Robust Contrastive Graph Clustering with Adaptive Local-Global Integration

Lei Zhang[1] , **Fubo Sun**[1] , **Haipeng Yang**[1] , **Zhong Guan**[2] and **Likang Wu**[2*]
[1]School of Computer Science and Technology, Anhui University
[2]College of Management and Economics, Tianjin University
zl@ahu.edu.cn, e24201024@stu.ahu.edu.cn, haipengyang@126.com, kerkersit21@gmail.com, wulk@tju.edu.cn

Abstract

Graph clustering is essential in graph analysis for revealing structural patterns and node communities. Despite recent advances in self-supervised contrastive learning that have improved clustering via structural and attribute signals, existing methods still struggle to flexibly capture high-order local structures and often overlook global semantics in complex graphs. These limitations lead to suboptimal node representations, especially in real-world graphs with fragmented structures and ambiguous cluster boundaries. To address these limitations, a contrastive graph clustering framework is proposed to jointly integrate multi-scale local structures with global semantics via attention mechanisms. At the local level, GNN-based topological signals extracted from multiple propagation depths are adaptively fused through attention-based weighting to capture multi-scale neighborhood features. At the global level, semantic prototypes derived from dynamically evolving cluster centers are adaptively aggregated through attention to guide node representations and enhance inter-cluster separability. The model is trained under a dual-view contrastive learning paradigm with a hybrid objective that combines instance-level and structure-aware losses to improve representation robustness and discrimination. Experiments on eight real-world graph datasets demonstrate that our method achieves competitive clustering performance. Code is available at https://github.com/vege12138/w2.

1 Introduction

Graph clustering is a fundamental task in graph analysis ever, aiming to partition nodes into clusters or communities based on structural attributes or connectivity patterns, which is essential for uncovering underlying patterns and relationships among nodes. This task has been studied and applied in domains including social network analysis [Qiu *et al.*, 2018], knowledge graphs [Bellomarini *et al.*, 2022], and recommender systems [Fan *et al.*, 2019; Wu *et al.*, 2025a]. As a long-standing research topic, graph clustering continues to evolve.

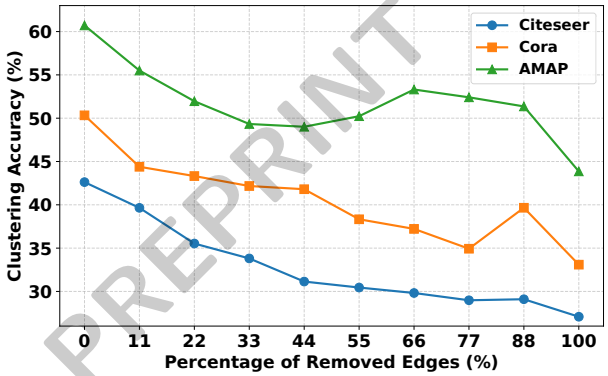

Figure 1: Clustering accuracy (%) versus the percentage of removed high betweenness centrality edges (up to 2700 edges) on three datasets: Cora, Citeseer, and AMAP.

To advance deep graph clustering, Graph Convolutional Networks (GNNs) [Kipf and Welling, 2016] have been employed to learn node representations through message passing. Building on this, self-supervised graph representation learning has emerged as a promising paradigm, leveraging intrinsic structural and attribute information as supervisory signals to guide similar nodes toward compact embeddings. Existing approaches typically follow generative or contrastive paradigms. Generative methods reconstruct graph topology and features via an encoder–decoder architecture, with enhancements combining MLPs and GCNs to capture semantic and structural cues [Tu *et al.*, 2021]. In contrast, contrastive learning constructs positive and negative sample pairs to maximize mutual information between positives while minimizing it for negatives [Peng *et al.*, 2024].

Despite significant advances in deep graph clustering, most existing methods focus primarily on local neighborhood aggregation while overlooking the integration of high-order local structures and global context. This often results in unstable and suboptimal node embeddings, especially in real-world graphs where cluster boundaries are ambiguous. Nodes situated near community intersections or in fragmented subgraphs face particular challenges: they require adaptive fusion of multi-scale neighborhood information across shal-

*Corresponding Author.

low and deep receptive fields, but conventional message passing either ignores this heterogeneity or introduces noise from distant nodes. Moreover, many real-world graphs are fragmented, leaving isolated or weakly connected components with limited access to high-level contextual signals and making cross-community discrimination unreliable. Without global awareness, long-range dependencies become fragile and embeddings for such nodes may collapse. In this work, a basic clustering pipeline is adopted in which GNN-based encoders produce node embeddings that are clustered by k-means; however, as high-betweenness edges are removed, clustering performance further deteriorates (see Figure 1).

To address these limitations, this paper proposes a novel graph clustering framework that jointly integrates multi-scale local structural information and global semantic prototypes in an adaptive manner. At the local level, node representations with different receptive fields are fused through a purpose-built multi-head attention mechanism rather than a generic combination of GNN and attention. Specifically, features from different propagation layers are aggregated via max-pooling and scaled dot-product attention, enabling each node to selectively absorb relevant shallow or deep neighborhood cues. This design directly alleviates the over-smoothing and noise accumulation problems in rigid message-passing schemes, while improving the ability to capture hierarchical structural patterns. At the global level, dynamically updated cluster centers act as semantic prototypes. Through a cluster-guided attention mechanism, nodes selectively integrate global signals that evolve with the latent clustering structure. Unlike static global aggregation, this dynamic strategy adapts to structural changes and is particularly beneficial for nodes in sparse or fragmented regions, addressing the unreliability of long-range dependencies in real-world graphs. This dual-level integration enables the model to handle both local structural ambiguity and global fragmentation.

In addition, the framework adopts a contrastive learning paradigm by generating two augmented graph views to enable label-free representation learning. This is particularly important in unsupervised graph clustering, where the absence of ground-truth labels often leads to collapsed node representations. By enforcing consistency between different views of the same node (via instance-level contrastive loss) and leveraging graph topology to align structural neighbors (via neighbor-supervised loss), the model mitigates collapse and enhances structural discrimination. Consequently, this hybrid training objective yields embeddings that are more robust and consistent, and generalize better across diverse graphs. The main contributions of this work are summarized as follows:

- We propose RCLG, a robust contrastive graph clustering framework that integrates multi-scale local structural cues with global semantic context via attention mechanisms under a dual-view self-supervised paradigm.
- We introduce two complementary components for local-global integration: a multi-head attention-based local fusion module that adaptively weights multi-depth representations, and a prototype-guided global aggregation module that uses dynamic cluster centers to inject global context and improve inter-cluster separability in sparse or fragmented graphs.
- The proposed model demonstrates strong robustness and stability. Without the need for dataset-specific hyperparameter tuning, it achieves consistently competitive clustering performance across diverse graph datasets with varying structures, highlighting its practical utility and deployment-friendliness.

