Neurocomputing 511 (2022) 426–436

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Low-rank constraint bipartite graph learning

Qian Zhou, Haizhou Yang, Quanxue Gao *

State Key laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, Shaanxi, China

ARTICLE INFO

Article history:
Received 14 March 2022
Revised 12 June 2022
Accepted 4 September 2022
Available online 13 September 2022
Communicated by Zidong Wang

Keyword:
Multi-view clustering
Graph clustering
Unsupervised learning

ABSTRACT

Bipartite Graph-based multi-view clustering has become an active topic recently due to its efficiency in tackling large scale multi-view data. However, most existing bipartite graph-based multi-view clustering methods have the following disadvantages: 1) the clustering performance heavily depends on the predefined bipartite graph; 2) they fail to explore the complementary information embedded in multiple bipartite graphs and spatial low-rank structure hidden in each bipartite graph. To address these issues, we propose an efficient multi-view clustering method, **L**earning **C**onstraint **B**ipartite **G**raphs for Multi-view Clustering (LCBG). LCBG adaptively learns each graph $\mathbf{S}^{(v)} \in \mathbb{R}^{N\times M}$ such that it well characterizes the relationship between M ($M \ll N$) anchors and N samples in the v-th view, and takes both the intra-view and inter-view spatial low-rank structures of the learned bipartite graphs into account by minimizing tensor Schatten p-norm and nuclear norm, respectively. Finally, an efficient algorithm, which scales linearly with the data size, is proposed to solve LCBG. Extensive experimental results on five benchmark datasets indicate our proposed LCBG is superior to the state-of-the-art methods.

1. Introduction

Multi-view clustering have become increasingly popular during the last decade [1–5], due to the fact that multi-view data ubiquitous in practical applications can provide supplementary information embedded in different views to improve clustering performance. Among numerous MVC methods, graph-based multi-view clustering (GMVC) is appealing due to its efficiency in characterizing both the complex structure of data and relationship between data [6–9].

The core of GMVC is how to construct a high-quality graph. To date, many graph construction methods have been proposed. Two of the most representative methods are sparse representation (SR) and low-rank representation (LRR) based methods [10–12]. SR based methods utilize the inherent sparsity of sparse representation to construct a similarity graph for clustering, *e.g.*, ℓ_1-norm. LRR based methods imposes low-rank constraint, *e.g.*, nuclear norm on graphs to construct robust graph. Although these methods achieve good performance, they need $\mathcal{O}\left(\mathrm{N}^2\right)$ computational complexity for $\mathrm{N}\times\mathrm{N}$ graph construction, which makes them ineffective in analyzing large-scale data, where N is the number of samples.

To address the aforementioned problem, bipartite graph based clustering method [13] has been proposed. The main idea is building the relationship between N data points and $M(\mathrm{M} \ll \mathrm{N})$ anchors. On the basis of this, [14] proposed a new method to learn a bipartite graph with constrained Laplacian rank such that the learned graph has K-connected components, but its performance heavily depends on the predefined bipartite graph. To improve the stableness of the method, [15] learned the bipartite graph by integrating the constrained Laplacian rank and adaptive anchors updated strategy into a unified framework such that the clustering results can be directly obtained according to the connected components of the learned bipartite graph. Although preliminary results are good, they only consider the single-view feature, and cannot be directly applied to multi-view data.

For bipartite graph-based MVC (BGMVC), [16] proposed to accelerate the classic multi-view spectral clustering via constructing multiple bipartite graphs (MVSC). However, the clustering results are very dependent on the pre-defined bipartite graphs, which limits the performance. To tackle this problem, [17] first learn view-specific graph based on self-expression model, then they designed a multiple graphs fusion mechanism to construct a view-consensus graph for clustering. Although they are effective, they still need extra post-processing operation to get the clustering results, resulting in unstable performance. In order to reasonably integrate these predefined bipartite graphs and obtain an optimal consensus one, [18] proposed to learn an optimal weight for each bipartite graph automatically without introducing an additive parameter (BiMVCC).

* Corresponding author.
E-mail address: qxgao@xidian.edu.cn (Q. Gao).

https://doi.org/10.1016/j.neucom.2022.09.002
0925-2312/

To this end, [19] proposed scalable and parameter-free multi-view clustering (SFMC), which aims to seek for a joint bipartite graph compatible across multiple views via a self-weighted scheme. Although achieving promising clustering performance, they still have the following drawbacks: (1) The performance heavily depends on the predefined bipartite graph, which fails to capture the non-convex data distribution. (2) None of them simultaneously considers the similarity of the inter-view, which well exploits the complementary information embedded in graphs of views, and the similarity of intra-view which exploits the cluster structure of data.

In this article, we prefer to retain the advantages of the bipartite graph, which is suitable for large scale clustering, and fully explore the complementary information and spatial structure embedded in different views. Combining the tensor Schatten-p nuclear norm, which is beneficial to exploit low-rank structural information of tensor [20,1], we propose a new clustering method, namely learning constraint bipartite graphs for multi-view clustering (LCBG). The main contributions are as follows:

1. The proposed method employs tensor Schatten p-norm minimization and nuclear norm minimization to simultaneously explore the low-rank structure embedded in inter-view graphs and intra-view graphs.
2. The learned view-consensus graph has clear connected components. Thus, the proposed method directly gets the clustering results according to the connected components without any post-processing.
3. We present an efficient optimization algorithm to solve LCBG. The computational complexity of the proposed algorithm is linear to the number of samples. Meanwhile, the proposed method achieves superior performance compared with the state-of-the-art methods on benchmark datasets.

Notations: In this article, we use bold calligraphy letters for third-order tensors, *e.g.*, $\mathscr{D} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, bold upper case letters for matrices, e.g., $\mathbf{D}$, bold lower case letters for vectors, e.g., $\mathbf{d}$, and lower case letters such as $d_{ijk}$for the entries of $\mathscr{D}$. The i-th frontal slice of $\mathscr{D}$ is $\mathscr{D}^{(i)}$. $\overline{\mathscr{D}}$ is the discrete Fast Fourier Transform (FFT) of $\mathscr{D}$ along the third dimension, i.e., $\overline{\mathscr{D}} = fft(\mathscr{D}, [], 3)$. Thus, $\mathscr{D} = ifft(\overline{\mathscr{D}}, [], 3)$. $tr(\mathbf{D})$ is the trace of matrix $\mathbf{D}$. $\mathbf{I}$ is an identity matrix.

2. Methodology

2.1. Objective

Bipartite graph-based clustering methods usually get the common shared graph based on the predefined graphs $\mathbf{S}^{(v)} \in \mathbb{R}^{N \times M}$ $(v = 1, 2, \cdots, V)$ by minimizing the mean squared errors, where $\mathbf{S}^{(v)}$ describes the relationship between M ($M \ll N$) anchors and N samples in the v-th view, V is the number of views. The anchor point selection strategy of our proposed method is the same as that of SFMC, named directly alternate sampling method (DAS). Fig. 1 shows the explanation of bipartite graph learning. There are three disadvantages: (1) the clustering performance is sensitive to the quality of $\mathbf{S}^{(v)}$; (2) these methods ignore the spatial low-rank structure embedded in $\mathbf{S}^{(v)}$ due to the fact that mean squared error is an element-based measurement model; (3) these methods cannot well explore the complementary information embedded in different views.

To tackle the aforementioned three challenges, we aim to adaptively learn an effective graph $\mathbf{S}^{(v)} \in \mathbb{R}^{N \times M}$ such that it well characterizes the relationship between M ($M \ll N$) anchors and N samples in the v-th view, and simultaneously get an implicit view-consensus graph which well explores the complementary information embedded in different views and has exactly K connected components, where K denotes the number of clusters.

To formulate our objective, we first introduce the Lemma 1 [21].

Lemma 1. The multiplicity K of the eigenvalue zeros of Laplacian matrix $\widetilde{\mathbf{L}}_{\mathbf{Q}}$ is equal to the number of connected components in the graph associated with $\mathbf{Q}$.

Lemma 1 indicates that, if $\text{rank}\left(\widetilde{\mathbf{L}}_{\mathbf{Q}}\right) = N + M - K$, then the corresponding graph has K connected components. We aim to learn graph $\mathbf{S}^{(v)}$ such that it can well characterize the relationship between M anchors and N data points in the v-th view and has good cluster structure. Moreover, considering the fact that different views have different contribution for clustering, we adaptively assign weight $\frac{1}{\beta^{(v)}}$ for the Laplacian matrix of the v-th view such that the weighted Laplacian matrix satisfies $\text{rank}\left(\widetilde{\mathbf{L}}_{\mathbf{Q}}\right) = N + M - K$. In this case, the implicit view-consensus graph $\mathbf{C} = \left(\sum_{v=1}^{V} \frac{\mathbf{S}^{(v)}}{\beta^{(v)}}\right) / \left(\sum_{v=1}^{V} \frac{1}{\beta^{(v)}}\right)$ has the K-connected components. Then, we can directly get the final clustering labels based on the connectivity of $\mathbf{C}$ without extra post-processing. To avoid the case that some rows of $\mathbf{S}^{(v)}$ are all zeros, we further constrain the learned $\mathbf{S}^{(v)}$. Thus, our clustering objective is

$$\min_{\mathbf{S}^{(v)}, \beta^{(v)}} \underbrace{\lambda_1 \sum_{v=1}^{V} \left\|\mathbf{S}^{(v)}\right\|_*}_{\text{intra-view}} + \lambda_2 \underbrace{\|\mathcal{S}\|_{\mathscr{S}p}^{p}}_{\text{inter-view}} \tag{1}$$

where $\sum_{v=1}^{V} \beta^{(v)} = 1, \beta^{(v)} \geqslant 0, \mathscr{S} \in \mathbb{R}^{N \times V \times M}, \mathscr{S}(:, v, :) = \mathbf{S}^{(v)}$ (See Fig. 2), $\mathbf{S}^{(v)} \geqslant 0$ and $\mathbf{S}^{(v)}\mathbf{1} = \mathbf{1}; \widetilde{\mathbf{L}}_{\mathbf{Q}} = \sum_{v=1}^{V} \frac{1}{\beta^{(v)}} \widetilde{\mathbf{L}}^{(v)}$ and $\text{rank}\left(\widetilde{\mathbf{L}}_{\mathbf{Q}}\right) = N + M - K, \widetilde{\mathbf{L}}^{(v)} = \mathbf{I} - \mathbf{D}_{(v)}^{-\frac{1}{2}} \mathbf{Z}^{(v)} \mathbf{D}_{(v)}^{-\frac{1}{2}}$ is the normalized

