

# ROBUST GRAPH DATA LEARNING VIA LATENT GRAPH CONVOLUTIONAL REPRESENTATION

**Anonymous authors**

Paper under double-blind review

## ABSTRACT

Graph Convolutional Representation (GCR) has achieved impressive performance for graph data representation. However, existing GCR is generally defined on the input fixed graph which may restrict the representation capacity and also be vulnerable to the structural attacks and noises. To address this issue, we propose a novel Latent Graph Convolutional Representation (LatGCR) for robust graph data representation and learning. Our LatGCR is derived based on reformulating graph convolutional representation from the aspect of graph neighborhood reconstruction. Given an input graph  $\mathbf{A}$ , LatGCR aims to generate a flexible latent graph  $\tilde{\mathbf{A}}$  for graph convolutional representation which obviously enhances the representation capacity and also performs robustly w.r.t graph structural attacks and noises. Moreover, LatGCR is implemented in a self-supervised manner and thus provides a basic block for both supervised and unsupervised graph learning tasks. Experiments on several datasets demonstrate the effectiveness and robustness of LatGCR.

## 1 INTRODUCTION

Graph Convolutional Networks (GCNs) have been widely studied for graph data representation and learning (Defferrard et al., 2016; Kipf & Welling, 2017; Veličković et al., 2018a;b; Chen et al., 2020; Jiang et al., 2021). Graph convolutional representation (GCR) is the core operation powering GCNs. The aim of GCR is to generate context-aware embeddings for graph nodes by aggregating the messages from their neighbors via some differentiable aggregation functions (Hamilton et al., 2017; Geisler et al., 2020). For example, Kipf et al. (Kipf & Welling, 2017) propose a graph convolution operation by exploring the first-order approximation of graph Laplacian spectral filter. Hamilton et al. (Hamilton et al., 2017) present Graph Sample and Aggregate (GraphSAGE) for inductive graph representation and learning by using graph sampling and aggregating techniques. Petar et al. (Veličković et al., 2018b) propose Deep Graph Infomax (DGI) to learn node’s representation in an unsupervised manner. Klicpera et al. (Johannes Klicpera & Günnemann, 2019) propose Personalized Propagation of Neural Predictions (PPNP) which combines GCN and PageRank together for graph semi-supervised learning. Zhu et al. (Zhu et al., 2019) propose Robust GCN (RGCN) by adopting robust learnable Gaussian distributions for message propagation.

However, the above GCRs are generally defined on the input fixed graph which may restrict the representation capacity and also be very vulnerable to the structural attacks and noises (Zügner & Günnemann, 2019; Jin et al., 2020b; Geisler et al., 2020). To address this issue, one kind of popular way is to incorporate graph learning modules into GNNs’ training via optimizing a general joint loss function, i.e., ‘graph learning loss + GNN’s loss’. For example, Wu et al. (Wu et al., 2019) propose GCN-Jaccard to denoise input graph by deleting the edges with low similarities. Li et al. (Li et al., 2018) propose AdaGCN to learn an optimal graph for GNN learning. Jiang et al. (Jiang et al., 2019) propose Graph Learning Convolutional Network (GLCN) by integrating graph learning module into GCN architecture for semi-supervised learning problem. Yang et al. (Yang et al., 2019) propose Topology Optimization based Graph Convolutional Networks (TO-GCN) for semi-supervised learning by jointly refining the graph structure and learning the parameters of GCN. Jin et al. (Jin et al., 2020b) develop Pro-GNN to adaptively train a more optimal graph for GNN’s learning via designing a joint loss function.

In this paper, we propose a novel Latent Graph Convolutional Representation (LatGCR) for *robust* graph data representation and learning. Our LatGCR is derived based on reformulating GCR from

the aspect of graph neighborhood reconstruction. The main difference between LatGCR and previous related works is that LatGCR gives a basic block which can be used within many GNN architectures by replacing traditional graph convolution layer with LatGCR block. We will discuss the detailed differences between LatGCR and previous related works including recent Pro-GNN (Jin et al., 2020b) and GeCN (Jiang et al., 2021) in §4. Specifically, given an input observed graph  $\mathbf{A}$ , LatGCR aims to generate a flexible latent graph  $\tilde{\mathbf{A}}$  for GCR in a self-supervised manner which enhances the representation capacity and also obviously performs robustly w.r.t graph structural attacks and noises.

Overall, we summarize the main contributions of this paper as follows:

- We propose a novel self-supervised Latent Graph Convolutional Representation (LatGCR) based on the reformulation of GCR from graph neighborhood reconstruction.
- LatGCR can be efficiently implemented via a simple recurrent architecture, i.e., LatGCR block, which provides a general basic block for GNNs.
- Based on the proposed LatGCR block, we propose an end-to-end LatGC neural network (LatGCN) for robust graph data representation and learning.

Experimental results on both semi-supervised classification and unsupervised clustering tasks demonstrate the effectiveness and robustness of the proposed LatGCR and LatGCN.

## 2 REVISITING GRAPH CONVOLUTIONAL REPRESENTATION

As the main aspect of Graph Convolutional Networks (GCNs) for graph data representation and learning, Graph Convolutional Representations (GCRs) have been widely studied in recent years (Defferrard et al., 2016; Kipf & Welling, 2017; Hamilton et al., 2017; Veličković et al., 2018a; Geisler et al., 2020). The aim of GCRs is to generate context-aware representations for graph nodes by aggregating the representations from their neighbors via some specific aggregation functions.

