GRAPH LEARNING FOR CLUSTERING MULTI-VIEW DATA

Sravanthi Gurugubelli and Sundeep Prabhakar Chepuri

Indian Institute of Science, Bangalore, India

ABSTRACT

In this paper, we focus on graph learning from multi-view data of shared entities for clustering. We can explain interactions between the entities in multi-view data using a multi-layer graph with a common vertex set representing the shared entities. The edges on different layers capture the relationships of the entities. Assuming a smoothness data model, we estimate the graph Laplacian matrices of the individual graph layers by constraining their ranks to obtain multi-component graph layers for clustering. We also learn low-dimensional node embeddings, common to all the views, that assimilate the complementary information present in the views. We propose an efficient solver based on alternating minimization to solve the proposed multi-layer multi-component graph learning problem. Numerical experiments on synthetic and real datasets demonstrate that the proposed algorithm outperforms state-of-the-art multi-view clustering techniques.

Index Terms— Clustering, graph learning, multi-layer graphs, multi-view data, representation learning.

1. INTRODUCTION

We often observe complementary information about a common source from multiple modalities or through different feature subsets in data analysis. Various aspects of interactions underlying multi-view datasets can be represented using multi-layer graphs, in which each graph layer represents a different view. All the graph layers in a multi-layer graph share the same set of nodes representing the shared entities. However, the edges on distinct layers represent the interactions between the entities in that view [1]. For example, in a social network graph, we can think of people’s interests like favorite sport, affiliation, and hobbies as multiple views of people in the network [2]. Each of these views can be represented by a graph giving rise to a multi-layer graph with each layer consisting of the same set of nodes representing people. Also, one can have data from multiple modalities. For example, in brain imaging and analysis, data from different modalities, like fMRI, sMRI and DTI [3], may be considered as signals on a multi-layer graph. The edges in different graph layers represent interactions between the same set of brain regions observed in different modalities [4,5].

The underlying graph structure is leveraged to solve several signal processing and machine learning tasks like denoising, clustering, and dimensionality reduction. However, the underlying graph may not always be readily available. Learning graphs from data is an ill-posed problem. Although nearest neighbor graphs, correlation graphs, or Gaussian similarity kernels learned from data are simple and commonly used choices, they are sensitive to noise or missing samples. Hence, several graph learning solutions that carefully model the data (e.g., using a smoothness or probabilistic graphical model) and incorporate prior information (such as sparsity, product, or multi-component structure) about the resulting topology have been proposed [6–12]. Although the existing graph learning methods can be directly used to learn the individual graph layers of a multi-layer graph, the information contained in all the views cannot be captured by any of the individual graph layers. Therefore, we propose multi-layer graph learning from multi-view data in this paper.

Low-dimensional embeddings of the nodes of a graph that encode the structural information about the graph are useful in graph-based clustering and other graph analysis tasks [13,14]. A common technique to compute these low-dimensional node embeddings is to perform a partial eigenvalue decomposition of the graph underlying the data. When dealing with multi-layer graphs, it is useful to find low-dimensional embeddings of the nodes common to all the views that best capture the complementary information that might be available in different views. Assuming that the individual graph layers of a multi-layer graph are available, there are subspace learning methods that learn a common subspace by merging the multiple subspaces computed from the individual graph layers [15,16]. In the context of multi-view canonical correlation analysis (MCCA), [17] learns a common subspace that is smooth on a known graph from multi-view data.

Instead, in this paper, we jointly learn a multi-layer graph and common low-dimensional node embeddings from multi-view data for clustering. Specifically, given M views of data, we estimate a graph with M layers having K components. The respective views of data are smooth on the graph topology of individual layers that we learn. The graph is restricted to a K-component graph to cluster data in K groups by imposing rank constraints on the graph Laplacian matrices of the individual graph layers. We propose an efficient solver based on alternating minimization, each subproblem of which is solved optimally. We evaluate the performance of the proposed method for clustering on synthetic and real-world datasets and compare them with state-of-the-art techniques for multi-view clustering [15,17]. The results show that our algorithm outperforms single-view clustering and the state-of-the-art methods.

