Double Graphs Regularized Multi-view Subspace Clustering

Longlong Chen¹, Yulong Wang¹,* , Youheng Liu¹, Yutao Hu¹, Libin Wang¹

¹College of Informatics, Huazhong Agricultural University, Wuhan 430062, China
wangyulong6251@gmail.com

Abstract
Recent years have witnessed a growing academic interest in multi-view subspace clustering. In this paper, we propose a novel Double Graphs Regularized Multi-view Subspace Clustering (DGRMSC) method, which aims to harness both global and local structural information of multi-view data in a unified framework. Specifically, DGRMSC firstly learns a latent representation to exploit the global complementary information of multiple views. Based on the learned latent representation, we learn a self-representation to explore its global cluster structure. Further, Double Graphs Regularization (DGR) is performed on both latent representation and self-representation to take advantage of their local manifold structures simultaneously. Then, we design an iterative algorithm to solve the optimization problem effectively. Extensive experimental results on real-world datasets demonstrate the effectiveness of the proposed method.

Introduction
With the development of information technology, increasing amounts of data are obtained from multiple views. For example, an image can be depicted by different features, such as LBP (Ojala, Pietikainen, and Maenpaa 2002), HOG (Dalal and Triggs 2005), and SIFT (Deng et al. 2009). A piece of news can be reported in different languages. Meanwhile, in the real world, the data collected is often unlabeled. Consequently, it is of significance to study multi-view clustering.

To process multi-view data, an intuitive approach is to directly use single-view clustering algorithms on the concatenated features of all views. Typically, FeatConcat employs standard spectral clustering (SPC) (Ng, Jordan, and Weiss 2001) on the concatenated features of all views. ConcatPCA firstly applies principle component analysis (PCA) (Abdi and Williams 2010) method to extract the low-dimensional representation and then employs SPC to obtain the final results. However, this naive strategy neglects the consistency and complementary information among multiple views. In the past decades, a large number of multi-view clustering methods have been proposed (Zhao, Ding, and Fu 2017; Zhang et al. 2018; Xie et al. 2019; Li et al. 2021). Among them, Multi-view Subspace Clustering (MSC) is popular since it can effectively capture the global structure and the complementary information of multi-view data. For instance, Diversity-induced Multi-view Subspace Clustering (DiMSC) (Cao et al. 2015) utilizes the Hilbert Schmidt Independence Criterion (HSIC) as a diversity term to explore the complementary of multi-view data. Considering both the consistency and the diversity of different views, Luo et al. proposed Consistent and Specific Multi-view Subspace Clustering (Luo et al. 2018). Lv et al. proposed a partition fusion strategy to enhance the robustness of multi-view clustering (Lv et al. 2021).

In real applications, each view is often insufficient and may be contaminated by noise. It is worth noting that the methods mentioned above all perform data reconstruction on the original features, which may lead to them being extremely sensitive to the quality of the original data. To deal with the problem, some multi-view clustering algorithms based on latent subspace have been proposed and achieved sufficiently good performance (Zhang et al. 2017; Li et al. 2019; Yin, Huang, and Gao 2020; Chen et al. 2020). For example, Latent Multi-view Subspace Clustering (LMSC) (Zhang et al. 2017) clusters data points with latent representation and simultaneously explores underlying complementary information from multiple views. Chen et al. proposed a novel Multi-view Clustering in Latent Embedding Space (MCLES) (Chen et al. 2020) method, which simultaneously learns the global structure and the indicator matrix in a unified framework.

However, most latent space based MSC methods do not leverage the manifold structure information within data throughout the data flow. In this paper, we propose a novel Double Graphs Regularized Multi-view Subspace Clustering (DGRMSC) method. It integrates latent representation learning, self-representation learning, and double graphs regularization into a unified framework. The overall flowchart of DGRMSC is shown in Fig. 1. First, a latent representation $Y$ is learned to exploit the global complementary information of multiple views. Then, DGRMSC learns its self-representation $Z$ to explore the global cluster structure. Meanwhile, Double Graphs Regularization (DGR) is performed to preserve the local geometric structure in each manifold throughout the data flow. In summary, the main contributions of this paper are as follows:

• We propose a novel Double Graphs Regularized Multi-
view Subspace Clustering (DGRMSC) method, which employs both global and local structural information of multi-view data simultaneously in a unified framework.