2 Related Work

This section reviews the literature from two key perspectives. The first part surveys the landscape of *Deep Graph Clustering*, providing foundational context for this research. The second part examines a key challenge in this domain: effective capture of *Long-range Dependencies in Graphs*, analyzing the capabilities and limitations of existing approaches.

2.1 Deep Graph Clustering

Deep graph clustering aims to partition graph nodes into meaningful communities by leveraging deep neural networks to jointly capture both node attributes and structural dependencies [Zhao *et al.*, 2025]. With the introduction of GNNs, this task has witnessed significant progress, as GNNs offer a natural way to propagate and fuse feature information over graph topology. By extending the principles of deep clustering, graph-based approaches no longer treat each sample in isolation but instead emphasize the relational structure among nodes, enabling more coherent clustering decisions.

Graph contrastive learning (GCL) [Xu *et al.*, 2021; Xia *et al.*, 2022] is a self-supervised learning method that has been widely studied in various research areas [Zhang *et al.*, 2022; Zheng *et al.*, 2025], such as node classification, recommendation, and clustering. Inspired by the success of GCL in self-supervised representation learning, several works have adopted GCL-based techniques to enhance cluster separability in an unsupervised manner [Devvrit *et al.*, 2022; Ma and Zhan, 2023; Yang *et al.*, 2023; Peng *et al.*, 2025]. Models such as SCGC and RGC initiate clustering by employing shallow graph convolutions to encode local interactions, followed by contrastive learning to capture relational patterns among samples. However, their shallow architectures limit the expressiveness of node embeddings, especially in capturing complex dependencies, which hinders clustering performance.

To address this, methods like SDCN [Bo *et al.*, 2020] and AGCN [Peng *et al.*, 2021] introduce deeper GNN architectures and attribute-structure fusion mechanisms to refine representation quality. These approaches demonstrate that deeper propagation can enrich the semantic content of node embeddings and better capture latent cluster structures. Nevertheless, they still suffer from limitations such as rigid neighborhood aggregation and inadequate handling of higher-order local information. Specifically, they often apply uniform message passing schemes across the graph, which can lead to noise accumulation or oversmoothing, particularly for nodes near community boundaries or in fragmented subgraphs.

2.2 Long-range Dependencies in Graphs

Capturing long-range dependencies is critical in graph learning, particularly for clustering tasks where nodes within the

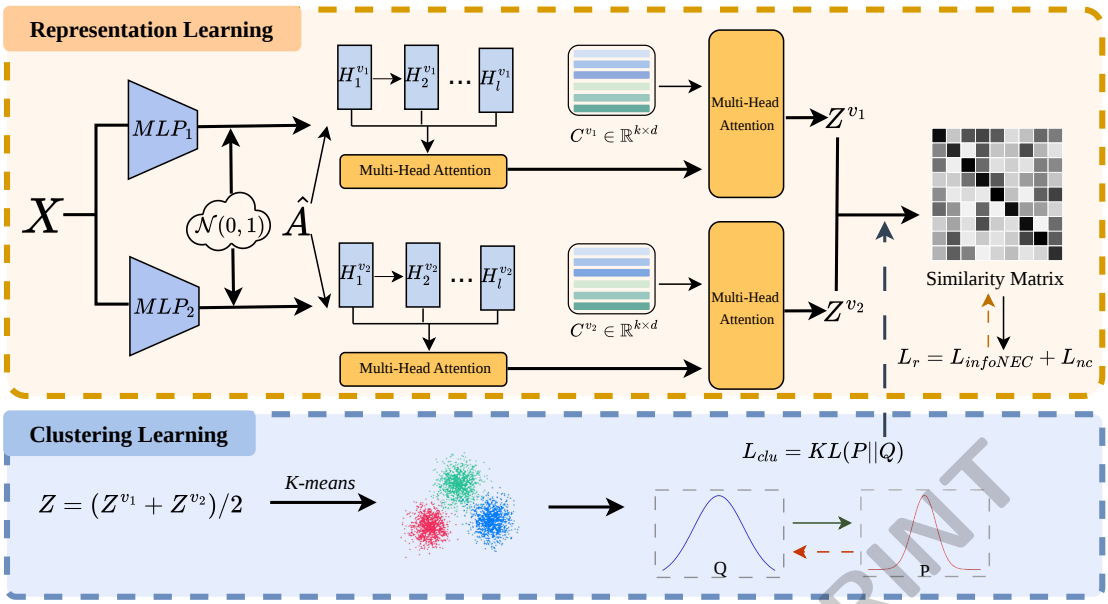

Figure 2: Architecture of the RCLG Model. Within the contrastive learning framework, the model first encodes node features X with noise perturbations. Subsequently, it employs multi-head attention mechanisms to integrate local information through the combination matrix $\hat{A}$ and global information through the cluster centers C^{v_e}. The model then utilizes both representation learning loss and clustering loss L_r to achieve graph clustering.

same community may be distant in the graph topology. Early methods, such as JKNet [Xu *et al.*, 2018] and DAGNN [Liu *et al.*, 2020], attempt to aggregate information using layer-wise skip connections or adaptive weighted mechanisms to capture local features. BM-GCN [He *et al.*, 2022] and HOG-GCN [Wang *et al.*, 2022] consider all neighbors to aggregate global information. However, these methods rely on the original graph topology, which makes them vulnerable to the influence of structural noise or sparsity in the graph.

Recently, MGCN [Li *et al.*, 2024] demonstrates the effectiveness of adaptively fusing multi-layer receptive field information for graph clustering. By incorporating deeper neighborhood contexts, it enables the aggregation of higher-order local structural information, beneficial for capturing nuanced intra-cluster relationships. However, these approaches typically rely on the assumption that the input graph is well-connected. In real-world scenarios, graphs are often fragmented, exhibiting isolated or weakly connected components. In such fragmented structures, traditional message-passing mechanisms either fail to transmit meaningful global signals or suffer from excessive noise introduced by irrelevant paths. As a result, embeddings of nodes—especially those in sparse regions—become unstable and less discriminative. Although DGAC [Xie *et al.*, 2025] is capable of capturing high-order semantic information, it can only marginally handle the issue of missing edges in the original graph. Therefore, effective local integration of higher-order structural information and strong global semantic guidance are both crucial for robust clustering, especially under sparse or fragmented graph connectivity.

