Abstract—With the rise of contrastive learning, unsupervised graph representation learning (GRL) has shown strong competitiveness. However, existing graph contrastive models typically either focus on the local view of graphs or take simple considerations of both global and local views. This may cause these models to overemphasize the importance of individual nodes and their ego networks, or to result in poor learning of global knowledge and affect the learning of local views. Additionally, most GRL models pay attention to topological proximity, assuming that nodes that are closer in graph topology are more similar. However, in the real world, close nodes may be dissimilar, which makes the learned embeddings incorporate inappropriate messages and thus lack discrimination. To address these issues, we propose a novel unsupervised GRL model by contrasting cluster assignments, called graph representation learning model via contrasting cluster assignment (GRCCA). To comprehensively explore the global and local views, it combines multiview contrastive learning and clustering algorithms with an opposite augmentation strategy. It leverages clustering algorithms to capture fine-grained global information and explore potential relevance between nodes in different augmented perspectives while preserving high-quality global and local information through contrast between nodes and prototypes. The opposite augmentation strategy further enhances the contrast of both views, allowing the model to excavate more invariant features. Experimental results show that GRCCA has strong competitiveness compared to state-of-the-art models in different graph analysis tasks.

Index Terms—Contrastive learning, graph data mining, graph representation learning (GRL), unsupervised learning.

I. INTRODUCTION

FROM understanding the influence of social interactions in social networks to modeling neural connectivity in the human brain, and optimizing multiagent interactions in robotics, graph and network as powerful tools for representing complex systems play pivotal role in advancing the study of development and cognition in natural and artificial systems. For instance, sociologists utilize graphs to cognize the patterns of human behavior in social networks [1], neuroscientists employ them to enhance understanding of cognitive processes and brain development [2], and in the realm of robotics, graphs facilitate the optimization of coordination strategies and task allocation within multiagent systems [3]. However, these graphs often embody intricate structures that require sophisticated tools to cognize and analyze. Therefore, graph representation learning (GRL) as an effective method [4] is employed for mining the inherent structures and patterns within the graphs. Traditional GRL models generally try to capture the topological structures of graphs through designing handcraft features [5] or using matrix decomposition [6], which requires expensive implementation costs. Random walk techniques [7] offer a flexible and stochastic solution to address the problem of model scalability, but they ignore node or graph attributes. To solve this problem, graph neural networks (GNNs) [8], [9], [10], [11], [12] take not only graph topology but also attributes into account. However, most of them are supervised models that still need labels for guidance. This requires massive sample annotations that are time consuming and may have human errors, as well as results in the weak generalization of the learned representations. Therefore, increasing attention has been attracted by unsupervised GRL.

Recent years have witnessed the development of unsupervised GRL, such as auto-encode-based [13] and random walk-based models [14]. With the rise of contrastive learning [15], [16], [17], [18], [19], unsupervised GRL has achieved impressive results. DGI [20] first employs contrastive learning to learn the node embeddings through maximizing mutual information between global and local feature vectors. However, it is limited by the choice of readout function, which results in the poor exploration of the global view and even may influence the learning of the local view. Based on DGI, MVGRL [21] uses a dual-path contrastive framework with graph augmentations to explore multiview features of graphs and employs a new readout function that incorporates node embeddings in each layer of graph convolution network (GCN), thus exploring the global view more effectively. However, this is also a coarse-grained global exploration, that is, it still simply condenses the whole node features into a global feature vector, which may crush some global information. To address this issue, GIC [22] leverages not only a global feature vector but also multiple clustering prototypes to characterize fine-grained global information. Other successful works include GRACE [23], MERIT [24], and GCA [25]. They employ multiview contrastive learning and pay more attention to local view, i.e., feature alignments between the
identical nodes that have the same identity but exist in different augmented perspectives.

Although existing graph contrastive learning models have made great progress, even outperforming the supervised models, they lack enough comprehensive exploration of both global and local views. These models usually either focus on the local view, emphasizing the importance of individual nodes and their ego networks, or take simple considerations in exploring both views, which tends to explore the global view in a coarse-grained way, leading to poor learning of global and even local knowledge. However, both global and local views are critical for graph analysis. The former reflects the overall characteristics and structure of networks, and the latter describes the individual attributes and roles of each node in a network. Besides, existing GRL models typically follow topological proximity assumption [26], a foundational inductive bias in GRL, i.e., the nodes that are connected or close within the network topology share more similarities. As a result, node representations of the closer nodes to be similar (appearing closely in the hidden feature space). However, the closer nodes may be dissimilar in real-world networks. As shown in Fig. 1, the neighbor nodes have different labels from the target node in a citation network (Cora), suggesting that their representations should be distinct from that of the target node. However, the topological proximity assumption indiscriminately makes these representations to be similar. This may affect their discriminative ability and consequently impact the performance of downstream tasks, such as node classification.