- To boost the performance of clustering, we propose a novel Double Graphs Regularization (DGR) strategy. It is capable of preserving the local geometric structures in each manifold throughout the data flow.
- We design an iterative algorithm to deal with the optimization problem. Experiments on real-world datasets validate the effectiveness of our method.

The Proposed Approach

In this section, we will describe the proposed method DGRMSC in detail, including the formulation, optimization, and computational complexity analysis.

In this paper, vectors and matrices are represented with bold lowercase and uppercase letters respectively. For a matrix $X$, $x_i$ and $x_{ij}$ represent its $i$-th column and $ij$-th element. \( \| \cdot \|_F \) and \( \| \cdot \|_2 \) denote the Frobenius norm and the nuclear norm of a matrix respectively.

### Latent Representation Learning

Given multi-view data with $n$ samples from $V$ views, we denote the $v$-th view as $X^v = [x^v_1, x^v_2, \cdots, x^v_n] \in \mathbb{R}^{d_v \times n}$, where $v \in \{1, 2, \cdots, V\}$, $d_v$ is the feature dimension of $v$-th view. Inspired by (Xu, Tao, and Xu 2013), it is assumed that each individual view is insufficient for capturing complete information while all the views together can encode complementary information. Building on this assumption, we consider all different views $\{X^v\}_{v=1}^V$ are originated from a shared latent representation $Y$. To be specific, each sample from different views can be reconstructed by their corresponding mapping model \( \{W^v\}_{v=1}^V \) with the shared latent representation $Y = [y_1, y_2, \cdots, y_n] \in \mathbb{R}^{m \times n}$, i.e. $x^v_i = W^v y_i$, $i = 1, \cdots, n$, $v = 1, \cdots, V$. Considering the noise, the latent representation learning model can be formulated as

\[
X = WY + E_L, \quad \text{s.t.} \quad WW^T = I,
\]

\[
X = \begin{bmatrix} X^1 & \cdots & X^V \end{bmatrix} \quad \text{and} \quad W = \begin{bmatrix} W^1 & \cdots & W^V \end{bmatrix},
\]

where $E_L \in \mathbb{R}^{d \times n}$ denotes the reconstruction error associated with latent representation learning, where $d = \sum_{v=1}^V d_v$. The constraint of $W$ is to prevent $Y$ from being pushed arbitrarily close to zero while scaling (Zhang et al. 2017). By learning the latent representation, complementary information from multi-view can be effectively explored.

| Notation | Description |
|----------|-------------|
| $X^v \in \mathbb{R}^{d_v \times n}$ | Matrix of the $v$-th view |
| $W^v \in \mathbb{R}^{d_v \times m}$ | Mapping matrix of the $v$-th view |
| $Y \in \mathbb{R}^{m \times n}$ | Latent representation |
| $Z \in \mathbb{R}^{m \times n}$ | Self-representation |
| $A \in \mathbb{R}^{n \times n}$ | Affinity matrix |
| $I$ | Identity matrix |
| $n, V$ | Number of samples, views |
| $d_v, m$ | Dimension of $v$-th view, latent representation |
| $d$ | Total dimension of all views |

Table 1: Key notations used throughout the paper
Self-representation Learning

After acquiring the latent representation $Y$, a naive approach to obtain the final clustering results is to perform the SPC directly. However, the restored latent representation $Y$ ignores the global cluster structure of data which plays a significant role in clustering, and hence may fail in exploring cluster relationship. Thus we learn a self-representation $Z \in \mathbb{R}^{m \times n}$ from the latent representation $Y$. This can be realized by leveraging the self-expression property of samples (Ren and Sun 2020). It is assumed that each data point can be expressed by a linear combination of other data points. Based on the learned latent representation $Y$, the self-representation learning model is written as

$$Y = YZ + E_S,$$

(2)

where $E_S \in \mathbb{R}^{m \times n}$ denotes the reconstruction error associated with self-representation learning.