3 Methodology

3.1 Overview

We use the following notations throughout this paper. An undirected graph is defined as $G = (V, E)$, where $V = \{v_1, v_2, \ldots, v_N\}$ denotes the set of N nodes and $E \subseteq V \times V$ is the edge set. The node features are represented by the matrix $X \in \mathbb{R}^{N\times d}$, where d is the input feature dimension. The structure of the graph is described by the adjacency matrix $A \in \{0,1\}^{N\times N}$, where $A_{ij} = 1$ if an edge exists between nodes v_i and v_j, and $A_{ij} = 0$ otherwise. The degree matrix is denoted as $D \in \mathbb{R}^{N\times N}$, where each diagonal element D_{ii} equals the degree of node v_i. Incorporating self-loops involves defining $\tilde{A} = A + I$ and $\tilde{D} = D + I$, with I representing the identity matrix. Given an unlabeled graph $G = (V, E)$ with node features X, graph clustering aims to partition the nodes into K disjoint clusters, denoted by an assignment vector $y \in \{1, \ldots, K\}^N$, such that nodes within the same cluster are highly similar while nodes from different clusters are well separated in the learned representation space.

In this section, the proposed model framework for addressing the above problem is introduced. As shown in Figure 2, the method adopts a contrastive learning framework composed of two structurally identical views. For each view, input node features are first encoded, followed by local and global information integration through multi-head attention. This design captures multi-scale structural information and enhances representation robustness. Unless otherwise specified, the following subsections describe the node representation learning process from the perspective of a single view, denoted as view e (where $e = 1$ or 2). In what follows, the framework is introduced component by component starting

from view construction at the input stage, then detailing local and global fusion modules, and finally presenting the overall training objective.

3.2 Noise-based View Augmentation for Robust Contrastive Learning

To improve the robustness of node representations and enhance contrastive learning, Gaussian noise is injected during feature encoding. Given the input node feature matrix X, view-specific embeddings are generated by applying an MLP followed by noise addition:

$$Z^{v_e} = \text{MLP}_e(X) + \alpha \cdot \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, I), \tag{1}$$

where $Z^{v_e} \in \mathbb{R}^{N \times d}$ denotes the embeddings for view e, α controls the noise level, and ε is standard Gaussian noise. This formulation preserves structural consistency while introducing feature perturbations, promoting harder positives and preventing collapse.

Unlike graph-specific augmentations such as edge dropping or diffusion—which are often complex and dataset-sensitive—this feature-space augmentation is simple, efficient, and broadly applicable [Liu *et al.*, 2023]. It offers a lightweight yet effective means of generating distinct but semantically consistent views, facilitating robust contrastive learning in subsequent fusion stages.

3.3 Multi-level Local Feature Fusion via Attention

To capture multi-scale local structural information, repeated message passing is performed over the noisy node embeddings Z^{v_e} using the normalized adjacency matrix $\hat{A}$. Specifically, for each propagation step t, the corresponding layer-wise representation H_t is computed, and all such representations are stacked to obtain a multi-layer feature tensor S:

$$\begin{aligned} H_t &= \hat{A}^t Z^{v_e}, \quad t = 1, 2, \ldots, l \\ S &= \text{stack}(H_1, H_2, \ldots, H_l) \in \mathbb{R}^{N \times l \times d}, \end{aligned} \tag{2}$$

where $\hat{A} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}$ is the symmetrically normalized adjacency matrix, $Z^{v_e} \in \mathbb{R}^{N \times d}$ denotes the initial node embeddings in view v_e, H_t is the node representation after t-hop propagation, and S is the resulting stacked tensor comprising multi-scale representations for each node.

To adaptively select the most informative layer-wise features for each node, max-pooling is applied across all H_t, followed by a linear transformation to obtain the query matrix for attention weighting.

$$Q_L = \left(\max_{1 \le t \le l} H_t \right) W_L^q \in \mathbb{R}^{N \times d}, \tag{3}$$

where $(\max_{1 \le t \le l} H_t)$ denotes the maximum across all layers for each dimension, and $W_L^q \in \mathbb{R}^{d \times d}$ is a learnable matrix transforming pooled features into query vectors for attention.

A scaled dot-product attention mechanism is employed to compute the attention weights between Q_L and each layer in S, and the resulting scores are used to aggregate the layer-wise representations. A residual connection and layer normalization are applied to obtain the locally fused embeddings $Z_L^{v_e}$:

$$\begin{aligned} Z_L^{v_e} &= \text{softmax}\left(\frac{Q_L S^\top}{\sqrt{d_h}} \right) S \\ Z_L^{v_e} &= \text{LayerNorm}\left(Z_L^{v_e} + Z^{v_e} \right), \end{aligned} \tag{4}$$

where d_h is hidden dimension for scaled attention, and Z^{v_e} is the residual input for stabilization. The function LayerNorm$(\cdot)$ normalizes the output to improve training.

This adaptive fusion strategy enables each node to selectively integrate signals from various receptive field depths, improving the quality and robustness of local representations.

3.4 Global Semantic Aggregation

To supplement local representations with semantic-level global context, we introduce a cluster-aware global aggregation mechanism. A set of cluster centers, as global semantic prototypes, is dynamically updated. These centers are computed by applying a clustering algorithm (e.g., K-Means [Hartigan and Wong, 1979] or spectral clustering [Von Luxburg, 2007]) to fused node representations $Z_L^{v_e}$:

$$C_e = \Phi(Z_L^{v_e}) \in \mathbb{R}^{k \times d}, \tag{5}$$

where C_e denotes the set of k cluster centers derived from the current view e, Φ denotes a clustering algorithm and each center represents a global semantic anchor corresponding to a latent cluster.

To incorporate global semantics into node-level representations, a multi-head attention mechanism [Vaswani *et al.*, 2017] is used. In this process, node embeddings and global context are linearly projected and reshaped into multiple attention heads, forming the query, key, and value matrices:

$$\begin{aligned} Q_G &= \text{reshape}(Z_L^{v_e} W_G^q) \in \mathbb{R}^{n \times h \times d_h} \\ K = V &= \text{reshape}(C_e) \in \mathbb{R}^{k \times h \times d_h}, \end{aligned} \tag{6}$$

where $W_G^q \in \mathbb{R}^{d \times d}$ is a learnable projection matrix for transforming node embeddings into queries, and $Z_L^{v_e} \in \mathbb{R}^{n \times d}$ represents the locally fused node embeddings from view v_e.

Scaled dot-product attention is then applied between each node and the cluster centers, enabling nodes to selectively attend to global semantic prototypes:

$$G = \text{softmax}\left(\frac{Q_G K^\top}{\sqrt{d_h}} \right) V, \tag{7}$$

where Q_G, K, and V are the query, key, and value matrices as defined in the previous step; d_h denotes the dimensionality of each attention head.

Subsequently, a residual connection is formed between the global aggregated output and the locally fused representation, followed by layer normalization:

$$Z_G^{v_e} = \text{LayerNorm}\left(\beta \cdot G + (1 - \beta) \cdot Z_L^{v_e} \right), \tag{8}$$

where $\beta \in [0, 1]$ is a coefficient controlling the trade-off between global and local signals.

This global semantic aggregation mechanism strengthens node embeddings by incorporating high-level structural prototypes from the entire graph. Such enhancement is especially advantageous for nodes situated in weakly connected or isolated components, where local message passing is often insufficient for capturing informative global context.