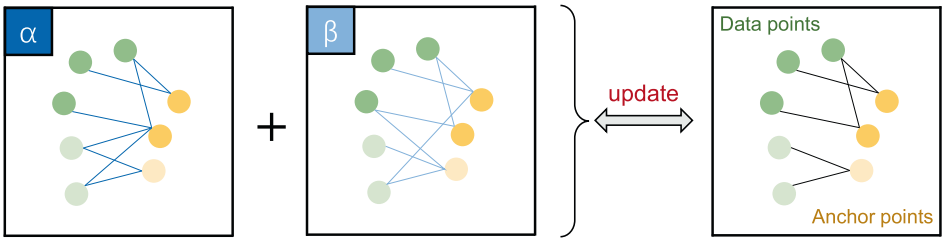


Fig. 1. Illustration of the joint structured optimal graph across two single-view bipartite graphs. The weights of the two single bipartite graphs are α and β, respectively. When a rank constraint is imposed on the Laplacian matrix of the joint bipartite graph, all weighted single-view graphs are integrated to be structured optimal with exactly k-connected components ($k = 2$ here) among data points and anchor points.

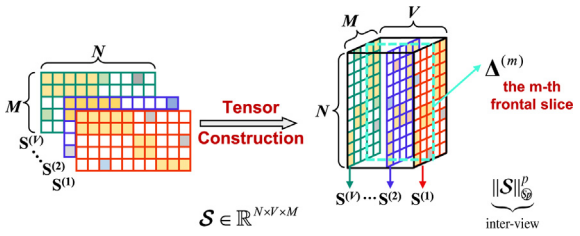


Fig. 2. Construction of tensor $\mathcal{S} \in \mathbb{R}^{N\times V\times M}$. $\mathbf{\Delta}^{(m)}$ denotes the m-th frontal slice of $\mathcal{S}$ ($m \in \{1, 2, \cdots, M\}$).

Laplacian matrix of bipartite graph $\mathbf{Z}^{(v)} \in \mathbb{R}^{(N+M)\times(N+M)}$ with $\mathbf{Z}^{(v)} = \begin{bmatrix} \mathbf{0} & \mathbf{S}^{(v)} \\ \mathbf{S}^{(v)^{\mathrm{T}}} & \mathbf{0} \end{bmatrix}$. $\mathbf{D}^{(v)}$ is a diagonal matrix whose diagonal elements are $\mathbf{D}^{(v)}(i,i) = \sum_{j=1}^{N+M} \mathbf{Z}^{(v)}(i,j)$. $\|\bullet\|_{\circledast_p}$ is the tensor Schatten p-norm (see Definition 1); $\|\bullet\|_*$ is the matrix nuclear norm; λ_1 and λ_2 are trade-off parameters.

Definition 1 [20]. Given $\mathcal{S} \in \mathbb{R}^{N\times V\times M}$, $h = \min(N, V)$, tensor Schatten p-norm of $\mathcal{S}$ is defined as

$$\begin{aligned} \|\mathcal{S}\|_{\circledast_p} &= \left(\sum_{i=1}^{M} \left\| \overline{\mathcal{S}}^{(i)} \right\|_{\circledast_p}^{p} \right)^{\frac{1}{p}} \\ &= \left(\sum_{i=1}^{M} \sum_{j=1}^{h} \sigma_j \left(\overline{\mathcal{S}}^{(i)} \right)^{p} \right)^{\frac{1}{p}} \end{aligned} \tag{2}$$

where $p \in (0, 1]$, $\sigma_j(\overline{\mathcal{S}}^{(i)})$ is the j-th singular value of $\overline{\mathcal{S}}^{(i)}$.

Remark 1. Note that when $p = 1$, the tensor Schatten p-norm of $\mathcal{S} \in \mathbb{R}^{N\times V\times M}$ is the tensor nuclear norm: $\|\mathcal{S}\|_* = \sum_{i=1}^{M}\sum_{j=1}^{h}\sigma_j(\overline{\mathcal{S}}^{(i)})$. Consider $\mathbf{\Delta}^{(m)} \in \mathbb{R}^{N\times V}$ and the singular values of $\mathbf{S}^{(v)}$, denoted by $\sigma_1, \ldots, \sigma_h$ in the descending order. Then for $p > 0$, we may consider $\|\mathbf{\Delta}^{(m)}\|_{\circledast_p}^{p} = \overset{g}{\sigma}_1^p + \cdots + \overset{h}{\sigma}_h^p$. If we let $p \to 0$, one can see $\lim_{p\to 0} \|\mathbf{\Delta}^{(m)}\|_{\circledast_p}^{p} = \#\{i : \overset{h}{\sigma}_i \neq 0\} = rank(\mathbf{\Delta}^{(m)})$. Hence, in literature, for $0 < p \leqslant 1$, the Schatten p-norm is introduced for the rank approximation.

Remark 2. In objective (1), adjacency matrix $\mathbf{S}^{(v)} \in \mathbb{R}^{N\times M}$ indicates the degree of approximation between M anchors and N sample points in the v-th view, if the same cluster contains i-th sample and j-th anchor, then $S_{ij}^{(v)}$ is high, otherwise $S_{ij}^{(v)}$ is low or zero. Moreover, we also hope $\mathbf{S}^{(v)}$ well characterizes the cluster structure of the v-th view. It indicates that the learned each graph $\mathbf{S}^{(v)}$ tends to be of low-rank. Thus, we leverage nuclear norm regularization (See the first term in objective (1)) to explore the spatial low-rank structure embedded in intra-view graphs. The second term in our objective (1) is used to explore the complementary information embedded in inter-views graphs $\mathbf{S}^{(v)}$. As shown in Fig. 2, for tensor $\mathcal{S}$, the m-th frontal slice $\mathbf{\Delta}^{(m)}$ describes the similarity between N sample points and the m-th anchor in different views. Considering the fact that different cluster structures are usually represented in different views, we introduce tensor Schatten-p norm minimization [20], i.e. tensor multi-rank minimization constraint on $\mathcal{S}$ which can make each $\mathbf{\Delta}^{(m)}$ has a low-rank structure in space. After it happens, the supplemental information embedded in inter-views can be well characterized by $\mathbf{\Delta}^{(m)}$.

2.2. Optimization

In the problem (1), it is hard to tackle for the rank constraint is a non-convex problem. Inspired by the Ky Fan's Theorem [21], the constraint $\mathrm{rank}\left(\widetilde{\mathbf{L}}_{\mathbf{Q}}\right) = N + M - K$ can be approximated by solving the following problem:

$$\min_{\mathbf{F}^{\mathrm{T}}\mathbf{F}=\mathbf{I}} tr\left(\mathbf{F}^{\mathrm{T}}\widetilde{\mathbf{L}}_{\mathbf{Q}}\mathbf{F}\right) \tag{3}$$

Fig. 3 shows the flowchart of LCBG. To avoid constructing the $N \times N$ graph, for each view, we construct a bipartite graph whose size is $N \times M (M \ll N)$, where M denotes the number of anchors, and then minimize divergence between anchor graphs of different views by the tensor Schatten p-norm and nuclear norm. And finally the V constructed $N \times M$ graphs $\mathbf{S}^{(v)}$ are stacked and rotated to obtain the final tensor $\mathcal{S}$, which will be updated by using tensor multi-rank minimization based on t-SVD. As shown in Fig. 3, we have that, for $\|\mathcal{S}\|_{\circledast_p}^{p}$, we can fully explore the complementary information embedded in inter-view graphs, and for $\sum_{v=1}^{V}\left\|\mathbf{S}^{(v)}\right\|_*$, we can explore the spatial low-rank structure embedded in inter-view graphs.

Currently, the problem (1) is approximated to solve the following problem:

$$\min_{\substack{\mathbf{S}^{(v)},\,\mathbf{F},\\ \beta^{(v)}}} \lambda_1 \underbrace{\sum_{v=1}^{V}\left\|\mathbf{S}^{(v)}\right\|_*}_{\text{intra-view}} + \lambda_2 \underbrace{\|\mathcal{S}\|_{\circledast_p}^{p}}_{\text{inter-view}} + \lambda_3 tr\left(\mathbf{F}^{\mathrm{T}}\widetilde{\mathbf{L}}_{\mathbf{Q}}\mathbf{F}\right) \tag{4}$$

Remark 3. In Eq. (4), λ_3 is a hidden parameter, which can be adaptively updated as follows. When initializing, we set λ_3 to a small value. It is then updated based on the number of eigenvalues of zero for the $\widetilde{\mathbf{L}}_{\mathbf{Q}}$ after each iteration. For the number is less than K, λ_3 is multiplied by 2, or for it is greater than $K + 1$, λ_3 is divided by 2, otherwise we terminate the iterations.

