Beyond Implicit Constraint: Explicit Low-Rank Structured Subspace Learning for Fast Attributed Graph Clustering

Yaoming Cai[1,2], **Song Liu**[1], **Zijia Zhang**[3,*], **You Wu**[1], **Xiaobo Liu**[4], **Yao Ding**[5], **Fei Li**[1]
[1]School of Information Engineering, Zhongnan University of Economics and Law, Wuhan, China
[2]Engineering Research Center of Natural Resource Information Management and Digital Twin Engineering Software (Ministry of Education); Hubei Key Laboratory of Intelligent Geo-Information Processing, China University of Geosciences, Wuhan, China
[3]School of Artificial Intelligence, Hubei University; Key Laboratory of Intelligent Sensing System and Security (Ministry of Education), Hubei University, Wuhan, China
[4]School of Automation, China University of Geosciences, Wuhan, China
[5]School of Optical Engineering, Xi'an Research Institute of Hi-Tech, Xi'an, China
caiyaom@zuel.edu.cn, 18571121467@163.com, zijiazhang@hubu.edu.cn, 2384084706@qq.com, xbliu@cug.edu.cn, dingyao.88@outlook.com, lifei@zuel.edu.cn

Abstract

Attributed graph clustering has achieved remarkable success by synergistically integrating topological structures and node attributes. While subspace learning has emerged as a dominant paradigm for node partitioning, most existing methods rely on *implicit* low-rank constraints, which often fail to capture complex nonlinear manifolds and suffer from prohibitive computational overhead on large-scale graphs. In this paper, we propose **ELSS** (**E**xplicit **L**ow-rank **S**tructured **S**ubspace learning), a scalable and robust framework that transcends implicit formulations. Specifically, ELSS learns an explicit and nonlinear low-rank subspace within a graph-structured embedding space, effectively uncovering latent cluster structures. To effectively mitigate the pervasive oversmoothing issue, we introduce a homophily-aware adaptive graph filter, which dynamically calibrates smoothing intensity to preserve discriminative ego-information. Furthermore, to ensure linear scalability, we develop a PageRank-guided structural sampling strategy for anchor-based approximation, which identifies pivotal landmarks based on their global topological prestige. Theoretical analysis guarantees that ELSS effectively mitigates spectral collapse while maintaining a linear complexity. Extensive experiments on diverse benchmarks demonstrate that ELSS consistently delivers superior clustering accuracy over state-of-the-art methods.

1 Introduction

Attributed Graph Clustering (AGC) is a fundamental task in unsupervised graph learning that aims to partition nodes into semantically consistent clusters by integrating two complementary sources of topological structure and node attributes [Li *et al.*, 2025]. This dual-source fusion is critical for addressing real-world problems such as community detection in social networks [Xian *et al.*, 2025], research field categorization in citation networks [Tang *et al.*, 2024], and functional module identification in biological networks [Su *et al.*, 2025]. With the exponential growth of graph data, e.g., social networks with billions of nodes and edges, AGC confronts substantial challenges in simultaneously achieving representational expressiveness, computational scalability, and clustering effectiveness.

The rapid development of deep clustering [Zhou *et al.*, 2024b] has witnessed increasing progress in AGC. Existing methods typically incorporate Graph Convolutional Networks (GCNs) [Wu *et al.*, 2019; Klicpera *et al.*, 2019] to learn graph embeddings, coupled with self-supervised clustering heads for label assignment. Despite their powerful representation learning capabilities, these methods heavily rely on pretext tasks, such as contrastive learning [Yang *et al.*, 2023; Liu *et al.*, 2024; Deng *et al.*, 2025] and attribute reconstruction [Peng *et al.*, 2021; Cheng *et al.*, 2021], which often require complex hyperparameter tuning and sensitive data augmentation. Furthermore, most deep AGC frameworks are computationally inefficient on full-scale graphs due to the overparameterized neural operations. While topology-guided subgraph sampling has been widely adopted to facilitate training [Xie *et al.*, 2025; Wu *et al.*, 2025], it frequently results in the loss of global structural dependencies, leading to suboptimal clustering boundaries.

Recently, Attributed Graph Subspace Clustering (AGSC) [Cai *et al.*, 2023] has gained significant attention by integrating the classic self-expression principle [Miao *et al.*, 2025] into the graph embedding space, achieving state-of-the-art performance. Mathematically, AGSC aims to learn a low-dimensional coefficient matrix that captures a refined latent structure, which typically exhibits superior discrimina-

*Corresponding author.

tive power compared to the raw graph topology. However, vanilla AGSC formulations suffer from quadratic space and cubic time complexity, severely limiting their applicability to large-scale graphs [Lin *et al.*, 2025]. To address this bottleneck, Fettal *et al.* [Fettal *et al.*, 2023] proposed the SAGSC model, which utilizes simplified graph convolution and orthogonal factorization to achieve linear computational complexity.

In this paper, we theoretically demonstrate that SAGSC is a specialized instance of explicit low-rank representation restricted to a linear kernel. This observation reveals several critical deficiencies in existing scalable models. First, these methods predominantly operate in linear feature spaces, hindering their capacity to uncover the complex nonlinear manifold structures ubiquitous in real-world graphs. Second, existing methods rely heavily on fixed-weight graph filters to aggregate long-range structural dependencies; as the filter depth increases, these mechanisms inevitably trigger the oversmoothing phenomenon, causing node representations to homogenize and lose discriminative ego-information. Finally, while anchor-based approximation facilitates scalability, it frequently resorts to uniform sampling that tends to overlook pivotal structural landmarks and topological hubs, leading to biased and suboptimal subspace approximations.

To address these challenges, we propose the Explicit Low-rank Structured Subspace Learning framework (termed ELSS), a scalable and robust solution for attributed graph partitioning. By extending explicit low-rank factorization to a graph-smoothed nonlinear kernel space, ELSS enables precise mapping of nodes onto a latent manifold where cluster structures are inherently more distinguishable. To alleviate the loss of discriminative node characteristics caused by excessive smoothing, we incorporate an adaptive filtering mechanism that adapts to local homophily, thus preserving critical discriminative features. For computational efficiency, we leverage the global topological prestige of nodes through a PageRank-based sampling strategy. This facilitates the selection of high-quality structural anchors, enabling linear-time optimization while retaining the graph's intrinsic structural properties.

2 Related Work

2.1 Notations

Throughout this paper, we use boldface lowercase italics (e.g., $\boldsymbol{x}$), boldface uppercase roman (e.g., $\mathbf{X}$), and regular italics (e.g., x_{ij}) to denote vectors, matrices, and scalars, respectively. Calligraphic symbols (e.g., $\mathcal{V}$) are used for sets. An attributed graph is defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A}, \mathbf{X})$, where $\mathcal{V}$ is the set of N nodes, $\mathcal{E}$ is the set of edges, $\mathbf{A} \in \{0,1\}^{N\times N}$ is the adjacency matrix, and $\mathbf{X} \in \mathbb{R}^{N\times D}$ represents the node attribute matrix. We define the augmented adjacency as $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$. The diagonal degree matrix is denoted by $\mathbf{D}$, where $D_{ii} = \sum_j \tilde{A}_{ij}$. The symmetric normalized graph Laplacian is defined as $\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1/2}\tilde{\mathbf{A}}\mathbf{D}^{-1/2}$.