Double Graphs Regularization

In spite of empirical performance in various applications, both latent space learning and self-representation learning fail to consider the manifold structures in the corresponding manifolds. To alleviate this deficiency, one may naturally hope that, if any two multi-view data samples $\{x_i^v\}_{v=1}^V$ and $\{x_j^v\}_{v=1}^V$ are close in the corresponding multi-view manifolds, the corresponding latent representations $y_i$ and $y_j$ should also close or similar too. To preserve such local manifold structure in the original space, a latent graph regularization term is designed, as follows

$$\frac{1}{2V} \sum_{v=1}^V \sum_{i,j=1}^n \|y_i - y_j\|_2^2 s_{ij}^v = \frac{1}{V} \sum_{v=1}^V \text{Tr} (YLY^T),$$

(3)

where $L^v = D^v - S^v$ is the graph Laplacian matrix of the $v$-th view, and $D^v$ is a diagonal matrix in which $D^v_{ii} = \sum_j s_{ij}^v$. $S^v$ is the similarity matrix of the $v$-th view. The $ij$-th element of $S^v$ is assigned as $s_{ij}^v = \exp \left(-\|x_i^v - x_j^v\|_2^2/2\sigma^2\right)$, where $\sigma$ controls the width of the neighborhoods.

Analogously, as shown in Fig. 1 we also want to make the self-representation manifold preserve the local geometric structure of the latent space. To this end, we devise the self-representation graph regularization termed as

$$\frac{1}{2} \sum_{i,j=1}^n \|z_i - z_j\|_2^2 s_{ij}^Y = \text{Tr} (LYZ^T),$$

(4)

where $L_Y$ is the graph Laplacian matrix of latent representation $Y$. Similarly, $s_{ij}^Y = \exp \left(-\|y_i - y_j\|_2^2/2\sigma^2\right)$.

Further, we assume that the errors corresponding to the latent representation $Y$ and the self-representation $Z$ are sample-specific. The learned self-representation $Z$ is low-rank. Incorporating Eq. (1)-(4) into a unified framework, we have

$$\min_{\Theta} \|E\|_{2,1} + \lambda \|Z\|_+ + \beta \text{Tr} (YLY^T) + \gamma \text{Tr} (ZLY^T)$$

s.t. $X = WY + E_L$, $Y = YZ + E_S$, $E = [E_L; E_S]$ and $WW^T = I$.

(5)

with $\Theta := \{W, Y, Z, E_L, E_S\}$, where $\lambda > 0$, $\beta > 0$, and $\gamma > 0$ are three non-negative regularization parameters for balancing these four terms and $L = \frac{1}{V} \sum_{v=1}^V L^v$.

Optimization

In this section, we design an algorithm based on the augmented Lagrange multiplier (ALM) with alternating direction minimizing (ADM) framework (Lin, Liu, and Sun 2011). To make objective function separable, we introduce one auxiliary variable $Q$ and the objective function can be reformulated as

$$\min_{\Theta, Q} \|E\|_{2,1} + \lambda \|Q\|_+ + \beta \text{Tr} (YLY^T) + \gamma \text{Tr} (ZLY^T)$$

s.t. $X = WY + E_L$, $Y = YZ + E_S$, $E = [E_L; E_S]$, $WW^T = I$ and $Q = Z$.

(6)

The Eq. (6) can be solved by minimizing the following problem

$$\mathcal{L}(\Theta, Q) = \|E\|_{2,1} + \lambda \|Q\|_+ + \beta \text{Tr} (YLY^T) + \gamma \text{Tr} (ZLY^T) + \Phi(A_1, X - WY - E_L) + \Phi(A_2, Y - YZ - E_S) + \Phi(A_3, Q - Z)$$

(7)

s.t. $WW^T = I$,

with the definition $\Phi(A, D) = \frac{\mu}{2} \|D\|^2_F + \langle A, D \rangle$, where $\langle \cdot, \cdot \rangle$ indicates the matrix inner product and $\mu$ is a positive penalty parameter. $A_1$, $A_2$ and $A_3$ are Lagrange multipliers. Specifically, we update each variable when fixing the others. The specific optimization process is as follows.

1. **W-subproblem**: By fixing the other variables, it is equivalent to solving the following problem

$$W^* = \arg \min_W \frac{\mu}{2} \left\| (A_1/\mu + X - E_L)^T - Y^TW^T \right\|_F^2$$

s.t. $WW^T = I$.

