Information Sciences 507 (2020) 339–355

Contents lists available at ScienceDirect

Information Sciences

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

Link-aware semi-supervised hypergraph

Taisong Jin[a,b], Liujuan Cao[a,*], Feiran Jie[b], Rongrong Ji[c]

[a] Media Analytics and Computing Lab, Department of Computer Science, School of Informatics, Xiamen University, 361005, China
[b] Science and Technology on Electro-optic Control Laboratory, Luoyang, 471023, China
[c] Media Analytics and Computing Lab, Department of Artificial Intelligence, School of Informatics, Xiamen University, 361005, China

ARTICLE INFO

Article history:
Received 9 March 2019
Revised 25 July 2019
Accepted 29 July 2019
Available online 22 August 2019

Keywords:
Hypergraph
Hyperedge
Link
Ridge regression

ABSTRACT

Hypergraph learning has been widely applied to various learning tasks. To ensure learning accuracy, it is essential to construct an informative hypergraph structure that effectively modulates data correlations. However, existing hypergraph construction methods essentially resort to an unsupervised learning paradigm, which ignores supervisory information, such as pairwise links/non-links. In this article, to exploit the supervisory information, we propose a novel link-aware hypergraph learning model, which modulates high-order correlations of data samples in a semi-supervised manner. To construct a hypergraph, a coefficients matrix of the entire dataset is first calculated by solving a linear regression problem. Then, pairwise link constraints are exploited and propagated to the unconstrained samples, upon which the coefficients matrix is adjusted accordingly. Finally, the adjusted coefficients are used to generate a set of the hyperedges, as well as calculate the corresponding weights. We have validated the proposed link-aware semi-supervised hypergraph model on the problem of image clustering. Superior performance over the state-of-the-art methods demonstrates the effectiveness of the proposed hypergraph model.

1. Introduction

Modeling of high-order data correlations is crucial in various learning tasks such as content-based image or video retrieval [1,3,8,36,43], analysis of Magnetic Resonance Imaging [4], human pose recover from an image [10] and image annotation [7,25]. Compared with a simple graph, a hypergraph, which captures more complex relations among multiple samples by introducing hyperedges, is therefore far more robust, yet discriminative [2,48].

Different from a simple graph that uses each edge to model the relationship between two samples, a hypergraph adopts a hyperedge to connect any number of samples, which enables the modeling of more general data correlations with high flexibility. For example, for a set of articles, a undirect graph can be constructed by linking any two articles with the common author. However, we don't know whether someone is the author of three or more articles or not. Such information loss is unexpected, which is used for many learning tasks. However, hypergraph can effectively preserve the higher relationship among multiple articles. Specifically, we observe that three articles have an common author by sharing one common vertex shown in Fig. 1.

For learning discriminative hypergraph models, it is fundamental to modulate the topological structure among samples. Most existing hypergraph learning methods use a neighborhood-based approach to represent high-order data correlations.

* Corresponding author.
E-mail address: caoliujuan@xmu.edu.cn (L. Cao).

https://doi.org/10.1016/j.ins.2019.07.095
0020-0255/

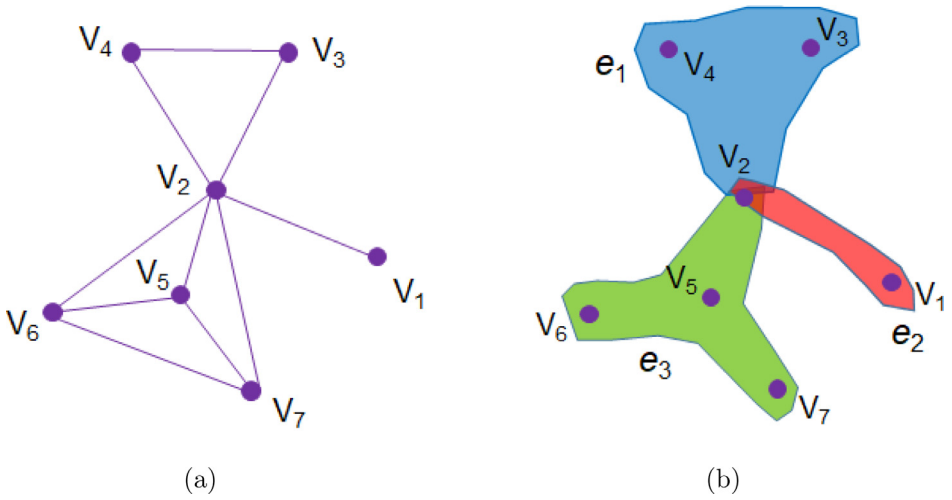


Fig. 1. Simple graph versus hypergraph.(a) Simple graph. (b) Hypergraph. Given a set of the articles $V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$, simple graph and hypergraph are used to model the relationship among the samples, where $E = \{e_1, e_2, e_3\}$ is a set of the constructed hyperedges. A simple graph links two articles together by an edge if there is at least one author in common. This graph cannot tell us whether the same person is the author of three or more articles or not. A hypergraph completely models the high-order relationships among the authors and articles. [47].

In particular, to generate a hyperedge, each sample is modeled as a centroid vertex, which is linked to its neighbors with a fixed number. However, performance of hypergraph learning is sensitive in two ways: (1) Neighborhood size. A small neighborhood size separates the samples from the same cluster; whereas a large neighborhood size groups together the samples from different clusters; (2) Noise. Hypergraph learning performance is usually sensitive to noise. To that effect, various types of noises such as Gaussian noise, sparse noise, and sample-specific corruptions, easily contaminate real-world data.

Alternatively, the representation-based approach has recently been proposed to modulate the high-order relationship among samples [24,37,46] Such an approach generates each hyperedge by solving a linear regression problem. For instance, ℓ_1-Hypergraph [37] leverages sparse regression to generate a noise-resistant hyperedge set. The work by [46] further extended the ℓ_1-Hypergraph to remove redundant hyperedges by jointly learning feature selection and hypergraph construction. Elastic-net Hypergraph [24] incorporates the elastic-net regularization into a linear regression framework, which effectively connects correlated samples to generate a set of hyperedges. ℓ_2-Hypergraph [12] leverages ridge regression to generate a set of the hyperedges. Since the noise components of the data are separated from the original data, such an approach has achieved considerable performance gains with noisy data.

Beyond the topic of hypergraph learning, recent works have well advocated exploiting supervisory information in graph construction. For instance, in [34], exploitation of available labeled data improves performance of graph learning. In [18], Gaussian Process (GP) is used to model label distribution, upon which graph hyper-parameters were learned through expectation maximization. To be robust to noise and offer datum-adaptive neighborhoods, the labels were explicitly incorporated into the representation-based graph, such as low-rank representation [49]. Beyond class labels, pairwise links are also modulated as constraints to construct a neighborhood-based graph [28].

Note that there exist two types of widely available supervisory information, which can be used to help enhance learning performance: class labels and pairwise link constraints [20]. Compared with class labels, pairwise link constraints, which can be automatically obtained without human intervention, are more *cheaply available*. In addition, pairwise link information can be derived from labeled data, but not vice versa. Thus, modeling graph learning with pairwise link constraints has been recently well investigated in many image processing applications [15,19,35,41].

Because of promising performance in applications, supervised graph modeling has received more and more attention from the researchers. However, how to exploit supervisory information, especially widely used pairwise link constraints, to modulate high-order relationships of data is an unsolved open problem. Specifically, existing hypergraph models essentially belong to an unsupervised learning paradigm. It is crucial to develop a novel supervised hypergraph model to more effectively leverage the *cheaply available* link constraints to generate a link-aware hyperedge set.