2. MULTI-LAYER GRAPHS

Consider a $M$-layer weighted, and undirected graph $G = \{V, W_m\}, m = 1, 2, \ldots, M$, where $V = \{v_1, \ldots, v_N\}$ denotes the common vertex (or node) set with $N = |V|$ nodes and $W_m \in \mathbb{R}^{N \times N}$ denotes the weighted adjacency matrix of the $m$th graph layer $G_m$. The $(i, j)$th element of $W_m$ contains a positive edge weight if two nodes $v_i$ and $v_j$ are connected in $G_m$ and is zero otherwise. The adjacency matrix $W_m$ is symmetric as $G_m$ is undirected. The degree of the nodes in $G_m$ is defined as $\deg_m$. The combinatorial graph Laplacian matrix $L_m \in \mathbb{R}^{N \times N}$ for $G_m$ is defined as $L_m = \text{diag}[W_m] - W_m$. By construction, $L_m$ is a symmetric, positive semidefinite matrix, and has a zero row sum. The space of all the valid combinatorial graph Laplacian matrices of

This work is supported in part by the SERB SRG/2019/000619 grant and the Pratiksha Trust Fellowship.
The number of connected components in a graph is given by the multiplicity of the zero eigenvalue of its graph Laplacian matrix \([13]\). Thus the rank of the graph Laplacian matrix of a \(K\)-component graph with \(N\) nodes is \(N - K\). The eigenvectors of the graph Laplacian matrix corresponding to the \(K\) zero eigenvalues preserve the nodal connectivity information and hence are used as low-dimensional node embeddings.

A signal (or dataset) indexed using the nodes of a graph is referred to as a graph signal (or data). Consider \(M\) datasets \(X_m \in \mathbb{R}^{N \times D_m}\), \(m = 1, 2, \ldots, M\), obtained from \(M \geq 2\) views of a common source. We assume that the \(m\)th row of \(X_m\), that contains \(D_m\) features of the \(m\)th entity resides on the \(m\)th node of \(G_m\). That is, we interpret the \(M\)-view dataset as a graph signal defined on the multi-layer graph \(G\) with each layer representing a different view of the dataset.

We use smoothness to measure how well a signal matches the underlying graph. A signal is said to be smooth over a graph if the signal values residing on adjacent nodes having large edge weights are similar. For the signal \(X_m \in \mathbb{R}^{N \times D_m}\) residing on \(G_m\) with the graph Laplacian matrix \(L_m\), the smoothness is measured using the total variation with respect to the graph \(G_m\), and is given by \(\text{tr}(X_m^T L_m X_m) = \text{tr}(L_m S_m)\). Here, \(S_m = X_m X_m^T\) is the sample data covariance matrix.

### 3. PROBLEM STATEMENT

In this work, we propose a rank-constrained multi-layer graph learning technique for clustering nodes given observations defined on a multi-layered graph. Specifically, we estimate the graph Laplacian matrices \(\{L_m\}_{m=1}^M\) that best explain the \(M\)-view dataset \(\{X_m\}_{m=1}^M\) by assuming that each view of the dataset is smooth on the graph of that layer. By constraining the rank of \(L_m\), \(m = 1, 2, \ldots, M\) to \(R = N - K\), we get \(K\)-component graph layers. Since the entities in the multi-view dataset correspond to the same nodes in all the views, there exists a common low-dimensional representation of the nodes. To compute the common low-dimensional node embeddings, we propose to use the following joint diagonalization of \(\{L_m\}_{m=1}^M\):

\[
L_m = U \text{diag}(\lambda_m) U^T = \begin{bmatrix} 0 & \star \end{bmatrix} \begin{bmatrix} 0 \cr \star \end{bmatrix} = \begin{bmatrix} 0 \cr \star \end{bmatrix} \begin{bmatrix} 0^T \cr \star^T \end{bmatrix},
\]