According to the Augmented Lagrange Multiplier (ALM) method, we rewrite the (4) by introducing auxiliary variables $\mathcal{J}$ and $\mathbf{P}^{(v)}$. Thus, we have

$$\begin{aligned} \min_{\substack{\mathbf{S}^{(v)},\,\mathbf{F},\,\beta^{(v)},\\ \mathbf{P}^{(v)},\,\mathcal{J}}} & \lambda_1 \sum_{v=1}^{V}\left\|\mathbf{P}^{(v)}\right\|_* + \lambda_2\|\mathcal{J}\|_{\circledast_p}^{p} + \\ & \lambda_3 tr\left(\mathbf{F}^{\mathrm{T}}\widetilde{\mathbf{L}}_{\mathbf{Q}}\mathbf{F}\right) + \frac{\rho_1}{2}\left\|\mathcal{J} - \left(\mathcal{S} + \frac{\mathcal{W}}{\rho_1}\right)\right\|_F^2 + \\ & \frac{\rho_2}{2}\sum_{v=1}^{V}\left\|\mathbf{P}^{(v)} - \left(\mathbf{S}^{(v)} + \frac{\mathbf{H}^{(v)}}{\rho_2}\right)\right\|_F^2 \end{aligned} \tag{5}$$

where tensor $\mathcal{W}$ and matrix $\mathbf{H}^{(v)}$ are Lagrange multiplier, ρ_1 and ρ_2 are the penalty parameters. The optimization process could be separated into following steps:

• **Solving F with fixed $\mathcal{J}, \mathbf{P}^{(v)}, \mathbf{S}^{(v)}$ and $\beta^{(v)}$.** In this case, the optimization of the problem (5) is simplified into solving (3), which can be solved by eigenvalue decomposition on $\widetilde{\mathbf{L}}_{\mathbf{Q}}$. However, the

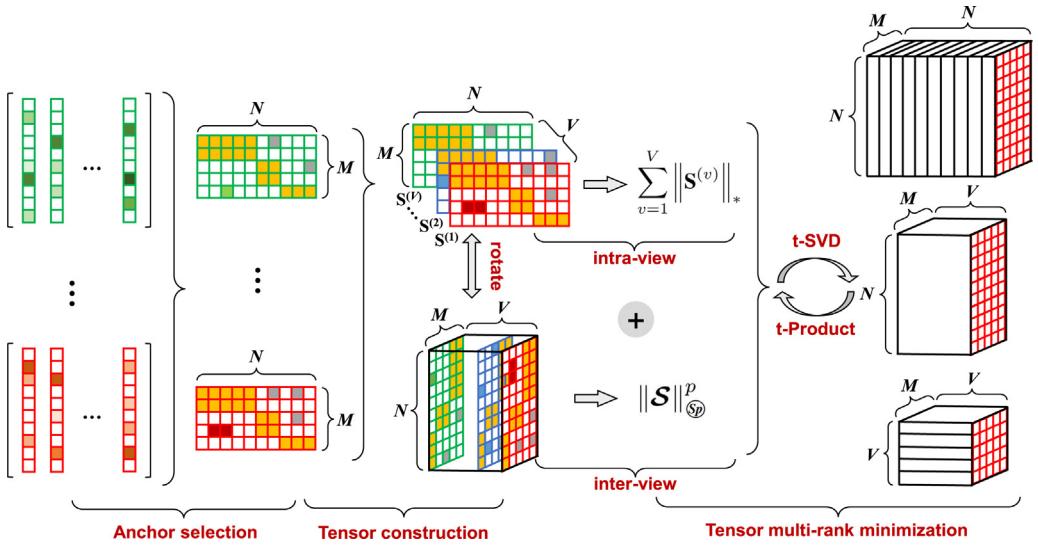


Fig. 3. The flowchart of LCBG.

computational complexity required to solve it directly is $\mathcal{O}\left((N+M)^2K\right)$, which is not suitable for large-scale data. To this end, we present an valid algorithm here. By simple algebra, the problem (3) is equivalent to the following problem:

$$\arg\max_{\mathbf{F}_N^{\mathrm{T}}\mathbf{F}_N+\mathbf{F}_M^{\mathrm{T}}\mathbf{F}_M=\mathbf{I}} tr\left(\mathbf{F}_N^{\mathrm{T}}\mathbf{E}\mathbf{F}_M\right) \tag{6}$$

where $\mathbf{E}=\mathbf{D}_N^{-\frac{1}{2}}\mathbf{Q}\mathbf{D}_M^{-\frac{1}{2}};\mathbf{F}=\begin{bmatrix}\mathbf{F}_N^{\mathrm{T}} & \mathbf{F}_M^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}};\mathbf{Q}=\sum_{v=1}^{V}\frac{\mathbf{S}^{(v)}}{\beta^{(v)}}$, where $\mathbf{F}_N\in\mathbb{R}^{N\times K}$ is the first N rows of $\mathbf{F}$ and $\mathbf{F}_M\in\mathbb{R}^{M\times K}$ is the remaining M rows of $\mathbf{F};\mathbf{D}_N\in\mathbb{R}^{N\times N}$ and $\mathbf{D}_M\in\mathbb{R}^{M\times M}$ are diagonal matrices whose diagonal elements are $\mathbf{D}_N(i,i)=\sum_{j=1}^{M}\mathbf{S}(i,j)$ and $\mathbf{D}_M(j,j)=\sum_{i=1}^{N}\mathbf{S}(i,j)$.

To solve the problem (6), we first introduce Theorem 1.

Theorem 1. Given $\mathbf{E}\in\mathbb{R}^{N\times M},\mathbf{F}_N\in\mathbb{R}^{N\times K},\mathbf{F}_M\in\mathbb{R}^{M\times K}$. The optimal solutions of

$$\arg\max_{\mathbf{F}_N^{\mathrm{T}}\mathbf{F}_N+\mathbf{F}_M^{\mathrm{T}}\mathbf{F}_M=\mathbf{I}} tr\left(\mathbf{F}_N^{\mathrm{T}}\mathbf{E}\mathbf{F}_M\right) \tag{7}$$

are $\mathbf{F}_N=\frac{\sqrt{2}}{2}\mathbf{U}_1,\mathbf{F}_M=\frac{\sqrt{2}}{2}\mathbf{V}_1$, where $\mathbf{U}_1$ and $\mathbf{V}_1$ are the leading K left and right singular vectors of $\mathbf{E}$, respectively.

Proof 1. According to Eq. (7), we have

$$tr\left(\mathbf{F}_N^{\mathrm{T}}\mathbf{E}\mathbf{F}_M\right)=\frac{1}{2}\left(tr\left(\mathbf{F}_N^{\mathrm{T}}\mathbf{E}\mathbf{F}_M+tr\left(\mathbf{F}_M^{\mathrm{T}}\mathbf{E}^{\mathrm{T}}\mathbf{F}_M\right)\right)=\frac{1}{2}tr\left(\begin{bmatrix}\mathbf{F}_N\\ \mathbf{F}_M\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\mathbf{0} & \mathbf{E}\\ \mathbf{E}^{\mathrm{T}} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{F}_N\\ \mathbf{F}_M\end{bmatrix}\right) \tag{8}$$

Then, the problem (7) is equal to

$$\arg\max_{\mathbf{F}_N,\mathbf{F}_M}\frac{1}{2}tr\left(\begin{bmatrix}\mathbf{F}_N\\ \mathbf{F}_M\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\mathbf{0} & \mathbf{E}\\ \mathbf{E}^{\mathrm{T}} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{F}_N\\ \mathbf{F}_M\end{bmatrix}\right) \tag{9}$$

where $[\mathbf{F}_N\ \mathbf{F}_M]^T[\mathbf{F}_N\ \mathbf{F}_M]=\mathbf{I}$. So the optimal solution of the problem (9) can be solved by

$$\frac{1}{2}\begin{bmatrix}\mathbf{0} & \mathbf{E}\\ \mathbf{E}^{\mathrm{T}} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{F}_N\\ \mathbf{F}_M\end{bmatrix}=\begin{bmatrix}\mathbf{F}_N\\ \mathbf{F}_M\end{bmatrix}\mathbf{\Lambda} \tag{10}$$

where $\mathbf{\Lambda}$ is diagonal matrix of which elements are constituted of eigenvalues of $\frac{1}{2}\begin{bmatrix}\mathbf{0} & \mathbf{E}\\ \mathbf{E}^{\mathrm{T}} & \mathbf{0}\end{bmatrix}$.

By simple algebra, we have

$$\begin{cases}\frac{1}{2}\mathbf{E}\mathbf{F}_M=\mathbf{F}_N\mathbf{\Lambda}\\ \frac{1}{2}\mathbf{E}^{\mathrm{T}}\mathbf{F}_N=\mathbf{F}_M\mathbf{\Lambda}\end{cases} \tag{11}$$

Then,

$$\begin{cases}\left(\frac{\sqrt{2}}{2}\mathbf{E}\right)^{\mathrm{T}}\left(\frac{\sqrt{2}}{2}\mathbf{E}\right)\mathbf{F}_M=\mathbf{F}_M\left(\sqrt{2}\mathbf{\Lambda}\right)^2\\ \left(\frac{\sqrt{2}}{2}\mathbf{E}\right)\left(\frac{\sqrt{2}}{2}\mathbf{E}\right)^{\mathrm{T}}\mathbf{F}_N=\mathbf{F}_N\left(\sqrt{2}\mathbf{\Lambda}\right)^2\end{cases} \tag{12}$$

Based on model (12), $\mathbf{F}_N$ and $\mathbf{F}_M$ consist of the first K left and right singular vectors of $\frac{\sqrt{2}}{2}\mathbf{E}$. $\mathbf{U}_1$ and $\mathbf{V}_1$ denote the first K left and right singular vectors of $\mathbf{E}$, respectively. Based on this, we derive $\mathbf{F}_M=\frac{\sqrt{2}}{2}\mathbf{V}_1,\mathbf{F}_N=\frac{\sqrt{2}}{2}\mathbf{U}_1$. □

Thus, according to Theorem 1, the optimal $\mathbf{F}^*=\frac{\sqrt{2}}{2}\begin{bmatrix}\mathbf{U}_1^{\mathrm{T}} & \mathbf{V}_1^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}$. Solving the problem (6) takes the computational complexity $\mathcal{O}\left(VNM+M^2N\right)$, which is much more efficient than directly solving the problem (3).