One popular formulation of GCR is to employ the weighted mean aggregation function for neighbor’s information aggregation (Hamilton et al., 2017; Jiang & Zhang, 2021). Let  $G(\mathbf{A}, \mathbf{Z})$  denotes the input graph where  $\mathbf{A} \in \mathbb{R}^{n \times n}$  denotes the adjacency matrix with  $\mathbf{A}_{ii} = 1$  and  $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2 \cdots \mathbf{z}_n) \in \mathbb{R}^{n \times d}$  denotes the collection of node features. Then, the weighted mean-type GCR (Hamilton et al., 2017) can be formulated as

$$\mathbf{h}'_i = \frac{1}{\mathbf{d}_i} \sum_{j \in \mathcal{N}_i \cup i} \mathbf{A}_{ij} \mathbf{z}_j \mathbf{W} \quad (1)$$

where  $\mathbf{d}_i = \sum_{j \in \mathcal{N}_i \cup i} \mathbf{A}_{ij}$  and  $\mathcal{N}_i$  represents the neighbor set of node  $i$ . Matrix  $\mathbf{W} \in \mathbb{R}^{d \times d'}$  denotes the graph convolutional parameter which is learned adaptively based on the specific downstream task. The output  $\mathbf{H}' = (\mathbf{h}'_1, \mathbf{h}'_2 \cdots \mathbf{h}'_n) \in \mathbb{R}^{n \times d'}$  provides the graph convolutional representations of graph nodes. Comparing with input  $\mathbf{Z}$ ,  $\mathbf{H}'$  involves more context information encoded in  $\mathbf{A}$  and also provides task-relevant representations for nodes via learned  $\mathbf{W}$ .

## 3 LATENT GRAPH CONVOLUTION REPRESENTATION

The above GCR is defined on the fixed input graph  $\mathbf{A}$  which may restrict the representation capacity and also has been demonstrated to be very vulnerable to the structural attacks and noises in  $\mathbf{A}$  (Zügner & Günnemann, 2019; Jin et al., 2020b;a). To address this issue, we present a novel Latent Graph Convolution Representation (LatGCR) for robust graph data representation and learning. Our LatGCR is motivated based on the reformulation of the above GCR (Eq.(1)) from the aspect of neighborhood reconstruction (Geisler et al., 2020; Jiang & Zhang, 2021). Specifically, Eq.(1) provides the optimal solution to the following node reconstruction problem,

$$\mathbf{h}'_i = \frac{1}{\mathbf{d}_i} \sum_{j \in \mathcal{N}_i \cup i} \mathbf{A}_{ij} \mathbf{z}_j \mathbf{W} = \arg \min_{\mathbf{h}_i} \sum_{j \in \mathcal{N}_i \cup i} \mathbf{A}_{ij} \|\mathbf{h}_i - \mathbf{z}_j \mathbf{W}\|^2 \quad (2)$$

where  $\|\cdot\|$  denotes Frobenius norm function.

### 3.1 LATGCR MODEL FORMULATION

Let  $G(\mathbf{A}, \mathbf{Z})$  be the input observed graph with adjacency matrix  $\mathbf{A} \in \mathbb{R}^{n \times n}$  ( $\mathbf{A}_{ii} = 1$ ) and node features  $\mathbf{Z} \in \mathbb{R}^{n \times d}$ . The aim of LatGCR is to estimate a latent and flexible graph  $\tilde{\mathbf{A}}$  to better support GCR. Based on the reformulation of GCR (Eq.(2)), we present our LatGCR model by integrating *latent graph estimation* and *graph convolutional representation* jointly as

$$\begin{aligned} \{\tilde{\mathbf{A}}', \mathbf{H}'\} &= \arg \min_{\tilde{\mathbf{A}}, \mathbf{H}} \|\mathbf{A} - \tilde{\mathbf{A}}\|^2 + \lambda \sum_i \sum_{j \in \mathcal{N}_i \cup i} \tilde{\mathbf{A}}_{ij} \|\mathbf{h}_i - \mathbf{z}_j \mathbf{W}\|^2 \\ &s.t. \quad \tilde{\mathbf{A}}_{ij} \geq 0 \end{aligned} \quad (3)$$

where  $\tilde{\mathbf{A}}'$  denotes the estimated latent graph and  $\mathbf{H}' = (\mathbf{h}'_1, \mathbf{h}'_2 \cdots \mathbf{h}'_n) \in \mathbb{R}^{n \times d'}$  denotes the output Latent GCRs for graph nodes. Matrix  $\mathbf{W} \in \mathbb{R}^{d \times d'}$  denotes the graph convolutional parameter and parameter  $\lambda > 0$  is the trade-off hyper-parameter balancing two terms. The first term in Eq.(3) represents the latent graph estimation/reconstruction while the second term denotes the GCR. Note that, when  $\lambda \rightarrow 0$ , the first term is penalized very largely and we can have  $\tilde{\mathbf{A}}' = \mathbf{A}$ . In this case, LatGCR degenerates to standard GCR (Eq.(1)). Overall, there are four main aspects of the above LatGCR model.