(8)

Based on (Huang, Nie, and Huang 2013), the optimal solution to problem (8) is $W^* = UV^T$, where $U \Sigma V^T$ are the Singular Value Decomposition (SVD) of $Y (A_1/\mu + X - E_L)^T$.

2. **Y-subproblem**: By fixing all the variables except $Y$, the optimization problem in Eq. (7) is transformed into

$$Y^* = \arg \min_Y \phi(Y, X - WY - E_L) + \beta \text{Tr}(YLY^T)$$

(9)

Taking the derivative of Eq. (9) with respect to $Y$ and setting it to zero, we have

$$L_1 Y + YR_1 = C_1$$

with $L_1 = \mu W^T W$, $R_1 = \mu (ZZ^T - Z - Z^T + I) + \beta (L + L^T)$,

$$C_1 = W^T A_1 + A_2 (Z^T - I) + \mu (W^T X + E_S - W^T E_L - E_S Z^T).$$
Similarly, the above equation can be efficiently solved by utilizing the Bartels-Stewart algorithm (Bartels and Stewart 1972).

Z-subproblem: Updating Z by fixing the other variables is equivalent to solving the following problem

\[
Z^* = \arg \min_Z \Phi(A_2, Y - YZ - E_S) + \Phi(A_3, Q - Z) + \gamma \text{Tr}(ZL_Y Z^T).
\]

Taking the derivative of Eq. (11) with respect to Z and setting it to zero, we have

\[
L_2Z + 2ZR_2 = C_2
\]

with \(L_2 = \mu(Y^T Y - I), R_2 = \gamma(L_Y + L_Y^T), C_2 = \mu(Y^T Y + Q - Y^T E_S) + A_3 + Y^T A_2.\)

(12)

Similarly, the above equation can be efficiently solved by the Bartels-Stewart algorithm (Bartels and Stewart 1972).

E-subproblem: Fix the other variables, the optimization problem in Eq. (7) can be written as

\[
E^* = \arg \min_E \frac{1}{\mu} \|E\|_{Z, 1.2} + \frac{1}{\mu} \|E - G\|_F^2,
\]

with \(G = [X - WY + A_1/\mu; Y - YZ + A_2/\mu].\) The Lemma 3.2 in (Liu et al. 2012) can be utilized to solve this problem.

Q-subproblem: Fix all the variables except Q, the optimization problem in Eq. (7) becomes

\[
Q^* = \arg \min_Q \frac{\lambda}{\mu} \|Q\|_\infty + \frac{1}{\mu} \|Q - (Z - A_3/\mu)\|_F^2.
\]

The singular value thresholding operator (Cai, Candès, and Shen 2010) can be used to solve this problem.

Updating Multipliers: The multipliers can be simply updated through

\[
\begin{align*}
\Lambda_1 &= \Lambda_1 + \mu(X - WY - E_L) \\
\Lambda_2 &= \Lambda_2 + \mu(Y - YZ - E_S) \\
\Lambda_3 &= \Lambda_3 + \mu(Q - Z).
\end{align*}
\]

(15)

Algorithm 1: DGRMSC

Input: Multi-view data \(X,\) hyperparameters \(\lambda, \beta\) and \(\gamma,\) and the dimension \(m\) of the latent representation \(Y.\)

Initialize: \(W = 0, E_L = 0, E_S = 0, Q = Z = 0, A_1 = 0, A_2 = 0, A_3 = 0, \mu = 10^{-4}, \rho = 1.2, \epsilon = 10^{-5}, \max_{\mu} = 10^6;\) Initialize \(Y\) with random values.

1: while not convergence do
2: Update \(W, X, Z, E_L, E_S\) and \(Q\) by solving subproblems 1-5;
3: Update \(A_1, A_2\) and \(A_3\) according to subproblem 6;
4: Update \(\mu = \min(\mu, \max_{\mu});\)
5: Check the convergence conditions: \(\|X - WY - E_L\|_\infty < \epsilon, \|Y - YZ - E_S\|_\infty < \epsilon\) and \(\|Q - Z\|_\infty < \epsilon.\)
6: end while

Output: \(Y, Z, W\) and \(E.\)

For clarity, the algorithm of the proposed DGRMSC method is summarized in Algorithm 1. After acquiring the self-representation \(Z,\) affinity matrix \(\mathbf{A} \in \mathbb{R}^{n \times n}\) is constructed with \(\mathbf{A} = [Z] + [Z^T].\) Finally, based on the affinity matrix \(\mathbf{A},\) standard spectral clustering algorithm is performed for the final results.