2.2 Subspace Learning

Subspace representation learning operates on the premise that data points are sampled from a union of multiple low-dimensional linear subspaces [Miao *et al.*, 2025]. This paradigm is traditionally formulated through the self-expression property, where each data point is represented as a linear combination of other points in the same subspace:

$$\min_{\mathbf{Z}} \ \|\mathbf{X} - \mathbf{Z}\mathbf{X}\|_F^2 + R(\mathbf{Z}), \tag{1}$$

where $\mathbf{Z} \in \mathbb{R}^{N\times N}$ denotes the self-expressive coefficient matrix, which ideally exhibits a block-diagonal structure indicating the cluster assignments. The regularization term $R(\mathbf{Z})$ imposes specific structural constraints, such as sparsity via $\|\mathbf{Z}\|_1$ in Sparse Subspace Clustering (SSC) [Yu *et al.*, 2023], or low-rankness via the nuclear norm $\|\mathbf{Z}\|_*$ in Low-Rank Representation (LRR) [Ding *et al.*, 2025]. A classic formulation of LRR is given by:

$$\min_{\mathbf{Z}} \|\mathbf{Z}\|_* \quad \text{s.t.} \quad \mathbf{X} = \mathbf{Z}\mathbf{X}. \tag{2}$$

The clustering result is obtained by conducting spectral clustering on $\frac{1}{2}(|\mathbf{Z}| + |\mathbf{Z}^\top|)$.

Despite its theoretical elegance, traditional self-expression models face significant challenges in the context of large-scale attributed graphs. Specifically, the implicit nature of $\mathbf{Z}$ necessitates solving for an $N \times N$ matrix, leading to at least quadratic or even cubic computational complexity, which is prohibitive for massive graphs. Although anchor-based approximation methods offer a scalable alternative [Cai *et al.*, 2025], they predominantly rely on uniform sampling, which ignores the non-uniform importance of nodes. This leads to a fidelity gap where structural hubs are undersampled, resulting in inaccurate spectral approximations.

2.3 Attributed Graph Clustering

Given $\mathcal{G}$, AGC aims to partition nodes into k disjoint clusters $\{\mathcal{C}_1, \ldots, \mathcal{C}_k\}$ by unsupervisedly integrating topological structures and node attributes. This dual-source approach is indispensable for applications such as social network analysis, citation networks, and bioinformatics, where neglecting either source leads to suboptimal clustering performance. Consequently, research in this field predominantly focuses on graph embedding learning or cluster structure optimization. Existing methods can be categorized into two-stage and end-to-end approaches. Representative two-stage methods, such as SAGSC [Fettal *et al.*, 2023] and MCGC [Pan and Kang, 2021], first project the raw data into a latent embedding space via graph filtering, followed by a separate clustering algorithm. In contrast, end-to-end methods integrate differentiable clustering heads with established techniques like graph convolutions [Zhu *et al.*, 2024], attention mechanisms [Zhou *et al.*, 2024a], or contrastive learning [Liu *et al.*, 2024], enabling joint optimization of embeddings and cluster assignments.

Despite their success, the computational overhead of processing large-scale graphs remains a critical bottleneck. In particular, many deep subspace clustering methods that rely on the self-expression of node embeddings, such as MCGC [Pan and Kang, 2021], suffer from $\mathcal{O}(N^3)$ complexity due to the prohibitive eigenvalue decomposition of the refined affinity matrix. To alleviate this, recent research has shifted toward scalable AGC architectures. Within the deep learning

community, methods such as [Xie *et al.*, 2025; Wu *et al.*, 2025] employ topology-guided subgraph sampling to facilitate efficient mini-batch training. However, such localized sampling strategies are inherently suboptimal, as they often fragment the graph, thereby failing to capture long-range structural dependencies and global manifold structures. Parallel to these efforts, a class of lightweight attributed subspace clustering (AGSC) methods, exemplified by SAGSC [Fettal *et al.*, 2023] and S^2AGC [Lin *et al.*, 2025], has emerged. These methods prioritize the design of efficient yet effective linear-time operators, bypassing the heavy neural architectures or costly matrix decompositions of their predecessors.

3 Method

Inspired by the efficiency of SAGSC, we propose ELSS, a scalable framework that generalizes explicit low-rank structured representations into a nonlinear kernel space, while integrating an adaptive graph filter and structural sampling to ensure both clustering effectiveness and linear complexity.

3.1 Homophily-aware Adaptive Graph Filtering

To capture the structural dependencies inherent in the graph, we map the raw attribute matrix $\mathbf{X}$ into a smoothed embedding space $\mathbf{H}$ through non-parametric graph filtering. Conventional approaches typically employ a L-order linear propagation [Wu *et al.*, 2019], i.e., $\mathbf{H} = \mathbf{P}^L\mathbf{X}$, where $\mathbf{P} = \mathbf{I} - \mathbf{L}$ denotes the normalized propagation matrix. However, such a simplified mechanism is prone to oversmoothing, a phenomenon where node representations converge toward a common subspace dominated by the graph's stationary distribution, thereby depleting discriminative capacity as L increases.

To mitigate this, we adopt a homophily-aware adaptive graph filter scheme inspired by APPNP [Klicpera *et al.*, 2019]. The core is to a teleport probability $\alpha \in (0, 1]$ to balance local smoothing and feature preservation. However, we argue that a fixed global probability overlooks the structural heterophily across different nodes. To this end, we dynamically calibrate the filtering intensity based on local homophily. For each node v_i, we measure the cosine similarity between its attributes $\boldsymbol{x}_i$ and the aggregated features of its neighbors $\bar{\boldsymbol{x}}_i = \frac{1}{|\mathcal{N}_i|}\sum_{j\in\mathcal{N}_i}\boldsymbol{x}_j$. The node-wise coefficient α_i is computed as:

$$\alpha_i = \left(\frac{1}{2}\left(1 - \cos(\boldsymbol{x}_i, \bar{\boldsymbol{x}}_i)\right)\right)^{\gamma}, \tag{3}$$

The hyperparameter $\gamma > 0$ controls the nonlinearity of the homophily-to-teleportation mapping. We empirically set $\gamma = 2$, which provides a balanced quadratic scaling that appropriately amplifies homophily differences without over-penalizing moderate heterophily.Further, we define our filter as the following recursive form

$$\mathbf{H}^l = (\mathbf{I} - \boldsymbol{\alpha})\mathbf{P}\mathbf{H}^{l-1} + \boldsymbol{\alpha}\mathbf{X},\ \mathbf{H}^0 = \mathbf{X}, \tag{4}$$

where $\boldsymbol{\alpha} = \mathrm{diag}([\alpha_i])$ is a teleport probability matrix that preserves the fidelity of the original attributes for each node. The weighting scheme in Eq. (4) ensures that nodes in high-homophily regions ($\cos \approx 1$) receive stronger smoothing ($\alpha_i \to 0$), while nodes in heterophilous or outlier regions ($\cos \approx -1$) retain more ego-information ($\alpha_i \to 1$). Based on the Theorem 1, our adaptive graph filtering effectively mitigates oversmoothing in high-order propagation.