To address these issues, we propose a novel unsupervised graph representation learning model via contrasting cluster assignments (GRCCAs), inspired by SwAV [19]. Through combining clustering algorithms and multiview contrastive learning with an opposite augmentation strategy, it provides a comprehensive insight into global and local views, as well as captures potential associations between nodes rather than just topological proximity. Specifically, clustering provides more fine-grained global information, instead of the coarse-grained one of using a global feature vector as in DGI and MVGRL. It also helps the model to capture the potential associations between nodes in different augmented perspectives, so as to explore more powerful patterns of graphs. The proposed GRCCA performs contrastive learning between nodes and clustering prototypes in another augmented perspective, which ensures it can achieve the comprehensive exploration of global and local views. Furthermore, the opposite augmentation strategy is developed to further enhance the contrast of both views, allowing the model to excavate more invariant features. It is worth noting that while GIC also takes the combination of clustering algorithm and contrastive learning into account, it differs from the proposed GRCCA in several ways. First, GIC does not employ multiview contrastive learning, whereas the proposed GRCCA does. This enables GRCCA to capture more abundant patterns of graphs by exploring the potential associations between nodes from different perspectives. Second, GIC considers contrasting between cluster summaries and nodes, which ignores the dissimilarity between nodes from different clusters. In contrast, the proposed GRCCA regards other prototypes that target nodes do not belong to as negative samples. This design is beneficial to distinguish the target nodes from the dissimilar nodes that belong to different clusters, enabling models to capture potential associations between nodes better. Finally, GIC leverages shuffling to construct a corrupt graph, which cannot help the contrastive learning to enhance explorations for global and local views. GRCCA employs the opposite augmentation strategy to achieve this. In general, our contributions can be summarized as follows.

1) We propose a novel unsupervised GRL model via contrasting cluster assignments called GRCCA. It aims to provide comprehensive insight into global and local views, as well as capture potential associations between nodes rather than just topological proximity.

2) To better combine the clustering algorithm and contrastive learning, the proposed GRCCA considers multiview contrastive learning and makes good use of all clustering prototypes. This facilitates the contrastive effects and further mines the relevance between nodes. Besides, the opposite augmentation strategy further enhances explorations of global and local views and allows the model to learn more invariant features of graphs.

3) Sufficient experimental results demonstrate that GRCCA has strong competitiveness in three popular graph analysis tasks, especially in node classification tasks, it outperforms all state-of-the-art baselines. This confirms its effectiveness and generalization.

The remainder of this article is organized as follows. Section II provides an overview of related works in unsupervised GRL and clustering-based unsupervised learning. In Section III, we describe the proposed GRCCA in detail. The empirical results and analysis are presented in Section IV. Finally, Section V concludes this article and analyzes the limitations of our work, as well as outlines directions for future work.

II. RELATED WORK

In recent years, supervised learning has made great progress. However, it heavily relies on large labeled data, leading to
some limitations, such as expensive annotation costs and poor generalization. To overcome these issues, increasing efforts have been made in unsupervised learning, such as autoencoder [27] and contrastive learning [16], [17]. With their development, the gap of model performance between supervised and unsupervised methods is gradually narrowing. Inspired by the successes of unsupervised learning, various unsupervised GRL models have been proposed. Graph contrastive learning is currently one of the most popular models and can be roughly grouped into three categories. The first focuses on the contrast of local–global relationships [20], [21], [22], and the second emphasizes feature alignments between different perspectives at the node level [23], [24], [25]. The third performs contrastive learning between the target node and its neighbors [26]. The proposed GRCCA further explores local–global relationships through combining multiview contrastive learning and clustering algorithms with the opposite augmentation strategy. This design enhances the model ability to comprehensively explore global and local views and allows the model to learn potential associations between nodes beyond topological connections. Note that GRCCA can be seen as using a balanced contrastive way between MVGRL and MERIT, where the former treats all nodes as a single cluster, while the latter treats each node as a separate cluster.

The clustering algorithm, a traditional unsupervised method, has been revived through combination with deep learning techniques [28], [29], [30], [31], [32]. By using clustering, clustering-based models can generate pseudo-labels for training samples, so as to get rid of label dependencies [28]. This provides a new way for unsupervised learning. Clustering algorithms have also been applied in contrastive learning to generate more diverse contrastive materials and provide the insight of potential associations between samples [19], [30], [32]. For example, samples that belong to the same cluster can be regarded as positive samples, while those from different clusters are used as negative samples. Recently, clustering algorithms have been used for unsupervised GRL. However, existing models mainly rely on simple guidance from clustering pseudo-labels [33]. Additionally, although GIC [22] attempts to combine clustering algorithm and contrastive learning, it has some limitations as described in Section I. Therefore, we make an attempt to further combine the two methods to explore global and local views more comprehensively and mine more invariant features from the topology and attributes of graphs.

III. METHODOLOGY

In this section, we introduce the proposed GRCCA in detail. The overview of GRCCA is illustrated in Fig. 2, including the overall framework and learning algorithm. The overall framework includes the opposite augmentation strategy and the network architecture of GRCCA. The contrastive mechanism of GRCCA is described in the learning algorithm.

A. Preliminary

For a graph $G$, it can be represented by a tuple $(V, E)$, where $V$ is the set of nodes and $E$ indicates the set of edges. Each edge defines a relation between two nodes in $V$. If the edge is unidirectional, the graph is a directed graph; otherwise, if the edge is bidirectional, it is an undirected graph. In this work, we focus on undirected graphs. In practice, a graph is typically described by an adjacency matrix $A \in \{0, 1\}^{N \times N}$, where $N$ is the number of nodes and $A_{ij} = 1$ denotes $(v_i, v_j) \in E$. We also take into account node attributes, which can be given by an attribute matrix $X \in \mathbb{R}^{N \times F}$. $F$ is the dimension of attributes, and $x_v \in \mathbb{R}^F$ denotes the attribute vector of node $v_i$.

In this work, we focus on unsupervised GRL. It aims at learning a graph encoder $f_G : \mathbb{R}^{N \times F} \times \mathbb{R}^{N \times N} \rightarrow \mathbb{R}^{N \times F'}$ without using labels to map network topology $A$ and node attributes $X$ into low-dimension node embeddings $H = f_G(X, A) \in \mathbb{R}^{N \times F'}$. 