- **Self-supervised joint learning:** In LatGCR, the estimation of latent graph is conducted in a *self-supervised* way. LatGCR conducts both latent graph generation and GCR jointly to boost their respective representation ability. Therefore, LatGCR can be potentially used in both supervised and unsupervised learning tasks.
- **Robustness:** When  $\mathbf{A}$  contains some noises/errors, i.e.,  $\mathbf{A} = \tilde{\mathbf{A}} + \mathbf{E}$  where  $\mathbf{E}$  denotes the noises/errors. Then Eq.(3) can be re-formulated as

$$\begin{aligned} \{\tilde{\mathbf{A}}', \mathbf{H}', \mathbf{E}'\} &= \arg \min_{\tilde{\mathbf{A}}, \mathbf{H}, \mathbf{E}} \|\mathbf{E}\|^2 + \lambda \sum_i \sum_{j \in \mathcal{N}_i \cup i} \tilde{\mathbf{A}}_{ij} \|\mathbf{h}_i - \mathbf{z}_j \mathbf{W}\|^2 \\ &s.t. \quad \mathbf{A} = \tilde{\mathbf{A}} + \mathbf{E}, \tilde{\mathbf{A}}_{ij} \geq 0 \end{aligned} \quad (4)$$

That is, LatGCR acts as recovering/generating a latent ‘clear’ graph  $\tilde{\mathbf{A}}$  from the input noisy graph  $\mathbf{A}$  for GCR and thus performs robustly w.r.t. graph noises and attacks. This is one important property of LatGCR and will be validated in Experiments in detail.

- **Sparsity:** The estimated graph  $\tilde{\mathbf{A}}'$  inherits the same sparse pattern from input graph  $\mathbf{A}$ , i.e., we can easily prove that if  $\mathbf{A}_{ij} = 0$ , then we have  $\tilde{\mathbf{A}}'_{ij} = 0$ , as also seen from Eq.(5) below.
- **Efficient implementation:** In LatGCR, both graph estimation and graph convolutional representation are implemented via simple one-step update rules which thus can be computed very efficiently, as discussed in §3.2.

### 3.2 LATGCR IMPLEMENTATION.

The optimum  $\tilde{\mathbf{A}}'$  and  $\mathbf{H}'$  can be obtained via a simple update algorithm which alternatively conducts the following Latent Graph Estimation (LGE) and Graph Convolutional Representation (GCR) steps.

**LGE-step:** Solving  $\tilde{\mathbf{A}}$  while fixing  $\mathbf{H}$ , the problem becomes

$$\tilde{\mathbf{A}}' = \arg \min_{\tilde{\mathbf{A}}} \|\mathbf{A} - \tilde{\mathbf{A}}\|^2 + \lambda \sum_i \sum_{j \in \mathcal{N}_i \cup i} \tilde{\mathbf{A}}_{ij} \|\mathbf{h}_i - \mathbf{z}_j \mathbf{W}\|^2 \quad s.t. \quad \tilde{\mathbf{A}}_{ij} \geq 0$$

which is equivalent to

$$\tilde{\mathbf{A}}'_{ij} = \arg \min_{\tilde{\mathbf{A}}_{ij}} \left( (\mathbf{A}_{ij} - \frac{\lambda}{2} \|\mathbf{h}_i - \mathbf{z}_j \mathbf{W}\|^2) - \tilde{\mathbf{A}}_{ij} \right)^2 \quad s.t. \quad \tilde{\mathbf{A}}_{ij} \geq 0$$

It has a simple closed-form solution which is given as

$$\tilde{\mathbf{A}}'_{ij} = \max \left\{ \left( \mathbf{A}_{ij} - \frac{\lambda}{2} \|\mathbf{h}_i - \mathbf{z}_j \mathbf{W}\|^2 \right), 0 \right\} \quad (5)$$

**GCR-step:** Solving  $\mathbf{H}$  while fixing  $\tilde{\mathbf{A}}$ , the problem becomes to the standard GCR Eq.(2). The exact optimal solution is

$$\mathbf{H}' = \tilde{\mathbf{D}}^{-1} \tilde{\mathbf{A}} \mathbf{Z} \mathbf{W} \quad (6)$$

where  $\tilde{\mathbf{D}}$  is the diagonal matrix with  $\tilde{\mathbf{D}}_{ii} = \sum_{j \in \mathcal{N}_i \cup i} \tilde{\mathbf{A}}_{ij}$ .

**Remark.** (1) As discussed before, since the optimal  $\tilde{\mathbf{A}}'$  inherits the sparse pattern of  $\mathbf{A}$ , in implementation of Eq.(5), we only need to compute the element  $\tilde{\mathbf{A}}'_{ij}$  with  $j \in \mathcal{N}_i \cup i$  which is efficient, as further analyzed in section **Complexity analysis**.

(2) In real implementation, instead of using Eq.(5), we use the following update rule to avoid the possible numerical issue when  $\lambda$  is large, i.e.,

$$\tilde{\mathbf{A}}'_{ij} = \max \left\{ \left( \mathbf{A}_{ij} - \frac{\lambda}{2} \|\mathbf{h}_i - \mathbf{z}_j \mathbf{W}\|^2 \right), \epsilon \right\}, \quad j \in \mathcal{N}_i \cup i \quad (7)$$

where  $\epsilon$  is a very small positive number. That is, when  $\lambda$  is large enough, we have  $\tilde{\mathbf{A}}_{ij} = \epsilon, j \in \mathcal{N}_i \cup i$ . In this case, the above GCR (Eq.(6)) becomes to the unweighed mean aggregation which also gives a reasonable solution.