for \(m = 1, 2, \ldots, M\), where the columns of \(U\) contain the joint eigenvectors (common factors) shared by all the graph layers and the vector \(\lambda_m = [\lambda_1(L_m), \lambda_2(L_m), \ldots, \lambda_N(L_m)]^T \in \mathbb{R}_+^N\) contains the eigenvalues of \(L_m\). Here, \(\lambda_i(L_m)\) is the \(i\)th eigenvalue of \(L_m\), where we assume that \(0 = \lambda_1(L_m) \leq \cdots \leq \lambda_N(L_m)\). The common eigenvectors corresponding to the \(K\) zero eigenvalues of \(\{L_m\}_{m=1}^M\) are collected in \(Q \in \mathbb{R}^{N \times K}\).

To estimate the graph Laplacian matrices corresponding to the graphs in each layer, we propose the following rank-constrained multi-layer graph learning (RMLG) optimization problem:

\[
\begin{align*}
\text{minimize} & \quad \sum_{m=1}^M \text{tr}(L_m S_m) + \alpha_m \|L_m\|_F^2 \\
\text{subject to} & \quad L_m \in \mathcal{L}, \quad \text{tr}(L_m) = N, \quad \text{rank}(L_m) = R, \quad m = 1, \cdots, M,
\end{align*}
\]

where recall that the set \(\mathcal{L}\), defined in (1), is the set of all the valid combinatorial graph Laplacian matrices. The first term in the objective function promotes smoothness of \(X_m\) with respect to the graph corresponding to the \(m\)th layer. The second term in the objective function with the tuning parameter \(\alpha_m > 0\) allows us to control the sparsity (i.e., the number of nonzero entries) of \(W_m\).

4. PROPOSED SOLVER

In this section, we solve the problem in (5) by alternatingly minimizing it with respect to \(\{L_m\}_{m=1}^M\) and \(Q\), while keeping the other variable fixed.

**4.1. Update of \(\{L_m\}_{m=1}^M\)**

Given \(Q\), the problem in (5) simplifies to the following convex optimization problem:

\[
\begin{align*}
\text{minimize} & \quad \sum_{m=1}^M \text{tr}(L_m R_m) + \alpha_m \|L_m\|_F^2 \\
\text{subject to} & \quad L_m \in \mathcal{L}, \quad \text{tr}(L_m) = N, \quad m = 1, \cdots, M,
\end{align*}
\]

where we have introduced the \(N \times N\) matrix \(R_m = S_m + \beta_m Q Q^T\). It can be seen that the parameter \(\beta_m\) regularizes the data with the eigenvectors of the graph Laplacian. Since \(L_m\) is a symmetric matrix, this quadratic program can be solved very efficiently as described next.

Let us define a duplication matrix \(D \in \mathbb{R}^{N^2 \times V}\) with \(V = N(N + 1)/2\) as \(D^T = \sum_{i,j} \delta_{ij} \text{vec}^T(\Theta_{ij})\), where the vector \(\delta_{ij} \in \mathbb{R}^{V}\) has 1 at position \((j - 1)N + i - \frac{1}{2}j(j - 1)\) and zero elsewhere. The matrix \(\Theta_{ij} \in \mathbb{R}^{N \times N}\) has \(-1\) at the positions \((i,j)\) and \((j,i)\) 1 at \((i,i)\), and zero elsewhere. Let us collect the absolute
values of the $V$ nonduplicated entries of $L_m$ in $I_m \in \mathbb{R}^V$. Then we have $\text{vec}(L_m) = DL_m$, \quad m = 1, 2, \ldots, M$, where $\text{vec}(\cdot)$ denotes the matrix vectorization operator.