• **Solving $\mathcal{J}$ with fixed $\mathbf{F},\mathbf{P}^{(v)},\mathbf{S}^{(v)}$ and $\beta^{(v)}$.** In this case, $\mathcal{J}$ can be solved by

$$\arg\min_{\mathcal{J}}\lambda_2\|\mathcal{J}\|_{\circledS p}+\frac{\rho_1}{2}\left\|\mathcal{J}-(\mathcal{S}+\frac{\mathcal{W}}{\rho_1})\right\|_F^2 \tag{13}$$

We use the Theorem 2 [20] to obtain the optimal solution of tensor $\mathcal{J}$ in the problem (13).

Theorem 2. Suppose $\mathcal{S}\in\mathbb{R}^{n_1\times n_2\times n_3},h=\min(n_1,n_2)$, let $\mathcal{S}=\mathcal{U}*\mathcal{A}*\mathcal{V}^T$. For the following model:

$$\arg\min_{\mathcal{X}}\frac{1}{2}\|\mathcal{X}-\mathcal{S}\|_F^2+\tau\|\mathcal{X}\|_{\circledS p}^p \tag{14}$$

the optimal solution $\mathcal{X}^*$ is

$$\mathcal{X}^*=\Gamma_{\tau\cdot n_3}(\mathcal{S})=\mathcal{U}*ifft\left(\mathbf{C}_{\tau\cdot n_3}(\overline{\mathcal{S}})\right)*\mathcal{V}^T \tag{15}$$

where, $\mathbf{C}_{\tau\cdot n_3}(\overline{\mathcal{S}})$ is a tensor, $\mathbf{C}_{\tau\cdot n_3}(\overline{\mathcal{S}}^{(i)})$ is the i-th frontal slice of $\mathbf{C}_{\tau\cdot n_3}(\overline{\mathcal{S}})$.

Accordingly, the solution of the problem (13) is

$$\mathscr{J}^* = \Gamma_{\frac{\lambda_2}{\rho_1}}\left(\mathscr{S} + \frac{\mathscr{W}}{\rho_1}\right) \tag{16}$$

• **Solving $\mathbf{P}^{(v)}$ with fixed $\mathbf{F}, \mathscr{J}, \mathbf{S}^{(v)}$ and $\beta^{(v)}$.** In this case, $\mathbf{P}^{(v)}$ can be solved by

$$\min_{\mathbf{P}^{(v)}} \lambda_1 \sum_{v=1}^{V} \left\|\mathbf{P}^{(v)}\right\|_* + \frac{\rho_2}{2} \sum_{v=1}^{V} \left\|\mathbf{P}^{(v)} - \left(\mathbf{S}^{(v)} + \frac{\mathbf{H}^{(v)}}{\rho_2}\right)\right\|_F^2 \tag{17}$$

Thus, $\mathbf{P}^{(v)}$ can be solved by the singular value thresholding [22] shrinkage operator as

$$\mathbf{P}^{(v)} = \Theta_{\frac{\lambda_1}{\rho_2}}\left(\mathbf{S}^{(v)} + \frac{\mathbf{H}^{(v)}}{\rho_2}\right) \tag{18}$$

where Θ denotes the SVT shrinkage operator.

• **Solving $\mathbf{S}^{(v)}$ with fixed $\mathbf{F}, \mathscr{J}, \mathbf{P}^{(v)}$ and $\beta^{(v)}$.** The optimization w.r.t $\mathbf{S}^{(v)}$ in (5) changes into

$$\begin{aligned}\arg\min_{\mathbf{S}^{(v)}} \lambda_3 \sum_{v=1}^{V} \frac{1}{\beta^{(v)}} tr\left(\mathbf{F}^{\mathrm{T}} \tilde{\mathbf{L}}^{(v)} \mathbf{F}\right) + \frac{\rho_1}{2} \left\|\mathscr{J} - \left(\mathscr{S} + \frac{\mathscr{W}}{\rho_1}\right)\right\|_F^2 \\ + \frac{\rho_2}{2} \sum_{v=1}^{V} \left\|\mathbf{P}^{(v)} - \left(\mathbf{S}^{(v)} + \frac{\mathbf{H}^{(v)}}{\rho_2}\right)\right\|_F^2\end{aligned} \tag{19}$$

By transforming the tensors $\mathscr{J}, \mathscr{S}$ and $\mathscr{W}$ into matrix in Eq. (19), we can obtain

$$\begin{aligned}\arg\min_{\mathbf{S}^{(v)}} \lambda_3 \sum_{v=1}^{V} \frac{1}{\beta^{(v)}} tr\left(\mathbf{F}^{\mathrm{T}} \tilde{\mathbf{L}}^{(v)} \mathbf{F}\right) + \frac{\rho_2}{2} \sum_{v=1}^{V} \left\|\mathbf{J}^{(v)} - \left(\mathbf{S}^{(v)} + \frac{\mathbf{W}^{(v)}}{\rho_1}\right)\right\|_F^2 \\ + \frac{\rho_2}{2} \sum_{v=1}^{V} \left\|\mathbf{P}^{(v)} - \left(\mathbf{S}^{(v)} + \frac{\mathbf{H}^{(v)}}{\rho_2}\right)\right\|_F^2\end{aligned} \tag{20}$$

By simple algebra, the first term in Eq. (20) can be rewritten as

$$Const - 2\lambda_3 \sum_{v=1}^{V} tr\left(\mathbf{S}^{(v)} \mathbf{T}^{(v)}\right) \tag{21}$$

where $\mathbf{T}^{(v)} = \frac{1}{\beta^{(v)}} \mathbf{D}_{(v)}^{-\frac{1}{2}} \mathbf{F}_M \mathbf{F}_N^T \mathbf{D}_{(v)}^{-\frac{1}{2}}$.

Substituting $\mathbf{U}^{(v)} = \mathbf{J}^{(v)} - \frac{\mathbf{W}^{(v)}}{\rho_1}, \mathbf{V}^{(v)} = \mathbf{P}^{(v)} - \frac{\mathbf{H}^{(v)}}{\rho_1}$, the rest of Eq. (20) can be rewritten as

$$\begin{aligned}Const + \frac{\rho_1}{2} \sum_{v=1}^{V} \left\{tr\left(\mathbf{S}^{(v)} \mathbf{S}^{(v)^{\mathrm{T}}}\right) - 2tr\left(\mathbf{S}^{(v)} \mathbf{U}^{(v)^{\mathrm{T}}}\right)\right\} \\ + \frac{\rho_2}{2} \sum_{v=1}^{V} \left\{tr\left(\mathbf{S}^{(v)} \mathbf{S}^{(v)^{\mathrm{T}}}\right) - 2tr\left(\mathbf{S}^{(v)} \mathbf{V}^{(v)^{\mathrm{T}}}\right)\right\}\end{aligned} \tag{22}$$

Then substituting Eq. (21) and Eq. (22) into Eq. (20), and by simple algebra, Eq. (20) becomes

$$\begin{aligned}\arg\min_{\mathbf{S}^{(v)}} Const - 2\lambda_2 tr\left(\mathbf{S}^{(v)} \mathbf{T}^{(v)}\right) + \frac{\rho_1 + \rho_2}{2} \sum_{v} tr\left(\mathbf{S}^{(v)} \mathbf{S}^{(v)^{\mathrm{T}}}\right) \\ - \rho_1 \sum_{v=1}^{V} tr\left(\mathbf{S}^{(v)} \mathbf{U}^{(v)^{\mathrm{T}}}\right) - \rho_2 \sum_{v=1}^{V} tr\left(\mathbf{S}^{(v)} \mathbf{V}^{(v)^{\mathrm{T}}}\right)\end{aligned} \tag{23}$$

By simple algebra, (19) can be rewritten as

$$\arg\min_{\mathbf{S}^{(v)}} \frac{\rho_1 + \rho_2}{2} \sum_{v=1}^{V} \left\|\mathbf{S}^{(v)} - \frac{\mathbf{A}^{(v)}}{\rho_1 + \rho_2}\right\|_F^2 \tag{24}$$

where $\mathbf{S}^{(v)} \geqslant 0, \mathbf{S}^{(v)} \mathbf{1} = \mathbf{1}, \mathbf{A}^{(v)} = \rho_1 \mathbf{U}^{(v)^{\mathrm{T}}} + \rho_2 \mathbf{V}^{(v)^{\mathrm{T}}} + 2\lambda_3 \mathbf{T}^{(v)}$.

In (24), all $\mathbf{S}^{(v)} (v = 1, \cdots, V)$ are independent. For each $\mathbf{S}^{(v)}$, the closed-form solution $\mathbf{S}^{(v)^*}$ is $\mathbf{S}^{(v)^*}(i,:) = \left(\frac{\mathbf{A}^{(v)}(i,:)}{\rho_1 + \rho_2} + \gamma \mathbf{1}\right)_+$ [23], where γ is the Lagrangian multiplier.

• **Solving $\beta^{(v)}$ with fixed other variables.** In this condition, the optimization w.r.t $\beta^{(v)}$ in (5) changes into

$$\arg\min_{\beta^{(v)}} \sum_{v=1}^{V} \frac{\tau^{(v)}}{\beta^{(v)}} \tag{25}$$

where $\sum_{v=1}^{V} \beta^{(v)} = 1, \beta^{(v)} \geqslant 0, \tau^{(v)} = tr\left(\mathbf{F}^{\mathrm{T}} \tilde{\mathbf{L}}^{(v)} \mathbf{F}\right)$.