3.5 Representation Learning and Clustering Objective

To obtain discriminative and semantically rich node representations, the representation learning loss L_r consists of an instance-level contrastive term L_{infoNEC} and a neighbor-supervised term L_{ns}.

The instance-level contrastive loss follows the InfoNCE principle, aiming to pull together representations of the same node across two augmented views while pushing apart representations of other node pairs. Given embeddings from the two augmented views $Z_G^{v_e}$ and $Z_G^{v_f}$, the similarity between two nodes is defined as:

$$\theta\left(i^{v_e}, j^{v_f}\right) = Z_{G_i}^{v_e} \cdot Z_{G_j}^{v_f}, \tag{9}$$

where $Z_{G_i}^{v_e}$ and $Z_{G_j}^{v_f}$ are the normalized embeddings of nodes i and j in views v_e and v_f, respectively.

The instance-level InfoNCE loss is then given by:

$$L_{\text{infoNEC}} = \frac{1}{2N} \sum_{e=1}^{2} \sum_{i=1}^{N} L_i^{v_e}, \tag{10}$$

$$L_i^{v_e} = -\log \frac{e^{\theta(i^{v_e}, i^{v_f})}}{e^{\theta(i^{v_e}, i^{v_f})} + \sum_{j \neq i} \left(e^{\theta(i^{v_e}, j^{v_e})} + e^{\theta(i^{v_e}, j^{v_f})}\right)} \tag{11}$$

where $L_i^{v_e}$ denotes the contrastive loss for node i in view v_e, and N is the number of nodes.

To further incorporate structural signals, L_{ns} is introduced. This term enforces similar embeddings for neighboring nodes based on the binarized normalized adjacency matrix $\tilde{A}$:

$$L_{\text{ns}} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \Big[- \tilde{a}_{ij} \log\big(\theta(i^{v_e}, j^{v_f})\big) - (1 - \tilde{a}_{ij}) \log\big(1 - \theta(i^{v_e}, j^{v_f})\big) \Big], \tag{12}$$

where $\tilde{a}_{ij}$ indicates the presence of an edge between nodes i and j.

The combined representation learning loss is defined as:

$$L_{\text{r}} = L_{\text{infoNEC}} + L_{\text{ns}}. \tag{13}$$

Beyond improving embedding quality, a clustering alignment loss L_{clu} is adopted to enhance clustering performance, inspired by previous works [Guo *et al.*, 2017; Wang *et al.*, 2019a; Xie *et al.*, 2016]. This term aligns the soft cluster assignments q with a sharpened target distribution p, and is typically optimized using KL divergence.

Combining all components, the final training objective is:

$$L = L_{\text{r}} + \gamma L_{\text{clu}}, \tag{14}$$

where γ is a hyperparameter controlling the influence of the clustering objective.

Dataset	#Nodes	#Edges	Edge Density	#Features	#Classes
UAT	1,190	13,599	0.0192	239	4
Wiki	2,405	8,261	0.0029	4,973	17
Cora	2,708	5,278	0.0014	1,433	7
ACM	3,025	13,128	0.0029	1,870	3
Citeseer	3,327	4,552	0.0008	3,703	6
DBLP	4,057	3,528	0.0004	334	4
AMAP	7,650	119,043	0.0041	745	8
COCS	18,333	81,894	0.0005	6,805	15

Table 1: Dataset statistics of real-world datasets.

4 Experiments

This section presents a comprehensive evaluation of the proposed model from multiple perspectives. It begins by introducing benchmark datasets, evaluation metrics, and experimental setup. The clustering results of RCLG are compared with state-of-the-art baselines, followed by analyses of RCLG variants, including ablation studies and hyperparameter sensitivity experiments. Experiments were conducted on a workstation with an NVIDIA RTX 4090 GPU.

4.1 Benchmark Datasets and and Evaluation Metrics

The evaluation is conducted on eight widely used real-world graph datasets with diverse structures and domains, covering air-traffic networks (UAT [Mrabah *et al.*, 2022]), word co-occurrence graphs (Wiki), citation networks (Cora [Cui *et al.*, 2020] and Citeseer [Cui *et al.*, 2020]), e-commerce co-purchase graphs (AMAP [Liu *et al.*, 2022]), and academic co-authorship networks (ACM [Wang *et al.*, 2019b], DBLP [Wang *et al.*, 2019b], and COCS [Shchur *et al.*, 2018]). Dataset statistics are summarized in Table 1. Clustering performance is evaluated using three widely adopted metrics: Accuracy (ACC), Normalized Mutual Information (NMI), and macro F1-score (F1).

4.2 Baselines and Experimental Setup

Eight representative graph clustering methods are selected for comparison, including traditional structure modeling approaches and contrastive learning-based methods. Specifically, AGC [Zhang *et al.*, 2019] employs adaptive graph convolution to model global structural information. SDCN [Bo *et al.*, 2020] integrates node attributes and structure in a unified clustering framework. FGC [Kang *et al.*, 2022] proposes a fine-grained strategy for attribute graph clustering. VGAER [Qiu *et al.*, 2022] utilizes variational inference to jointly model topology and attributes. AGC-DRR [Gong *et al.*, 2022] designs a redundancy-reduced representation learning method. In addition, recent contrastive learning-based methods are considered, including SCGC [Liu *et al.*, 2023] which focuses on efficient neighbor contrast, MAGI [Liu *et al.*, 2024] which leverages modularity, MCGC [Wu *et al.*, 2025b] addressing motif-based contrastive clustering, and WGCN [Wang *et al.*, 2026] which adopts a representation-then-augmentation framework.

The model is trained with learning rate $\text{lr} = 5 \times 10^{-4}$ for 500 epochs. The noise perturbation coefficient is set to $\alpha = 0.01$, and the clustering update interval is $T = 5$. The global semantic fusion coefficient is $\beta = 0.2$, and the cluster-