Fig. 2. Overview of the proposed GRCCA. First, an opposite augmentation strategy is used to generate two augmented graphs $G'$ and $G''$ from the input graph. Then, the two augmented graphs are fed into a graph encoder $f_G$ and an MLP projector $p_G$. The obtained feature vectors $Z_G$ and $Z_{G''}$ are clustered, and their prototypes can be represented by $C_G$ and $C_{G''}$. Different colors are used to distinguish their clustering assignments, i.e., which prototypes the nodes belong to. Moreover, different colored lines connecting nodes with the same prototypes reveal their potential associations rather than topological associations shown by the slight gray lines. Finally, for any pair of identical nodes $u_i$ and $v_i$, clustering cluster assignments are used to maximize the cluster-level consistency between them by a cross-entropy loss. This is equivalent to contrastive learning between nodes and prototypes in another augmented perspective, where the prototypes their identical nodes belong to are regarded as positive samples, and the other prototypes are used as negative samples.
where $F' \ll F$. The learned embeddings can be used in different downstream tasks, such as node classification, link prediction, and community detection.

### B. Overall Framework

Data augmentations are widely used in image data to increase the diversity of samples and generate multiview information. However, devising an effective strategy for graph-structural data is still an open problem. Existing graph contrastive models have devised some strategies to generate abundant contrastive perspectives, but these strategies generally do not consider how to promote the exploration of global and local views. To make up for this, we employ an opposite augmentation strategy that focuses on the global view in one perspective and the local view in another perspective as shown in Fig. 2. For the global view, we employ graph diffusion (GD) that takes a larger range of topological neighborhoods into account instead of a one-hop neighborhood [34]. It not only provides an augmented perspective with more global information but also alleviates the problem of inherent noise in real graphs as an equivalent polynomial filter, which is demonstrated in [34]. Formally, a generalized GD matrix can be defined as

$$S = \sum_{k=0}^{\infty} \theta_k T^k$$

where $T \in \mathbb{R}^{N \times N}$ is the generalized transition matrix, and $\theta_k$ is the weighting coefficient that determines the explored intensity of global information. According to the empirical results of [21], we use the personalized PageRank (PPR) kernel as the instantiation of GD. Formally, given an adjacency matrix $A \in \mathbb{R}^{N \times N}$ and a diagonal degree matrix $D \in \mathbb{R}^{N \times N}$, PPR kernel can be given by

$$S = \sigma \left( I - (1 - \alpha)D^{-1/2}AD^{-1/2} \right)^{-1}$$

where $I \in \mathbb{R}^{N \times N}$ is the identity matrix, and $\alpha \in (0, 1)$ is the teleport probability of random walk. For the local view, we employ removing edges (REs) to create a sparse topological structure, which makes the model emphasize local associations between nodes. Specifically, given an adjacency matrix $A$ and removal probability $P_{re}$

$$\tilde{A}_{ij} = \begin{cases} 1, & b_{ij} > P_{re} \text{ and } A_{ij} = 1 \\ 0, & \text{otherwise} \end{cases}$$

where $b_{ij}$ represents a random number sampled from uniform distribution $U(0, 1)$. Intuitively, this strategy enhances the difference between contrastive perspectives to capture the more invariant features and assists the model to learn more knowledge from global and local views. To further enhance the effects of contrastive learning, we use masking node features (MNFs) as a supplementary technique. It performs data augmentation of node attributes by randomly masking them, which can be given by

$$\tilde{X}_i = X_i \odot M$$

where $X_i \in \mathbb{R}^F$ is the attribute vector of the $i$th node, and $M \in \{0, 1\}^F$ is a random vector controlled by a mask rate $P_{mof}$. $P_{mof}$ determines the probability that each dimension of the attribute vector is masked.

After data augmentation, the original graph is transformed into two augmented graphs, as shown in Fig. 2. The two augmented graphs are fed into a graph encoder $f_0$ and a nonlinear projector $g_0$. Graph encoders are crucial for GRL, which get the input of network topology and node attributes and output the node embeddings. Typically, existing graph encoders employ layer-by-layer neighbor message aggregation to explore multiple hop neighborhoods of target nodes, which can be defined as

$$h^k_v = \sigma \left( E \cdot h_{v'}^{k-1} + W \cdot \sum_{v' \in N(v)} h_{v'}^{k-1} / |N(v)| \right)$$

where $E$ and $W$ are learnable parameters, $\sigma$ is the activation function, and $N(v)$ represents neighborhood nodes of target node $v$. Due to the conciseness and efficiency of GCN [9], it is used as the graph encoder in this work. It can be defined as

$$H = \sigma \left( \tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2} \chi \Theta \right)$$

where $\tilde{D}$ denotes the degree matrix of adjacency matrix with self-loop $\tilde{A} = A + I_N$, $X$ represents the attributes matrix, and $\Theta$ represents the learnable parameters. In order to enhance the expressive ability of contrastive measure [35], we further leverage a nonlinear projector, i.e., an MLP, to transfer node embedding into a metric space $Z = g_0(H) \in \mathbb{R}^{N \times F'}$. This is commonly used in contrastive learning models to improve the model performance [17]. Its success may be attributed to its adaptive generalization of metrics and filtering effect on the information of pretext tasks.