Computational Complexity Analysis

In this section, the computational complexity analysis will be presented, which consists of constructing similarity matrix and solving Eq. (7). Specifically, the similarity matrix is constructed in \(k\)-nearest neighbors fashion by heat kernel strategy (Belkin and Niyogi 2001), which costs \(O(n^2d)\). Updating \(W\) takes \(O(m^2d + d^3).\) The Bartels-Stewart algorithm (Bartels and Stewart 1972) is used to update \(Y\) and \(Z,\) which leads to \(O(m^2\nu).\) The complexity of updating \(Q\) is \(O(n^3).\) The main complexity of updating \(E\) and the multipliers is the matrix multiplication, which is \(O(mn^2 + mn)\). In summary, the overall computational complexity is \(O(n^2d) + O((m^2d + d^3 + n^3 + 2m^3 + mn^2 + m^2n)\)).

Experiments

In this section, extensive experiments are constructed to evaluate the effectiveness of the proposed method on six real-world multi-view datasets.

Datasets: In the following experiments, six benchmark datasets are adopted to evaluate the performance of our method, including MSRCV1, COIL-20, Yale, BBCSport, 100leaves, and BBC1. The detailed characteristics of the datasets are summarized in Table 2.

| Dataset   | Samples | Number of Views | Clusters |
|-----------|---------|-----------------|----------|
| MSRCV1    | 210     | 6               | 7        |
| COIL-20   | 1440    | 3               | 20       |
| Yale      | 165     | 3               | 15       |
| BBCSport  | 544     | 2               | 5        |
| 100leaves | 1600    | 3               | 100      |
| BBC       | 685     | 4               | 5        |

| Dataset | Samples | Number of Views | Clusters |
|---------|---------|-----------------|----------|

Table 2: Characteristics of the datasets

1. http://vision.ucsd.edu/content/yale-face-database
2. https://archive.ics.uci.edu/ml/datasets/One-hundred+plant+species+leaves+data+set
3. http://mlg.ucd.ie/datasets/segment.html
| Datasets | Methods | NMI | ACC | F-measure | AR | Recall | Precision |
|---------|---------|-----|-----|-----------|----|--------|-----------|

**Evaluation Metrics**: For evaluation metrics, six widely used metrics including F-measure, Precision, Recall, Normalized Mutual Information (NMI), adjusted rand index (AR), and accuracy (ACC) are adopted [Li et al., 2021]. For all metrics, higher values indicate better clustering results.

**Parameter Settings**: For the compared methods, the best results are reported through tuning the parameters as suggested in the original paper. For our method, the dimension of latent representation \( m \) is set to 100 and the parameters \( \lambda, \beta, \gamma \) are selected from \( \{0.001, 0.01, 0.1, 1, 10, 100, 1000\} \) for all datasets. The number of nearest neighbors \( k \) is set to 5 when constructing \( k \)-nearest neighbors graph. For each method, the mean performance is reported.
Table 4: Performance comparison of clustering on BBC and BBCSport. The best results are highlighted in bold. The underlined ones are our methods. The clustering results of ConcatPCA are not presented, since the features of both datasets are too sparse to run PCA.

Figure 2: Visualization of different views and different latent representations learned by LMSC, MCLES, GRMSC, and DGRMSC with t-SNE on the BBC dataset.

values and standard deviations of all evaluation metrics are reported after running 30 times with the optimal parameters.

Experimental Results

Comparison Results. The experimental results on the six datasets are presented in Table 3 and Table 4. For the results, the following observations are obtained.

- In most cases, the multi-view methods are always better than the single-view methods, which show that it is of significance to study how to utilize the multi-view information effectively.

  - Simply concatenating multiple views and then directly using the single-view methods to deal with it can not achieve effective results, and sometimes even make the results worse.