**LatGCR block.** From network architecture aspect, the above LatGCR implementation can be designed via a recurrent architecture, i.e., LatGCR block, which consists of Latent Graph Estimation (LGE) and Graph Convolutional Representation (GCR) submodules, as shown in Figure 1. As shown in Fig. 1(B), LGE and GCR are alternatively conducted in LatGCR. On the contrary, the graph  $\mathbf{A}$  is fixed in traditional GCR (shown in Fig. 1(A)). Therefore, LatGCR is more flexible than GCR. More importantly, LatGCR block performs more robustly than GCR, as demonstrated in Experiments.

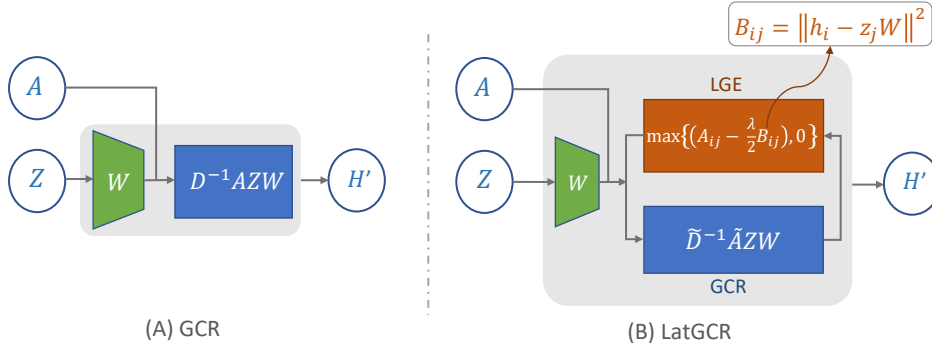


Figure 1: Architectures of GCR and our LatGCR block.

**Complexity analysis.** The whole computation complexity of the proposed LatGCR block involves three parts, i.e., linear projection, LGE and GCR. The complexity of projection  $\mathbf{Z}\mathbf{W}$  is  $O(ndd')$ . The computation cost of LGE-step mainly focuses on computing  $\|\mathbf{h}_i - \mathbf{z}_j \mathbf{W}\|^2$  with  $j \in \mathcal{N}_i \cup i$  and the complexity is  $O(|\xi|d')$  where  $|\xi|$  denotes edge number and  $d'$  denotes dimension of feature vector  $\mathbf{h}'_i$ . The complexity of GCR-step is  $O(|\xi|d')$ . In summary, the whole LatGCR complexity is  $O(ndd') + O(2r|\xi|d')$  where  $r$  is the recurrent time for alternatively conducting LGE and GCR and set to 3 in experiments. Note that, LatGCR does not bring very high complexity when comparing with GCR (Eq.(1)) whose main complexity is  $O(ndd') + O(|\xi|d')$ .

## 4 LATGC NEURAL NETWORKS

LatGCR gives a basic block which can be used within many GNN architectures (Veličković et al., 2018b; Hamilton et al., 2017; Kipf & Welling, 2017) by replacing the traditional graph convolution layer with LatGCR module. Here, we adopt the GNN architecture utilized in traditional GCN (Kipf & Welling, 2017) and design an end-to-end multi-layer neural network architecture, named Latent Graph Convolutional Network (LatGCN) for graph data learning. Concretely, the proposed LatGCN contains one input layer, several hidden propagation layers and one final perceptron layer, as shown in Figure 2. For each hidden propagation layer, it takes features  $\mathbf{H}^{(l)}$  and initial graph  $\mathbf{A}$  as input and

outputs features  $\mathbf{H}^{(l+1)}$  by using LatGCR module with parameter  $\mathbf{W}^{(l)}$ . LatGCN can be used in many graph learning tasks. For example, when applying LatGCN for semi-supervised node classification tasks, the last perceptron layer outputs the final label predictions  $\mathbf{P}$  for all nodes. The convolutional parameter  $\mathbf{W}^{(l)}$  of each hidden layer can be learned via minimizing the cross-entropy function on the labelled nodes, as discussed in work (Kipf & Welling, 2017).

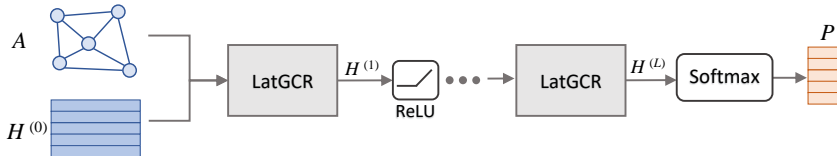


Figure 2: The architecture of LatGCN.