Using this transformation, we express the two equality constraints in (6), namely, $tr(L_m) = \text{vec}(I)^T \text{vec}(L_m) = N$ and $L_l = (I^T \otimes I)\text{vec}(L_m) = 0$ as $C \text{diag}(L_m) = d$, where $C = [D^T \text{vec}(I), D^T (I \otimes I)]^T \in \mathbb{R}^{N + 1 \times V}$ and $d = [N, 0]^T \in \mathbb{R}^{N+1}$. Here, $\otimes$ denotes the Kronecker product. We express the first term in the objective function of (6) as $tr(L_m^2) = \text{vec}^T(R_m)\text{vec}(L_m) = r_m^T L_m$. The second term in the objective function of (6) simplifies to $\alpha_m ||L_m||_F^2 = \alpha_m \text{vec}(L_m)^T \text{vec}(L_m) = \frac{1}{4} r_m^T \text{diag}(p_m) L_m$, where we have used the fact that $D^T D$ is a (positive definite) diagonal matrix to obtain $\text{diag}(p_m) = 2\alpha_m D^T D$. Now, we can simplify (6) as

$$
\begin{align*}
\text{minimize} & \quad \sum_{m=1}^{M} \left( \frac{1}{2} I_m^T \text{diag}(p_m) L_m + r_m^T L_m \right) \quad \text{subject to} \quad C I_m = d, \quad I_m \succeq 0, \quad m = 1, \ldots, M. \\
\end{align*}
$$

This is a special form of a convex quadratic program in which the matrix associated with the quadratic term is diagonal. It is computationally efficient and equivalent to solve for each $I_m$, \quad $m = 1, 2, \ldots, M$ separately as this convex program is separable in these variables.

The Lagrangian function for the problem (7) associated to the variable $I_m$ is given by

$$
\mathcal{J}(I_m, \lambda_m, \mu_m) = \frac{1}{2} I_m^T \text{diag}(p_m) L_m + r_m^T L_m + \mu_m^T (d - C I_m) - \lambda_m^T L_m,
$$

for $m = 1, 2, \ldots, M$, where $\mu_m \in \mathbb{R}^{N+1}$ and $\lambda_m \in \mathbb{R}^V$ are the Lagrange multipliers corresponding to the equality and inequality constraints, respectively. The Karush-Kuhn-Tucker (KKT) conditions are given by

$$
\text{diag}(p_m) I_m^* + r_m = C^T \mu^* - \lambda_m = 0, \quad C I_m^* = d, \quad I_m^* \succeq 0, \quad \lambda_m^* \succeq 0, \quad \lambda_m^* \odot I_m^* = 0,
$$

where $\odot$ denotes the elementwise Hadamard product. Eliminating the variable $\lambda_m^*$ and solving for $I_m^*$, we get $I_m^*(\mu^*) = \left\{ \text{diag}^{-1}(p_m)[C^T \mu_m^* - r_m] \right\}_+$, where $\{\cdot\}_+$ denotes the elementwise projection onto the nonnegative orthant. Using $I_m^*(\mu^*)$ in the second KKT condition, we can compute $\mu^*_m$ iteratively as

$$
\begin{align*}
I_m^{(k)} &= \left\{ \text{diag}^{-1}(p_m)[C^T \mu_m^{(k)} - r_m] \right\}_+, \\
\mu_m^{(k+1)} &= \mu_m^{(k)} - \rho[C I_m^{(k)} - d],
\end{align*}
$$

where $\rho > 0$ is the step size. We initialize the iterations with $\mu^{(0)}$. The computational complexity of these iterations is dominated by the matrix-vector multiplication. Since $C$ is sparse with $N(N+1)$ non-zero entries, the above iterative procedure approximately costs order $N^2$ flops. For $M$ variables, the total computation cost is approximately order $MN^2$ flops.