We argue that the previous methods cannot exploit the supervised information of a small number of samples to model high-order data correlations. Specifically, the pairwise link constraints, which are *cheaply available*, cannot be effectively exploited. Thus, the previous methods cannot fully explore the noisy data, which only achieve limited performance improvements on noisy data. To tackle the above issues, in this paper, we propose a novel hypergraph learning approach, Link-Aware Semi-Supervised Hypergraph, which effectively leverages widely-available pairwise link constraints to learn a robust hypergraph in a semi-supervised manner (See Fig. 2). The contributions of the proposed model are as follows:

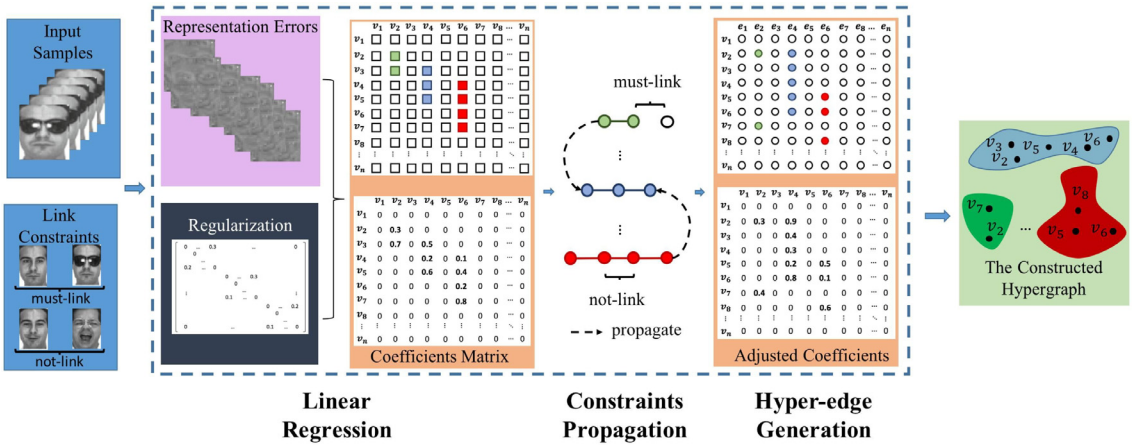


Fig. 2. The flowchart of our proposed hypergraph. For the given dataset (some samples are initially imposed on pairwise link constraints), each sample is firstly represented as the linear combination of the rest of the samples, resulting in a coefficients matrix; Furthermore, the initial link constraints are propagated to the entire dataset, and the coefficients matrix is adjusted accordingly; Finally, the adjusted coefficients are used to generate a link-aware hyperedge set.

1. We propose a hypergraph model, which constructs an informative hypergraph in a semi-supervised manner. Specifically, cheaply available pairwise link constraints in real applications are leveraged to generate a set of link-aware hyperedges. To the best of our knowledge, this is the first attempt to incorporate link constraints in hypergraph learning to modulate the high-order data correlations.
2. The proposed hypergraph model belongs to the representation-based hypergraph construction scope, where each hyperedge is generated by solving a linear regression problem. Initial link constraints are further exploited and propagated to the entire dataset, and the resulting adjusted coefficients are insensitive to noise. Furthermore, the samples with large coefficients are used to generate a set of noise resistant hyperedges.
3. Based on the proposed hypergraph model, two novel hypergraph construction methods: Elastic-Net Semi-Supervised Hypergraph and Nuclear-induced Semi-Supervised Hypergraph, are proposed, which are not only link-aware, but also robust to noise and data corruptions. The experimental results on the noisy datasets demonstrate that the proposed two methods are superior to the existing methods.

The rest of this paper is organized as follows: Section 2 introduces the related work. Section 3 presents a hypergraph learning model. Section 3.1 presents two novel hypergraph construction methods. Section 3.2 reports the experimental results. Finally, we conclude the paper and discuss our future work in Section 3.3.

2. Related work

A hypergraph is defined as $G = (V, E, \mathbf{W})$, which contains the vertex set V, the hyperedge set E, and the hyperedge weight matrix $\mathbf{W}$. A $|V| \times |E|$ incidence matrix $\mathbf{H}$ indicates whether a vertex belongs to a hyperedge, which has the following elements: $\mathbf{H}(v, e) = \begin{cases} 1, \text{if } v \in e \\ 0, \text{otherwise} \end{cases}$. The degree of each vertex v is defined as: $d(v) = \sum_{e \in E} w(e)\mathbf{H}(v, e)$, and the degree of hyperedge e is defined as $\delta(e) = \sum_{v \in V} \mathbf{H}(v, e)$. We use $\mathbf{D}_v$, $\mathbf{D}_e$ and $\mathbf{W}$ to denote the diagonal matrices of vertex degrees, hyperedge degrees, and hyperedge weights, respectively, based on which the normalized hypergraph Laplacian is formulated as $\mathbf{L} = \mathbf{I} - \mathbf{\Theta}$, where $\Theta = \mathbf{D}_v^{-\frac{1}{2}}\mathbf{H}\mathbf{W}_e\mathbf{D}_e^{-1}\mathbf{H}^T\mathbf{D}_v^{-\frac{1}{2}}$.

For much data analysis, modeling of the data relationship is crucial to obtain useful information. Hypergraph, an effective tool to model data relationship, has been used to solve various problems in image analysis. For multi-modal feature integration, a hypergraph is used to capture the high-order correlation among multi-modal features [45]. Hypergraph matching modulates the complex relations of feature set to robustly match image features [44]. Hypergraph-based ranking is used to explore the photos-relevance to a given event [26], and solve the problems of image retrieval [14] and object tracking [27]. Hypergraph ranking is used to solve the problem of multi-objective e-commerce recommendations [29]. Hypergraph-based image classification [42] generates a set of hyperedges and automatically learns the hyperedge weights for classification. Hypergraph [40] is used to multi-view based image classification. Context-Aware hypergraph construction method [23] is designed for robust spectral clustering. Hypergraph hierarchical reduction [22] is considered a macro-vertex in a reduced hypergraph corresponding to a hyperedge. Hypergraph-based tracking [5] exploits high-order geometric relations among multiple correspondences of object parts in different frames. Hypergraphs are incorporated into sparse coding [17] and low-rank matrix factorization [16], which shows promising results with real-world datasets. Recently, the fuzzy hypergraph model is used to represent granular structure [38].

Table 1
List of important notations.

Notation	and description
X	The set of data samples
ML	The must-link set
CL	The not-link set
$\mathbf{x}_i$	The feature vector of the i-th data sample
$\mathbf{e}_i$	The data errors vector of the i-th sample
$\mathbf{X}_i$	The pixels matrix of the i-th image sample
$\mathbf{E}_i$	The data errors matrix of the i-th image sample
$\bar{\mathbf{C}}$	The normalized coefficients matrix
$\widehat{\mathbf{C}}$	The adjusted coefficients matrix
$\mathbf{Z}$	The initial link constraints matrix
$\mathbf{F}$	The exhaustive constraints matrix
f_{ij}	the confidence score of $(\mathbf{x}_i, \mathbf{x}_j)$
$\mathbf{H}$	The incidence matrix
$\mathbf{L}$	Hypergraph Laplacian
N	The number of data samples

All these approaches demonstrate that learning performance can be substantially enhanced if the high-order relationship among the data samples is effectively modulated. However, existing hypergraph construction methods ignore supervisory information. The aim of this work is to integrate *cheaply available* link supervisory information into hypergraph construction to generate a noise resistant hyperedges.

3. Link-aware semi-supervised hypergraph

To construct an informative hypergraph, the key issue is to find *multiple* correlated samples of each sample to depict correct high-order data correlations. Here, we make an attempt to leverage the commonly-used pairwise link constraints to generate a link-aware hyperedge set.

Given a set of N data samples $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$, where $\mathbf{x}_i$ is the feature vector of the i-th sample, two types of pairwise link constraints are usually imposed on a small number of samples, which specify whether a pair of samples belong to the same class (must-link) or different classes (cannot-link). For convenience, the must-link set is defined as $ML = \{(\mathbf{x}_i, \mathbf{x}_j)|\ell_i = \ell_j\}$ and the cannot-link set is defined as $CL = \{(\mathbf{x}_i, \mathbf{x}_j)|\ell_i \neq \ell_j\}$, where ℓ_i is the class label of x_i. Table 1 lists the important notations in this article.