**Comparison with related works.** Exploiting graph learning for GNNs has been studied in recent years. The main differences between LatGCN and previous graph learning guided GCNs including Pro-GNN (Jin et al., 2020b), TO-GCN (Yang et al., 2019) and GLCN (Jiang et al., 2019) are follows. First, LatGCN is designed based on our proposed new LatGCR block (Figure 1 (B)). In contrast, previous works generally incorporate graph learning into GNN’s training via designing a joint loss function. Second, LatGCR is derived based on the joint reconstruction framework (Eq.(3)) which is implemented in a self-supervised manner. This makes LatGCR be a general block which can be used within many GNN’s architectures to derive various kinds of LatGCNs. LatGCR is also different from recent GeC and GeCN (Jiang et al., 2021). (1) LatGCR aims to generate a flexible latent graph for graph convolution while GeC incorporates neighborhood selection into graph convolution. (2) LatGCR is derived based on neighborhood reconstruction while GeC is designed based on graph Laplacian regularization. (3) LatGCR implements both graph estimation and GCR via simple one-step update rules while GeC (Jiang et al., 2021) adopts  $T$ -step update rules for both neighborhood selection and graph convolution operation. Thus, the implementation of LatGCR is generally more efficient than GeC (Jiang et al., 2021).

## 5 EXPERIMENT

To verify the effectiveness and robustness of LatGCR block and LatGCN, we test it on both semi-supervised node classification and unsupervised clustering tasks on three standard benchmark datasets, i.e., Cora, Citeseer and Pubmed (Sen et al., 2008; Jin et al., 2020b).

### 5.1 SEMI-SUPERVISED NODE CLASSIFICATION

#### 5.1.1 EXPERIMENTAL SETTING

Similar to the architecture of GCN (Kipf & Welling, 2017), LatGCN consists of one input layer, two LatGCR layers and one final perceptron layer. The skip-connection strategy is also utilized in LatGCN, as suggested in work (Kipf & Welling, 2017; He et al., 2016). We optimize the network weight matrices of all LatGCR modules by minimizing the cross-entropy loss function (Kipf & Welling, 2017). For fair comparison, we use the same attacked data setting used in work (Jin et al., 2020b) and employ two types of attacks, i.e., Metattack (Zügner & Günnemann, 2019) and Random Attack (Jin et al., 2020a). For Metattack (Zügner & Günnemann, 2019), we utilize the most destructive attack variant ‘Meta-Self’ and set the perturbation level from 0 to 25% with step 5%. For Pubmed dataset, we use the approximate version ‘A-Meta-Self’ as used in work (Jin et al., 2020b) For Random Attack (Jin et al., 2020a), we apply the variant ‘Add’ and set the perturbation level from 0 to 100% with step 20%. Following the experimental setup in previous works (Kipf & Welling, 2017; Veličković et al., 2018a), we set the number of units in each hidden layer to 16 and train our LatGCN by using Adam algorithm (Kingma & Ba, 2015) with learning rate of 0.001. The recurrent time of each LatGCR block is set to 3 and the hyper-parameter  $\lambda$  is determined based on validation set. We provide additional experiments across different settings of parameter  $\lambda$  in §5.3.

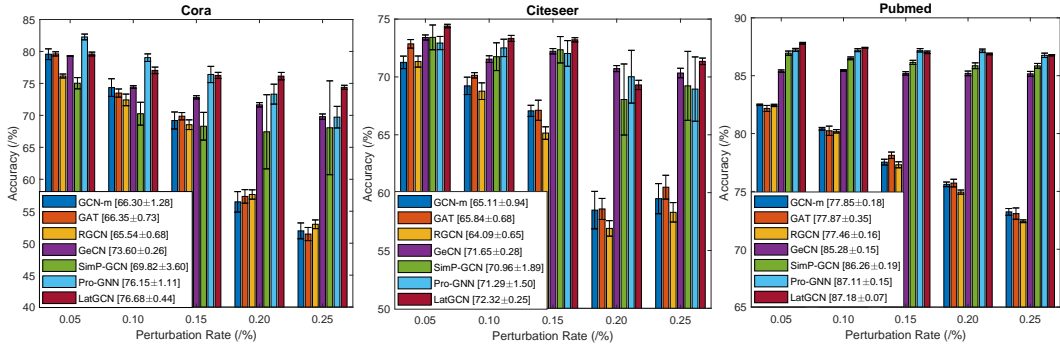


Figure 3: Semi-supervised classification performance under Metattack (Zügner & Günnemann, 2019).

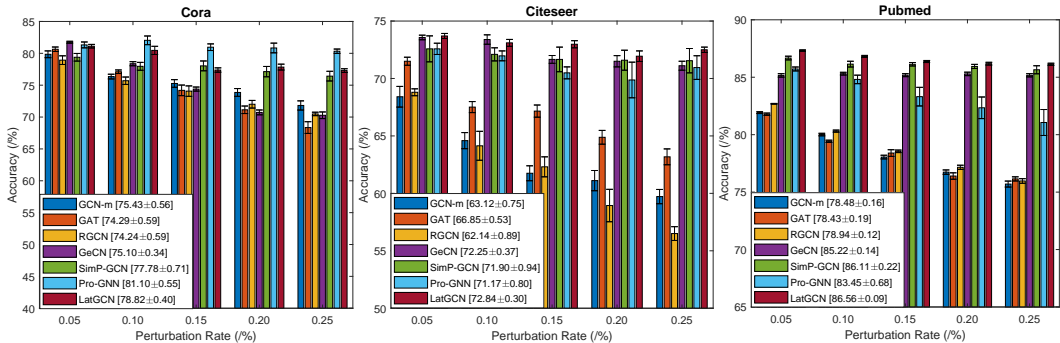


Figure 4: Semi-supervised classification performance under Random Attack (Jin et al., 2020a).