### 4.2. Update of $Q$

Given $\{L_m\}_{m=1}^M$, (5) reduces to the following eigenvalue problem

$$
\begin{align*}
\text{minimize} & \quad tr(Q^T L Q) \\
\text{subject to} & \quad Q^T Q = I_N \quad \text{and} \quad L = \sum_{m=1}^{M} \beta_m L_m. \\
\end{align*}
$$

The optimal $Q$ is given by the $K$ eigenvectors corresponding to the $K$ smallest eigenvalues of $L = \sum_{m=1}^{M} \beta_m L_m$. Computing this partial eigendecomposition approximately costs $KN^2$ flops [19].

The complete alternating minimization procedure is summarized as Algorithm 1. The proposed algorithm approximately costs order $(M+K)N^2$ flops per iteration.

#### 5. NUMERICAL EXPERIMENTS

In this section, we evaluate our framework in terms of clustering performance on a synthetically generated dataset and two real-world datasets. We use normalized mutual information (NMI) as a measure to assess the clustering performance. We compare the clustering performances of our proposed method with the following existing approaches. The first is the graph learning method based on signal representation method (GL-SigRep) in [9]. We learn the graph layers independently for each of the views available using GL-SigRep. Spectral clustering is then performed to find the clusters using the eigenvectors of the estimated graph Laplacian matrices. Next, we compare our method with graph-regularized dual multi-view CCA (GD-MCCA) [17], which learns a low-dimensional representation of the data common to all the views by minimizing the distance between the canonical variables and the common low-dimensional representations while leveraging a graph structure of the common source. Since this method requires a common graph, we compute the results by taking each of the affinity matrices $D_m^T X_m X_m^T$ as the common graph as in [17]. Here, $D_m$ is the degree matrix of the similarity graph $X_m^T X_m$. We finally consider the common graph giving the best results for comparison. We also compare our method with multi-view spectral clustering methods SC-GED [13] and SC-ML [16], which obtain the common node

---

**Algorithm 1** Rank-constrained multi-layer graph learning

1: function RMGL($\{p_m, r_m\}_{m=1}^M$, C, D, d, K, Tol, $\rho$, MaxIter)
2: Initialize $k \leftarrow 0$
3: while $k < \text{MaxIter}$ do
4: for $m = 1$ to $M$ do
5: Initialize $\mu_m \leftarrow 0$
6: while $||C I_m - d||_2 < \text{Tol}$ do
7: $I_m \leftarrow \{\text{diag}^{-1}(p_m)[C^T \mu_m - r_m]\}_+$
8: $\mu_m \leftarrow \mu_m - \rho[C I_m - d]$
9: $L_m \leftarrow \text{mat}(D I_m)$
10: $L \leftarrow \sum_{m=1}^{M} \beta_m L_m$
11: $Q \leftarrow \text{eigs}(L, K)$
12: $k \leftarrow k + 1$
13: return $\{L_m\}_{m=1}^M$ and $Q$
embeddings by computing a joint spectrum via a generalized eigen-

decomposition and merging the node embeddings from multiple

graph layers, respectively. We give the independent graph layers

estimated from as input to SC-GED and SC-ML to find the

common node embeddings and perform spectral clustering.

To generate synthetic data, we construct a 3-layer graph, as

shown in Fig. 1(a). Each of the graph layers consists of \( N = 100 \)

nodes with an equal number of samples from four classes, indicated

by different colors. We generate data \( \mathbf{X}_m \in \mathbb{R}^{100 \times 15}, m = 1, 2, 3 \)

by taking the eigenvectors corresponding to the 15 smallest eigen-

values of the individual graph Laplacian matrices and corrupting

it with zero-mean additive white Gaussian noise of variance 0.01.

The graph layers are so constructed that they carry complementary

information that none of the individual graphs provide. The denser

within-class connections than the inter-class connections present in

all the views convey that the nodes with different colors belong to

different classes. The estimated graphs layers with \( K \)-components

from the proposed method \( \text{RMGL} \) is shown in Fig. 1(b). From

Fig. 1(c), we can see that the embeddings obtained using the three

estimated graph layers using GL-SigRep fail to incorporate this

information. In contrast, RMGL assigns the same node representations

to the points belonging to one class and hence are better clustered

by our proposed method than the representations obtained using

SC-GED, SC-ML, and GD-MCCA.