### C. Learning Algorithm

Inspired by SwAV, the proposed GRCCA combines multiview contrastive learning and clustering algorithms by contrasting cluster assignments between the identical nodes from two augmented perspectives. Specifically, the $k$-means algorithm is used to conduct clustering for feature vectors of two augmented perspectives $Z_a$ and $Z_u$, respectively, because of its simplicity and effectiveness. Their clustering prototypes can be represented by $C_a \in \mathbb{R}^{K \times F'}$ and $C_u \in \mathbb{R}^{K \times F'}$. They reveal the associations between nodes that may not have topological connections in different augmented perspectives, where one focuses on the global view and another emphasizes the local view. $K$ is the number of prototypes and $F'$ indicates the dimension of prototypes. The cluster assignments $Q_a \in \mathbb{R}^{N \times K}$ and $Q_u \in \mathbb{R}^{N \times K}$ indicate which prototypes nodes belong to. For example, if $v_i$ belongs to the $k$th prototype $C_a^k$, the $Q_a^i$ can be defined as a one-hot vector, where the value of the $k$th dimension is 1 and other ones are 0. Then, the proposed GRCCA enforces the identical nodes in different perspectives to recognize the cluster assignments each other, and thus maximize their cluster-level consistency. In other words, we expect that the target node has the same cluster assignment as its identical node in another perspective when computing the similarity between the target node embedding and the clustering prototypes of another perspective. Given a target node $v_i$, $v_i$,
we can define the cluster-level consistency between it and its identical node $v_i$ as

$$P_{v_i} = \text{Softmax}(Z_{v_i}C_u^T / \tau)$$  \tag{7}$$

where $\tau$ is a temperature parameter. Through minimizing the cross-entropy loss, the cluster-level consistency between the two nodes is maximized, which can be given by

$$\mathcal{L}(Q_u, P_v) = -\frac{1}{N} \sum_{i=1}^{N} Q_u \log P_{v_i}$$  \tag{8}$$

where $N$ is the number of nodes. Therefore, the contrastive loss can be defined as

$$\mathcal{L}_c = \mathcal{L}(Q_v, P_u) + \mathcal{L}(Q_u, P_v)$$  \tag{9}$$

Notably, this loss implicitly drives the node embeddings to approach their corresponding prototypes, that is, the prototypes in another perspective that their identical nodes belong to, and set them apart from other prototypes. In fact, it is equivalent to maximizing the mutual information between the node embeddings and the corresponding prototypes, which is discussed in the Appendix. This design not only explores the potential associations between nodes in different perspectives but also promotes a comprehensive understanding of global and local views. This is because the two augmented perspectives focus on global and local views, respectively, and the node embeddings in the global view require learning the clustering associations from the local view, and vice versa.

Inspired by the multihead attention mechanism [36], GRCCA employs a multiclustering strategy to further enhance the diversity of contrastive samples. Specifically, we perform clustering for each perspective $h$ times to generate the pairwise contrastive materials $\{M_1, M_2, \ldots, M_M\}$, where $M$ is a tuple of $(C_v, C_u, Q_v, Q_u)$ and $h$ denotes the number of contrastive material sets. For each clustering process, we randomly select the initial $K$ prototypes independently. Therefore, the final loss can be given by

$$\mathcal{L}_{mc} = \frac{1}{h} \sum_{i=1}^{h} \mathcal{L}_c$$  \tag{10}$$

The learning algorithm is summarized in Algorithm 1. First, we use the opposite augmentation strategy for the original graph to generate two augmented graphs $\tilde{G}_g$ and $\tilde{G}_l$, where $t_g$ consists of $GD$ and $MNF$, and $t_l$ includes $RE$ and $MNF$. Then, the two augmented graphs with adjacency matrix and node attribute matrix are fed into a graph encoder $f_0$ and a nonlinear projector $g_0$. For the output feature vectors, we employ $k$-means with multiclustering strategy $K_m$ to generate multiple pair contrastive materials. After that, we minimize the contrastive loss in (10) to maximize the cluster-level consistency between the identical nodes from different perspectives. In this work, we try two different cluster assignment schemes: asynchronous and synchronous versions. The asynchronous version uses the representation matrices from the previous epoch to generate cluster assignments, while the synchronous version resorts to the current representation matrices. Notably, the asynchronous version needs to initialize a memory bank $B$

### Algorithm 1: Learning Algorithm for GRCCA

**Input:** graph $G = [X, A]$, parameters $\Theta$, learning rate $\beta$ and maximum epoch $T$

**Result:** node embeddings $H$

1. **node embeddings $H$**
   - $\tilde{G}_g = t_g(G)$; $\tilde{G}_l = t_l(G)$; $H_v = f_0(\tilde{G}_g)$; $H_u = f_0(\tilde{G}_l)$;
   - $B \leftarrow H_v, H_u$: // initialize memory bank

2. **for epoch $= 1: T$ do**
   - // asynchronous cluster assignments
     - $\{M_1, M_2, \ldots, M_M\} \leftarrow K_m(B)$;
     - $\tilde{G}_g \leftarrow t_g(G)$; $\tilde{G}_l \leftarrow t_l(G)$; $H_v \leftarrow f_0(\tilde{G}_g)$; $H_u \leftarrow f_0(\tilde{G}_l)$;
     - $Z_v \leftarrow g_0(H_v)$; $Z_u \leftarrow g_0(H_u)$;
     - $B \leftarrow Z_v, Z_u$: // update memory bank

3. **end**

4. **return** $H$;

**TABLE I**

| Dataset        | Nodes | Edges | Features | Class |
|----------------|-------|-------|----------|-------|
| Cora           | 2,708 | 5,429 | 1,433    | 7     |
| Citeseer       | 3,327 | 4,732 | 3,703    | 6     |
| Pubmed         | 19,717| 44,338| 500      | 3     |
| Amazon-Photo   | 7,650 | 119,081| 745     | 8     |
| Coauthor-CS    | 18,333| 81,894| 6,805    | 15    |
| Amazon-Computers | 13,752| 245,861| 767     | 10    |

and update it with each round of representation. Finally, the node embeddings learned from the graph encoder $f_0$ are used for downstream tasks.