Dataset	Metrics	AGC	SDCN	FGC	VGAER	AGC-DRR	SCGC	MAGI	MCGC	WGCN	RCLG
UAT	ACC	48.72±0.40	43.85±2.28	45.65±2.70	48.02±0.80	41.38±3.73	**59.74±0.82**	53.72±0.38	54.14±0.38	58.04±0.91	59.58±0.79
	NMI	23.74±1.17	13.59±2.19	23.20±1.87	24.44±0.88	10.58±2.79	29.32±1.06	21.74±0.85	27.20±0.26	29.50±1.13	**30.17±0.34**
	F1	44.57±0.93	40.72±2.37	40.17±5.00	49.13±0.62	37.22±5.29	57.33±1.47	52.04±0.46	55.61±0.20	57.50±1.04	**58.12±0.98**
Wiki	ACC	38.01±7.28	41.58±3.07	49.57±5.76	41.53±1.16	43.75±6.86	57.07±0.43	52.21±1.13	58.21±0.68	58.44±0.84	**59.13±0.70**
	NMI	36.26±6.95	36.05±1.48	49.01±3.79	36.37±0.45	41.54±6.94	52.80±0.60	52.49±0.57	54.40±0.74	55.06±0.97	**55.16±0.60**
	F1	28.06±8.61	26.85±2.34	40.15±3.47	36.56±0.96	33.17±4.73	46.27±1.42	43.88±1.18	48.71±0.56	48.66±0.82	**49.98±0.77**
Cora	ACC	60.98±2.92	57.00±4.42	58.43±10.66	67.53±1.25	59.74±4.22	73.67±0.51	**75.09±0.88**	73.83±0.44	74.92±0.74	75.06±0.38
	NMI	47.00±2.88	36.44±3.33	38.99±11.68	47.79±0.73	48.16±3.07	55.92±0.58	56.97±0.75	56.77±0.64	58.52±0.96	**59.18±0.72**
	F1	51.33±5.09	50.86±2.62	53.62±11.17	66.27±0.78	54.10±2.43	70.43±1.37	**72.78±1.19**	69.28±0.38	71.38±1.64	72.61±0.76
ACM	ACC	83.50±0.00	74.50±2.13	81.55±1.76	51.04±0.49	87.50±10.03	89.91±0.32	89.58±0.37	89.34±0.31	90.46±0.23	**90.56±0.38**
	NMI	55.17±0.06	44.23±1.65	51.94±3.84	18.56±0.59	63.11±17.05	66.75±0.64	65.58±0.64	66.10±0.93	68.47±0.54	**68.90±0.96**
	F1	83.60±0.00	74.26±2.39	81.01±1.92	41.31±0.64	87.45±10.21	89.87±0.34	89.65±0.35	89.30±0.24	90.44±0.23	**90.55±0.39**
Citeseer	ACC	68.43±0.05	62.03±1.83	61.20±3.36	47.46±0.99	67.71±1.35	71.14±0.91	69.80±0.57	70.97±0.70	**72.49±0.39**	71.04±0.71
	NMI	42.61±0.03	35.55±2.14	35.34±3.98	25.16±0.80	42.67±1.43	44.78±1.08	43.96±0.52	44.87±0.30	**46.76±0.59**	46.16±0.59
	F1	63.80±0.00	58.16±1.55	57.76±2.65	46.26±1.36	63.64±1.20	61.95±0.97	65.03±0.70	61.89±0.54	63.34±0.49	**65.92±0.57**
DBLP	ACC	62.38±3.51	56.31±0.94	73.65±0.98	44.54±1.88	61.43±2.34	68.23±0.58	78.86±0.51	72.33±0.21	62.05±1.52	**80.32±1.14**
	NMI	32.15±3.20	23.64±1.78	42.51±1.69	11.57±1.26	32.77±2.29	39.86±0.57	49.72±0.52	43.75±1.35	31.34±1.72	**50.76±1.44**
	F1	62.12±3.53	55.74±1.20	73.31±0.97	43.35±1.81	59.13±2.07	68.02±0.60	78.55±0.48	71.97±0.47	62.62±1.53	**79.98±1.10**
AMAP	ACC	61.51±5.31	57.77±3.77	45.22±10.15	71.41±1.77	66.79±2.80	78.21±0.62	79.01±0.47	77.81±0.65	78.76±0.62	**83.66±2.17**
	NMI	56.48±3.13	44.09±4.08	33.88±14.69	58.56±1.95	56.02±5.20	65.51±1.05	70.23±0.67	65.23±0.31	68.65±0.76	**75.03±2.18**
	F1	56.82±5.53	42.88±3.40	38.23±9.19	66.95±1.22	60.51±3.80	74.18±2.75	72.58±0.35	75.39±0.35	72.76±0.59	**83.20±2.32**
COCS	ACC	65.76±4.59	60.07±3.00	75.38±2.93	61.12±1.77	60.37±9.92	76.21±2.00	75.36±0.12	71.45±0.26	76.58±1.27	**81.85±1.89**
	NMI	64.77±2.53	60.10±1.22	77.16±1.34	61.47±2.27	61.81±10.78	71.35±1.74	77.88±0.07	65.84±0.23	78.44±1.26	**82.90±0.56**
	F1	53.14±5.34	37.64±2.33	75.71±3.90	57.60±2.68	43.10±7.60	69.08±3.26	74.08±0.08	55.02±0.32	70.71±0.69	**76.14±1.70**

Table 2: Clustering performance comparison on eight datasets (Mean ± Std over 10 runs; Bold = best; Underline = second best).

ing alignment loss is weighted by $\gamma = 10$. The hidden representation dimensionality is 256, and the number of graph propagation steps is $l = 6$. For baselines, the best configuration is individually selected for each dataset, whereas our method uses a single fixed configuration across all datasets.

4.3 Clustering Performance Comparisons

As shown in Table 2, the proposed method exhibits particularly notable improvements on larger-scale datasets. Specifically, on the AMAP dataset, a relative accuracy improvement of approximately 5.88% is achieved compared to the second-best approach, while on the COCS dataset, the improvement reaches 7.40%. These results indicate that integrating local and global information becomes increasingly effective as the dataset scale increases. Moreover, the method is robust to varying sparsity. Table 1 reports that Wiki and ACM are relatively denser (0.0029), whereas DBLP (0.0004) and COCS (0.0005) are among the sparsest datasets; nevertheless, strong performance is consistently achieved, indicating reliable clustering under sparse or fragmented connectivity.

4.4 Ablation Study

To analyze the contribution of each key component and mechanism, ablation studies are conducted on different model variants. Specifically, w/o A denotes removing all attention mechanisms: attention is replaced with mean fusion in the local component, and equal weights are assigned to all cluster centers in the global component. w/o L&G denotes the variant that performs simple mean fusion of multi-scale representations without the global information integration module. w/o G removes only the global information integration component, while RCLG represents the full model. As shown in Table 3, on the Wiki dataset, the integration of local information improves clustering performance with a relative enhancement of 7.1%. On the Cora dataset, incorporating global information yields a relative improvement of 1.7%. These findings demonstrate that both local and global information integration modules substantially enhance the overall framework's effectiveness.