To be robust to data noise and corruption, deployed over the representation-based hypergraph construction is the proposed model, which exploits the self-expressive property of data to represent each sample as a linear combination of the remaining samples in the entire dataset. Coefficients of a linear regression naturally characterize how other samples contribute to represent the target sample, which is crucial to discover the grouping and discriminative structures.

The regression coefficients are directly used to generate hyperedges [24]. However, the coefficients matrix does not consider the effects of link constraints, which prohibits improving the performance. Thus, it is important to leverage the *cheaply available* link information to adjust the regression coefficients, which are used to generate a link-aware hyperedge set.

3.1. Derivation of coefficients matrix using linear regression

Based on the "data self-expression" property, we reconstruct data by itself, formulated as

$$\begin{aligned} &\min_{\mathbf{C},\mathbf{E}} Metric(\mathbf{E}) + \mu_1 Regularization(\mathbf{C}) + \mu_2 Regularization(\mathbf{E}) + \mu_3 Grouping - Regularization(\mathbf{C}) \\ &\quad s.t.\ \mathbf{E} = \mathbf{X} - \mathbf{X}\mathbf{C},\ the\ other\ \ constraints, \end{aligned} \tag{1}$$

where $\mathbf{C} = [\mathbf{c}_1\mathbf{c}_2 \cdots \mathbf{c}_N] \in \mathbb{R}^{N\times N}$ is the coefficients matrix of the data, and $\mathbf{c}_i$ is the coefficients vector of the i-th sample; $\mathbf{E} = [\mathbf{e}_1\mathbf{e}_2 \cdots \mathbf{e}_N] \in \mathbb{R}^{M\times N}$ is the data errors matrix and $\mathbf{e}_i \in \mathbb{R}^M$ is the data errors vector of the i-th sample; $Metric(\cdot)$ is the metric to measure the data errors; $Regularization(\cdot)$ is the regularization term to derive the desired structures of the data and $Group$ - $Regularization(\mathbf{C})$ is the group regularization, aiming to make the coefficients reflect the grouping structures of the data. $Constraint(\mathbf{C})$ are the constraints imposed on the coefficients or data errors to introduce prior knowledge. $\mu_1 > 0$, $\mu_2 > 0$ and $\mu_3 > 0$ are the regularization parameters to tradeoff the regularization and representation errors.

The coefficients matrix characterizes how data samples contribute to data reconstruction, where date errors are measured by the metric and the desired data structures are reflected by the regularization terms and the imposed constraints. Thus, the coefficients matrix is used to generate a set of hyperedges. In fact, the recently proposed ℓ_1-Hypergraph [37], ℓ_2-Hypergraph [12] and Elastic-net Hypergraph [24] are considered as the special cases of the representation-based hypergraphs.

Worth emphasizing is that the group regularization, which has the group selection effects, is used in the linear regression framework. This property, it is suitable to choose a group of correlated samples together for hyperedge generation, is crucial for hypergraph construction.

3.2. Exploiting pairwise link constraints

For many real world problems, there exists available data, where some data samples are imposed on pair link constraints. It is natural to use link information as the constraints and incorporate them into the hypergraph learning framework. To leverage the representation-based hypergraph learning framework, the important issue is to make the representation coefficients reflect the link information.

A straightforward approach for using link information is to directly incorporate link constraints into a linear regression framework. For example, we force the coefficient of the not-link samples to be zero, i.e., the not-link samples are not involved in representing the response sample. By this way, the regression coefficients matrix reflects the effects of the link constraints. However, because link constraints are imposed initially only a small number of samples in the entire dataset, the effects of the link information on the linked samples are limited. It is better to propagate the initial link constraints to the unconstrained samples, and correspondingly significantly adjust the coefficients matrix.

To exploit the link information imposed on a small number of samples, we adopt the constraint propagation strategy [28], which determines the constraint relationship of two unconstrained samples by classifying the relationship to class labeled as +1 or class labeled as -1.

Note that the i-th column of the normalized coefficients matrix actually provides the initial configuration of a two-class semi-supervised learning problem with respect to x_i, where class 1 contains the samples close to x_i and class -1 contains the samples far away from x_i. The other columns of coefficients matrix are handled similarly to decompose constraint propagation in vertical direction into N independent label propagation sub-problems. Likewise, we can perform horizontal constraint propagation by considering coefficients matrix row by row. Thus, by vertical constraint propagation and the horizontal constraint propagation, constraint propagation is performed efficiently using the label propagation technique [47], where the iterative approach is used to solve the optimization problem (See Steps 3 and 4 in Algorithm 1).

Algorithm 1: Link constraints propagation.

Input: The initial constraints matrix $\mathbf{Z}$ and the coefficients matrix $\mathbf{C}$.

Output: The exhaustive constraints matrix $\mathbf{F}^*$ and the adjusted coefficients matrix $\widehat{\mathbf{C}}$.

1 Normalize $\mathbf{C}$ by $\mathbf{C}^*_{ij} = \frac{|\mathbf{C}_{ij}|}{\max_i |\mathbf{C}_{ij}|}$ and set $\overline{\mathbf{C}} = \frac{(\mathbf{C}^* + (\mathbf{C}^*)^T)}{2}$ to ensure that the normalized coefficients matrix $\overline{\mathbf{C}}$ is symmetric.

2 Compute the matrix $\overline{\mathbf{L}} = \mathbf{D}_{-1/2}\overline{\mathbf{C}}\mathbf{D}_{-1/2}$, where $\mathbf{D}$ is a diagonal matrix with its i-th diagonal element being $\sum_j \overline{\mathbf{C}}_{ij}$.

3 Iterate $\mathbf{F}_v(t+1) = \alpha\mathbf{L}\mathbf{F}_v(t) + (1-\alpha)\mathbf{Z}$ for the vertical constraint propagation until convergence, where α is a parameter in the range (0,1).

4 Iterate $\mathbf{F}_h(t+1) = \alpha\mathbf{F}_h(t)\mathbf{L} + (1-\alpha)\mathbf{F}^*_v$ for the horizontal constraint propagation until convergence, where $\mathbf{F}^*_v$ is the limit of $\{\mathbf{F}_v(t)\}$.

5 $\mathbf{F}^* = \mathbf{F}^*_h$ is the exhaustive constraint matrix being propagated, where $\mathbf{F}^*_h$ is the limit of $\{\mathbf{F}_h(t)\}$.

6 Derive the adjusted coefficient matrix$\widehat{\mathbf{C}} = \{\widehat{\mathbf{C}}_{ij}\}_{N\times N}$ by using the exhaustive constraint matrix, $\mathbf{F}^*$, defined as

$$\widehat{\mathbf{C}}_{ij} = \begin{cases} 1 - \left(1 - f^*_{ij}\right)\left(1 - \overline{\mathbf{C}}_{ij}\right) & \text{if } f^*_{ij} \geq 0 \\ (1+f^*_{ij})\overline{\mathbf{C}}_{ij} & otherwise \end{cases}$$

First, the initial constraints matrix $\mathbf{Z} = \{\mathbf{Z}_{ij}\}_{N\times N}$ is defined as

$$\mathbf{Z}_{ij} = \begin{cases} 1 & \text{if } (\mathbf{x}_i, \mathbf{x}_j) \in ML \\ -1 & \text{if } (\mathbf{x}_i, \mathbf{x}_j) \in CL \\ 0 & \text{otherwise} \end{cases}. \tag{2}$$

Furthermore, the propagated exhaustive constraints matrix $\mathbf{F} = \{f_{ij}\}_{N\times N}$ is defined, where $|f_{ij}| \leq 1$ indicates the confidence score of $(\mathbf{x}_i, \mathbf{x}_j)$ being a must-link (or cannot-link) constraint. In other words, $f_{ij} > 0$ indicates that two samples $\mathbf{x}_i$ and $\mathbf{x}_j$ is imposed on a must-link constraint; $f_{ij} < 0$ indicates that two samples $\mathbf{x}_i$ and $\mathbf{x}_j$ is imposed on a not-link constraint.