  - The proposed method significantly outperforms the other...
The reasons are as follows. Firstly, we cluster on the latent representation rather than operate directly on the original data like other methods, which can effectively reduce the impact of noise and obtain more sufficient information. Based on the learned latent representation, the self-representation property can better explore the global cluster structure so that better results can be obtained. This can be seen from the results obtained by LMSC and MCLES. However, they ignore the manifold structure information within data throughout the data flow, which leads to their results are often suboptimal. To alleviate the effect, the proposed method performs DGR on both latent representation and self-representation to take advantage of their local manifold structures simultaneously. Thus our method obtains better clustering results. We also report the performance of our method with $\gamma = 0$ (GRMSC) as an ablation comparison. It is observed that GRMSC achieves better performance on most datasets. DGRMSC further outperforms GRMSC, which validates the importance of DGR.

To be more intuitive, we visualize different views and the learned latent representations by different methods with t-Distributed Stochastic Neighbor Embedding (t-SNE) (Van der Maaten and Hinton 2008) for the dataset BBC as shown in Fig. 3. It is observed that the latent representations learned by different methods are superior to each single view. Fig. 3(g)-(h) demonstrates the learned latent representation by our GRMSC and DGRMSC can well explore the complementary information within multi-view data, compared to MCLES and LMSC. It is consistent with the performance of clustering in Table 4. Further, Fig. 3 demonstrates the affinity matrices based clustering learned by LMSC, MCLES, GRMSC, and DGRMSC. Obviously, our methods GRMSC and DGRMSC can obtain better results.

Sensitivity Analysis

Specifically, we take the 100leaves dataset as an example to conduct parameter sensitivity analysis and the results are shown in Fig. 4. It can be seen that our method is relatively parameter insensitive and can obtain good results in a large range. Meanwhile, we also analyze the effect of dimensionality $m$ of the latent representation and $k$ nearest neighbors value on the results, as shown in Fig. 5. It can be seen from Fig. 5(a) that our method can achieve good results at lower dimensionality, which further illustrates the effectiveness of latent representation learning. Fig. 5(b) shows that our method can achieve a better result when $k = 5$. Other datasets have similar properties, so for simplicity, we set $m = 100$ and $k = 5$ in all experiments, respectively.
Conclusion

In this paper, a novel Double Graph Regularized Latent Multi-view Subspace Clustering (DGRMSC) method is proposed, which exploits both global and local structural information of multi-view data in a unified framework. With double graphs regularized strategy, the local geometric information in different manifolds throughout the data flow can be well preserved. Accordingly, the learned multi-view latent representation and the affinity matrix based clustering are well improved simultaneously. Further, we design an iterative algorithm to solve the optimization problem effectively. The experimental results further validate the effectiveness of the proposed method.

References

Ojala, T.; Pietikainen, M.; and Maenpaa, T. 2002. Multiresolution gray-scale and rotation invariant texture classification. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(7): 971–987.

Dalal, N.; and Triggs, B. 2005. Histograms of oriented gradients for human detection. In CVPR, volume 1, 886–893.

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. ImageNet: A large-scale hierarchical image database. In CVPR, 248–255.

Ng, A.; Jordan, M.; and Weiss, Y. 2001. On spectral clustering: Analysis and an algorithm. In NeurIPS, 849–856.

Abdi, H.; and Williams, L. J. 2010. Principal component analysis. Wiley interdisciplinary reviews: computational statistics, 2(4): 433–459.

Cao, X.; Zhang, C.; Fu, H.; Liu, S.; and Zhang, H. 2015. Diversity-induced multi-view subspace clustering. In CVPR, 586–594.

Luo, S.; Zhang, C.; Zhang, W.; and Cao, X. 2018. Consistent and specific multi-view subspace clustering. In AAAI, volume 32, 3730–3737.

Lv, J.; Kang, Z.; Wang, B.; Ji, L.; and Xu, Z. 2021. Multi-view subspace clustering via partition fusion. Information Sciences, 560: 410–423.

Zhang, C.; Hu, Q.; Fu, H.; Zhu, P.; and Cao, X. 2017. Latent multi-view subspace clustering. In CVPR, 4279–4287.

Li, R.; Zhang, C.; Hu, Q.; Zhu, P.; and Wang, Z. 2019. Flexible multi-view representation learning for subspace clustering. In IJCAI, 2916–2922.

Yin, M.; Huang, W.; and Gao, J. 2020. Shared generative latent representation learning for multi-view clustering. In AAAI, volume 34, 6688–6695.