According to the method of Lagrange multipliers, the (25) can be rewritten as

$$\arg\min_{\beta^{(v)}, \eta} \sum_{v=1}^{V} \frac{\tau^{(v)}}{\beta^{(v)}} - \eta\left(\sum_{v=1}^{V} \beta^{(v)} - 1\right) \tag{26}$$

where η is the Lagrange multiplier.

Demand (26) on the $\beta^{(v)}$ and η partial derivatives, we can get $\beta^{(v)} = \sqrt{\frac{\tau^{(v)}}{\eta}}$. With simple algebraic calculation, the optimal $\beta^{(v)}$ is

$$\beta^{(v)} = \frac{\sqrt{\tau^{(v)}}}{\sum_{v=1}^{V} \left(\sqrt{\tau^{(v)}}\right)} \tag{27}$$

Update Variables $\mathscr{W}, \mathbf{H}^{(v)}, \rho_1$ and ρ_2. The four variables are updated as follows

$$\mathscr{W} = \mathscr{W} + \rho_1 (\mathscr{S} - \mathscr{J}) \tag{28}$$

$$\mathbf{H}^{(v)} = \mathbf{H}^{(v)} + \rho_2 \left(\mathbf{S}^{(v)} - \mathbf{P}^{(v)}\right) \tag{29}$$

$$\rho_{1,2} = \min\left(pho\_\rho \times \rho_{1,2}, max\_\rho\right) \tag{30}$$

where max_ρ and pho_ρ are constants.

The complete algorithm is shown in Algorithm 1.

Algorithm 1: Low-rank Constraint Bipartite Graph Learning

Input: Data matrices: $\left\{\mathbf{X}^{(v)}\right\}_{v=1}^{V} \in \mathbb{R}^{N \times d_v}$, anchors number M and cluster number K
Parameter: $\mathscr{S}, \mathscr{W}, \mathscr{J}, \mathbf{H}^{(v)}, \mathbf{P}^{(v)}, \mathbf{F}, \rho_1, \rho_2, \beta^{(v)}, \lambda_1, \lambda_2$ and λ_3
Output: Graph $\mathbf{C}$ with K-connected components
1: Let $\mathscr{W} = \mathscr{J} = \mathbf{0}, \mathbf{H}^{(v)} = \mathbf{P}^{(v)} = \mathbf{0}$, $max\_\rho = 10^{10}$, $pho\_\rho = 1.1, \beta^{(v)} = \frac{1}{V}$, $\lambda_1 = \frac{1}{n}, \lambda_2 = \frac{1}{\sqrt{n \times V}}$, $\lambda_3 = 10^{-3}, \rho_1 = \rho_2 = 0.1, anchor\_rate = 0.5, p$ varies for different datasets (see 3.2.3).
2: **while**not converg **do**
3: Update $\mathbf{F}$ by solving (6);
4: Update $\mathscr{J}$ by using (16);
5: Update $\mathbf{P}^{(v)}$ by using (18);
6: Update $\mathbf{S}^{(v)}$ by solving (24);
7: Update $\beta^{(v)}$ by using (27);
8: Update $\mathscr{W}, \mathbf{H}^{(v)}$ by using (28) and (29), respectively;
9: Update ρ_1 and ρ_2 by using (30);
10: Update λ_3 adaptively;
11: **end while**
12: Directly achieve the K clusters based on the connectivity of $\mathbf{C} = \left(\sum_{v=1}^{V} \frac{\mathbf{S}^{(v)}}{\beta^{(v)}}\right) / \left(\sum_{v=1}^{V} \frac{1}{\beta^{(v)}}\right)$;
13: **return** Clustering results

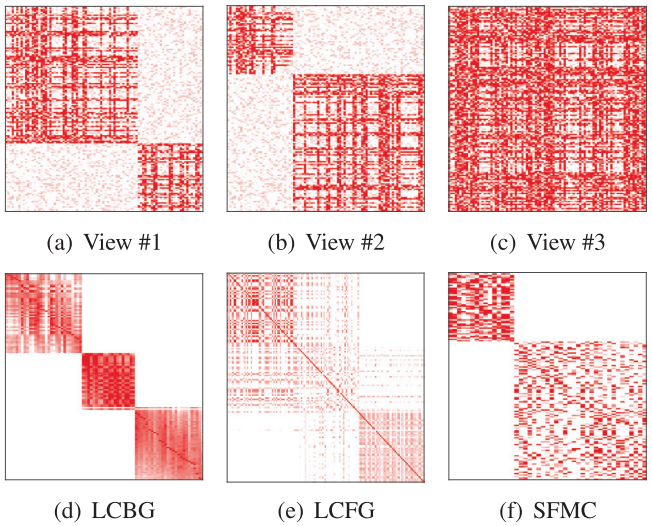
2.3. Complexity

The computational complexity of LCBG consists of two main components: construction of bipartite graphs $\left\{\mathbf{S}^{(v)}\right\}_{v=1}^{V}$ and optimization by iteratively solving expression (5). Due to our anchor selection method is the same as SFMC, constructing bipartite graphs takes $\mathcal{O}(VNMd + VNM\log(M))$ time, where $d = \sum_{v=1}^{V} d_v, V, M$ and N are the number of views, anchors and samples, respectively. Solving expression (5) mainly pays attention to four variables $\mathbf{S}^{(v)}, \mathcal{J}, \mathbf{P}^{(v)}$ and $\mathbf{F}$. For $\mathcal{J} \in \mathbb{R}^{N\times M\times V}$ and $\mathbf{P}^{(v)}(v = 1, 2, \cdots, V)$, the computational complexity is the same, this subproblem involves calculating the 3D FFT and 3D inverse FFT of an $N \times V \times M$ tensor and N SVDs of $M \times V$ matrices in the Fourier domain, and that the complexity in both of them are $\mathcal{O}(2VNM\log(VM))$ and $\mathcal{O}\left(V^2MN\right)$. Thus, updating four variables iteratively has the complexity of $\mathcal{O}(VNM(K+1) + VNM\log(M))$, $\mathcal{O}(2VNM\log(VM) + V^2MN)$, $\mathcal{O}\left(2VNM\log(VM) + V^2MN\right)$ and $\mathcal{O}\left(VNM + M^2N\right)$,respectively, where K and t are the number of clusters and iteration, respectively. Due to $M \ll N$, the main complexity in this stage is $\mathcal{O}\left(M^2Nt + 2VNMt\log(VM)\right)$, where t is the number of iterations. Therefore, the main computational complexity of our method is actually $\mathcal{O}\left(M^2Nt + VNMd\right)$, which is linear to N. Thus, LCBG is more efficient for large-scale data.

3. Experiments

3.1. Evaluations on Synthetic Data

For features from different views are embedded with different data features, graphs constructed from different views can even reflect different clustering structures. In view of this, we designed three different bipartite graphs as the input of the single-view graph, as shown in Fig. 4(a-c), where Fig. 4a) and and Fig. 4(b) have two diagonal blocks embodying different cluster distributions, Fig. 4(c) fills the Gaussian noise without any cluster structures. And we add random noise uniformly to all of these single-view graphics. The joint graph learned by LCBG (see Fig. 4(d)) has exact 3 connected components, and information in both View #1 and View #2 is absorbed by the final clusters jointly. We cannot see a clear connected component in the joint graph learned by LCFG (Low-rank Constraint Full Graph Learning, i.e. LCBG without bipartite graph, see Fig. 4(e)). This shows that the use of bipartite graphs can effectively reduce the effect of useless noise data. There are only exact 2 connected components in the joint graph learned by SFMC (see Fig. 4(f)), which are similar to View #2. The reason may be that LCBG fully explores the complementary information embedded in inter-view and intra-view graphs, while SFMC not.



(a) View #1 (b) View #2 (c) View #3

(d) LCBG (e) LCFG (f) SFMC

Fig. 4. Experiments on synthetic multi-view data.

From Fig. 4, we can conclude that: our model adaptively assign the weighted parameters for different views and explores not only informative structure within each view but also the compatible structure between multiple views, thus, the severely ill-views cannot degrade the result. In conclusion, our method can perform well when dealing with data with high amounts of noise.

3.2. Evaluations on Real Datasets

3.2.1. Experimental Setup

In this subsection, we will present the datasets, comparisons, and metrics used for evaluation.

Datasets: The five multi-view datasets selected to demonstrate the superiority of our proposed method are as follows:

Handwritten4 [24] includes 10 digits with 2,000 images generated from UCI machine learning repository. 76-D FOU, 216-D FAC, 47-D ZER and 6-D MOR features are employed as 4 views.

Cal101-20 is a subsets of Caltech101 [25] dataset, it includes 20 categories with 2,386 images. 48-D GABOR, 40-D WM, 254-D CENT, 1,984-D HOG, 512-D GIST and 928-D LBP features are employed as 6 views.

MSRC [26] includes 7 kinds of objects with 210 images. We choose 24-dimension (D) CM feature, 576-D HOG feature, 512-D GIST feature, 256-D LBP feature, 254-D CENT feature as 5 views.

Mnist [27] includes 4 categories handwritten digits, i.e., from digit 0 to digit 3, with 4,000 images. We utilize 30-D ISO feature, 9-D LDA feature and 30-D NPE feature as 3 views.

Reuters [28] consists of 18,757 documents from 6 categories. 21,531-D EN, 24,892-D FR, 24,892-D GR, 15,506-D IT and 11,547-D SP features are employed as 5 views.

NUS-WIDE [29] consists of 30,000 object images from 31 classes. 64-D CH feature, 225-D CM feature, 144-D CORR feature, 73-D EDH feature and 128-D WT feature are employed as 5 views.