Algorithm 1 presents the propagation procedure of the link constraints, which produces an adjusted coefficients matrix $\widehat{\mathbf{C}}$ that has the following properties:

1. $\widehat{\mathbf{C}}$ is nonnegative and symmetric, $\widehat{\mathbf{C}}_{ij} \in [0, 1]$.
2. $\widehat{\mathbf{C}}_{ij} \geq \overline{\mathbf{C}}_{ij}$ (or $< \overline{\mathbf{C}}_{ij}$) if $f^*_{ij} \geq 0$ (or < 0).
   This property means the adjusted coefficients matrix $\widehat{\mathbf{C}}$ is actually derived from the coefficients matrix $\overline{\mathbf{C}}_{ij}$ by increasing $\overline{\mathbf{C}}_{ij}$ for the must-link constraints with $f^*_{ij} \geq 0$ and decreasing $\overline{\mathbf{C}}_{ij}$ for the cannot-link constraints with $f^*_{ij} < 0$.
3. $\partial\widehat{\mathbf{C}}_{ij}/\partial\overline{\mathbf{C}}_{ij} = 1 - f^*_{ij}$.
   This property shows that both must-link and cannot-link constraints play the important roles (with the same derivatives) in adjusting coefficients if they have the same pairwise constraint values.
4. $\widehat{\mathbf{C}}_{ij} \geq f^*_{ij}$ (or $\leq f^*_{ij}$) if $f^*_{ij} \geq 0$ (or < 0).

This property ensures $\widehat{\mathbf{C}}_{ij}$ is set to be large when f_{ij}^{*} takes a large positive value, even if $\overline{\mathbf{C}}_{ij}$ is initially very small. On the other hand, $\widehat{\mathbf{C}}_{ij}$ is set to be small when f_{ij}^{*} takes a large negative value, even if $\overline{\mathbf{C}}_{ij}$ is initially very large.

It is worth emphasizing that the used link propagation scheme is different from the constraint propagation [28] method. The work in [28] propagates link constraints by using the neighborhood-based graph, where the edge weights are adjusted using the exhaustive link constraints. In contrast, our learning model directly adjusts the regression coefficient matrix by the propagated exhaustive constraints.

3.3. Hyperedge generation

The coefficients matrix obtained by solving the aforementioned liner regression problem effectively reflects the grouping and discriminant structures of the data. After the link constraints imposed on a small number of samples are propagated, the coefficients matrix is changed accordingly. Thus, the obtained coefficients are link-aware. Based on the properties of the adjusted regression coefficients, the samples with the same class have much larger values, while the samples from different classes have much smaller values.

Since the samples with large non-zero coefficients in the adjusted coefficients matrix are most closely related to $\mathbf{x}_i$, we combine such samples together to generate a link-aware hyperedge. Furthermore, an incidence matrix $\mathbf{H} \in \mathbf{R}^{N\times N}$ is defined, whose incidence vector indicates the relationship between the hyperedge and its vertices. The incidence vector associated with the centroid vertex $\mathbf{x}_i$ is formulated as

$$\mathbf{H}(v_i, e_j) = \begin{cases} 1, & \text{if } \widehat{\mathbf{C}}_{ij} > \theta \\ 0, & \text{otherwise} \end{cases}, \tag{3}$$

where θ is a threshold. $\widehat{\mathbf{C}}_{ij}$ is the j-th adjusted regression coefficient of $\mathbf{x}_i$. We take each data sample as centroid vertex and link a group of data samples lying on or close to the same linear subspace as centroid vertex to generate a hyperedge and the corresponding noise and data corruption are detected and removed. Thus, the high-order relationship between centroid vertex and its most correlated vertices chosen from the remaining data samples are modulated.

We then turn to decide the hyperedge weights, which are crucial for the hypergraph-based applications. Since the real-world dataset is usually contaminated by noise, it is not suitable to directly measure the similarity between two samples. Instead, we take the regression coefficients vector [24] to measure the similarity between two samples as follows:

$$\mathbf{A}(i, j) = \begin{cases} \dfrac{(\widehat{\mathbf{C}}_i)^T \widehat{\mathbf{C}}_j}{\|\widehat{\mathbf{C}}_i\| \|\widehat{\mathbf{C}}_j\|}, & \text{if (i} \neq \text{j)} \\ 0, & otherwise \end{cases} \tag{4}$$

After a hypergraph is constructed from the incidence matrix, $\mathbf{H}$, the hyperedge weight $w(e_i)$ is calculated by [10]

$$w(e_i) = \sum_{v_j \in e_i} \mathbf{H}(v_i, e_j)\mathbf{A}(i, j), \tag{5}$$

4. The proposed hypergraph construction

In this section, we present two hypergraph construction methods, which serve as the implementation of the proposed hypergraph model. Except for the computation of the coefficients matrix, the two hypergraph construction methods share the same mechanism.

4.1. Elastic-net semi-supervised hypergraph

Real-world data may contain sample-specific corruptions. For instance, the faces in images may be occluded by sunglasses or a scarf [31]. To be robust to such sample-specific corruptions, the recent Elastic-net Hypergraph model [24] generates each hyperedge by solving the following linear regression problem:

$$\begin{aligned} &\min_{\mathbf{C},\mathbf{E}} \lambda_1 \|\mathbf{C}\|_1 + \lambda_2 \|\mathbf{C}\|_F + \|\mathbf{E}\|_{21} \\ &s.t.\ \mathbf{X} = \mathbf{X}\mathbf{C} + \mathbf{E}, diag(\mathbf{C}) = 0, \end{aligned} \tag{6}$$

where $\mathbf{E}$ denotes the error matrix and the ℓ_{21}-norm based matrix is used to measure the data errors; $\mathbf{C}$ is the coefficients matrix; $\lambda_1 > 0$ and $\lambda_2 > 0$ are the regularization parameters. The diagonal elements of $\mathbf{C}$ are set as zeros, ensuring that each image is represented by the remaining images, excluding itself.

Eq. (6) adopts an Elastic-net regularized linear representation to compute the representation coefficients of each sample, which favors the selection of multiple correlated samples together to represent the response sample. We remove the sample $\mathbf{x}_i$ from the sample matrix to decompose the $\mathbf{x}_i$ as the elastic-net regularized based linear regression, formulated as

$$\min_{\mathbf{C}_i'} \lambda_1 \left\|\mathbf{C}_i'\right\|_1 + \lambda_2 \left\|\mathbf{C}_i'\right\|_2 + \left\|\mathbf{x}_i - \mathbf{D}_i\mathbf{C}_i'\right\|_{21} \tag{7}$$

where the dictionary $\mathbf{D}_i = [\mathbf{x}_1, \mathbf{x}_1, \cdots, \mathbf{x}_{n-1}]$ is designed for removing the constraint $\mathbf{C}_{ii} = 0$.

To compute the entire elastic-net regularization path of Eq. (7), the LARS-EN algorithm [50] can be used to solve the optimization problem, where the Elastic-net linear regression problem is transformed into an equivalent problem of a lasso regression on the augmented data set for employing the LARS algorithm [6].

After obtaining the coefficients matrix $\mathbf{C} \in \mathbb{R}^{N\times N}$, the coefficients of Elastic-net regularized linear regression are adjusted by propagating the initial link constraints to the unconstrained samples (See Algorithm 1). Finally, the adjusted coefficients are used to generate link-aware hyperedges. For clarification, we refer to the hypergraphs constructed in our way as EN-sHG in the rest of the article.

4.2. Nuclear-induced semi-Supervised hypergraph

In various image analysis and classification tasks, often used to measure the representation errors is the pixel-based errors model which assumes that the pixel-wised errors are independent and identically distributed (i.i.d). However, such i.i.d assumption does not hold in regions with occlusions or large illumination changes, which instead have approximated low-rank structures [33,39].