5.1.2 COMPARISON RESULTS

We first compare LatGCN with some traditional baseline methods including weighted mean-type GCN (GCN-m) (Kipf & Welling, 2017; Hamilton et al., 2017), Graph Attention Networks (GAT) (Veličković et al., 2018a). To demonstrate the robustness of LatGCN, we also compare LatGCN with some recent robust GNNs including Robust Graph Convolutional Networks (RGCN) (Zhu et al., 2019), SimP-GCN (Jin et al., 2021), Property Graph Neural Networks (Pro-GNN) (Jin et al., 2020b) and Graph elastic Convolutional Networks (GeCN) (Jiang et al., 2021).

**Effectiveness analysis.** Figure 3 and 4 summarize the comparison results across different levels of Metattack (Zügner & Günnemann, 2019) and Random Attack (Jin et al., 2020a), respectively. For each attack level, the results are the average performance of 10 runs with different network initializations. The overall average performance of comparison methods for all attack levels are reported in the legend of each Figure. Here, we can observe that (1) Traditional GCN-m (Kipf & Welling, 2017; Hamilton et al., 2017) and GAT (Veličković et al., 2018a) are vulnerable to the structural attacks and noises. Comparing with them, LatGCN obtains obviously better performance under various attacks and noises. This clearly demonstrates the effectiveness and robustness of the proposed LatGCR module for robust graph data learning. (2) LatGCN obtains better performance than some recent robust GNN methods including RGCN (Zhu et al., 2019) and SimP-GCN (Jin et al., 2021), which indicates the more robustness of the proposed LatGCN w.r.t graph structural attacks and noises. (3) Comparing with some recent graph learning guided GNNs including Pro-GNN (Jin et al., 2020b) and GeCN (Jiang et al., 2021), LatGCN generally obtains the best average learning performance on attacked graph data. This further demonstrates the effectiveness of the proposed self-supervised latent graph estimation in LatGCR for noisy graph data representation.

**Efficiency analysis.** Figure 5 shows the average running time of each epoch in training LatGCN on semi-supervised node classification tasks on the attacked datasets used in work (Jin et al., 2020b). All methods are implemented by PyTorch on NVIDIA A6000. We can note that (1) The methods

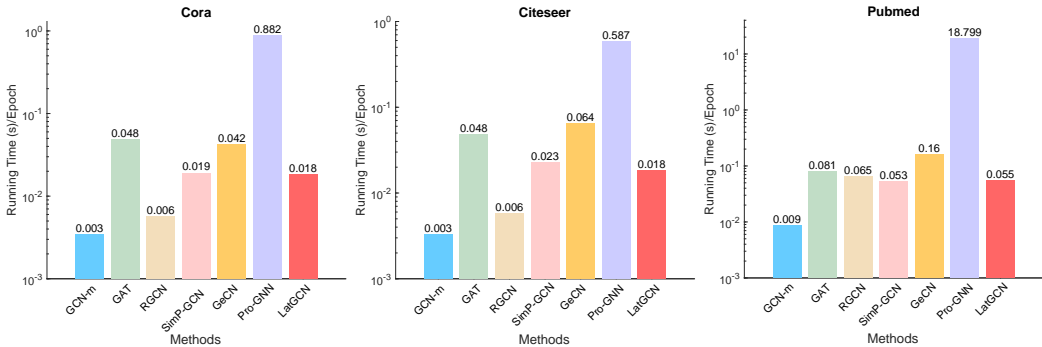


Figure 5: The empirical average running time in each epoch of different methods.

using fixed graph GNNs, such as GCN-m, SimP-GCN (Jin et al., 2021) and RGCN (Zhu et al., 2019), generally run faster than graph learning guided GNN methods. (2) Our LatGCN performs obviously faster than most of graph learning guided methods including GAT (Veličković et al., 2018a), GeCN (Jiang et al., 2021) and Pro-GNN (Jin et al., 2020b), especially on the larger dataset Pubmed (Sen et al., 2008). It demonstrates the efficiency of the proposed LatGCN on conducting robust graph data learning.

## 5.2 UNSUPERVISED CLUSTERING

To evaluate the effectiveness of the proposed self-supervised LatGCR, we further test it on unsupervised clustering tasks. Following the experimental setting in previous work (Zhang et al., 2019), we first use Singular Value Decomposition (SVD) to replace projection step to obtain low-dimensional embeddings for graph nodes. Then, we utilize LatGCR to obtain context-aware representations for graph nodes and employ K-means clustering algorithm (Charu & Chandan, 2014) to obtain the final clustering results (Zhang et al., 2019). We set the recurrent time and parameter  $\lambda$  to 1 and 0.03 respectively. Similar to work (Zhang et al., 2019), we adopt three widely used performance measurements (Charu & Chandan, 2014), i.e., clustering accuracy (Acc), normalized mutual information (NMI) and macro F1-score (F1) for evaluation.

We compare our LatGCR with some other popular clustering approaches including Graph Variational Autoencoder (VGAE) (Kipf & Welling, 2016), Marginalized Graph Autoencoder (MGAE) (Wang et al., 2017), Adversarially Regularized Variational Graph Autoencoder (ARVGE) (Pan et al., 2018) and Attributed Graph Clustering (AGC) (Zhang et al., 2019). Table 1 summarizes the comparison results on all original datasets (Sen et al., 2008). The results of these comparison methods have been reported in work (Zhang et al., 2019) and we use them directly. From Table 1, we can note that comparing with some other clustering methods, the proposed LatGCR can obtain the best average performance in most cases, which further indicates the effectiveness of the proposed LatGCR on conducting unsupervised clustering tasks.