Theorem 1. *The iteration of Eq.* (4) *is a contraction mapping that converges to a unique fixed point* $\mathbf{H}^\infty = (\mathbf{I} - (\mathbf{I} - \boldsymbol{\alpha})\mathbf{P})^{-1}\boldsymbol{\alpha}\mathbf{X}$. *For any node pair* (v_i, v_j)*, the Euclidean distance satisfies:*

$$\|\boldsymbol{h}_i^\infty - \boldsymbol{h}_j^\infty\|_2 \geq \left|\|\alpha_i\boldsymbol{x}_i - \alpha_j\boldsymbol{x}_j\|_2 - \|(\Phi_i^\infty - \Phi_j^\infty)\|_2\right|,$$

where $\Phi^\infty = (\mathbf{I} - \boldsymbol{\alpha})\mathbf{P}\mathbf{H}^\infty$.

3.2 Low-Rank Property of ASGC

Here, we theoretically analyze the properties of existing SAGSC [Fettal *et al.*, 2023], a representative AGGC model. SAGSC reformulates the node partitioning problem through an explicit self-representation framework. Specifically, given the graph-smoothed features $\mathbf{H} \in \mathbb{R}^{N\times D}$, it seeks an optimal orthogonal basis $\mathbf{U} \in \mathbb{R}^{N\times r}$ ($r \ll N$) by minimizing the following reconstruction residual:

$$\min_{\mathbf{U}^\top\mathbf{U}=\mathbf{I}} \left\|\mathbf{H} - \mathbf{U}\mathbf{U}^\top\mathbf{H}\right\|_F^2, \tag{5}$$

where the self-expressive coefficient matrix is explicitly factored as $\mathbf{Z} = \mathbf{U}\mathbf{U}^\top$. By expanding Eq. (5), the optimization is equivalent to the principal component pursuit:

$$\max_{\mathbf{U}^\top\mathbf{U}=\mathbf{I}} \mathrm{tr}\left(\mathbf{U}^\top\mathbf{H}\mathbf{H}^\top\mathbf{U}\right), \tag{6}$$

where the optimal solution $\mathbf{U}^*$ consists of the top-r left singular vectors of $\mathbf{H}$. It should be noted that the procedure requires only linear complexity. Further analyzing this formulation, we establish the following two lemmas.

Lemma 1. *The reconstruction objective in Eq.* (5) *is equivalent to maximizing the spectral energy of a linear kernel* $\mathbf{K} = \mathbf{H}\mathbf{H}^\top$*, where* $K_{ij} = \langle\boldsymbol{h}_i, \boldsymbol{h}_j\rangle$*. Furthermore, the projection operator* $\mathbf{Z} = \mathbf{U}\mathbf{U}^\top$ *is the analytical optimal solution to the implicit LRR problem under the explicit hard-rank constraint rank*$(\mathbf{Z}) = r$.

Lemma 1 reveals the spectral duality between explicit factorization and implicit low-rank regularization. It demonstrates that while the explicit framework achieves linear scalability by bypassing the cubic complexity of nuclear norm minimization, its expressive capability remains fundamentally limited by the linearity of subspace.

3.3 Explicit Low-Rank Structured Representation

To overcome the inherent linearity and structural limitations of existing scalable AGSC models, we propose a generalized explicit low-rank representation framework. By mapping the filtered features $\mathbf{H}$ into a Reproducing Kernel Hilbert Space (RKHS) $\mathcal{F}$ via an implicit nonlinear mapping $\phi : \mathcal{H} \to \mathcal{F}$, the model is formulated to capture complex nonlinear manifold structures. The objective function is defined as:

$$\min_{\mathbf{U}^\top\mathbf{U}=\mathbf{I}} \left\|\phi\left(\mathbf{H}\right) - \mathbf{U}\mathbf{U}^\top\phi\left(\mathbf{H}\right)\right\|_F^2 + \lambda\mathrm{tr}\left(\mathbf{U}^\top\mathbf{L}\mathbf{U}\right), \tag{7}$$

where λ is a smoothing coefficient. In this formulation, $\mathbf{U}$ serves as the explicit orthogonal basis that spans the principal subspace in $\mathcal{F}$, effectively representing the data coordinates in a r-dimensional latent space. The second term,

$\mathrm{tr}(\mathbf{U}^\top \mathbf{L}\mathbf{U})$, acts as a manifold regularizer that enforces spectral smoothness, ensuring that the learned representation preserves the intrinsic structured relationships of the graph.

By expanding the Frobenius norm in Eq. (7) and invoking the kernel trick, the objective is reformulated into a nonlinear generalization of the problem in Eq. (6):

$$\max_{\mathbf{U}^\top\mathbf{U}=\mathbf{I}} \mathrm{tr}\left(\mathbf{U}^\top\left(\mathbf{K}-\lambda\mathbf{L}\right)\mathbf{U}\right), \tag{8}$$

where $\mathbf{K} = \phi(\mathbf{H})\phi(\mathbf{H})^\top$ denotes the kernel matrix (Gaussian kernel is used in the paper). Let $\mathbf{M} = \mathbf{K} - \lambda\mathbf{L}$ be the target operator. According to the Rayleigh-Ritz theorem, the optimal $\mathbf{U}^*$ is comprised of the r eigenvectors associated with the largest eigenvalues of $\mathbf{M}$.

While this formulation offers enhanced nonlinear expressivity, the direct eigendecomposition of the $N \times N$ matrix $\mathbf{M}$ incurs cubic complexity $\mathcal{O}(N^3)$, which undermines the linear scalability inherent in the original SAGSC. Crucially, if $\mathbf{M}$ can be factorized as $\mathbf{M} = \mathbf{B}\mathbf{B}^\top$ where $\mathbf{B} \in \mathbb{R}^{N\times m}$ ($m \ll N$), linear complexity can be recovered. To this end, we introduce a scalable and efficient optimization strategy in the following subsection.

3.4 Structural Anchor Sampling for Scalability

Theorem 2. *Let $\mathbf{M} = \mathbf{K} - \lambda\mathbf{L}$, where $\mathbf{K} \succeq 0$ and $\mathbf{L} \succeq 0$. For a Nyström approximation $\tilde{\mathbf{M}}_m$ with m sampled columns, if $\lambda \leq \frac{\sigma_{m+1}(\mathbf{K})}{\lambda_{\max}(\mathbf{L})}$, then $\|\mathbf{M} - \tilde{\mathbf{M}}_m\|_2 = O(\sigma_{m+1}(\mathbf{K}))$.*

Based on Theorem 2, the potential operator $\mathbf{M}$ can be approximated by using Nyström method. This allows us to solve Eq. (8) in linear scalability. Unlike standard Nyström approaches that rely on uniform random sampling, which may overlook critical topological hubs in sparse graphs, we propose a structure-aware sampling strategy to select informative anchors. Specifically, we first compute the PageRank centrality vector $\boldsymbol{\pi} \in \mathbb{R}^N$ by solving for the steady-state distribution of a random walk on the graph, formulated as:

$$\boldsymbol{\pi} = \beta\hat{\mathbf{P}}^\top\boldsymbol{\pi} + \frac{1-\beta}{N}\mathbf{1}, \tag{9}$$

where $\hat{\mathbf{P}}$ denotes the transition matrix and β is the damping factor. This process identifies landmark nodes that possess significant structural importance or reside in dense cluster centers. We then select a set of m anchors ($m \ll N$) based on the probability distribution defined by $\boldsymbol{\pi}$.