In such a case, the linear matrix regression, which doesn't require converting each image to a long vector, is more suitable to represent each image. To achieve that effect, the pixels matrix $\mathbf{X}_i$ of the i-th image ($i = 1, \cdots, N$) is directly represented as

$$\mathbf{X_i} = \mathbf{c_{i1}X_1} + \cdots + \mathbf{c_{ii-1}X_{i-1}} + \mathbf{c_{ii+1}X_{i+1}} + \cdots + \mathbf{c_{iN}X_N} + \mathbf{E_i}, \tag{8}$$

where $\mathbf{c}_i = [\mathbf{c}_{i1}\mathbf{c}_{i2}, \mathbf{c}_{ii-1}, \mathbf{c}_{ii+1} \cdots \mathbf{c}_{iN}]^T \in \mathbb{R}^{N-1}$ is the coefficients vector and $\mathbf{E}_i$ is the representation error, corresponding to an error image.

Defining $\mathbf{A}(\mathbf{c}_i) = \mathbf{c}_{i1}\mathbf{X}_1 + \cdots + \mathbf{c}_{ii-1}\mathbf{X}_{i-1} + \mathbf{c}_{ii+1}\mathbf{X}_{i+1} + \cdots + \mathbf{c}_{iN}\mathbf{X}_N$, Eq. (8) is re-written as

$$\mathbf{X}_i = \mathbf{A}(\mathbf{c}_i) + \mathbf{E}_i. \tag{9}$$

Furthermore, by adopting nuclear-norm based metric to measure the data errors and adding the ℓ_2-norm based group selection regularization, the linear matrix regression is rewritten as

$$\begin{aligned} &\min_{\mathbf{c}_i, \mathbf{E}_i} \|\mathbf{E}_i\|_* + \lambda_3 \|\mathbf{c}_i\|_2^2 \\ &s.t. \quad \mathbf{x}_i = \mathbf{A}(\mathbf{c}_i) + \mathbf{E}_i, \end{aligned} \tag{10}$$

where the second term is the ℓ_2-norm regularization for discovering the grouping structures of the data, $\lambda_3 > 0$ is the parameter to tradeoff the regularization and representation errors.

To measure the error image, Eq. (10) adopts a nuclear-norm based metric to discover the underlying low-rank structures of the data [6,39]. The objective function in Eq. (10) is non-convex, which is solved by using the Alternating Direction Method of Multipliers (ADMM) approach [9,39].

By incorporating the penalty term of the constraint into the objective function of Eq. (10), the augmented Lagrange function is

$$L(\mathbf{Z}, \mathbf{c}_i, \mathbf{E}_i) = \|\mathbf{E}_i\|_*^2 + \tfrac{\lambda}{2}\|\mathbf{c}_i\|_2^2 + tr(\mathbf{Z}^T(\mathbf{X}_i - \mathbf{A}_i(\mathbf{c}) - \mathbf{E}_i) + \tfrac{\rho}{2}\left\|\mathbf{X}_i - \mathbf{A}_i(\mathbf{c}) - \mathbf{E}_i\right\|_F^2 \tag{11}$$

where ρ is the augmented Lagrange parameter. $\mathbf{Z} \in \mathbb{R}^{N\times N}$ are the matrix of the Lagrange multipliers for the constraints. $tr(\cdot)$ denotes the trace of the matrix. ADMM effectively solves an optimization problem in two separate steps: primal variable updating and dual ascending; i.e., ADMM iteratively updates the primal variables ($\mathbf{c}_i$,$\mathbf{E}_i$),and the Lagrange multipliers ($\mathbf{Z}$) to obtain the local optimal solution.

For each sample, the coefficients vector is derived by solving Eq. (10). After all the samples have been selected once, a coefficients matrix $\widetilde{\mathbf{C}} = [\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_N]^T \in \mathbb{R}^{(N-1)\times N}$ can be derived. Then, the coefficients matrix $\mathbf{C} \in \mathbb{R}^{N\times N}$ is obtained by adding zeros to the diagonal elements. Furthermore, the regression coefficients of nuclear induced linear matrix regression are adjusted by propagating the initial link constraints to the unconstrained samples (See Algorithm 1). Finally, the adjusted coefficients are used to generate a set of link-aware hyperedges.

For clarification, we refer to the hypergraphs constructed in our way as Nuclear-sHG in the rest of the article.

Algorithm 2 overviews the proposed semi-supervised hypergraph construction to capture the high-order data correlations.

4.3. Computational complexity

The computational cost of our proposed hypergraph model consists of three components: representation coefficients computation, coefficients propagation and hyperedge generation. Note that two proposed semi-supervised hypergraph construction methods are the same except for the derivation of coefficients matrix. We assume that there exist N samples in the dataset. For coefficients propagation (See Algorithm 1), the computation cost is $O(pN^2)$, where p is the number of non-zero coefficients of centroid vertex. The computational cost of hyperedge generation is $O(N^2)$.

Algorithm 2: Link-aware semi-supervised hypergraph.

Input: A feature set of the data samples **X** and a set of link constraints
Output: A link-aware Semi-Supervised Hypergraph
1 Compute the coefficients matrix according to Eq. (6) or (10).
2 Adjust coefficients matrix by employing Algorithm 1.
3 **for** $i = 1, \cdots, N$ **do**
(a) Link centroid vertex, $\mathbf{x}_i$ to the vertices with $non - zero$ coefficients to generate a hyperedge, e_i, according to Eq. (3) ;
(b) Define the weight of the hyperedge, e_i. according to Eq. (5);
end

For the EN-sHG method, the computation cost of coefficients derivation of using the Least angle regression approach [6] is $O(N^3)$.

For the Nuclear-sHG method, the computation cost of the coefficients derivation using ADMM approach [9] is mainly determined by the nuclear-norm based minimization operator, where singular value decomposition of image matrix is involved. For convenience, we assume the image matrix is $p \times q$ sizes, the computation cost of coefficients derivation is $O(N \times \min(pq^2, qp^2))$.

5. Experimental results and discussions

In this section, we evaluate the proposed two hypergraph construction methods in the task of image clustering.

5.1. Experimental setting and compared methods

Clustering via hypergraph is typically performed from a spectral perspective. First, the K- largest eigenvectors of hypergraph Laplacian are computed. Then, the K-means clustering is employed on the rows of the eigenvector matrix to obtain the final clustering result, where the cluster number K is set as the number of data classes. In our experiments, we adopted two commonly used metrics to measure the performance, *i.e.*, the accuracy (AC) and the normalized mutual information (NMI) [16]. Accuracy is the percentage of correct clustering labels, while normalized mutual information measures how similar different cluster sets are.

We compared the performance of the proposed methods (EN-SHG and Nuclear-SHG) with the following baselines:

KNN – HG [13]. KNN-HG is a hypergraph construction method, where each hyperedge is generated by using the neighborhood-based approach and a Gaussian kernel is used to compute the edge weight.

ℓ_2 **– HG** [12]. ℓ_2-HG adopts ridge regression to construct a hypergraph, where the ℓ_2-norm based metric is used to measure the representation errors.

EN – HG [24]. EN-HG is a hypergraph construction method, where each hyperedge is generated by solving the problem of elastic-net regularized linear regression. The ℓ_{21}-norm based metric is used to measure the representation errors.

KNN – sGr [28]. KNN-cGr is a semi-supervised graph construction method, where the pairwise constraints are propagated on the neighborhood-based graph, and the edge weights of a Gaussian kernel is adjusted accordingly.

Nuclear – Gr [39]. Nuclear-Gr is a graph construction method, where the nuclear-norm based metric is used to measure the representation errors and the ℓ_2-norm is used as the regularization.

To further know how link constraints can help enhance the hypergraph learning performance. Based on the linear regression based learning framework, we develop novel multiple constraints based graph construction methods. From the above, four graph or hypergraph construction methods are developed as follows:

Nuclear – cHG. Nuclear-cHG is a semi-supervised hypergraph construction method, where the link constraints are directly imposed on the nuclear-induced hypergraph learning framework.