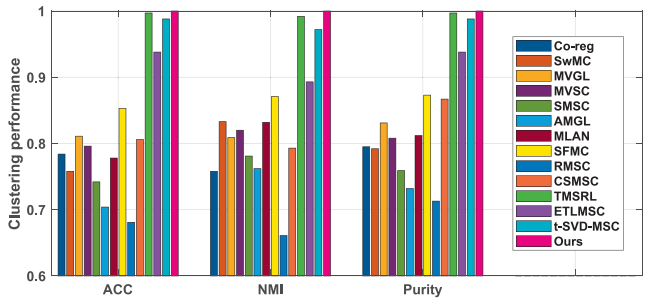
Comparisons and Metrics: We choose 13 methods as comparison methods including Co-reg [30], SwMC [31], MVGL [32], MVSC [16], SMSC [33], AMGL [34], MLAN [35], SFMC [19], RMSC [36], CSMSC [12], TMSRL [37], ETLMSC [6], t-SVD-MSC [38] and then evaluate performance by 3 indicators, including Accuracy (ACC); Normalized Mutual Information (NMI) and Purity. For all metrics, the higher value indicates the better clustering performance.

Hardware platform: Hardware platform used for the experiments: Window 10 PC with Intel Core i7-9700 K 3.60 GHz CPU, 64 GB RAM and MATLAB R2020a.

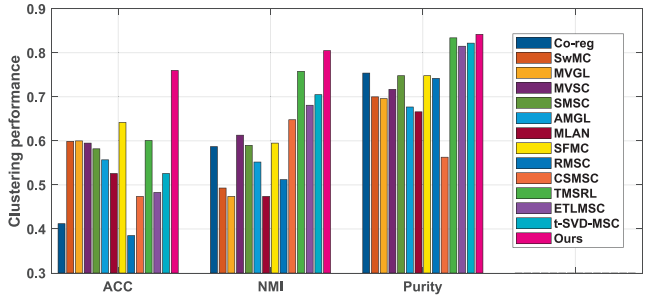
3.2.2. Comparisons with State-of-the-art Methods

The metrics comparison of the above methods on four small datasets is shown in Fig. 5. In order to verify the superiority of our algorithm on large-scale datasets, we test our algorithm on Reuters and NUS-WIDE datasets, and the relevant results are shown in Table 1. For the comparison methods that are not included in Table 1, they can't process this large-scale datasets. For each experiment, we independently repeat the involved methods 20 times and show the averages. From Fig. 5 and Table 1, we have the following interesting observations:

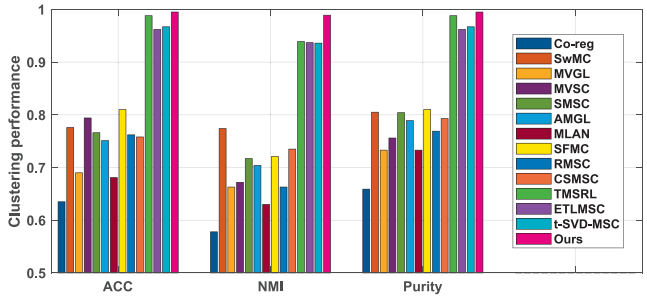
- First, TMSRL, ETLMSC, t-SVD-MSC and our method obtain the top four results, proving that the use of tensor nuclear norm



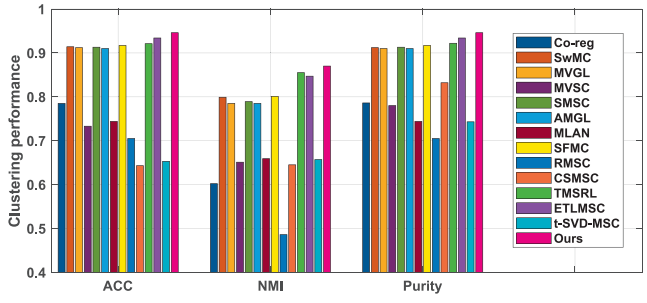
(a) Handwritten4



(b) Cal101-20



(c) MSRC



(d) Mnist

Fig. 5. The clustering performances on Handwritten4, Cal101-20, MSRC and Mnist datasets.

Table 1
The clustering performances on Reuters and NUS-WIDE datasets.

Dataset	Reuters			NUS-WIDE		
Metric	ACC	NMI	Purity	ACC	NMI	Purity
Co-reg	0.563	0.326	0.552	0.119	**0.114**	0.214
MVSC	0.596	0.347	0.574	0.150	0.075	0.184
SFMC	0.602	0.354	0.604	0.169	0.060	0.190
Ours	**0.640**	**0.484**	**0.686**	**0.210**	0.107	**0.219**

has a beneficial impact on the clustering results. The reason may be that the low-rank representation can well characterize the data relationships between data and use spatial structure.

- Second, the result of SFMC and MVSC are also good, proving that the anchor selection has a beneficial impact on the clustering results. This is because redundant information may be included in origin data, and anchor selection can just eliminate the interference of some redundant information, so as to obtain better results.
- Third, our method obtains good result on large dataset. Comparing with AMGL, which cannot handle large-scale dataset, our method scales linearly with the data size. Thus, our method can well handle large-scale dataset.
- Fourth, despite a anchor-based algorithm, our method is remarkably superior to SFMC. reason may be that our method well exploit the complementary information and spatial low-rank structure information embedded in graphs of different views, while SFMC does not.

3.2.3. Effect of parameter p

We perform an analysis of the clustering effect of p on four selected datasets. Specifically, the value of p is changed from 0.1 to 1.0 with 0.1 intervals, and then we use ACC, NMI, and Purity as evaluation criteria, as shown in Fig. 6. It can be observed that the results under different p are distinguishing mostly, and when $p = 0.9, 0.6, 0.5\text{and}1$, the best clustering results are obtained on Handwritten4, Cal101-20, MSRC and Mnist datasets, respectively. This demonstrates that p has a significant influence on the clustering results, and we need to take different value of p for different datasets. This is probably because that p exploits the significant difference between singular values. Another reason may be that tensor Schatten p-norm makes well exploits complementary information and spatial structure embedded in graphs of different views.

3.2.4. Effect of the number of anchors

We perform an analysis of the clustering impact of the number of anchors on Handwritten4 and Cal101-20 datasets. To do this, we change the proportion of anchors in the whole data points from 0.1 to 1.0 with 0.1 intervals, and then we show the three metrics in Fig. 7. It can be clearly observed that our method fluctuates greatly with the number of anchors. Our method achieves the best performance on these datasets when the proportion is set to 0.5. Furthermore, we find that the metrics curves w.r.t. anchors proportion are not monotonically increasing. This suggests that clustering with a large number of anchors is not required. Therefore, for all datasets, we set the anchors proportion to 0.5 for all data points.

3.2.5. Effect of tensor $\mathscr{S}$'s Schatten p-norm and nuclear norm

In order to verify the effect of tensor $\mathscr{S}$'s Schatten p-norm and $\mathbf{S}^{(v)}$'s nuclear norm in our model, we made the following modifications on the basis of this model: taking $\lambda_1, \lambda_2 = 0$, respectively and simplifying the update steps of our model to get a new model without tensor $\mathscr{S}$'s Schatten p-norm (LCBG w.o. inter-view) and one without nuclear norm (LCBG w.o. intra-view). The clustering performances of three models are shown in Fig. 8. It can be seen that the clustering results of the model without tensor $\mathscr{S}$'s Schatten p-norm are the worst of the three models and the model without nuclear norm get the suboptimal consequences. We can easily observe that the Schatten p-norm can effectively exploit the view-similar between graphs of different views, the nuclear norm can learn the local structure information embedded in each graph of different views and the influence of Schatten p-norm is greater than that of nuclear norm.

3.2.6. Graph visualization analysis

Taking Handwritten4 dataset as an example, we draw the input graphs and the learned view-consensus graph of our method on

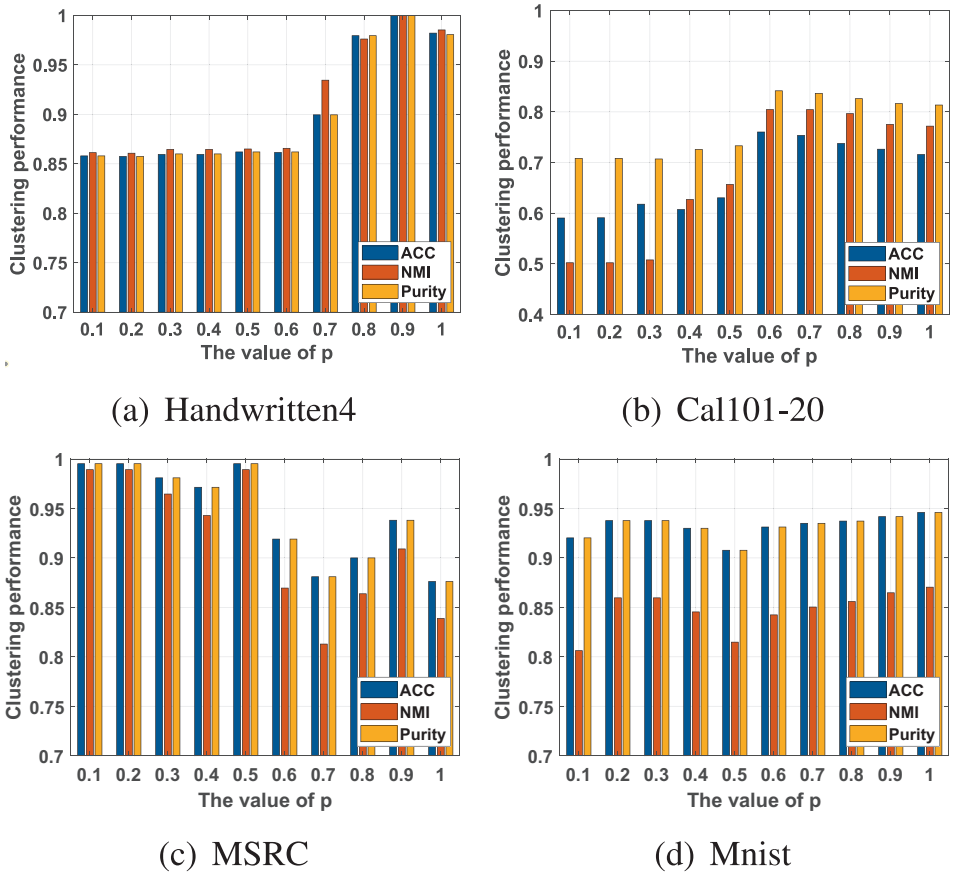


Fig. 6. The clustering performances of our method with the varying value of p on selected datasets.