Let $\mathbf{C} \in \mathbb{R}^{N\times m}$ be the sub-matrix capturing the affinities between all nodes and the anchors, and $\mathbf{W} \in \mathbb{R}^{m\times m}$ be the inner-potential matrix among the anchors. The Nyström approximation is then formulated as:

$$\hat{\mathbf{M}} = \mathbf{C}\mathbf{W}^\dagger\mathbf{C}^\top = (\mathbf{C}\mathbf{V}\boldsymbol{\Sigma}^{-1/2})(\mathbf{C}\mathbf{V}\boldsymbol{\Sigma}^{-1/2})^\top = \mathbf{B}\mathbf{B}^\top, \tag{10}$$

where $\mathbf{W}^\dagger = \mathbf{V}\boldsymbol{\Sigma}^{-1}\mathbf{V}^\top$ denotes the Moore-Penrose pseudoinverse derived from the eigendecomposition of $\mathbf{W}$. Note that $\mathbf{B} \in \mathbb{R}^{N\times m}$ is a tall-and-skinny matrix. Consequently, the optimal embedding $\mathbf{U}^*$ can be efficiently obtained by extracting the top-r left singular vectors of $\mathbf{B}$, i.e., $\mathbf{U}^* = \mathrm{SVD}(\mathbf{B}, r)$, maintaining linear computational complexity with respect to N.

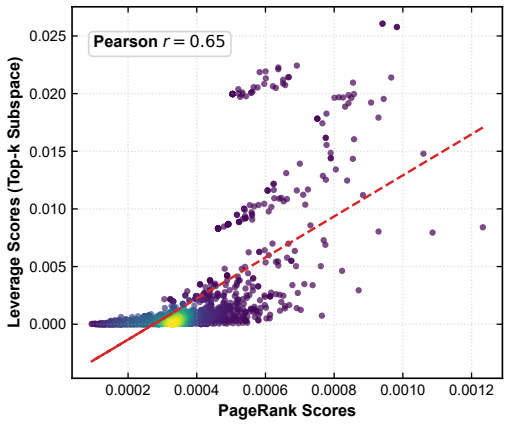


Figure 1: Strong positive correlation ($r = 0.65$) between PageRank scores and the ground-truth leverage scores of the graph embedding for the ACM dataset.

Theorem 3. *Let $\ell_i = \|\mathbf{U}_{i,:}\|^2$ be the statistical leverage scores of $\mathbf{M}$. The PageRank vector $\boldsymbol{\pi}$ is an asymptotically optimal sampling distribution for Nyström approximation such that $p_i \propto \pi_i$ minimizes the approximation error $\|\mathbf{M} - \hat{\mathbf{M}}\|_2$ by satisfying the structural coherence constraint: $\pi_i \geq C \cdot \ell_i \cdot \delta(\lambda_{min})$, where $\delta(\cdot)$ is a spectral filter that biases sampling toward the principal manifold.*

From Theorem 3, we observe that by integrating PageRank centrality into the Nyström sampling framework, the approximation becomes inherently structure-aware. This strategy effectively biases the low-rank reconstruction toward the principal manifold, where the latent clustering signals are most concentrated. As evidenced in Figure 1, the PageRank scores of an attributed graph exhibit a strong empirical correlation with the matrix leverage scores.

Algorithm 1 Explicit Low-rank Structured Subspace clustering (ELSS)

Input: Attributed graph $\mathcal{G} = (\mathbf{A}, \mathbf{X})$, number of clusters k, propagation order L, number of anchors m, regularization coefficient λ.
Output: Clustering assignments $\{\mathcal{C}_1, \ldots, \mathcal{C}_k\}$.
Step 1: Homophily-aware Adaptive Graph Filtering
Compute adaptive teleport coefficients α_i via Eq. (3);
Compute graph embedding $\mathbf{H}$ using Eq. (4);
Step 2: Structural Anchor Sampling
Compute PageRank scores $\boldsymbol{\pi}$ via Eq. (9);
Sample m anchor according to distribution $p_i \propto \pi_i$;
Step 3: Explicit Low-rank Structured Approximation
Compute eigendecomposition: $\mathbf{W} = \mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^\top$;
Form low-rank basis: $\mathbf{B} = \mathbf{C}\mathbf{V}\boldsymbol{\Sigma}^{-1/2}$;
Extract top-k left singular vectors of $\mathbf{B}$ as $\mathbf{U}^*$;
Step 4: Postprocessing
Apply non-negative mapping on $\mathbf{U}$ and perform truncated SVD;
Conduct K-Means to obtain cluster assignment;

3.5 Postprocessing and Analysis

We adopt a postprocessing strategy similar to the SAGSC model [Fettal *et al.*, 2023] to conduct implicit spectral segmentation. The complete learning procedure of our ELSS is summarized in Algorithm 1. The primary computational bottlenecks stem from the construction of the kernelized anchor graph ($\mathcal{O}(NDm)$) and the eigendecomposition of the $m \times m$ affinity matrix ($\mathcal{O}(m^3)$). Since the number of anchors m is a fixed hyperparameter independent of the graph size, the cubic term $\mathcal{O}(m^3)$ is negligible for large N. Due to the sparseness of adjacent matrix, the graph filtering requires $\mathcal{O}(L|\mathcal{E}|D)$. Consequently, the overall time complexity is $\mathcal{O}(NDm + L|\mathcal{E}|D + m^3)$, and the space complexity is $\mathcal{O}(N(D + m) + |\mathcal{E}|)$. This analysis confirms that ELSS achieves linear scalability with respect to the number of nodes and edges.

Dataset	#Vertices	#Edges	#Attributes	#Clusters
ACM	3025	16153	1870	3
Wiki	2405	14001	4973	17
DBLP	4057	2502276	334	4
Computers	13381	259159	767	10
PubMed	19717	64041	500	3
OGBN-ArXiv	169343	1327142	128	40

Table 1: Dataset statistics.

4 Experiments

In this section, we evaluate the performance of the proposed ELSS on six benchmark datasets. Due to space limitations, we defer additional experimental results to the Appendix B.

4.1 Datasets and Metrics

To evaluate the performance of the proposed models, we conduct experiments on six widely used benchmark datasets. These include four citation networks: ACM, DBLP [Wang *et al.*, 2019], PubMed [Sen *et al.*, 2008], and Wiki [Yang *et al.*, 2015], as well as the Amazon Computers sales dataset [Shchur *et al.*, 2018] and the large-scale OGBN-ArXiv dataset [Hu *et al.*, 2020]. Detailed summary statistics for these datasets are provided in Table 1. For performance assessment, we adopt three standard clustering evaluation metrics [Cai *et al.*, 2021]: Clustering Accuracy (ACC), Normalized Mutual Information (NMI), and the Adjusted Rand Index (ARI).