### IV. Experiments

In this section, we evaluate the effectiveness and generalization of the proposed GRCCA compared to the baseline models across three different graph analysis tasks. We also perform the ablation study to investigate the impact of different components of GRCCA. To further illustrate the conciseness and convergence of our proposed model, we provide a complexity and convergence analysis. Our experiments are implemented by PyTorch and performed on a workstation with an Intel 10900X CPU, two RTX3060 GPUs, and 32-GB memory. The source code is available at https://github.com/MrYaoH/GRCCA.

**A. Data Sets**

To evaluate the performance of GRCCA, we use six popular data sets from different domains, including Cora, Citeseer, Pubmed, Amazon-Photo, Amazon-Computers, and Coauthor-CS. The relevant statistics of data sets are shown in Table I.
models and adopt their hyperparameter settings described in their papers. Since some baseline models have not reported the performance in link prediction and community detection tasks, for fairness, we try our best to adjust their hyperparameters and report their best results. Meanwhile, we slightly adjust the hyperparameters of GRCCA to better adapt to these two tasks. In the link prediction task, $P_{re}$ is set to 0.2 for Amazon-photo. In the community detection task, $P_{nmf}^1$ and $\tau$ are set to 0.4 and 0.1, respectively, for Cora. Additionally, $P_{re}$ and $P_{nmf}^2$ are set to 0.3 and 0.1, respectively, for Pubmed.

### C. Node Classification

Node classification is a typical task used to evaluate the performance of GRL models. To demonstrate the effectiveness of the proposed GRCCA, we select six state-of-the-art unsupervised models, including DGI [20], GMI [26], MVGRL [21], GIC [22], GCA [25], and MERIT [24], and three classical GNN models, that is, ChebyshevGCN [8], GCN [9], and GAT [10] as baseline models. For unsupervised models, we train them without labels to generate node embeddings, and then employ an additional linear classifier to classify them. Otherwise, we perform training with end-to-end settings. To ensure the reliability of results, we employ their available codes and use the consistent data partitioning scheme as [24]. For the three citation networks we randomly sample 20 nodes per class as the training set and 1000 nodes as the testing set. For the other three data sets, we randomly select 30 nodes per class for training and validating, respectively, while the remaining nodes are used for testing. Furthermore, we run the experiments ten times independently and report the mean accuracy with standard deviation according to [24].

The classification accuracy of the proposed GRCCA and other baseline models across six data sets is shown in Table III, where $X$ represents the node attribute matrix, $A$ indicates the adjacency matrix, and $Y$ denotes the labels of nodes. Obviously, GRCCA outperforms both unsupervised models and the classical models with label supervision over all data sets, particularly by 2.3% and 3.5% than the suboptimal models on the two co-purchasing networks, i.e., Amazon-Photo and Amazon-Computers. We attribute the improvements to the fact that the proposed GRCCA effectively utilizes both global and local views, and explores potential associations between nodes beyond topological proximity assumption. Specifically, compared to DGI and MVGRL, the proposed GRCCA outperforms them. This is because DGI and MVGRL employ the coarse local–global exploration by contrasting node embeddings with a global feature vector generated by averaging all node embeddings. Moreover, this implicitly assumes fully connected relationships among nodes, which is a stronger condition than the topological proximity assumption. In contrast, the proposed GRCCA conducts contrastive learning between node embeddings and clusters implicitly, which considers a more refined global view and thereby enhances local–global exploration. Additionally, it also takes into account the potential associations between nodes based on feature-level similarity, which is beyond topological proximity. Therefore, it demonstrates superior performance compared to DGI and MVGRL.

| Dataset             | $P_{re}$ | $P_{nmf}^1$ | $P_{nmf}^2$ | $\tau$ | $N_{pt}$ |
|---------------------|---------|-------------|-------------|-------|---------|
| Cora                | 0.2     | 0.3         | 0.4         | 0.05  | 14      |
| Citeseer            | 0.2     | 0.3         | 0.4         | 0.21  | 21      |
| Pubmed              | 0.4     | 0.2         | 0.2         | 0.15  | 9       |
| Amazon-Photo        | 0.4     | 0.4         | 0.3         | 0.15  | 17      |
| Coauthor-CS         | 0.4     | 0.2         | 0.2         | 0.1   | 24      |
| Amazon-Computers    | 0.4     | 0.2         | 0.2         | 0.1   | 21      |
Compared to GMI, GCA, and MERIT, which emphasize contrastive learning within the local view, the proposed GRCCA also exhibits great improvements. It is concluded that only focusing on the exploration of the local view is not the best solution. For example, GMI performs contrastive learning between the target nodes and their neighbor nodes, which considers only the local view while ignoring the global view. Note that this is a typical manner induced by topological proximate assumption, which may lead to poor discrimination of node embeddings as described in Section I. These limitations cause it to underperform, while the proposed GRCCA overcomes these limitations through mining both local and global views and considering feature-level similarity rather than just topological proximity. Therefore, our model can achieve significant improvements.

Moreover, it is observed that the performance of MVGRL is comparable to GCA and MERIT, and even surpasses them in some cases, despite using a global vector to coarsely describe the global view. This also underscores the importance of exploring both global and local views. In addition, it is interesting to find that the gap between unsupervised and supervised models is gradually reduced, and unsupervised models can even surpass supervised models by using contrastive learning with graph augmentation strategies. This highlights the effectiveness of multiview contrastive learning. The proposed GRCCA, as a multiview graph contrastive learning model, shows great improvements compared to the supervised models, which further demonstrates this point.