Metrics	Methods	UAT	Wiki	Cora	ACM	Citeseer	DBLP	AMAP	COCS
ACC	w/o A	54.99	54.97	73.09	89.51	70.49	76.73	81.34	80.41
	w/o L&G	55.06	54.67	73.26	89.55	69.80	77.27	81.34	80.91
	w/o G	58.74	58.54	73.80	90.01	70.38	79.70	82.88	81.09
	RCLG	**59.58**	**59.13**	**75.06**	**90.56**	**71.04**	**80.32**	**83.66**	**81.84**
NMI	w/o A	25.90	48.92	57.74	66.19	44.75	45.85	73.72	79.38
	w/o L&G	25.93	48.52	58.12	66.19	44.67	46.75	73.77	79.43
	w/o G	30.08	54.68	57.89	67.35	45.38	50.10	75.85	82.83
	RCLG	**30.17**	**55.16**	**59.18**	**68.90**	**46.16**	**50.76**	**75.03**	**82.90**
F1	w/o A	54.40	46.22	70.76	89.49	65.54	76.47	82.31	77.73
	w/o L&G	54.49	46.37	71.38	89.52	65.37	76.95	81.66	76.18
	w/o G	57.23	49.53	71.83	90.02	65.32	79.36	82.38	**76.60**
	RCLG	**58.12**	**49.98**	**72.61**	**90.55**	**65.92**	**79.98**	**83.20**	76.14

Table 3: Comparison between RCLG and its variants in terms of ACC, NMI, and F1.

Metrics	Methods	UAT	Wiki	Cora	ACM	Citeseer	DBLP	AMAP	COCS
ACC	w/o L_r	50.08	54.15	24.85	69.49	25.62	57.88	66.16	69.69
	w/o $L_{InfoNEC}$	52.74	52.47	52.54	85.10	46.25	62.81	72.38	80.72
	w/o L_{ns}	58.47	55.92	74.16	89.65	69.20	68.78	82.37	78.98
	RCLG	**59.58**	**59.13**	**75.06**	**90.56**	**71.04**	**80.32**	**83.66**	**81.85**
NMI	w/o L_r	25.09	47.54	5.40	29.45	2.40	26.59	56.66	73.95
	w/o $L_{InfoNEC}$	25.73	50.09	33.43	57.71	23.22	31.62	60.66	77.90
	w/o L_{ns}	29.92	53.76	56.53	66.49	44.43	38.31	74.16	82.05
	RCLG	**30.17**	**55.16**	**59.18**	**68.90**	**46.16**	**50.76**	**75.03**	**82.90**
F1	w/o L_r	48.49	41.96	22.53	69.26	22.90	58.36	62.76	65.24
	w/o $L_{InfoNEC}$	50.47	44.43	51.56	85.06	42.67	62.77	68.35	69.01
	w/o L_{ns}	56.50	47.88	72.56	89.67	64.98	69.14	82.01	76.04
	RCLG	**58.12**	**49.98**	**72.61**	**90.55**	**65.92**	**79.98**	**83.20**	**76.14**

Table 4: Ablation Study on Instance-Level (L_{infoNEC}) and Neighbor-Supervised (L_{ns}) Contrastive Losses in the Full Representation Objective (L_r).

In addition, Table 4 presents the effect of the two components constituting the representation learning loss L_{R}. The results indicate that both the instance-level contrastive loss L_{infoNEC} and the neighbor-supervised contrastive loss L_{ns} contribute to improving clustering performance. Among them, L_{infoNEC} plays a more prominent role by ensuring fundamental discriminability between node embeddings, which is particularly beneficial for enhancing clustering quality.

4.5 Impact of Clustering Algorithm Choice

To examine the effect of the clustering algorithm, we assess whether the Global component is sensitive to the choice of

Metrics	Model	UAT	Wiki	Cora	ACM	Citeseer	DBLP	AMAP	COCS
ACC	SpecC	58.12	59.10	74.74	**91.22**	**71.06**	80.16	83.27	80.21
	RCLG	**59.58**	**59.13**	**75.06**	90.56	71.04	**80.32**	**83.66**	**81.85**
NMI	SpecC	29.77	**55.71**	58.50	**70.17**	**46.16**	**50.76**	**76.59**	82.45
	RCLG	**30.17**	55.16	**59.18**	68.90	**46.16**	**50.76**	75.03	**82.90**
F1	SpecC	57.06	49.37	70.80	**91.20**	**66.58**	79.77	81.47	73.68
	RCLG	**58.12**	**49.98**	**72.61**	90.55	65.92	**79.98**	**83.20**	**76.14**

Table 5: Performance comparison after replacing the clustering algorithm with Spectral Clustering (SpecC) method.

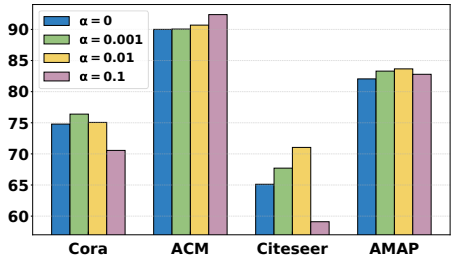

Figure 3: Sensitivity analysis of hyper-parameter noise coefficient α on clustering performance(ACC).

Dataset	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
UAT	59.19	58.86	59.58	**59.68**	59.19	57.36	51.92	50.19	48.66	48.64
Wiki	59.01	59.61	59.13	59.43	59.53	**59.89**	53.84	55.28	54.37	49.21
Cora	72.48	73.35	**75.06**	73.50	71.86	52.85	33.29	26.22	26.61	25.57
ACM	90.03	90.75	90.56	90.98	**92.05**	88.73	83.35	73.10	74.55	61.00
Citeseer	70.31	70.75	**71.04**	70.75	70.70	59.34	31.10	26.18	25.93	25.11
DBLP	79.83	79.66	**80.32**	80.02	79.66	80.17	75.75	58.66	57.75	50.87
AMAP	82.66	83.24	**83.66**	83.65	66.97	65.20	66.28	64.36	64.04	61.14
COCS	81.09	81.33	81.85	81.43	84.50	**87.09**	84.16	80.65	75.44	74.06

Table 6: Sensitivity analysis of global fusion coefficient β on clustering performance(ACC).

Dataset	1	2	3	4	5	6	7	8	9	10
UAT	57.51	58.77	58.71	59.63	59.10	59.58	**59.97**	58.71	59.49	58.97
Wiki	59.83	59.27	57.96	59.00	**60.36**	59.13	58.62	59.20	58.45	58.68
Cora	71.48	71.63	71.92	71.62	73.84	**75.06**	72.28	73.95	73.46	73.86
ACM	90.23	90.77	**91.01**	90.81	90.75	90.56	90.14	90.27	90.46	90.07
Citeseer	70.43	70.34	70.89	71.50	70.95	71.04	**71.38**	71.01	71.01	70.50
DBLP	80.09	79.66	80.26	80.32	79.48	**80.35**	79.88	79.32	78.26	79.73
AMAP	**84.62**	83.04	84.62	83.13	82.49	83.66	83.91	82.39	83.26	82.01
COCS	78.09	80.17	80.73	81.79	81.23	81.85	82.15	83.31	82.82	**83.55**

Table 7: Sensitivity of propagation steps l on clustering performance(ACC).

method. The Global module only requires reliable cluster centers to deliver global information via attention, and these centers are treated as learnable parameters updated jointly with the model, reducing dependence on random initialization and enhancing robustness.