4.2 Baseline Models

To validate the superior clustering performance of our method, we selected representative baseline methods, ranging from classical clustering algorithms to state-of-the-art deep graph clustering models. For consistent comparison, our evaluation includes classical methods (K-Means and Spectral Clustering (SC) and subspace learning approaches (LSR [Lu *et al.*, 2012], EnSC [You *et al.*, 2016a], SSC-OMP [You *et al.*, 2016b], and k-FSC [Fan, 2021]). We also compare with popular deep graph clustering methods, including SGC [Wu *et al.*, 2019], GIC [Mavromatis and Karypis, 2021], S^2GC [Zhu and Koniusz, 2021], and GCC [Fettal *et al.*, 2022]. Furthermore, we include recent state-of-the-art frameworks: CCGC [Yang *et al.*, 2023], HSAN [Liu *et al.*, 2023], MAGI [Liu *et al.*, 2024], DGCluster [Bhowmick *et al.*, 2024], S^2CAG [Lin *et al.*, 2025], FPGC [Xie *et al.*, 2025], and SAGSC [Fettal *et al.*, 2023].

4.3 Implementation Details

We conduct all experiments on a high-performance Linux workstation (Ubuntu) equipped with an Intel Core i9-14900K CPU and a NVIDIA GeForce RTX 5090 GPU with 32GB of GDDR7 memory. The system is supported by 128GB of RAM, and the deep learning environment is implemented using PyTorch with CUDA 12.4. To ensure statistical reliability, all experiments are repeated 5 times with different random seeds, and the mean values with standard deviations are reported. For consistent comparison, we utilize the official source codes provided by the respective authors for most baseline methods. We strictly adhere to a standardized experimental protocol. Furthermore, we set the maximum number of training epochs to 400 for all datasets. To ensure computational efficiency, we impose a runtime limit of 1 hour for standard datasets and 6 hours for large-scale graphs. Empirically, we configure the key parameters of our method as follows. We set $m \in [100, 300]$ for medium graphs and $m \in [500, 1000]$ for large graphs. We also adopt $L \in [10, 30]$, $r \approx k$, and $\lambda \in [10^{-5}, 10^{-4}]$.

4.4 Performance Evaluation

This set of experiments reports the ACC, NMI, and ARI scores achieved by our proposed ELSS and all representative baselines across six datasets ranging from small-scale to large-scale networks.

The quantitative results are summarized in Table 2 (for ACM, DBLP, Wiki) and Table 3 (for Amazon Computers, PubMed, OGBN-ArXiv). The best and second-best results are highlighted in blue, with darker shades indicating superior performance. The underlined entries denote the third-best results. From the reported statistics, we can observe that ELSS consistently achieves the highest or competitive ranking in terms of overall clustering quality among all evaluated methods. This observation validates the effectiveness of our proposed subspace learning framework for attributed graph clustering tasks.

Specifically, on small and medium-sized attributed graphs (Table 2), ELSS outperforms or matches the state-of-the-art baselines. For instance, on the ACM and DBLP datasets, ELSS achieves comparable high performance to the strong competitor S^2CAG, while significantly surpassing classical methods like K-Means and SSC-OMP. It is worth noting that while S^2CAG shows a slight advantage on the small Wiki dataset, ELSS demonstrates better robustness and generalization capabilities as the graph scale increases.

The superiority of ELSS is particularly pronounced on large-scale datasets with high feature dimensions and complex structures, as shown in Table 3. On Amazon Computers, ELSS improves upon the best baseline (S^2CAG) by a margin of 2.7% in ACC and 3.4% in NMI. Most importantly, on the massive OGBN-ArXiv dataset (*approx.* $170k$ nodes

Method	Input	ACM			DBLP			Wiki		
		ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
Kmeans	$\mathbf{X}$	87.8±0.9	61.7±1.5	67.4±2.1	67.9±0.0	37.3±0.0	31.5±0.1	47.6±1.4	48.6±0.2	26.6±0.2
LSR	$\mathbf{X}$	78.6±0.0	43.1±0.0	48.3±0.0	69.4±0.1	34.7±0.1	36.4±0.2	17.8±0.5	2.8±1.7	0.3±0.2
EnSC	$\mathbf{X}$	83.8±0.0	53.0±0.0	58.6±0.0	30.0±0.1	0.8±0.2	0.1±0.0	47.5±0.0	45.2±0.2	30.2±0.1
SSC-OMP	$\mathbf{X}$	82.1±0.0	49.4±0.1	55.3±0.0	29.4±0.1	0.4±0.1	-0.1±0.0	37.8±8.5	34.4±9.1	21.2±7.9
k-FSC	$\mathbf{X}$	59.7±7.2	25.2±7.1	27.2±7.2	51.3±11.1	17.4±7.3	17.3±9.6	38.2±5.1	35.6±3.9	17.7±4.4
SC	$\mathbf{A}$	36.5±0.2	1.0±0.2	0.7±0.1	91.0±0.0	73.0±0.1	78.3±0.1	30.7±1.1	24.0±0.8	6.0±0.1
SGC	$\mathbf{A}, \mathbf{X}$	83.7±0.0	55.7±0.0	58.8±0.0	88.8±0.0	69.5±0.0	73.2±0.0	51.9±0.8	49.6±0.2	28.6±0.1
GIC	$\mathbf{A}, \mathbf{X}$	90.1±0.3	68.2±0.6	73.2±0.6	90.2±0.2	72.4±0.4	77.4±0.3	48.0±0.7	48.4±0.3	31.0±0.3
S^2GC	$\mathbf{A}, \mathbf{X}$	84.1±0.1	56.8±0.1	59.6±0.2	88.3±0.0	69.2±0.0	71.9±0.0	52.1±1.0	52.2±0.1	33.0±0.4
GCC	$\mathbf{A}, \mathbf{X}$	91.3±0.0	71.2±0.1	76.0±0.1	91.8±0.0	74.5±0.0	80.5±0.0	53.7±1.4	53.5±0.5	31.6±1.1
CCGC	$\mathbf{A}, \mathbf{X}$	87.5±1.3	60.8±2.6	66.5±3.2	87.9±0.2	68.8±0.3	73.0±0.3	50.7±1.3	47.3±0.5	26.0±0.7
HSAN	$\mathbf{A}, \mathbf{X}$	90.1±0.0	67.3±0.0	72.9±0.0	87.6±0.0	68.5±0.0	72.1±0.0	54.4±0.0	49.1±0.0	33.6±0.0
MAGI	$\mathbf{A}, \mathbf{X}$	91.3±1.4	70.5±3.1	75.9±3.7	91.1±0.1	74.0±0.1	78.9±0.2	54.9±4.7	51.9±0.7	37.1±4.9
DGCluster	$\mathbf{A}, \mathbf{X}$	90.4±0.0	68.7±0.0	73.6±0.0	92.1±0.1	75.2±0.0	80.9±0.0	56.4±0.0	50.2±0.0	40.6±0.0
S^2CAG	$\mathbf{A}, \mathbf{X}$	93.5±0.0	75.4±0.0	81.4±0.0	93.5±0.0	78.8±0.0	84.3±0.0	64.4±0.0	55.1±0.0	44.9±0.0
FPGC	$\mathbf{A}, \mathbf{X}$	68.4±0.1	37.2±0.0	37.0±0.0	92.1±0.1	76.2±0.0	80.9±0.0	56.0±2.1	40.5±0.0	20.8±0.0
SAGSC	$\mathbf{A}, \mathbf{X}$	93.2±0.1	75.0±0.2	80.8±0.2	93.0±0.1	77.8±0.2	83.1±0.2	56.3±1.7	52.7±0.5	34.4±1.1
ELSS	$\mathbf{A}, \mathbf{X}$	93.5±0.0	75.7±0.0	81.5±0.0	93.6±0.0	79.2±0.0	84.4±0.0	61.4±0.5	54.1±0.3	35.8±1.1