**D. Link Prediction**

For link prediction task, it requires models to predict the interactive relationships between nodes. We compare the proposed GRCCA with the six unsupervised models used in the node classification across Cora, Amazon-Photo, and Coauthor-CS. According to [22], the training set is constructed by removing partial edges from the original graph topology. In this work, we remove 15% edges in the original graphs. The validation and test set contain 5% and 10% of edges from the original graph, respectively, and the same number of negative edges is randomly sampled from pairs of unconnected nodes. In this task, we train all models in an unsupervised manner to obtain the node embeddings, and compute the similarity, the inner product of the embeddings for a pair of nodes, to predict whether a link exists between them or not. The mean and variance of their area under the ROC curve (AUC) and average precision (AP) scores are used as evaluative criteria after five independent experiments over each data set.

As shown in Table IV, the proposed GRCCA also shows strong competitiveness in link prediction, which demonstrates its generalization to some extent. It has a great advantage over other models, although on Amazon-Photo it is slightly inferior to MERIT, which emphasizes the local view. Intuitively, this is because link prediction primarily concerns topological proximity between nodes under the local view. For example, in social networks, two users may be more likely to interact with each other if they share common friends. Otherwise, even though they have the same hobbies, establishing a friendship...
may be challenging due to their distant relationships within the network topology. However, the proposed GRCCA may focus on the hidden associations that do not exist in data sets, which may result in its performance slightly worse than MERIT, because our experiments are based on predicting the existing edges. Therefore, how to actually evaluate the performance of link prediction may be an open problem. Despite this, our model is comparable to MERIT, even surpassing it on Coauthor-CS. This demonstrates the effectiveness of local–global exploration of our model. Moreover, DGI and GIC that also focus on local–global exploration are still comparable to MERIT in most cases, further confirming the importance of local–global exploration. Besides, when comparing the proposed GRCCA with DGI and MVGRL, which consider the global view coarsely using a simple mean readout function, it is clear that our model significantly outperforms them. This suggests our model employs local–global exploration more effectively than DGI and GIC to some extent.

In conclusion, exploring the global view remains beneficial for link prediction, and designing an effective method, such as the proposed GRCCA, to account for fine-grained local–global exploration may be the key to improving the performance.

**E. Community Detection**

To further demonstrate the generalization of the proposed GRCCA, we compare it with other baseline models in the community detection task. In this task, three citation networks are used as benchmarks, and its target is to detect the research areas (communities) that each article belongs to, where nodes are in the same community if they have the same label. We still choose the six unsupervised models as baselines. Similar results are also observed with GIC, which may be because both of them employ cluster-level contrasting to gain deep insight into hidden associations between nodes according to feature-level similarity, thereby overcoming the limitation of topological proximity assumption. However, we notice that although the nodes with similar properties are likely to belong to the same community, this assumption does not always hold. Moreover, using the same clustering algorithm for all data sets may not be the best choice, due to different model preferences. As a result, they are slightly worse than other models on Cora. It is worthy that GMI, which strictly follows the topological proximity assumption, performs worse than all models. This also confirms the effectiveness of our motivation. Additionally, MVGRL, GIC, and the proposed GRCCA overall perform better than GCA and MERIT that emphasize the local view. We believe that this result is attributed to the community detection task actually requiring models to understand the hidden relationships among nodes from the global view. Interestingly, MERIT remains a strong competitiveness compared to GCA. This may be related to its consideration of the global view by data augmentations. This further highlights the importance of exploring both global and local views in GRL. Totally speaking, the proposed GRCCA, which takes more fine-grained local–global exploration into consideration and considers feature-level similarity beyond topological proximity assumption, shows significant improvements over state-of-the-art models. Thus, it can be an alternative unsupervised GRL model.

**F. Ablation Study**

To investigate the impacts of each component in the proposed GRCCA, we conduct the ablation study in node classification, as shown in Table VI. GRCCA/MC denotes GRCCA without multiclustering strategy. GRCCA/AM represents GRCCA without asynchronous mechanism, i.e., synchronous version mentioned in Section III-C, which applies current representations for contrastive clustering. GRCCA/GD is GRCCA without GD strategy, in which the augmentation strategy just focuses on the local view. From Table VI, it is observed that the multicloning strategy brings a slight improvement on two smaller citation networks and negligible effects on the larger data sets. This may be
because $k$-means can enrich the diversity of contrast materials for small-scale data sets. Intuitively, the asynchronous mechanism is used to make the training process more continuous, so as to improve the quality of node embeddings. Its effectiveness is demonstrated in [19]. Our experiments show its effectiveness again in GRL, particularly in Cora and Pubmed. Besides, it can be seen that the opposite augmentation strategy indeed enhances model performance through comparison between GRCCA and GRCCA/GD. This also highlights the importance of exploration of global views.

Notably, compared to MVGRL and MERIT that also employ GD, GRCCA/GD not only performs on par with these models across the three citation networks but also surpasses them on the other three data sets. Furthermore, even if GRCCA/GD employs contrasting clustering assignments with the naive augmentation strategy, it can still surpass GCA that leverages the improved strategy and local node-level contrastive learning. These results further validate the effectiveness of contrastive learning between local and global views. Even without a GD, the model can effectively utilize both global and local information to learn more robust embeddings.