Chen, M.-S.; Huang, L.; Wang, C.-D.; and Huang, D. 2020. Multi-view clustering in latent embedding space. In AAAI, volume 34, 3513–3520.

Xu, C.; Tao, D.; and Xu, C. 2015. Multi-view intact space learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(12): 2531–2544.

Lin, Z.; Liu, R.; and Su, Z. 2011. Linearized alternating direction method with adaptive penalty for low-rank representation. In NeurIPS, volume 24, 612–620.

Huang, J.; Nie, F.; and Huang, H. 2013. Spectral rotation versus k-means in spectral clustering. In AAAI, volume 27, 431–437.

Bartels, R. H.; and Stewart, G. W. 1972. Solution of the matrix equation AX+ XB= C. Communications of the ACM, 15(9): 820–826.

Liu, G.; Lin, Z.; Yan, S.; Sun, J.; Yu, Y.; and Ma, Y. 2012. Robust recovery of subspace structures by low-rank representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(1): 171–184.

Cai, J.-F.; Candès, E. J.; and Shen, Z. 2010. A singular value thresholding algorithm for matrix completion. SIAM Journal on Optimization, 20(4): 1956–1982.

Belkin, M.; and Niyogi, P. 2001. Laplacian eigenmaps and spectral techniques for embedding and clustering. In NeurIPS, volume 14, 585–591.

Xu, J.; Han, J.; and Nie, F. 2016. Discriminatively embedded k-means for multi-view clustering. In CVPR, 5356–5364.

Xia, R.; Pan, Y.; Du, L.; and Yin, J. 2014. Robust multi-view spectral clustering via low-rank and sparse decomposition. In AAAI, volume 28, 2149–2155.

Kumar, A.; Rai, P.; and Daume, H. 2011. Co-regularized multi-view spectral clustering. In NeurIPS, volume 24, 1413–1421.

Tan, C.; Liu, X.; Zhu, X.; Zhu, E.; Luo, Z.; Wang, L.; and Gao, W. 2020. CGD: Multi-view clustering via cross-view graph diffusion. In AAAI, volume 34, 5924–5931.

Kang, Z.; Shi, G.; Huang, S.; Chen, W.; Pu, X.; Zhou, J. T.; and Xu, Z. 2020. Multi-graph fusion for multi-view spectral clustering. Knowledge-Based Systems, 189(C): 105102.

Zhong, G.; and Pun, C.-M. 2021. Improved Normalized Cut for Multi-view Clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Van der Maaten, L.; and Hinton, G. 2008. Visualizing data using t-SNE. Journal of Machine Learning Research, 9(11): 2579–2605.

Li, Z.; Tang, C.; Liu, X.; Zheng, X.; Zhang, W.; and Zhu, E. 2021. Consensus graph learning for multi-view clustering. IEEE Transactions on Multimedia, 24: 2461–2472.

Zhao, H.; Ding, Z.; and Fu, Y. 2017. Multi-view clustering via deep matrix factorization. In AAAI, 2921—2927.

Zhang, C.; Fu, H.; Hu, Q.; Cao, X.; Xie, Y.; Tao, D.; and Xu, D. 2018. Generalized latent multi-view subspace clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(1): 86–99.

Xie, D.; Zhang, X.; Gao, Q.; Han, J.; Xiao, S.; and Gao, X. 2019. Multiview clustering by joint latent representation and similarity learning. IEEE Transactions on Cybernetics, 50(11): 4848–4854.

Cai, J.-F.; Candès, E. J.; and Shen, Z. 2010. A singular value thresholding algorithm for matrix completion. SIAM Journal on Optimization, 20(4): 1956–1982.

Belkin, M.; and Niyogi, P. 2001. Laplacian eigenmaps and spectral techniques for embedding and clustering. In NeurIPS, volume 14, 585–591.

Xu, J.; Han, J.; and Nie, F. 2016. Discriminatively embedded k-means for multi-view clustering. In CVPR, 5356–5364.

Xia, R.; Pan, Y.; Du, L.; and Yin, J. 2014. Robust multi-view spectral clustering via low-rank and sparse decomposition. In AAAI, volume 28, 2149–2155.