmented graphs. The model performance is further improved because through using low-rank reconstruction to obtain an augmented graph, SVD can effectively capture the global interactions,

which makes up for the limitations of random augmentations. Combined with contrastive learning, the embeddings of positive pairs are brought closer, while negative pairs are pushed apart, leading to more discriminative node representations.

4) CSG^2L s: On the basis of the SVD-aug module, after introducing the LGDL module with an adaptive reweighting strategy, the three baseline models achieve the best performance. By adaptively adjusting the weights of hard and easy sample pairs, the strategy improves the optimization process of InfoNCE loss and mitigates the effects of noisy and inaccurate contrastive signals. In addition, combined with the SVD-aug module and the LGDL module, CSG^2L further improves the robustness and effectiveness of node representations for classification while maintaining global information.

In summary, the ablation study results demonstrate the effectiveness and necessity of each component in CSG^2L . By utilizing the SVD-aug module and the LGDL module, CSG^2L significantly improves the performance of the basic GNN models, exhibiting strong generalization ability and robustness in node classification tasks.

V. CONCLUSION

In this paper, we proposed a novel graph contrastive signal generative framework, CSG^2L , designed to address the limitations of traditional graph learning methods in node classification tasks. CSG^2L contains two components, we build a SVD-aug module which uses low-rank SVD to generate augmented graph that captures global interactions. In addition, we design a LGDL module, which integrates local and global information, optimizing the contrastive learning loss by introducing an adaptive reweighting strategy to further improve the effect of contrastive learning, effectively mitigating the issues of insufficient global interaction extraction and inaccurate contrastive signals. In future work, we plan to integrate attribute information into SVD-aug module to further improve the accuracy of the augmented graph.

REFERENCES

- [1] Y. Wang, X. Hu, Q. Gan, X. Huang, X. Qiu, and D. Wipf, "Efficient link prediction via gnn layers induced by negative sampling," *IEEE Transactions on Knowledge and Data Engineering*, 2024.
- [2] L. Wei, H. Zhao, Z. He, and Q. Yao, "Neural architecture search for GNN-based graph classification," *ACM Transactions on Information Systems*, vol. 42, no. 1, pp. 1–29, 2023.
- [3] W. Ju et al., "Towards Graph Contrastive Learning: A Survey and Beyond," *arXiv preprint arXiv:2405.11868*, 2024.
- [4] Y. Liu, K. Ding, J. Wang, V. Lee, H. Liu, and S. Pan, "Learning strong graph neural networks with weak information," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023, pp. 1559–1571.
- [5] M. Welling and T. N. Kipf, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. on Learning Representations*, 2016.
- [6] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," *Advances in neural information processing systems*, vol. 30, 2017.
- [7] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," in *Proceedings of the International Conference on Learning Representations*, 2018.
- [8] F. Frasca, E. Rossi, D. Eynard, B. Chamberlain, M. Bronstein, and F. Monti, "Sign: Scalable inception graph neural networks," *arXiv preprint arXiv:2004.11198*, 2020.
- [9] X. Li et al., "Finding global homophily in graph neural networks when meeting heterophily," in *International Conference on Machine Learning*, PMLR, 2022, pp. 13242–13256.
- [10] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," *Advances in neural information processing systems*, vol. 29, 2016.
- [11] F. Wu, A. Souza, T. Zhang, C. Fifty, T. Yu, and K. Weinberger, "Simplifying graph convolutional networks," in *International conference on machine learning*, PMLR, 2019, pp. 6861–6871.
- [12] J. Gasteiger, S. Weißenberger, and S. Günnemann, "Diffusion improves graph learning," *Advances in neural information processing systems*, vol. 32, 2019.
- [13] J. Gasteiger, A. Bojchevski, and S. Günnemann, "Predict then Propagate: Graph Neural Networks meet Personalized PageRank," in *ICLR*, Feb. 2019.
- [14] Y. You, T. Chen, Y. Sui, T. Chen, Z. Wang, and Y. Shen, "Graph contrastive learning with augmentations," *Advances in neural information processing systems*, vol. 33, pp. 5812–5823, 2020.
- [15] Y. Zhu, Y. Xu, F. Yu, Q. Liu, S. Wu, and L. Wang, "Graph Contrastive Learning with Adaptive Augmentation," in *Proceedings of the Web Conference 2021*, Apr. 2021, pp. 2069–2080. doi: 10.1145/3442381.3449802.
- [16] Q. Sun et al., "Sugar: Subgraph neural network with reinforcement pooling and self-supervised mutual information mechanism," in *Proceedings of the web conference 2021*, 2021, pp. 2081–2091.
- [17] S. Thakoor, C. Tallec, M. G. Azar, R. Munos, P. Veličković, and M. Valko, "Bootstrapped representation learning on graphs," in *ICLR 2021 Workshop on Geometrical and Topological Representation Learning*, 2021.
- [18] J. Xia, L. Wu, G. Wang, J. Chen, and S. Z. Li, "ProGCL: Rethinking Hard Negative Mining in Graph Contrastive Learning," in *International Conference on Machine Learning*, PMLR, 2022, pp. 24332–24346.
- [19] G. Chu, X. Wang, C. Shi, and X. Jiang, "CuCo: Graph Representation with Curriculum Contrastive Learning," in *IJCAI*, 2021, pp. 2300–2306.
- [20] A. Rajwade, A. Rangarajan, and A. Banerjee, "Image denoising using the high order singular value decomposition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 4, pp. 849–862, 2012.
- [21] N. Halko, P.-G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM review*, vol. 53, no. 2, pp. 217–288, 2011.
- [22] H. Pei, B. Wei, K. C.-C. Chang, Y. Lei, and B. Yang, "Geom-gcn: Geometric graph convolutional networks," *arXiv preprint arxiv:2002.05287*, 2020.
- [23] B. Rozemberczki, C. Allen, and R. Sarkar, "Multi-scale attributed node embedding," *Journal of Complex Networks*, vol. 9, no. 2, p. cnab014, 2021.
- [24] P. Sen, G. Namata, M. Bilgic, L. Getoor, B. Galligher, and T. Eliassi-Rad, "Collective classification in network data," *AI magazine*, vol. 29, no. 3, pp. 93–93, 2008.
- [25] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?," *arXiv preprint arXiv:1810.00826*, 2018.
- [26] E. Chien, J. Peng, P. Li, and O. Milenkovic, "Adaptive universal generalized pagerank graph neural network," *arXiv preprint arxiv:2006.07988*, 2020.
- [27] S. Abu-El-Haija et al., "Mixhop: Higher-order graph convolutional architectures via sparsified neighborhood mixing," in *international conference on machine learning*, PMLR, 2019, pp. 21–29.
- [28] J. Zhu, Y. Yan, L. Zhao, M. Heimann, L. Akoglu, and D. Koutra, "Beyond homophily in graph neural networks: Current limitations and effective designs," *Advances in neural information processing systems*, vol. 33, pp. 7793–7804, 2020.
- [29] Y. Zheng, H. Zhang, V. Lee, Y. Zheng, X. Wang, and S. Pan, "Finding the missing-half: Graph complementary learning for homophily-prone and heterophily-prone graphs," in *International Conference on Machine Learning*, PMLR, 2023, pp. 42492–42505.
- [30] Y. Song, C. Zhou, X. Wang, and Z. Lin, "Ordered gnn: Ordering message passing to deal with heterophily and over-smoothing," *arXiv preprint arXiv:2302.01524*, 2023.
- [31] J. Guo, K. Huang, R. Zhang, and X. Yi, "ES-GNN: Generalizing graph neural networks beyond homophily with edge splitting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.