Table 2: Clustering performance of the different models over ACM, DBLP, and Wiki.

Method	Input	Amazon Computers			PubMed			OGBN-ArXiv		
		ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
Kmeans	$\mathbf{X}$	29.8±0.3	29.5±0.5	12.1±0.1	59.4±0.0	30.1±0.0	26.7±0.0	17.7±0.1	22.7±0.2	0.1±0.0
SGC	$\mathbf{A}, \mathbf{X}$	65.5±0.0	52.2±0.0	45.7±0.0	69.6±0.0	29.3±0.0	29.9±0.0	34.6±0.4	39.2±0.1	25.2±0.6
GIC	$\mathbf{A}, \mathbf{X}$	46.8±2.2	47.5±0.9	31.3±3.5	64.5±0.4	26.2±0.3	23.8±0.4	16.0±0.8	17.9±0.5	5.8±0.2
S^2GC	$\mathbf{A}, \mathbf{X}$	65.4±0.0	55.4±0.0	49.5±0.0	71.0±0.0	32.9±0.0	33.7±0.0	41.9±0.3	45.9±0.1	36.9±0.5
GCC	$\mathbf{A}, \mathbf{X}$	67.6±0.0	56.0±0.0	46.5±0.0	70.5±0.0	32.2±0.0	33.1±0.0	40.5±1.7	46.8±0.2	35.1±2.0
CCGC	$\mathbf{A}, \mathbf{X}$	58.9±0.4	57.0±0.3	42.4±0.6	63.7±1.9	29.3±1.1	27.1±0.8	OOM	OOM	OOM
HSAN	$\mathbf{A}, \mathbf{X}$	49.1±0.7	46.6±0.8	29.8±0.8	61.6±1.4	27.8±3.3	25.6±1.9	OOM	OOM	OOM
MAGI	$\mathbf{A}, \mathbf{X}$	62.0±0.2	59.2±0.2	46.2±0.3	60.6±0.1	18.6±0.1	17.3±0.1	38.8±0.1	46.9±0.1	31.0±0.1
DGCluster	$\mathbf{A}, \mathbf{X}$	51.2±0.0	54.5±0.0	51.2±0.0	41.4±0.0	34.7±0.0	24.4±0.0	33.9±0.0	43.2±0.0	29.2±0.0
S^2CAG	$\mathbf{A}, \mathbf{X}$	73.6±0.0	59.8±0.0	51.1±0.0	75.3±0.0	36.5±0.0	41.9±0.0	46.9±0.0	46.1±0.0	38.7±0.0
FPGC	$\mathbf{A}, \mathbf{X}$	55.0±2.9	52.3±1.3	38.0±4.1	70.2±0.9	33.3±2.6	24.4±2.2	18.3±0.3	28.1±1.7	10.3±1.1
SAGSC	$\mathbf{A}, \mathbf{X}$	68.9±2.5	58.1±1.8	46.3±4.3	71.1±0.0	32.8±0.0	34.0±0.0	47.3±2.0	46.7±0.5	38.8±1.1
ELSS	$\mathbf{A}, \mathbf{X}$	76.3±1.8	63.2±1.2	55.3±0.7	75.9±0.0	37.3±0.0	41.9±0.0	52.5±0.3	48.2±0.8	41.5±0.4

Table 3: Clustering performance of the SOTA models over the larger graphs: Amazon Computers, Pubmed, and OGBN-ArXiv. OOM: Out-of-memory.

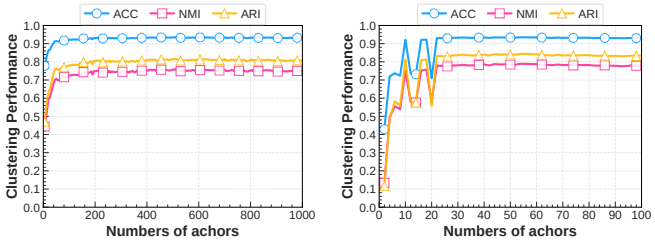
and 1.3M edges), where traditional deep clustering methods like CCGC and HSAN fail due to memory constraints, ELSS not only runs efficiently but also delivers a substantial performance gain. Compared to the second-best SAGSC, ELSS achieves a remarkable improvement of 4.7%, 1.1%, and 3.1% in ACC, NMI, and ARI, respectively. This confirms that ELSS effectively mitigates the scalability bottleneck while maintaining superior clustering accuracy on large-scale graphs.

4.5 Parameter Sensitivity

To evaluate model robustness and optimal configurations, we conduct sensitivity analysis on four key parameters (the number of anchor points (m), the feature propagation order (L), the dimension of the low-rank subspace (r), and the Laplacian regularization coefficient (λ)) across four datasets. Results for two datasets are deferred to Appendix B.2. In the following experiments, we vary one parameter at a time while fixing the others at default settings, with results visualized in Figures 2-5.

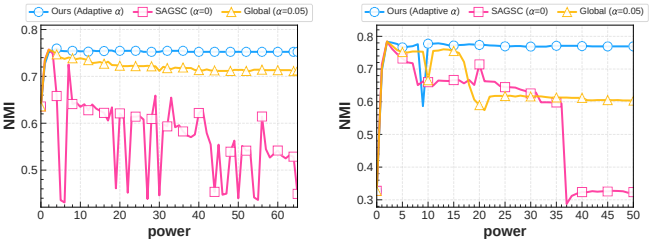
Impact of Anchor Number (m)

Figure 2 illustrates the performance trend as the anchor size varies. We observe a "saturation" pattern across all datasets: performance improves sharply as m increases from a small value, as an insufficient number of anchors fails to span the diverse data manifold. However, performance stabilizes after reaching a critical point (e.g., $m = 100$ for ACM). This observation aligns with our theoretical analysis in Theorem 3, indicating that selecting anchors proportional to PageRank scores allows ELSS to approximate the global graph structure with a small, representative subset, thereby avoiding unnecessary computational overhead.