To validate this, we replaced the default clustering algorithm with Spectral Clustering. As shown in Table 5, the replacement has little impact on performance. Across datasets, Spectral Clustering and our original RCLG achieve comparable results, confirming that the model is not tied to a specific clustering algorithm and remains robust and effective.

4.6 Sensitivity Analysis

To further assess model robustness, sensitivity analysis is performed on three hyperparameters: noise coefficient α, global fusion coefficient β, and propagation layers l.

As shown in Figure 3, setting the noise coefficient α to 0.01 yields stable and strong performance across multiple datasets,

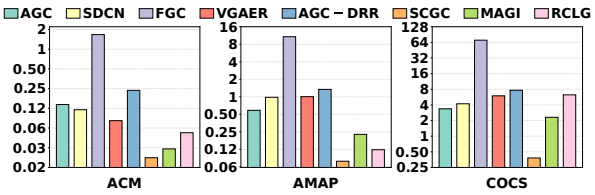

Figure 4: Per-epoch runtime comparison (s) of different methods across datasets.

demonstrating the effectiveness of introducing mild noise to enhance representation robustness. Likewise, Table 6 shows that within the range $\beta \in [0.2, 0.4]$, each dataset tends to achieve its optimal performance, indicating that the model remains robust and broadly applicable when balancing global and local information in this interval. When either α or β is set excessively high, clustering performance degrades on several datasets, suggesting that overly strong perturbations or global dependencies may overwhelm local learning. This confirms the necessity of maintaining a balanced integration between local and global information.

Furthermore, as summarized in Table 7, datasets with larger diameters generally reach best performance at slightly deeper propagation steps, reflecting the need for sufficient receptive field expansion to capture long-range dependencies. Considering both accuracy and efficiency, we fix the propagation depth to $l = 6$ in all subsequent experiments.

4.7 Computational Efficiency

To further assess the practicality of our method, we compared the per-epoch runtime of RCLG with several state-of-the-art baselines on three representative datasets, namely COCS (18,333 nodes), AMAP (7,650 nodes), and ACM (3,025 nodes). As summarized in Figure 4, the runtime of RCLG is of the same order of magnitude as other methods, indicating that the improved clustering performance does not incur excessive computational overhead. This suggests that our framework achieves competitive efficiency while maintaining superior clustering quality.

5 Conclusion

This paper presents a robust contrastive graph clustering framework that leverages both local and global structural signals within a dual-view architecture to address key limitations. A multi-level attention fusion aggregates neighborhood information from multiple convolutional layers, allowing each node to selectively absorb informative cues across depths and reducing noise from rigid message passing. Semantic prototypes from evolving cluster centers further provide global guidance, enabling nodes—especially in sparse or disconnected regions—to incorporate context beyond immediate neighbors, stabilizing long-range semantics and strengthening cross-community discrimination. Together, these mechanisms improve representation quality and promote more compact and separable clusters. The framework achieves strong performance across diverse benchmarks without dataset-specific hyperparameter tuning.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 61976001, 62502340), the Natural Science Foundation of Anhui Province (No. 2408085MF152, 2508085MF176) and the Natural Science Foundation of Tianjin (No. 24JCQNJC01560).

References

[Bellomarini *et al.*, 2022] Luigi Bellomarini, Davide Benedetto, Georg Gottlob, and Emanuel Sallinger. Vadalog: A modern architecture for automated reasoning with large knowledge graphs. *Information Systems*, 105:101528, 2022.

[Bo *et al.*, 2020] Deyu Bo, Xiao Wang, Chuan Shi, Meiqi Zhu, Emiao Lu, and Peng Cui. Structural deep clustering network. In *Proceedings of the Web Conference 2020*, pages 1400–1410, 2020.

[Cui *et al.*, 2020] Ganqu Cui, Jie Zhou, Cheng Yang, and Zhiyuan Liu. Adaptive graph encoder for attributed graph embedding. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 976–985, 2020.

[Devvrit *et al.*, 2022] Devvrit, Aditya Sinha, Inderjit S. Dhillon, and Prateek Jain. S3GC: scalable self-supervised graph clustering. In Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022.

[Fan *et al.*, 2019] Wenqi Fan, Yao Ma, Qing Li, Yuan He, Eric Zhao, Jiliang Tang, and Dawei Yin. Graph neural networks for social recommendation. In *The World Wide Web Conference*, pages 417–426, 2019.

[Gong *et al.*, 2022] Lei Gong, Sihang Zhou, Wenxuan Tu, and Xinwang Liu. Attributed graph clustering with dual redundancy reduction. In *IJCAI*, pages 3015–3021, 2022.

[Guo *et al.*, 2017] Xifeng Guo, Long Gao, Xinwang Liu, and Jianping Yin. Improved deep embedded clustering with local structure preservation. In *IJCAI*, volume 17, pages 1753–1759, 2017.

[Hartigan and Wong, 1979] John A Hartigan and Manchek A Wong. Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. series c (applied statistics)*, 28(1):100–108, 1979.

[He *et al.*, 2022] Dongxiao He, Chundong Liang, Huixin Liu, Mingxiang Wen, Pengfei Jiao, and Zhiyong Feng. Block modeling-guided graph convolutional neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 4022–4029, 2022.

[Kang *et al.*, 2022] Zhao Kang, Zhanyu Liu, Shirui Pan, and Ling Tian. Fine-grained attributed graph clustering. In *Proceedings of the 2022 SIAM International Conference on Data Mining (SDM)*, pages 370–378. SIAM, 2022.

[Kipf and Welling, 2016] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[Li *et al.*, 2024] Xiulai Li, Wei Wu, Bin Zhang, and Xin Peng. Multi-scale graph clustering network. *Information Sciences*, 678:121023, 2024.

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

[Liu *et al.*, 2022] Yue Liu, Wenxuan Tu, Sihang Zhou, Xinwang Liu, Linxuan Song, Xihong Yang, and En Zhu. Deep graph clustering via dual correlation reduction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 7603–7611, 2022.

[Liu *et al.*, 2023] Yue Liu, Xihong Yang, Sihang Zhou, Xinwang Liu, Siwei Wang, Ke Liang, Wenxuan Tu, and Liang Li. Simple contrastive graph clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[Liu *et al.*, 2024] Yunfei Liu, Jintang Li, Yuehe Chen, Ruofan Wu, Ericbk Wang, Jing Zhou, Sheng Tian, Shuheng Shen, Xing Fu, Changhua Meng, et al. Revisiting modularity maximization for graph clustering: A contrastive learning perspective. In *Proceedings of the 30th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1968–1979, 2024.

[Ma and Zhan, 2023] Yixuan Ma and Kun Zhan. Self-contrastive graph diffusion network. In Abdulmotaleb El-Saddik, Tao Mei, Rita Cucchiara, Marco Bertini, Diana Patricia Tobon Vallejo, Pradeep K. Atrey, and M. Shamim Hossain, editors, *Proceedings of the 31st ACM International Conference on Multimedia, MM 2023, Ottawa, ON, Canada, 29 October 2023- 3 November 2023*, pages 3857–3865. ACM, 2023.