Nuclear – sGr. Nuclear-sGr is a semi-supervised graph construction method, which is the same as Nuclear-Gr except that the coefficients obtained by link propagation are directly used to construct a graph instead of a hypergraph.

EN – cHG. EN-cHG is a semi-supervised hypergraph construction method, where the link constraints are directly imposed on the elastic-net-based hypergraph learning framework.

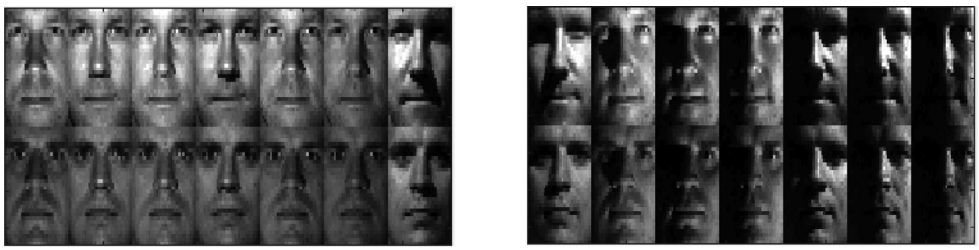
EN – sGr. EN-sGr is a semi-supervised graph construction method, which is the same as the proposed EN-sHG method except that the coefficients obtained by link propagation are directly used to construct a graph instead of a hypergraph.

Among these methods, Nuclear-cHG, EN-cHG directly impose the link constraints on the coefficients for reflecting the link effects. In practice, we define the constraint $c_{ij} = \mathbf{Q}_{ij}$, where $\mathbf{Q}_{ij}$ reflects the not-link and must-link constraints having the following formulation:

$$\mathbf{Q}_{ij} = \begin{cases} \alpha & \text{if } (\mathbf{x_i}, \mathbf{x_j}) \in \mathbf{ML} \text{ and } \mathbf{c_{ij}} < \theta \\ 0 & \text{if } (\mathbf{x_i}, \mathbf{x_j}) \in \mathbf{CL} \end{cases}, \tag{12}$$

Table 2
The compared methods.

Method	Graph type	Link constraint	Metric
KNN-sGr [28]	Graph	Constraint	Null
KNN-HG [13]	Hypergraph	Null	Null
ℓ_2-HG [12]	Hypergraph	Null	ℓ_2-norm
Nuclear-Gr [39]	Graph	Null	Nuclear-norm
Nuclear-sGr	Graph	Constraint Propagation	Nuclear-norm
Nuclear-cHG	Hypergraph	Constraint	Nuclear-norm
Nuclear-sHG	Hypergraph	Constraint Propagation	Nuclear-norm
EN-HG [24]	Hypergraph	Constraint	ℓ_{21}-norm
EN-cHG	Hypergraph	Constraint	ℓ_{21}-norm
EN-sGr	Graph	Constraint Propagation	ℓ_{21}-norm
EN-sHG	Hypergraph	Constraint Propagation	ℓ_{21}-norm



(a) (b)

Fig. 3. The image examples from Extended YaleB dataset. (a) Session 1. (b) Session 2.

where θ is the threshold parameter of choosing the vertex of hyperedge, ensuring the must-link sample to be chosen for hyperedge generation, while at the same time, ensuring the not-link sample isn't chosen for hyperedge generation. For our experiments, α is set equal to $1.2|c_{ij}|$.

For the KNN-sGr, Nuclear-sGr, EN-sGr and the proposed methods (EN-sHG and Nuclear-sHG), they adopt the constraint propagation approach to generate the link-aware edges or hyperedges. Thus, for the constraints involved methods, regardless of exploiting the link constraints or the constraints propagation, they construct a graph or hypergraph in a semi-supervised manner. For convenience, we list the compared methods in Table 2.

For the parameters in these methods, we conducted cross-validation experiments to set the optimal parameter values for data clustering. For the hypergraph-based methods, the θ is set equal to $0.8|c_{ij}|$.

For the Nuclear induced methods (Nuclear-Gr, Nuclear-HG, Nuclear-cHG, Nuclear-sGr and our proposed Nuclear-sHG), we use the pixels matrix of an image as the image feature. For all other methods, we convert the pixels matrix of an image into a long vector, which is used as the image feature.

For semi-supervised graph or hypergraph construction methods, we randomly label 1% of the link pairs from the entire dataset to simulate both must-link and cannot-link constraints, which are imposed on a small number of image samples. Since the link information doesn't necessarily contain the class label of the samples, the classification task is unsuitable for constraints involved methods. We conducted the clustering experiments on the real-world datasets to demonstrate our proposed two methods.

5.2. Face clustering with different illuminations

To handle the data with illumination changes, we conducted the face clustering experiments on the Extended Yale B face dataset [21]. The database contains 5760 single light source images of 10 subjects each seen under 576 viewing conditions (9 poses $\times$ 64 illumination conditions). For every subject in a particular pose, an image with ambient (background) illumination was also captured. Hence, the total number of images is 5760 + 90 = 5850. For this database, we simply use the cropped images and resize them to 32 $\times$ 32 pixels. This dataset has 38 individuals and around 64 near frontal images under different illuminations per individual.

We adopted two separated sessions: Session 1 (containing the images from Subset 1, 2 and 3 with slight illumination changes) and Session 2 (containing the images from Subset 1,4,5 with extreme illumination conditions). Fig. 3 shows the image examples on Extended Yale B. Table 3 lists the face clustering results.

As shown in Table 3, when the lighting change is large (Session 2), the clustering performance of all the methods is degraded; Compared with the neighborhood-based methods (KNN-sGr and KNN-HG), the representation-based methods are relatively robust to lighting changes.

Table 3
Clustering results on the Yale B dataset.

Different	AC(%)		NMI(%)	
illuminations	Session 1	Session 2	Session 1	Session 2
KNN-sGr [28]	29.9	20.2	43.0	35.3
KNN-HG[[13]	27.6	21.6	40.1	33.1
ℓ_2-HG [12]	39.6	34.5	48.8	44.3
Nuclear-Gr [39]	48.6	47.2	60.1	58.3
Nuclear-sGr	50.7	49.1	62.4	61.5
Nuclear-cHG	49.7	48.1	60.4	60.5
Nuclear-sHG	**53.3**	**49.8**	**66.7**	**62.1**
EN-HG [24]	44.3	41.2	57.6	53.1
EN-cHG	44.9	42.3	58.4	54.5
EN-sGr	45.8	44.1	65.6	57.4
EN-sHG	47.3	43.2	58.1	59.6

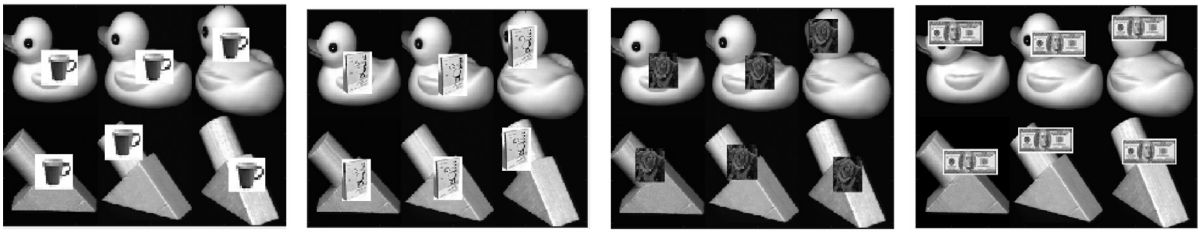


Fig. 4. The image examples from Coil20 Dataset. (a) The images occluded by the cups. (b) The images occluded by the books. (c) The images occluded by the flowers. (d) The images occluded by the dollars.