**G. Complexity Analysis**

To illustrate the conciseness and efficiency of the proposed GRCCA, we analyze its model complexity from two aspects: time complexity and model size. The time complexity of proposed GRCCA mainly includes three parts: 1) encoder (GCN); 2) clustering algorithm ($k$-means); and 3) contrastive loss. The time complexity of GCN is $O(E)$, where $E$ is the number of edges. Clustering has complexity $O(NK)$, where $N$ is the number of nodes and $K$ is the number of prototypes. Since $K$ is a constant, its time complexity can be $O(N)$. The contrastive loss costs $O(N)$, as it conducts contrasting between nodes and prototypes and $K$ can be ignored. Therefore, the total time complexity of GRCCA is $O(E)$. Due to all models using the same encoder, we mainly focus on their time complexity of contrastive loss computation, which is shown in Table VII. It is clear to observe that GRCCA and the models based on global and local views like DGI have the lowest time complexity, but the models focusing largely on the local view as GCA have the largest time complexity. This reflects the high efficiency of GRCCA to a certain extent.

We also summarize the model size, i.e., the number of learnable parameters, to show the conciseness of proposed GRCCA. As illustrated in Table VIII, GRCCA has minimal parameters over almost all data sets and reduces nearly half the parameters than the second lowest model, except for Pubmed and Amazon-Computer. We evaluate the actual training time per epoch and GPU memory usage of the proposed GRCCA and the baseline models with the same time complexity, including DGI, GIC, and MVGRL, in the node classification task on Cora. The actual training times for the proposed GRCCA, DGI, GIC, and MVGRL are 0.74, 0.05, 0.02, and 0.7 s, respectively. The GPU memory usages for the proposed GRCCA, DGI, GIC, and MVGRL are 1597, 3985, 1291, and 9565 MB, respectively. We can see that our model uses less GPU memory during training. Although it requires more time per epoch, this is within an acceptable range when it outperforms the baseline models on three downstream tasks in most cases. It is worthy that the actual training time depends on the specific implementation of the code and hardware. To sum up, the proposed GRCCA has good model complexity, and its concise design and fewer parameters enable it to achieve better scalability.

**H. Convergence Analysis**

To demonstrate the convergence of the proposed GRCCA, we show the curve of training loss over all data sets in Fig. 3. It can be seen that GRCCA has convergence across all data sets around 200 epochs. While the curves have slight fluctuations because we report the loss of each epoch without any smoothing process in order to reflect the real convergent cases, we can still observe that the loss curves eventually reach an almost stable value. Therefore, we can empirically conclude that the proposed GRCCA will converge.

**V. CONCLUSION**

To better integrate global and local information, we propose a novel unsupervised GRL model by contrasting cluster assignments called GRCCA. This model combines a clustering algorithm and multiview contrastive learning to get
Fig. 3. Curve of training loss.

insight into potential associations between nodes rather than just topological proximity and achieves the comprehensive exploration of global and local views. Furthermore, it shows a new avenue for unsupervised GRL and demonstrates that exploring cluster-level information is valuable. Experimental results also confirm the rationality of our motivation and the effectiveness of GRCCA.

While GRCCA shows promising performance in various graph analysis tasks, there are still some limitations to be addressed in future work. First, the number of prototypes is sensitive for different data sets. Therefore, how to adaptively select a proper number of prototypes will be one of our future works. Second, the GRCCA is not an end-to-end model, i.e., the clustering process is not contained in gradient optimization. Hence, reinforcement learning may be a promising solution.

For readability and completeness of proof, we restate (8) here

$$L = \frac{1}{N} \sum_{i=1}^{N} Q_{ui} \log P_{vi}$$

where $v_i$ is the target node, $P_{vi}$ is the cluster-level consistency defined as (7), and $Q_{ui}$ is the cluster assignment of its identical nodes $u_i$. Since $Q_{ui}$ is a $K$-dimension one-hot vector that indicates the node $u_i$ belongs to the $k$th prototype, that is, its $k$th dimension value is 1 and other ones is 0, we can rewrite (8) as

$$L = -\mathbb{E}_{v_i \in V} \log \frac{\exp(Z_{v_i} \cdot C^k_{ui} / \tau)}{\exp(Z_{v_i} \cdot C^k_{ui} / \tau) + \sum_{j \neq k} \exp(Z_{v_i} \cdot C^j_{ui} / \tau)}$$

where $C^k_{ui}$ represents the positive sample, that is, the $k$th prototype the identical node $u_i$ belongs to, and $C^j_{ui}$ are other prototypes that are regarded as negative samples. Obviously, (12) is related to the InfoNCE loss [38]. According to [38], the best optimal value $\exp(Z_{v_i} \cdot C^k_{ui} / \tau)$ is given by $\exp((p(C^k_{ui} | Z_{v_i}) / p(C^k_{ui})))$. Therefore, (12) can be rewritten as

$$L^* = -\mathbb{E} \log \left[ \frac{p(C^k_{ui} | Z_{v_i})}{p(C^k_{ui})} + \sum_{j \neq k} \frac{p(C^j_{ui} | Z_{v_i})}{p(C^j_{ui})} \right]$$

By moving the negative sign into the logarithmic function, we can transform (13) to

$$L^* = \mathbb{E} \log \left[ 1 + \frac{p(C^k_{ui})}{p(C^k_{ui} | Z_{v_i})} \sum_{j \neq k} \frac{p(C^j_{ui} | Z_{v_i})}{p(C^j_{ui})} \right]$$

and then we can construct an equal formula

$$L^* = \mathbb{E} \log \left[ 1 + \frac{p(C^k_{ui})}{p(C^k_{ui} | Z_{v_i})} (K - 1) \frac{1}{K - 1} \sum_{j \neq k} \frac{p(C^j_{ui} | Z_{v_i})}{p(C^j_{ui})} \right]$$