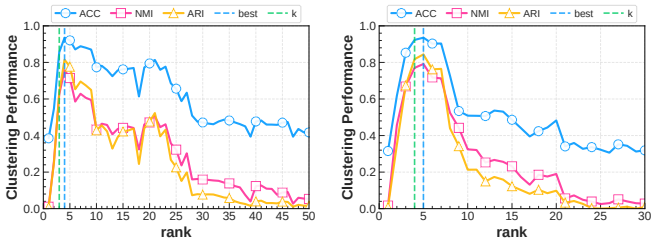
(a) ACM (b) DBLP

Figure 2: The impact of the number of anchor points on clustering results.



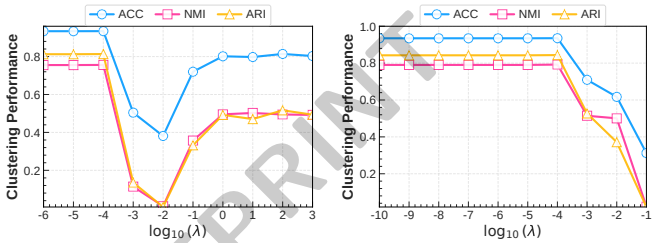
(a) ACM (b) DBLP

Figure 3: The impact of propagation order L on clustering performance (NMI).



(a) ACM (b) DBLP

Figure 4: The impact of Low-rank representation subspace dimension rank on clustering results.



(a) ACM (b) DBLP

Figure 5: The impact of Laplacian regularization constraint parameter λ on clustering results performance.

Sensitivity Analysis on Propagation Depth (L)

To validate our homophily-aware adaptive graph filtering's superiority in capturing long-range dependencies while mitigating oversmoothing, we evaluated clustering performance on two benchmarks with varying propagation orders L (Figure 3). Simple propagation without residuals shows severe degradation as L increases (especially on ACM), confirming oversmoothing-induced discriminative loss. Global filtering with fixed teleport improves stability but fails to adapt to structural heterogeneity, leading to suboptimal results. In contrast, our adaptive method achieves the highest accuracy and robust performance even for deep L, verifying that local homophily-based dynamic calibration effectively balances neighborhood aggregation and ego-feature preservation, alleviating noise accumulation and feature collapse in deep graph learning.

Impact of The Subspace Rank (r)

Figure 4 depicts the sensitivity to the subspace rank. A distinctive feature in these plots is the green dashed line, which represents the ground-truth number of clusters (k). We observe a clear "rise-and-fall" pattern. When $r < k$, the subspace is too compressed to encode the discriminative information for all clusters (underfitting). Interestingly, the optimal r is consistently located slightly to the right of the green line (i.e., $r \geq k$). For example, on Wiki, the optimal $r = 17$ is close to the cluster number $k = 17$. However, as r continues to grow significantly beyond k, performance degrades due to the introduction of noise and redundant information (overfitting). This confirms that a subspace dimension slightly larger than k is the ideal choice.

Impact of Graph Regularization (λ)

Finally, we analyze the effect of the regularization coefficient λ in Figure 5. The model demonstrates high robustness to variations in λ when it is small (e.g., $\lambda < 10^{-4}$), maintaining stable high performance. However, when λ exceeds a certain threshold (typically $> 10^{-4}$), the clustering metrics drop precipitously. This observation is consistent with our theoretical analysis in Theorem 2. This can be attributed to the fact that excessive regularization coefficient over-constrains the embedding space, forcing node representations to become overly smooth and losing the local discriminative variations necessary for accurate clustering. Therefore, a moderate λ (around 10^{-4}) is recommended to enforce manifold smoothness without compromising feature expressiveness.

5 Conclusion

This work presents the ELSS framework to address the fundamental challenge of balancing representational expressiveness, computational scalability, and robustness against oversmoothing in attributed graph clustering. By introducing nonlinear explicit low-rank modeling, homophily-adaptive graph smoothing, and structure-aware anchor sampling, ELSS effectively addresses key limitations of existing AGSC methods while providing theoretical guarantees for spectral stability and linear complexity. Extensive empirical results across diverse benchmarks validate its superior clustering performance and practical utility for large-scale data, thereby establishing a robust foundation for future research in unsupervised graph representation learning.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 62401632), the Hubei Provincial Natural Science Foundation (Nos. 2024AFB484 and 2026AFB758), the Natural Science Foundation of Wuhan (No. 2026040301020058), the China Postdoctoral Science Foundation (No. 2024M753661), the Open Project of the Key Laboratory of Intelligent Sensing System and Security, Hubei University (No. KLISSS202407), the Hubei Provincial Postdoctoral Project (No. 2024HBBHCXA100), the Open Fund of the Research Platforms, School of Computer Science, China University of Geosciences, Wuhan (No. PTLH2024-B-06), the Key Project of Hubei Provincial Department of Education, and the Fundamental Research Funds for the Central Universities of Zhongnan University of Economics and Law (No. 2722025BJ037).

References

[Bhowmick *et al.*, 2024] Aritra Bhowmick, Mert Kosan, Zexi Huang, Ambuj Singh, and Sourav Medya. Dgcluster: A neural framework for attributed graph clustering via modularity maximization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 11069–11077, 2024.

[Cai *et al.*, 2021] Yaoming Cai, Zijia Zhang, Zhihua Cai, Xiaobo Liu, Xinwei Jiang, and Qin Yan. Graph convolutional subspace clustering: A robust subspace clustering framework for hyperspectral image. *IEEE Transactions on Geoscience and Remote Sensing*, 59(5):4191–4202, 2021.

[Cai *et al.*, 2023] Yaoming Cai, Zijia Zhang, Pedram Ghamisi, Zhihua Cai, Xiaobo Liu, and Yao Ding. Fully linear graph convolutional networks for semi-supervised and unsupervised classification. *ACM Trans. Intell. Syst. Technol.*, 14(3), April 2023.

[Cai *et al.*, 2025] Yaoming Cai, Zijia Zhang, Xiaobo Liu, Yao Ding, Fei Li, and Jinhua Tan. Learning unified anchor graph for joint clustering of hyperspectral and lidar data. *IEEE Transactions on Neural Networks and Learning Systems*, 36(4):6341–6354, 2025.

[Cheng *et al.*, 2021] Jiafeng Cheng, Qianqian Wang, Zhiqiang Tao, Deyan Xie, and Quanxue Gao. Multi-view attribute graph convolution networks for clustering. In *Proceedings of the twenty-ninth international conference on international joint conferences on artificial intelligence*, pages 2973–2979, 2021.

[Deng *et al.*, 2025] Bowen Deng, Tong Wang, Lele Fu, Sheng Huang, Chuan Chen, and Tao Zhang. Thesaurus: contrastive graph clustering by swapping fused gromov-wasserstein couplings. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 16199–16207, 2025.

[Ding *et al.*, 2025] Meng Ding, Jing-Hua Yang, Xi-Le Zhao, Jie Zhang, and Michael K Ng. Nonconvex low-rank tensor representation for multi-view subspace clustering with insufficient observed samples. *IEEE Transactions on Knowledge and Data Engineering*, 2025.