[Mrabah *et al.*, 2022] Nairouz Mrabah, Mohamed Bouguessa, Mohamed Fawzi Touati, and Riadh Ksantini. Rethinking graph auto-encoder models for attributed graph clustering. *IEEE Transactions on Knowledge and Data Engineering*, 35(9):9037–9053, 2022.

[Peng *et al.*, 2021] Zhihao Peng, Hui Liu, Yuheng Jia, and Junhui Hou. Attention-driven graph clustering network. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 935–943, 2021.

[Peng *et al.*, 2024] Xin Peng, Jieren Cheng, Xiangyan Tang, Bin Zhang, and Wenxuan Tu. Multi-view graph imputation network. *Information Fusion*, 102:102024, 2024.

[Peng *et al.*, 2025] Tianhao Peng, Xuhong Li, Haitao Yuan, Yuchen Li, and Haoyi Xiong. Sola-gcl: Subgraph-oriented learnable augmentation method for graph contrastive learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 19875–19883, 2025.

[Qiu *et al.*, 2018] Jiezhong Qiu, Jian Tang, Hao Ma, Yuxiao Dong, Kuansan Wang, and Jie Tang. Deepinf: Modeling

influence locality in large social networks. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2110–2119, 2018.

[Qiu *et al.*, 2022] Chenyang Qiu, Zhaoci Huang, Wenzhe Xu, and Huijia Li. Vgaer: graph neural network reconstruction based community detection. *arXiv preprint arXiv:2201.04066*, 2022.

[Shchur *et al.*, 2018] Oleksandr Shchur, Maximilian Mumme, Aleksandar Bojchevski, and Stephan Günnemann. Pitfalls of graph neural network evaluation. *arXiv preprint arXiv:1811.05868*, 2018.

[Tu *et al.*, 2021] Wenxuan Tu, Sihang Zhou, Xinwang Liu, Xifeng Guo, Zhiping Cai, En Zhu, and Jieren Cheng. Deep fusion clustering network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 9978–9987, 2021.

[Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.

[Von Luxburg, 2007] Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17:395–416, 2007.

[Wang *et al.*, 2019a] Chun Wang, Shirui Pan, Ruiqi Hu, Guodong Long, Jing Jiang, and Chengqi Zhang. Attributed graph clustering: A deep attentional embedding approach. *arXiv preprint arXiv:1906.06532*, 2019.

[Wang *et al.*, 2019b] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. Heterogeneous graph attention network. In *The World Wide Web Conference*, pages 2022–2032, 2019.

[Wang *et al.*, 2022] Tao Wang, Di Jin, Rui Wang, Dongxiao He, and Yuxiao Huang. Powerful graph convolutional networks with adaptive propagation mechanism for homophily and heterophily. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 4210–4218, 2022.

[Wang *et al.*, 2026] Youqing Wang, Tianxiang Zhao, Mingliang Cui, Junbin Gao, Li Liang, and Jipeng Guo. Representation then augmentation: Wide graph clustering network with multi-order filter fusion and double-level contrastive learning. *IEEE/CAA Journal of Automatica Sinica*, 13(2):421–435, 2026.

[Wu *et al.*, 2025a] Xinyi Wu, Donald Loveland, Runjin Chen, Yozen Liu, Xin Chen, Leonardo Neves, Ali Jadbabaie, Mingxuan Ju, Neil Shah, and Tong Zhao. Graphhash: Graph clustering enables parameter efficiency in recommender systems. In *Proceedings of the ACM on Web Conference 2025, WWW 2025, Sydney, NSW, Australia, 28 April 2025- 2 May 2025*, pages 357–369. ACM, 2025.

[Wu *et al.*, 2025b] Xunlian Wu, Jingqi Hu, Yining Quan, Qiguang Miao, and Peng Gang Sun. Motif-based contrastive graph clustering with clustering-oriented prompt. *Information Processing & Management*, 62(5):104208, 2025.

[Xia *et al.*, 2022] Jun Xia, Lirong Wu, Jintao Chen, Bozhen Hu, and Stan Z Li. Simgrace: A simple framework for graph contrastive learning without data augmentation. In *Proceedings of the ACM Web Conference 2022*, pages 1070–1079, 2022.

[Xie *et al.*, 2016] Junyuan Xie, Ross Girshick, and Ali Farhadi. Unsupervised deep embedding for clustering analysis. In *International Conference on Machine Learning*, pages 478–487. PMLR, 2016.

[Xie *et al.*, 2025] Kun Xie, Renchi Yang, and Sibo Wang. Diffusion-based graph-agnostic clustering. In *Proceedings of the ACM on Web Conference 2025, WWW 2025, Sydney, NSW, Australia, 28 April 2025- 2 May 2025*, pages 1353–1364. ACM, 2025.

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

[Xu *et al.*, 2021] Dongkuan Xu, Wei Cheng, Dongsheng Luo, Haifeng Chen, and Xiang Zhang. Infogcl: Information-aware graph contrastive learning. *Advances in Neural Information Processing Systems*, 34:30414–30425, 2021.

[Yang *et al.*, 2023] Xihong Yang, Yue Liu, Sihang Zhou, Siwei Wang, Wenxuan Tu, Qun Zheng, Xinwang Liu, Liming Fang, and En Zhu. Cluster-guided contrastive graph clustering network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 10834–10842, 2023.

[Zhang *et al.*, 2019] Xiaotong Zhang, Han Liu, Qimai Li, and Xiao-Ming Wu. Attributed graph clustering via adaptive graph convolution. *arXiv preprint arXiv:1906.01210*, 2019.

[Zhang *et al.*, 2022] Yixin Zhang, Yong Liu, Yonghui Xu, Hao Xiong, Chenyi Lei, Wei He, Lizhen Cui, and Chunyan Miao. Enhancing sequential recommendation with graph contrastive learning. *arXiv preprint arXiv:2205.14837*, 2022.

[Zhao *et al.*, 2025] Peiyao Zhao, Xin Li, Zeyu Zhang, Mingzhong Wang, Xueying Zhu, and Lejian Liao. Robust deep signed graph clustering via weak balance theory. In *Proceedings of the ACM on Web Conference 2025, WWW 2025, Sydney, NSW, Australia, 28 April 2025- 2 May 2025*, pages 3819–3830. ACM, 2025.

[Zheng *et al.*, 2025] Bowen Zheng, Junjie Zhang, Hongyu Lu, Yu Chen, Ming Chen, Wayne Xin Zhao, and Ji-Rong Wen. Enhancing graph contrastive learning with reliable and informative augmentation for recommendation. In *Proceedings of the 31st ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2101–2112. ACM, 2025.