The real datasets that we test on are UCI hand-written digits [20]

and COIL-20 [21] datasets. The UCI dataset consists of 200 instan-
tces of images of the 10 digits 0 to 9. We consider six different

views of this data related to the local binary pattern, histogram of ori-
gines of each object captured at different angles. We construct three

views of this data as in [17] with \( N = 2000, D_1 = 216, D_2 = 76, D_3 = 64, D_4 = 6, D_5 = 240, \)

and \( D_6 = 47 \). The COIL-20 dataset is an image dataset consisting of 20 classes of objects with 72 im-
ages of each object captured at different angles. We construct three

views of this data related to the local binary pattern, histogram of ori-
ginated gradients, and pixel values of resized images with \( N = 1440, D_1 = 59, D_2 = 36, \)

and \( D_3 = 1024 \).

With the different views of data taken as input, we estimate

\( \{ \mathbf{L}_m \}_{m=1}^M \) and \( \mathbf{Q} \). The rows of \( \mathbf{Q} \) are then given as input to the

\( k \)-means algorithm for clustering. We also perform spectral clus-
tering on \( \{ \mathbf{L}_m \}_{m=1}^M \). The accuracy of the clusters obtained using these can be treated as a measure of the correctness of the estimated graph. The parameters involved in the different methods con-

sidered are chosen to obtain the best possible NMI. The values of

\( \{ \beta_m \}_{m=1}^M \) for RMGL can be chosen according to the importance of

the available views by choosing a higher value of \( \beta_m \) for a view \( m \)

that is more informative, such that the subspace \( \mathbf{Q} \) is learnt without getting corrupted by the insignificant (noisy/missing) information present in some view. Our model gives the best results for the following choice of parameters: \( \alpha_m = 100, m = 1, \ldots, 6, \beta_1 = \beta_2 = \beta_3 = 2, \) and \( \beta_4 = 20, \beta_5 = 7, \beta_6 = 1 \) for the UCI dataset. For the COIL-20 dataset, we use \( \alpha_m = 100, m = 1, 2, 3, \) and \( \beta_1 = 12, \beta_2 = 2, \beta_3 = 15 \). The NMI scores of the different methods in Fig. 2 suggest that using \( \mathbf{Q} \) from \( \text{RMGL} \) as an input to the \( k \)-means algorithm gives the best clustering accuracy when compared to the other considered methods (including performing spectral clustering on the estimated graph layers either from RMGL or GL-SigRep). Fig. 3 shows the node embeddings corresponding to the 2000 samples from the ten classes of data in the UCI dataset set. We show the two-dimensional node representations obtained from the last two columns of the common \( K \)-dimensional subspaces computed using RMGL, SC-GED, SC-ML, GD-MCCA. The embed-
ddings obtained from RMGL show better separability even in the two-
dimensional space.

6. CONCLUSIONS

We developed a framework for multi-view clustering, where we si-
multaneously learn a multi-layer graph with a common vertex set and a low-dimensional representation for the nodes common to all the views. We presented an efficient solver based on an alternating minimization procedure to solve the proposed non-convex optimization problem. We demonstrated via numerical experiments on synthetic and real datasets that the proposed method performs better than the existing multi-view clustering methods based on CCA or merging the node embeddings of the individual graph layers.
7. REFERENCES

[1] Y. Yang and H. Wang, “Multi-view clustering: A survey,” Big Data Mining and Analytics, vol. 1, no. 2, pp. 83–107, June 2018.

[2] D. Greene and P. Cunningham, “Producing a unified graph representation from multiple social network views,” in Proc. of the 5th annual ACM web science conf., Paris, France, May 2013.

[3] V. D. Calhoun and J. Sui, “Multimodal fusion of brain imaging data: A key to finding the missing link(s) in complex mental illness,” Biological psychiatry. Cognitive neuroscience and neuroimaging, vol. 1, no. 3, p. 230—244, May 2016.