[Fan, 2021] Jicong Fan. Large-scale subspace clustering via k-factorization. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 342–352, 2021.

[Fettal *et al.*, 2022] Chakib Fettal, Lazhar Labiod, and Mohamed Nadif. Efficient graph convolution for joint node representation learning and clustering. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, WSDM '22, pages 289–297, New York, NY, USA, 2022. Association for Computing Machinery.

[Fettal *et al.*, 2023] Chakib Fettal, Lazhar Labiod, and Mohamed Nadif. Scalable attributed-graph subspace clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 7559–7567, 2023.

[Hu *et al.*, 2020] Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. Open graph benchmark: Datasets for machine learning on graphs. *Advances in neural information processing systems*, 33:22118–22133, 2020.

[Klicpera *et al.*, 2019] Johannes Klicpera, Aleksandar Bojchevski, and Stephan Günnemann. Predict then propagate: Graph neural networks meet personalized pagerank. In *International Conference on Learning Representations*, 2019.

[Li *et al.*, 2025] Mengyao Li, Zhibang Yang, Xu Zhou, Yixiang Fang, Kenli Li, and Keqin Li. Clustering on attributed graphs: From single-view to multi-view. *ACM Computing Surveys*, 57(7):1–36, 2025.

[Lin *et al.*, 2025] Xiaoyang Lin, Renchi Yang, Haoran Zheng, and Xiangyu Ke. Spectral subspace clustering for attributed graphs. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1*, KDD '25, page 789–799, New York, NY, USA, 2025. Association for Computing Machinery.

[Liu *et al.*, 2023] Yue Liu, Xihong Yang, Sihang Zhou, Xinwang Liu, Zhen Wang, Ke Liang, Wenxuan Tu, Liang Li, Jingcan Duan, and Cancan Chen. Hard sample aware network for contrastive deep graph clustering. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(7):8914–8922, Jun. 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, Weiqiang Wang, and Liang Chen. Revisiting modularity maximization for graph clustering: A contrastive learning perspective. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '24, page 1968–1979, New York, NY, USA, 2024. Association for Computing Machinery.

[Lu *et al.*, 2012] Can-Yi Lu, Hai Min, Zhong-Qiu Zhao, Lin Zhu, De-Shuang Huang, and Shuicheng Yan. Robust and efficient subspace segmentation via least squares regression. In *European conference on computer vision*, pages 347–360. Springer, 2012.

[Mavromatis and Karypis, 2021] Costas Mavromatis and George Karypis. Graph infoclust: Maximizing coarse-grain mutual information in graphs. In *Advances in Knowledge Discovery and Data Mining*, pages 541–553. Springer International Publishing, 2021.

[Miao *et al.*, 2025] Jianyu Miao, Xiaochan Zhang, Tiejun Yang, Chao Fan, Yingjie Tian, Yong Shi, and Mingliang Xu. A comprehensive survey on subspace clustering: Methods and applications. *Artificial Intelligence Review*, 58(11):346, 2025.

[Pan and Kang, 2021] Erlin Pan and Zhao Kang. Multi-view contrastive graph clustering. *Advances in neural information processing systems*, 34:2148–2159, 2021.

[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.

[Sen *et al.*, 2008] Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher, and Tina Eliassi-Rad. Collective classification in network data. *AI Magazine*, 29(3):93, Sep. 2008.

[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.

[Su *et al.*, 2025] Xiaorui Su, Pengwei Hu, Dongxu Li, Bowei Zhao, Zhaomeng Niu, Thomas Herget, Philip S Yu, and Lun Hu. Interpretable identification of cancer genes across biological networks via transformer-powered graph representation learning. *Nature biomedical engineering*, 9(3):371–389, 2025.

[Tang *et al.*, 2024] Jiabin Tang, Yuhao Yang, Wei Wei, Lei Shi, Lixin Su, Suqi Cheng, Dawei Yin, and Chao Huang. Graphgpt: Graph instruction tuning for large language models. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 491–500, 2024.

[Wang *et al.*, 2019] 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.

[Wu *et al.*, 2019] Felix Wu, Amauri Souza, Tianyi Zhang, Christopher Fifty, Tao Yu, and Kilian Weinberger. Simplifying graph convolutional networks. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pages 6861–6871, 2019.

[Wu *et al.*, 2025] Benyu Wu, Yue Liu, Qiaoyu Tan, Xinwang Liu, Wei Du, Jun Wang, and Guoxian Yu. Dgcbench: A deep graph clustering benchmark. *Proc. NeurIPS*, pages 1–11, 2025.

[Xian *et al.*, 2025] Yantuan Xian, Pu Li, Hao Peng, Zhengtao Yu, Yan Xiang, and Philip S Yu. Community detection in large-scale complex networks via structural entropy game. In *Proceedings of the ACM on Web Conference 2025*, pages 3930–3941, 2025.

[Xie *et al.*, 2025] Xuanting Xie, Bingheng Li, Erlin Pan, Zhaochen Guo, Zhao Kang, and Wenyu Chen. One node one model: Featuring the missing-half for graph clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 21688–21696, 2025.

[Yang *et al.*, 2015] Cheng Yang, Zhiyuan Liu, Deli Zhao, Maosong Sun, and Edward Y Chang. Network representation learning with rich text information. In *IJCAI*, volume 2015, pages 2111–2117, 2015.

[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.

[You *et al.*, 2016a] Chong You, Chun-Guang Li, Daniel P Robinson, and René Vidal. Oracle based active set algorithm for scalable elastic net subspace clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3928–3937, 2016.

[You *et al.*, 2016b] Chong You, Daniel Robinson, and René Vidal. Scalable sparse subspace clustering by orthogonal matching pursuit. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3918–3927, 2016.

[Yu *et al.*, 2023] Shengju Yu, Suyuan Liu, Siwei Wang, Chang Tang, Zhigang Luo, Xinwang Liu, and En Zhu. Sparse low-rank multi-view subspace clustering with consensus anchors and unified bipartite graph. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[Zhou *et al.*, 2024a] Haicang Zhou, Tiantian He, Yew-Soon Ong, Gao Cong, and Quan Chen. Differentiable clustering for graph attention. *IEEE Transactions on Knowledge and Data Engineering*, 36(8):3751–3764, 2024.

[Zhou *et al.*, 2024b] Sheng Zhou, Hongjia Xu, Zhuonan Zheng, Jiawei Chen, Zhao Li, Jiajun Bu, Jia Wu, Xin Wang, Wenwu Zhu, and Martin Ester. A comprehensive survey on deep clustering: Taxonomy, challenges, and future directions. *ACM Computing Surveys*, 57(3):1–38, 2024.

[Zhu and Koniusz, 2021] Hao Zhu and Piotr Koniusz. Simple spectral graph convolution. In *International Conference on Learning Representations*, 2021.

[Zhu *et al.*, 2024] Pengfei Zhu, Qian Wang, Yu Wang, Jialu Li, and Qinghua Hu. Every node is different: Dynamically fusing self-supervised tasks for attributed graph clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 17184–17192, 2024.