The constraint involved methods consistently outperform corresponding unconstraint involved methods.(Nuclear-sGr, Nuclear-cHG and the proposed Nuclear-sHG $>$ Nuclear-Gr; EN-cHG, En-sGr and the proposed EN-sHG $>$ EN-HG, KNN-sGr is $>$ KNN-HG). The experimental results demonstrate that link constraints effectively enhance the hypergraph learning performance. For both two sessions, our proposed Nuclear-sHG *consistently and significantly* outperforms other compared methods and our proposed EN-sHG is also superior to EN-HG by a large margin.

These experimental results demonstrate that our proposed two methods effectively leverage the link information to generate a link-aware hyperedge set. Specifically, the linear matrix regression-based hypergraph construction (Nuclear-sHG) captures the underlying low-rank structures of the image, which is robust to the case of extreme lighting conditions.

5.3. Object clustering with contiguous occlusions

To handle the data with the contiguous occlusions, we conducted the image clustering experiments on the Coil20 dataset [32]. To simulate the contiguous occlusions, we replaced 30% of randomly chosen pixels from each selected image with different kinds of occlusions: cup, book, flower and monkey. Fig. 4 shows the image examples occluded by different objects. Table 4 lists the experimental results on Coil20 dataset.

As shown in Table 4, the link constraints involved methods outperform the corresponding unconstrained methods (Nuclear-sGr, Nuclear-cHG and the proposed Nuclear-sHG $>$ Nuclear-Gr; EN-cHG, En-sGr and the proposed EN-sHG $>$ EN-HG). Specifically, the constraint propagation-based methods outperform the methods using the link constraints (the proposed Nuclear-sHG $>$ Nuclear-cHG; the proposed EN-sHG $>$ En-cHG). The proposed Nuclear-sHG method achieves the best clustering results under all circumstances of different occluded objects. The proposed EN-sHG significantly outperforms the unsupervised EN-HG. These experimental results demonstrates that the proposed methods can leverage the link information to be more robust to the contiguous occlusions.

5.4. Face clustering with real-world occlusions

To handle the data with real-world occlusions, we conducted face clustering experiments on the AR dataset [30]. The AR dataset contains over 4000 face images of 126 people, which vary in expression, illumination and disguise (wearing sunglasses or scarves). Each subject has 26 images: fourteen clean, six with sunglasses, and six with scarves. First, we used a subset containing 1400 clean faces randomly selected from fifty male and fifty female subjects. Limited by computational capabilities (as was the case in [17]), we re-sized all the images from the original 165×120 to 60×43 and converted them to gray scale, we selected, from the AR dataset, selected 700 faces wearing sunglasses (occluded ratio is about 20%) and 700 faces wearing scarves (occluded ratio is about 40%).

Table 4
Clustering results on the Coil20 dataset.

Contiguous occlusion	AC(%)			
by occluded object	Cup	Book	Flower	Monkey
KNN-sGr [28]	36.0	32.9	37.0	33.9
KNN-HG [13]	39.5	33.2	39.1	34.3
ℓ_2-HG [12]	49.2	45.7	48.5	47.3
Nuclear-Gr [39]	60.8	55.4	59.7	57.2
Nuclear-sGr	62.3	57.9	61.6	59.8
Nuclear-cHG	61.7	56.4	60.4	58.5
Nuclear-sHG	**67.8**	**61.3**	**66.9**	**68.2**
EN-HG [24]	57.8	54.3	59.6	60.3
EN-cHG	58.3	55.1	60.4	60.9
EN-sGr	60.4	56.9	62.1	61.3
EN-sHG	61.4	57.6	63.1	62.8
Contiguous occlusion	NMI(%)			
by Occluded object	Cup	Book	Flower	Monkey
KNN-sGr [28]	43.5	39.4	44.5	46.4
KNN-HG [13]	45.1	37.0	46.5	40.8
ℓ_2-HG [12]	57.8	54.3	55.4	56.2
Nuclear-Gr [39]	69.7	68.3	72.1	76.2
Nuclear-sGr	71.4	70.6	73.1	75.8
Nuclear-cHG	70.3.7	69.1	72.7	77.1
Nuclear-sHG	**75.1**	**72.3**	**76.9**	**79.2**
EN-HG [24]	68.8	66.6	70.3	73.3
EN-cHG	69.2	67.1	70.4	73.9
EN-sGr	70.7	68.3	73.4	74.5
EN-sHG	70.3	67.2	72.1	75.6

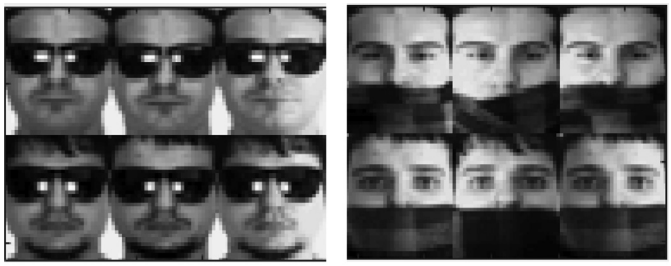


Fig. 5. The image examples from the AR Dataset. (a) The images occluded by the glasses. (b) The images occluded by a scarf.

For computation efficiency, a subset containing 2600 faces was used, which were randomly selected from fifty male and fifty female subjects. Each subject contains 26 images, *i.e.*, fourteen clean, six with sunglasses, and six with scarves. As the AR dataset contains two types of occlusions (the occlusion by the glasses and scarf), we conducted two clustering experiments corresponding to the sunglasses and scarf occlusions. For the sunglasses occlusion, we select 14 clean face images and 6 face images with sunglasses from each subject. For the scarf occlusion, we selected 14 clean face images and 6 face images with scarf from each subject. Fig. 5 shows the image examples occluded by the glasses or a scarf. Table 5 presents the experimental results on the AR datasets.

As shown in Table 5, the constraint propagation-based methods outperform the methods using the link constraints (the proposed Nuclear-sHG $>$ Nuclear-cHG; the proposed EN-sHG $>$ EN-cHG). The proposed nuclear-sHG achieves the best clustering results and the proposed EN-sHG is superior to the unsupervised EN-HG. Specifically, when the occlusion level is large (the occlusion is caused by a scarf), the proposed two methods achieve much better clustering results than the corresponding unsupervised methods (the proposed Nuclear-sHG significantly outperforms Nuclear-HG; the proposed EN-sHG outperforms EN-HG by a large margin). The experimental results demonstrate that the exploitation of link information makes the proposed methods robust to real-world occlusions.

5.5. Face clustering with uncontrolled setting

To handle the data with uncontrolled setting, we use the LFW face dataset [11] to conduct face clustering experiments. LFW is a large-scale unconstrained face dataset under the uncontrolled setting. Images in LFW were collected from the

Table 5
Clustering results on the AR dataset.

Real occlusion	AC(%)		NMI(%)	
by occluded object	Glasses	Scarf	Glasses	Scarf
KNN-sGr [28]	37.1	35.5	51.8	53.3
KNN-HG [13]	33.7	33.4	59.2	59.4
ℓ_2-HG [12]	47.2	46.8	62.3	61.7
Nuclear-Gr [39]	55.4	52.5	67.6	64.3
Nuclear-sGr	57.3	54.7	69.2	65.4
Nuclear-cHG	55.9	53.1	68.2	65.0
Nuclear-sHG	**59.3**	**57.7**	**70.2**	**66.5**
EN-HG [24]	49.8	47.1	63.2	62.4
EN-cHG	50.2	47.8	63.9	63.0
EN-sGr	52.8	49.4	64.9	64.4
EN-sHG	53.6	50.3	65.6	65.3

Table 6
Clustering results on the LFW dataset.