[4] F. Battiston, V. Nicosia, M. Chavez, and V. Latora, “Multilayer motif analysis of brain networks,” Chaos: An Interdisciplinary J. of Nonlinear Science, vol. 27, no. 4, Mar. 2017.

[5] M. Vaiana and S. F. Muldoon, “Multilayer brain networks,” J. of Nonlinear Science, vol. 30, no. 5, pp. 2147–2169, Jan. 2018.

[6] D. Xiaowen, T. Dorina, F. Pascal, and V. Pierre, “Learning laplacian matrix in smooth graph signal representations,” IEEE Trans. Signal Process., vol. 64, no. 23, pp. 6160–6173, Dec. 2016.

[7] D. Xiaowen, T. Dorina, R. Michael, and F. Pascal, “Connecting the dots: Identifying network structure via graph signal processing,” IEEE Signal Process. Mag., vol. 36, no. 3, pp. 44–63, May 2019.

[8] M. Gonzalo, S. Santiago, M. Antonio G, and R. Alejandro, “Connecting the dots: Identifying network structure via graph signal processing,” IEEE Signal Process. Mag., vol. 36, no. 3, pp. 16–43, Dec. 2019.

[9] M. Coutino, S. P. Chepuri, and G. Leus, “Sparsest network support estimation: A submodular approach,” in Proc. of IEEE Data Science Workshop (DSW), Lausanne, Switzerland, June 2018.

[10] S. P. Chepuri, M. Coutino, A. G. Marques, and G. Leus, “Distributed analytical graph identification,” in Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Process. (ICASSP).

[11] S. K. Kadambi and S. P. Chepuri, “Learning product graphs from multidomain signals,” in Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Process. (ICASSP), Barcelona, Spain, May 2020.

[12] S. P. Chepuri, S. Liu, G. Leus, and A. O. Hero, “Learning sparse graphs under smoothness prior,” in Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Process. (ICASSP), New Orleans, USA, Mar. 2017.

[13] W. L. Hamilton, R. Ying, and J. Leskovec, “Representation learning on graphs: Methods and applications,” arXiv preprint arXiv:1709.05584, Sep. 2017.

[14] U. Von Luxburg, “A tutorial on spectral clustering,” Stats. and comput., vol. 17, no. 4, pp. 395–416, Aug. 2007.

[15] D. Xiaowen, F. Pascal, V. Pierre, and N. Nikolai, “Clustering with multi-layer graphs: A spectral perspective,” IEEE Trans. Signal Process., vol. 60, no. 11, pp. 5820–5831, Nov. 2012.

[16] D. Xiaowen, F. Pascal, V. Pierre, , and N. Nikolai, “Clustering on multi-layer graphs via subspace analysis on Grassmann manifolds,” IEEE Trans. Signal Process., vol. 62, no. 4, pp. 905–918, Feb. 2014.

[17] C. Jia, W. Gang, and G. B. Giannakis, “Graph multiview canonical correlation analysis,” IEEE Trans. Signal Process., vol. 62, no. 4, pp. 905–918, June 2014.

[18] T. Tao, Topics in random matrix theory. American Mathematical Soc., 2012.

[19] L. K. Saul and S. T. Roweis, “An introduction to locally linear embedding,” unpublished. Available at: http://www.nyu.edu/˜roweis/tl/locally_linear_embedding.html, Jan. 2001.

[20] R. P. Duin, “UCI multiple features dataset of handwritten digits extracted from dutch utility maps,” 1998 (accessed Aug., 2020). [Online]. Available: [http://archive.ics.uci.edu/ml/datasets/Multiple+Features](http://archive.ics.uci.edu/ml/datasets/Multiple+Features)

[21] S. Nane, S. Nayar, and H. Murase, “Columbia object image library: Coil-20,” Dept. Comp. Sci., Columbia University, New York, Tech. Rep, Feb. 1996.