Table 1: Comparison results on clustering task. The best results are marked by bold.

Method	Cora			Citeseer			Pubmed		
	Acc%	NMI%	F1%	Acc%	NMI%	F1%	Acc%	NMI%	F1%
VGAE	55.95	38.45	41.50	44.38	22.71	31.88	65.48	25.09	50.95
MGAE	63.43	45.57	38.01	63.56	39.75	39.49	43.88	8.16	41.98
ARVGE	63.80	45.00	62.70	54.40	26.10	52.90	58.22	20.62	23.04
AGC	68.92	<b>53.68</b>	<b>65.61</b>	67.00	41.13	62.48	69.78	31.59	68.72
LatGCR	<b>69.19</b>	53.48	65.50	<b>67.83</b>	<b>42.07</b>	<b>63.31</b>	<b>70.52</b>	<b>32.77</b>	<b>69.71</b>

### 5.3 MODEL ANALYSIS

#### 5.3.1 VISUALIZATION RESULTS

To demonstrate the effectiveness of proposed LatGCR, we utilize 2D t-SNE (Van der Maaten & Hinton, 2008) visualization to show the feature representation ability of LatGCR comparing with baseline method GCN-m (Hamilton et al., 2017; Kipf & Welling, 2017). Figure 6 shows 2D t-SNE (Van der Maaten & Hinton, 2008) visualization results of the feature maps output by the first hidden layer of GCN-m and LatGCN on Cora and Citeseer (Sen et al., 2008; Jin et al., 2020b) datasets under 0.25% Metattack (Zügner & Günnemann, 2019). Different colors denote different classes. One can note that, LatGCR obtains clearer and compacter embeddings than baseline method GCN-m, which intuitively demonstrates that the proposed LatGCR can obtain more robust and discriminative feature representations for graph data with structural attacks.

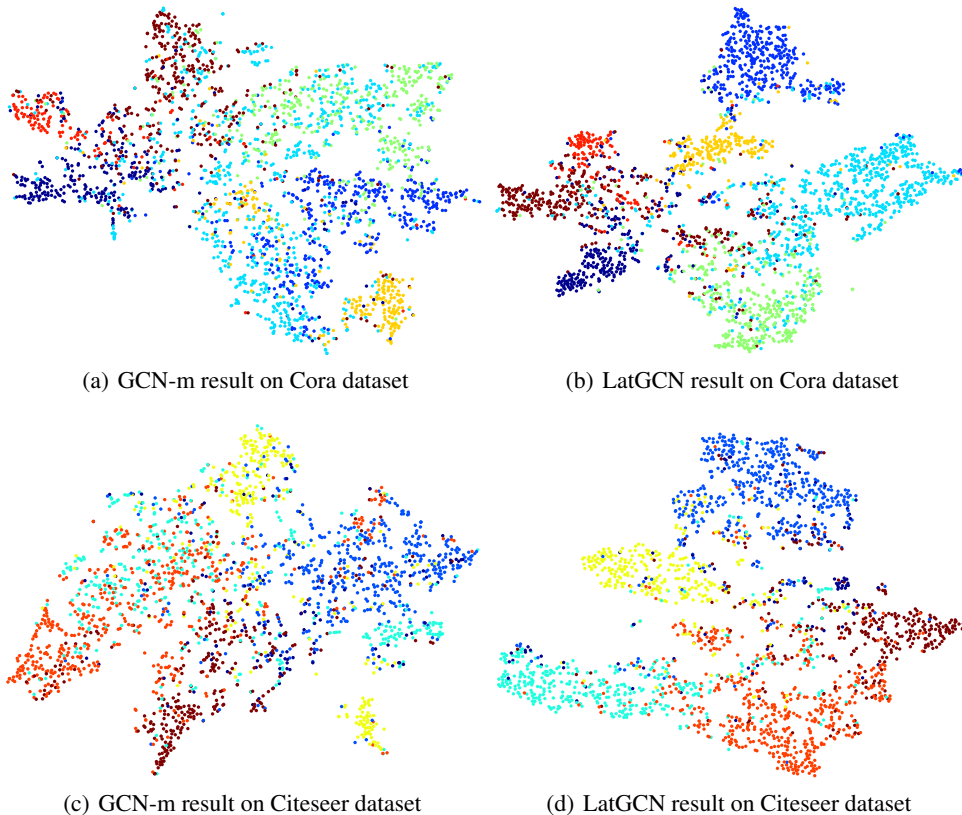


Figure 6: 2D t-SNE (Van der Maaten & Hinton, 2008) visualizations of the feature map output from the first layer of GCN-m and LatGCN. Different colors denote different classes.

Table 2: Results of LatGCN with different parameter  $\lambda$  values.