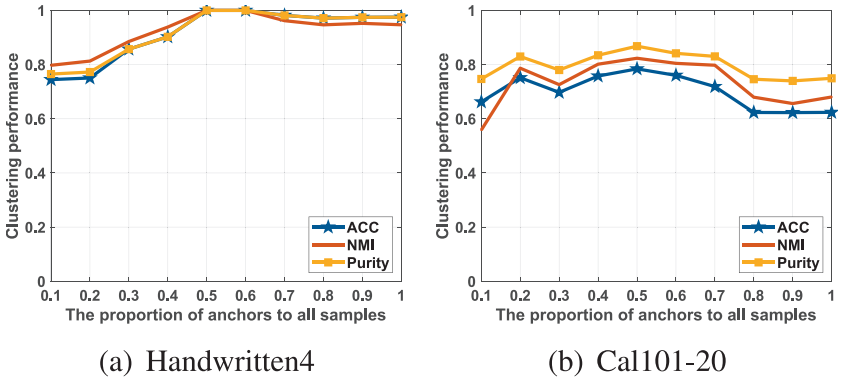


Fig. 7. The performances of our method with varying number of anchor points on Handwritten4 and Cal101-20 datasets.

Handwritten4 dataset in Fig. 9. The input graphs corresponding to all views on the dataset are shown in Fig. 9(a,c,e,g), Fig. 9(b,d,f,h) show learned graphs of each input graph and Fig. 9(i) is the view-consensus graph corresponding to Handwritten4 dataset. It can be seen that the connected components in the input graphs of all views are not clear. By employing our proposed method, we can observe that all the learned graphs have exact K-connected components($K = 10$ for Handwritten4 dataset) and the learned view-consensus graph is the clearest one. It demonstrates that our method can characterize the cluster structure using the hidden information embedded between different views. The experimental results demonstrate that our method helps make the rank of the learned view-consensus graph be similar to the target rank.

3.2.7. Ablation studies of bipartite graph

We conducted ablation experiments on four selected datasets about bipartite graph. The experimental results are shown in Fig. 10, where LCFG refers to LCBG without bipartite graph. From Fig. 10, we can see that LCBG outperforms LCFG in all evaluation metrics, so the use of bipartite graph can effectively improve the clustering performance of our algorithm. The reason may be that bipartite graph makes full use of the weights between the different views and reduces the effect of useless noise data.

3.2.8. Experimental analysis on large-scale datasets

None of the comparison algorithms not mentioned in Table 1 can handle Reuters and NUS-WIDE datasets, so the proposed

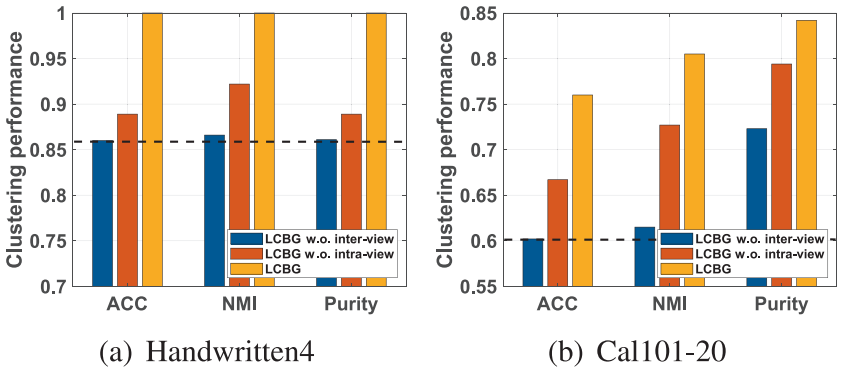


Fig. 8. Ablation Studies on Handwritten4 and Cal101-20 datasets.

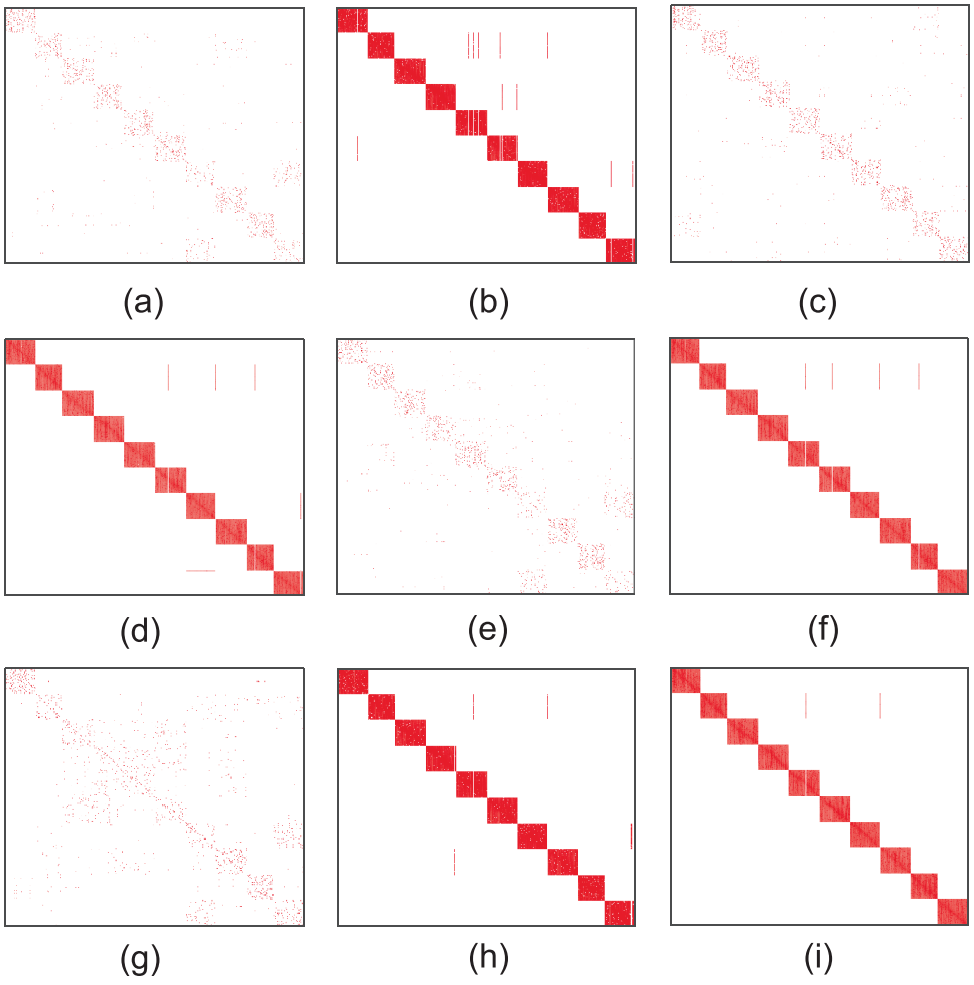


Fig. 9. The graphs visualizations on Handwritten4 dataset.

method has a great advantage in handling large-scale datasets compared to those comparison algorithms. Based on this, in order to verify the advantages of bipartite graph learning when dealing large scale datasets, we divide the Reuters dataset, setting the data scale values from 6,000 to 18,000 with the interval of 2000 (there are totally 18757 data in Reuters), then use LCFG and LCBG for clustering in order. The clustering results on Reuters datasets at different scales are shown in Table 2. In Table 2, we can see that when the data scale reaches 14,000, LCFG cannot handle this dataset. But LCBG can handle 18,757 data. So we can see the superiority of bipartite graph learning to handle the large-scale datasets.

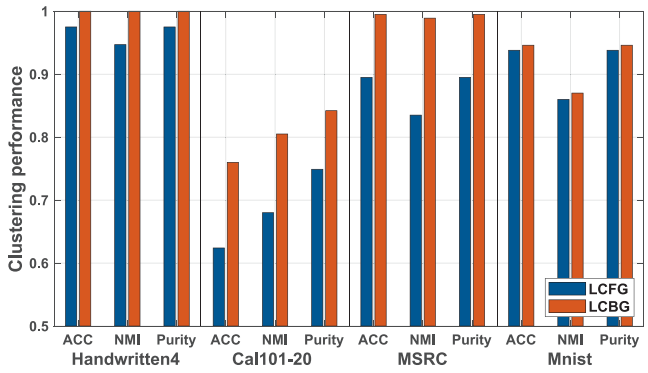


Fig. 10. Ablation Studies of bipartite graph.

Table 2
The clustering performances on different scale Reuters dataset.

Algorithm	LCFG			LCBG		
Dataset scale	ACC	NMI	Purity	ACC	NMI	Purity
6000	0.462	0.606	0.462	0.487	0.622	0.485
8000	0.596	0.622	0.596	0.595	0.603	0.599
10000	0.485	0.462	0.486	0.581	0.557	0.581
12000	0.405	0.321	0.405	0.574	0.554	0.578
14000	Out of memory			0.598	0.503	0.602
16000	Out of memory			0.630	0.485	0.630
18000	Out of memory			0.652	0.469	0.658
18757	Out of memory			0.640	0.484	0.686

4. Conclusion

Our method adaptively learns each small graph $\mathbf{S}^{(v)}$such that it well characterizes the relationship between M anchors and N samples in the v-th view, and takes both the intra-view and inter-view spatial low-rank structures of the learned bipartite graphs into account by minimizing tensor Schatten p-norm and nuclear norm, respectively. Moreover, the learned implicit view-consensus graph has exactly K connected components. Finally, an efficient algorithm, which scales linearly with the data size, is proposed to solve LCBG. Extensive experiments on real-world datasets indicate that our method outperforms the state-of-the-art competitors.