### APPENDIX

In this section, we provide a theoretical proof to demonstrate minimizing the loss in (8) is equivalent to maximizing the mutual information between the node embeddings and the corresponding prototypes of another augmented perspective. First of all, we give the definition of mutual information. The mutual information between two continuous random variables, $X$ and $Y$, is a measure of the degree of association between

$$I(X; Y) = \mathbb{E}_{X,Y}[\log \frac{p(X, Y)}{p(X)p(Y)}].$$

### Table VIII

| Model   | Cora   | Citeseer | Pubmed | Amazon-Photo | Coauthor-CS | Amazon-Computers |
|---------|--------|----------|--------|--------------|-------------|------------------|
| DGI     | 996,354| 2,158,594| 193,794| 644,098      | 3,746,818   | 655,362          |
| GMI     | 1,992,709| 4,054,532| 774,660| 1,025,540    | 7,230,980   | 1,048,068        |
| GIC     | 996,354| 2,158,594| 193,794| 644,098      | 3,746,818   | 655,362          |
| GCA     | 997,120| 2,159,361| 519,424| 546,368      | 3,747,584   | 256,896          |
| MVGRL   | 1,730,563| 4,055,043| 775,171| 1,026,051    | 7,231,491   | 1,048,579        |
| MERIT   | 14,089,733| 16,414,213| 13,134,341| 13,385,221 | 19,590,661  | 13,407,749        |
| GRCCA   | 499,201| 1,080,322| 260,353| 323,073      | 1,874,433   | 344,834          |
Therefore, an approximate estimation of (15) is calculated as

$$L^* \approx \mathbb{E} \log \left[ 1 + \frac{p(C^k_{ui})}{p(C^k_{ui}|Z_{vi})} (K - 1) \right]$$

(16)

where the expectation can be approximated by the mean value. Because $C^k_{ui}$ and $Z_{vi}$ are independent, $p(C^k_{ui}|Z_{vi})$ is equal to $p(C^k_{ui})$ and we have

$$L^* \approx \mathbb{E} \log \left[ 1 + \frac{p(C^k_{ui})}{p(C^k_{ui}|Z_{vi})} (K - 1) \right]$$

(17)

$$\geq \mathbb{E} \log \left[ \frac{p(C^k_{ui})}{p(C^k_{ui}|Z_{vi})} \right]$$

(18)

$$= \mathbb{E} \log \left[ \frac{p(C^k_{ui})p(Z_{vi})}{p(C^k_{ui}, Z_{vi})} \right] + \log K$$

(19)

$$= -I(C^k_{ui}, Z_{vi}) + \log K.$$  

(20)

It can be rearranged as

$$I(C^k_{ui}, Z_{vi}) \geq \log K - L^*.$$  

(21)

Therefore, minimizing (8) is equivalent to estimating the lower bound of $I(C^k_{ui}, Z_{vi})$, which means maximizing the mutual information between the node embeddings and their corresponding prototypes of another augmented perspective. Notably, (9) consists of the symmetrical loss function defined in (8), so the proof can be easily generalized for it.

REFERENCES

[1] Q. He et al., “Reinforcement-learning-based dynamic opinion maximization framework in signed social networks,” IEEE Trans. Cogn. Commun. Netw., vol. 15, no. 1, pp. 54–64, Mar. 2023.

[2] K. Pan, T. Jiang, R. Zheng, T. Wang, F. Gao, and J. Cao, “Graph theory based multi-level cortical functional connectivity developmental analysis,” IEEE Trans. Cogn. Dev. Syst., early access, Jun. 14, 2023, doi: 10.1109/TCDS.2023.3285771.

[3] Y.-J. Liu, S. Zhang, and L. Tang, “Distributed cognitive learning strategy for cooperative-competitive multi-agent systems,” IEEE Trans. Cogn. Dev. Syst., early access, Feb. 6, 2023, doi: 10.1109/TCDS.2023.3242370.

[4] W. Zhu, X. Wang, and P. Cui, “Deep learning for learning graph representations,” in Proc. Deep Learn. Concepts Architect., 2020, pp. 169–210.

[5] S. Bhagat, G. Cormode, and S. Muthukrishnan, “Node classification in social networks,” in Proc. Social Netw. Data Analy., 2011, pp. 115–148.

[6] M. Ou, P. Cui, J. Pei, Z. Zhang, and W. Zhu, “Asymmetric transitivity preserving graph embedding,” in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 2016, pp. 1105–1114.

[7] B. Perozzi, R. Al-Rfou, and S. Skiena, “DeepWalk: Online learning of social representations,” in Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 2014, pp. 701–710.

[8] M. Defferrard, X. Bresson, and P. Vandergheynst, “Convolutional neural networks on graphs with fast localized spectral filtering,” in Proc. Adv. Neural Inf. Process. Syst., 2016, pp. 3844–3852.

[9] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” in Proc. Int. Conf. Learn. Represent., 2017, pp. 1–14.

[10] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, “Graph attention networks,” in Proc. Int. Conf. Learn. Represent., 2018, pp. 1–12.

[11] Z.-P. Li et al., “Hierarchical graph pooling with self-adaptive cluster aggregation,” IEEE Trans. Cogn. Develop. Syst., vol. 14, no. 3, pp. 1198–1207, Sep. 2022.

[12] C. Ying et al., “Do transformers really perform bad for graph representation?” in Proc. Adv. Neural Inf. Process. Syst., 2021, pp. 28877–28888.