$\lambda$	0.05	0.1	0.5	1	5	10	20
Cora	76.06	77.11	76.36	75.96	72.79	70.82	67.25
Citeseer	71.09	71.27	71.68	71.5	72.16	70.44	69.73
Pubmed	84.87	85.36	86.03	86.07	86.42	86.66	86.40

#### 5.3.2 PARAMETERS ANALYSIS

One main hyper-parameter in LatGCR is the balanced parameter  $\lambda$  (Eq.(3)). Table 2 shows the semi-supervised classification results of LatGCN with different  $\lambda$  values on three datasets under



100% Random Attack (Jin et al., 2020a). We can see that our LatGCN can achieve relatively stable results in a certain range of parameter  $\lambda$  value which indicates that LatGCR is generally insensitive to the value of hyper-parameter  $\lambda$  in a certain range.

## 6 CONCLUSION

This paper proposes a novel Latent Graph Convolutional Representation (LatGCR) for robust graph data representation and learning. LatGCR is proposed based on a joint reconstruction framework, i.e., graph structure reconstruction + node’s feature reconstruction. It can estimate a latent and flexible graph for GCR in a self-supervised way and provides a general basic block for GNNs. The main advantage of LatGCR is that it can perform robustly w.r.t graph structural attacks and noises. Experiments on several benchmark datasets demonstrate the effectiveness and robustness of LatGCR. In our future work, we will extend LatGCR to address the data with multiple graphs and further apply it on some more applications and tasks, such as computer vision, recommendation, etc.

## REFERENCES

- Aggarwal Charu, C and Reddy Chandan, K. *Data Clustering: Algorithms and Applications*. CRC Press, Boca Raton, 2014.
- Yu Chen, Lingfei Wu, and Mohammed Zaki. Iterative deep graph learning for graph neural networks: Better and robust node embeddings. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, pp. 19314–19326, 2020.
- Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems*, pp. 3844–3852, 2016.
- Simon Geisler, Daniel Zügner, and Stephan Günnemann. Reliable graph neural networks via robust aggregation. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, pp. 13272–13284, 2020.
- William L Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems (NIPS)*, pp. 1024–1034, 2017.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.
- Bo Jiang and Ziyang Zhang. Incomplete Graph Representation and Learning via Partial Graph Neural Networks. *arXiv e-prints*, 2021.
- Bo Jiang, Ziyang Zhang, Doudou Lin, Jin Tang, and Bin Luo. Semi-supervised learning with graph learning-convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 11313–11320, 2019.
- Bo Jiang, Beibei Wang, Jin Tang, and Bin Luo. Gecns: Graph elastic convolutional networks for data representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP:1–1, 04 2021.
- Wei Jin, Yaxin Li, Han Xu, Yiqi Wang, and Jiliang Tang. Adversarial attacks and defenses on graphs: A review and empirical study. *arXiv e-prints*, pp. arXiv–2003, 2020a.
- Wei Jin, Yao Ma, Xiaorui Liu, Xianfeng Tang, Suhang Wang, and Jiliang Tang. Graph structure learning for robust graph neural networks. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 66–74, 2020b.
- Wei Jin, Tyler Derr, Yiqi Wang, Yao Ma, Zitao Liu, and Jiliang Tang. Node similarity preserving graph convolutional networks. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, pp. 148–156, 2021.
- Aleksandar Bojchevski Johannes Klicpera and Stephan Günnemann. Combining neural networks with personalized pagerank for classification on graphs. In *ICLR*, 2019.

- Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.
- Thomas N Kipf and Max Welling. Variational graph auto-encoders. *NIPS Workshop on Bayesian Deep Learning*, 2016.
- Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017.
- Ruoyu Li, Sheng Wang, Feiyun Zhu, and Junzhou Huang. Adaptive graph convolutional neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.
- Shirui Pan, Ruiqi Hu, Guodong Long, Jing Jiang, Lina Yao, and Chengqi Zhang. Adversarially regularized graph autoencoder for graph embedding. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pp. 2609–2615, 2018.
- Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher, and Tina Eliassi-Rad. Collective classification in network data. *AI magazine*, 29(3):93, 2008.
- Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.
- Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018a.
- Petar Veličković, William Fedus, William L Hamilton, Pietro Liò, Yoshua Bengio, and R Devon Hjelm. Deep graph infomax. In *International Conference on Learning Representations*, 2018b.
- Chun Wang, Shirui Pan, Guodong Long, Xingquan Zhu, and Jing Jiang. Mgae: marginalized graph autoencoder for graph clustering. In *CIKM'17 - Proceedings of the 2017 ACM Conference on Information and Knowledge Management*, pp. 889–898, 2017.
- Huijun Wu, Chen Wang, Yuriy Tyshetskiy, Andrew Docherty, Kai Lu, and Liming Zhu. Adversarial examples for graph data: Deep insights into attack and defense. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pp. 4816–4823. AAAI Press, 2019.
- Liang Yang, Zesheng Kang, Xiaochun Cao, Di Jin, Bo Yang, and Yuanfang Guo. Topology optimization based graph convolutional network. In *IJCAI*, pp. 4054–4061, 2019.
- Xiaotong Zhang, Han Liu, Qimai Li, and Xiao-Ming Wu. Attributed graph clustering via adaptive graph convolution. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, (IJCAI)*, pp. 4327–4333, 2019.
- Dingyuan Zhu, Ziwei Zhang, Peng Cui, and Wenwu Zhu. Robust graph convolutional networks against adversarial attacks. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1399–1407. Association for Computing Machinery, 2019.
- Daniel Zügner and Stephan Günnemann. Adversarial attacks on graph neural networks via meta learning. In *International Conference on Learning Representations (ICLR)*, 2019.