CRediT authorship contribution statement

Qian Zhou: Conceptualization, Methodology, Software, Investigation, Writing - original draft. **Haizhou Yang:** Validation, Formal analysis, Supervision, Writing - review & editing. **Quanxue Gao:** Conceptualization, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

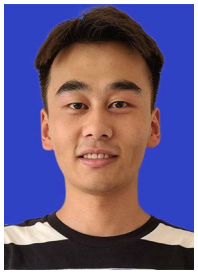
The authors would like to thank the anonymous reviewers and AE for their constructive comments and suggestions. This work was supported in part by National NSFC under Grant 62176203; in part by Natural Science Basic Research Plan in Shaanxi Province under Grant 2020JZ-19, in part by the Fundamental Research Funds for the Central Universities, the Innovation Fund of Xidian University; in part by Natural Science Foundation of Shandong Province under Grant ZR202102180986.

References

[1] W. Xia, X. Zhang, Q. Gao, X. Shu, J. Han, X. Gao, Multi-view subspace clustering by an enhanced tensor nuclear norm, IEEE Trans. Cybern. (2021), https://doi.org/10.1109/TCYB.2021.3052352.
[2] C. Zhang, H. Fu, Q. Hu, X. Cao, Y. Xie, D. Tao, D. Xu, Generalized latent multi-view subspace clustering, IEEE Trans. Pattern Anal. Mach. Intell. 42 (1) (2020) 86–99.
[3] J. Wen, Y. Xu, H. Liu, Incomplete multiview spectral clustering with adaptive graph learning, IEEE Trans. Cybern. 50 (4) (2020) 1418–1429.
[4] C. Tang, X. Liu, X. Zhu, E. Zhu, Z. Luo, L. Wang, W. Gao, CGD: multi-view clustering via cross-view graph diffusion, in: AAAI, 2020, pp. 5924–5931.
[5] S. Huang, Z. Kang, I.W. Tsang, Z. Xu, Auto-weighted multi-view clustering via kernelized graph learning, Pattern Recognit. 88 (2019) 174–184.
[6] J. Wu, Z. Lin, H. Zha, Essential tensor learning for multi-view spectral clustering, IEEE Trans. Image Process. 28 (12) (2019) 5910–5922.
[7] Z. Wan, H. Xu, Q. Gao, Multi-view clustering by joint spectral embedding and spectral rotation, Neurocomputing 462 (2021) 123–131.
[8] Q. Qiang, B. Zhang, F. Wang, F. Nie, Fast multi-view discrete clustering with anchor graphs, in: AAAI, 2021, pp. 9360–9367.
[9] S. Huang, I. Tsang, Z. Xu, J.C. Lv, Measuring diversity in graph learning: A unified framework for structured multi-view clustering, IEEE Transactions on Knowledge and Data Engineering (2021), https://doi.org/10.1109/TKDE.2021.3068461, 1–1.
[10] E. Elhamifar, R. Vidal, Sparse subspace clustering: Algorithm, theory, and applications, IEEE Trans. Pattern Anal. Mach. Intell. 35 (11) (2013) 2765–2781.
[11] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, Y. Ma, Robust recovery of subspace structures by low-rank representation, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 171–184.
[12] S. Luo, C. Zhang, W. Zhang, X. Cao, Consistent and specific multi-view subspace clustering, in: AAAI, 2018, pp. 3730–3737.
[13] D. Cai, X. Chen, Large scale spectral clustering via landmark-based sparse representation, IEEE Trans. Cybern. 45 (8) (2015) 1669–1680.
[14] F. Nie, X. Wang, C. Deng, H. Huang, Learning A structured optimal bipartite graph for co-clustering, in: NeurIPS, 2017, pp. 4129–4138.
[15] F. Nie, C. Wang, X. Li, K-multiple-means: A multiple-means clustering method with specified K clusters, in: ACM SIGKDD, 2019, pp. 959–967.
[16] Y. Li, F. Nie, H. Huang, J. Huang, Large-scale multi-view spectral clustering via bipartite graph, in: AAAI, 2015, pp. 2750–2756.
[17] Z. Kang, W. Zhou, Z. Zhao, J. Shao, M. Han, Z. Xu, Large-scale multi-view subspace clustering in linear time, in: AAAI, 2020, pp. 4412–4419.
[18] S. Huang, Z. Xu, I.W. Tsang, Z. Kang, Auto-weighted multi-view co-clustering with bipartite graphs, Inf. Sci. 512 (2020) 18–30.
[19] X. Li, H. Zhang, R. Wang, F. Nie, Multi-view clustering: A scalable and parameter-free bipartite graph fusion method, IEEE Trans. Pattern Anal. Mach. Intell. (2020), https://doi.org/10.1109/TPAMI.2020.3011148.
[20] Q. Gao, P. Zhang, W. Xia, D. Xie, X. Gao, D. Tao, Enhanced tensor rpca and its application, IEEE Trans. Pattern Anal. Mach. Intell. 43 (6) (2021) 2133–2140.
[21] K. Fan, On a theorem of weyl concerning eigenvalues of linear transformations i, Natl. Acad. Sci. USA 35 (11) (1949) 652–655.
[22] J.-F. Cai, E.J. Candès, Z. Shen, A singular value thresholding algorithm for matrix completion, SIAM J. Optimization 20 (4) (2010) 1956–1982.
[23] F. Nie, X. Wang, M.I. Jordan, H. Huang, The constrained laplacian rank algorithm for graph-based clustering, in: AAAI, 2016, pp. 1969–1976.
[24] D. Dua, C. Graff, UCI machine learning repository (2017). http://archive.ics.uci.edu/ml.
[25] L. Fei-Fei, R. Fergus, P. Perona, Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories, Comput. Vis. Image Understand. 106 (1) (2007) 59–70.
[26] J. Winn, N. Jojic, Locus: learning object classes with unsupervised segmentation, in: IEEE ICCV, 2005, pp. 756–763.
[27] L. Deng, The MNIST database of handwritten digit images for machine learning research, IEEE Signal Process. Mag. 29 (6) (2012) 141–142.
[28] C. Apté, F. Damerau, S.M. Weiss, Automated learning of decision rules for text categorization, ACM Trans. Inf. Syst. 12 (3) (1994) 233–251.
[29] T.-S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, Y. Zheng, Nus-wide: a real-world web image database from national university of singapore, in: ACM CIVR, 2009.
[30] A. Kumar, P. Rai, Co-regularized multi-view spectral clustering, in: NeurIPS, 2011, pp. 1413–1421.
[31] F. Nie, J. Li, X. Li, Self-weighted multiview clustering with multiple graphs, in: IJCAI, 2017, pp. 2564–2570.
[32] K. Zhan, C. Zhang, J. Guan, J. Wang, Graph learning for multiview clustering, IEEE Trans. Cybern. 48 (10) (2018) 2887–2895.
[33] Z. Hu, F. Nie, R. Wang, X. Li, Multi-view spectral clustering via integrating nonnegative embedding and spectral embedding, Inf. Fusion 55 (2020) 251–259.
[34] F. Nie, J. Li, X. Li, et al., Parameter-free auto-weighted multiple graph learning: A framework for multiview clustering and semi-supervised classification., in: IJCAI, 2016, pp. 1881–1887.
[35] F. Nie, G. Cai, J. Li, X. Li, Auto-weighted multi-view learning for image clustering and semi-supervised classification, IEEE Trans. Image Process. 27 (3) (2018) 1501–1511.

[36] R. Xia, Y. Pan, L. Du, J. Yin, Robust multi-view spectral clustering via low-rank and sparse decomposition, in: AAAI, 2014, pp. 2149–2155.
[37] C. Zhang, H. Fu, J. Wang, W. Li, X. Cao, Q. Hu, Tensorized multi-view subspace representation learning, Int. J. Comput. Vis. 128 (8) (2020) 2344–2361.
[38] Y. Xie, D. Tao, W. Zhang, Y. Liu, L. Zhang, Y. Qu, On unifying multi-view self-representations for clustering by tensor multi-rank minimization, Int. J. Comput. Vis. 126 (11) (2018) 1157–1179.

Qian Zhou received the B.Eng. degree in electronic Information engineering from Wuhan University of Technology, Wuhan, China, in 2021. She is currently pursuing the master's degree in electronic Information with Xidian University, Xi'an, China. And her research interests include deep learning and principal component analysis.



Haizhou Yang received the B.Eng. degree in Communication Engineering from Xidian University, Xi'an, China, in 2021. He is currently pursuing the Master degree in communication and information system in Xidian University, Xi'an, China. His research interests focus on pattern recognition and deep learning.

Quanxue Gao received the B.Eng. degree from Xi'an Highway University, Xi'an, China, in 1998, the M.S. degree from the Gansu University of Technology, Lanzhou, China, in 2001, and the Ph.D. degree from Northwestern Polytechnical University, Xi'an China, in 2005. He was an associate research with the Biometrics Center, The Hong Kong Polytechnic University, Hong Kong from 2006 to 2007. From 2015 to 2016, he was a visiting scholar with the department of computer science, The University of Texas at Arlington, Arlington USA. He is currently a professor with the School of Telecommunications Engineering, Xidian University, and also a key member of State Key Laboratory of Integrated Services Networks. He has authored around 80 technical articles in refereed journals and proceedings, including IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Transactions on Image Processing, IEEE Transactions on Neural Networks and Learning Systems, IEEE Transactions on Cybernetics, CVPR, AAAI, and IJCAI. His current research interests include pattern recognition and machine learning.