Clustering	AC(%)					NMI(%)				
number	12	22	32	42	52	12	22	32	42	52
KNN-sGr [28]	20.4	11.4	12.2	12.5	10.6	21.9	14.5	16.8	17.8	15.4
KNN-HG [13]	21.6	15.6	14.6	13.3	13.2	23.6	19.2	18.5	18.4	17.8
ℓ_2-HG [12]	27.3	20.2	18.6	18.2	17.4	25.4	20.1	19.8	20.1	18.1
Nuclear-Gr [39]	30.4	25.5	20.0	18.5	16.4	34.7	31.3	24.7	21.6	19.6
Nuclear-sGr	32.1	27.6	22.5	21.3	19.6	37.2	33.9	28.3	24.9	21.9
Nuclear-cHG	30.8.	26.1	20.5	19.0	16.9	35.3	31.8	25.2	22.1	20.3
Nuclear-sHG	**33.4**	**30.8**	**25.4**	**24.6**	**21.3**	**35.6**	**30.2**	**28.6**	**25.6**	**23.2**
EN-HG [24]	28.7	18.3	19.6	17.2	17.7	24.7	20.7	20.6	19.6	16.9
EN-cHG	29.5	19.1	20.4	17.9	17.8	25.3	21.3	21.0	20.2	17.2
EN-sGr	30.2	20.1	21.5	19.2	18.9	26.7	22.5	21.9	21.6	19.2
EN-sHG	30.1	24.9	22.1	19.0	18.9	30.3	28.4	26.8	20.3	21.6

Yahoo News, which contains 13,233 face images of 5749 subjects in uncontrolled environments. These images contain large variations in pose, illumination, expression, occlusion, and resolution.

For computational convenience, we only selected a subset of 2213 images, which only contains 52 subjects and each subject has more than 20 images and simply crop the face image to remove the background, the images are resized to 64×64 pixels. For computation convenience, we only select a subset of 2213 images, which contains 52 subjects, each of which has more than 20 images. We crop the face image to remove the background and resize the image to be 64×64 pixels. Fig. 6 shows the image examples of LFW dataset. Table 6 presents the clustering results.

As shown in Table 6, the uncontrolled setting makes face clustering become a challenging task. All the methods perform less promising, whereas the proposed Nuclear-sHG method still obtains the best clustering results, and the proposed EN-sHG is superior to EN-HG, EN-cGr and EN-CHG. These experimental results demonstrate that the proposed methods can generate link-aware hyperedges, which are relatively robust to handle such a uncontrolled setting.

5.6. The qualitative experiments on robustness

To illustrate the superiority of our proposed link-aware hypergraph model, specifically to handle the noisy data, we selected 20 images from each individual of the Extended Yale B dataset and the entire 200 images are selected to conduct

Fig. 6. The image examples from LFW Dataset.

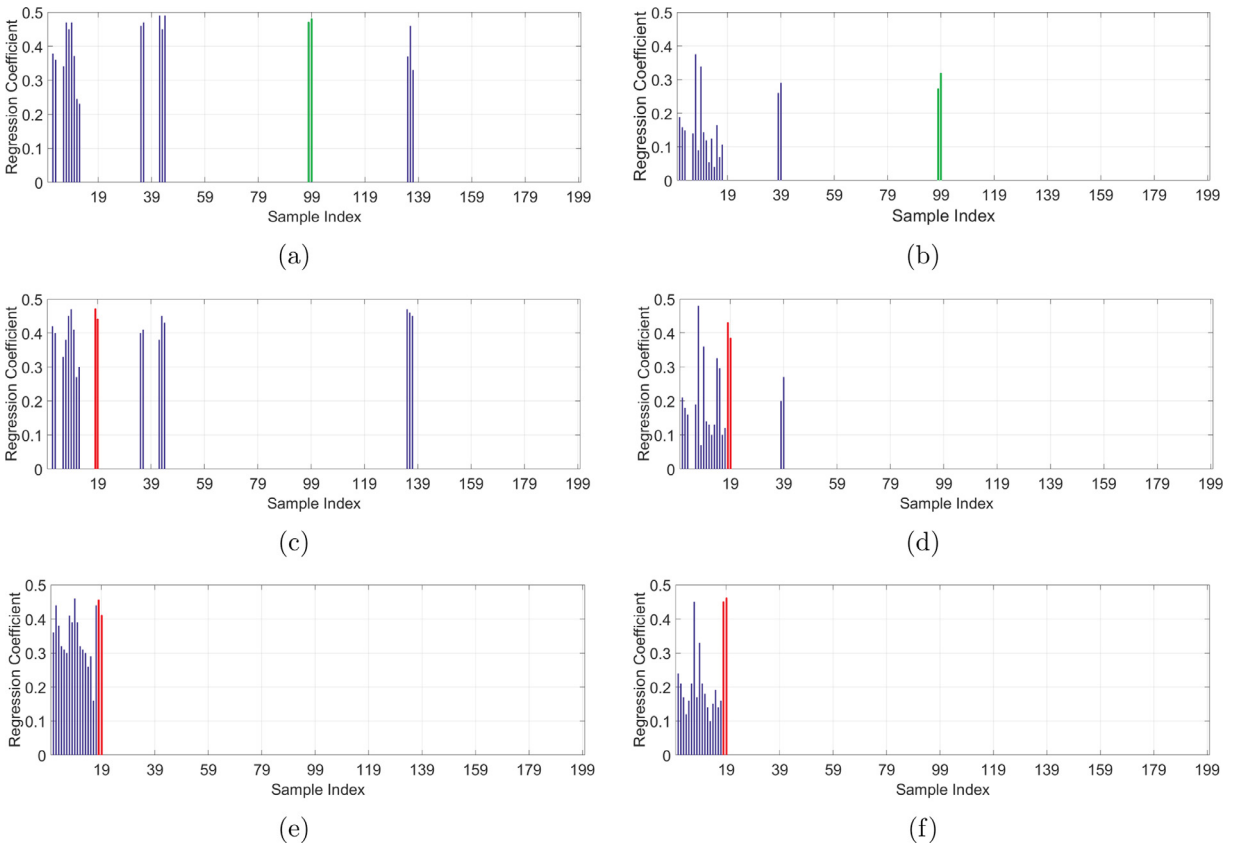


Fig. 7. The plot of the regression coefficients of the response image via different methods, where the response image is represented as the linear combination of the rest of the images and the red lines are the coefficients of must-linked samples and the green lines are the coefficients of not-linked samples. (a) EN-HG [24]. (c) EN-cHG. (e) Our proposed EN-sHG. (b) Nuclear-Gr [39]. (d) Nuclear-cHG. (f) Our proposed Nuclear-sHG. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

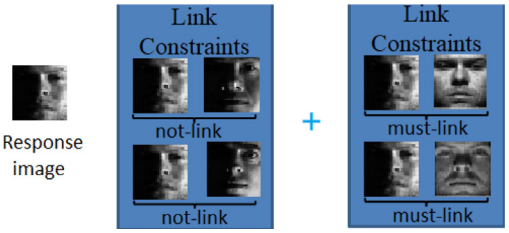


Fig. 8. The response image and the images imposed on the link constraints. For the selected 199 images using the dictionary, the 18-th and 19-th images are the must-linked samples and the 98-th and 99-th images are the not-linked samples.

the experiments. In this case, an image is used as the response image, while the remaining 199 images are used as the dictionary to represent this response image, where the must-link constraints are imposed on the 18-th and 19-th samples and not link constraints are imposed on the 98-th and 99-th samples (See Fig. 7 for the greater details). Quantitatively, only the 19 largest coefficients are retained. Fig. 7 shows the absolute values of the regression coefficients of response image.

As shown in Fig. 7, for the constraint involved methods (EN-cHG, Our proposed EN-sHG, Nuclear-cHG, our proposed Nuclear-sHG), the must-linked samples are used to represent the response image and not-linked samples are not involved in representing the response image. Compared with the methods using the link constraints (EN-cHG and Nuclear-cHG), our proposed two methods (EN-sHG and Nuclear-sHG) adopt the constraints propagation strategy, which can more effectively choose the samples having the same class to represent the response image. For the methods without using the link information (EN-HG and Nuclear-Gr), more samples of the other class are involved in representing the response image. Note that the regression coefficients of several none mask-linked samples or not-link image are big. The possible reason is that these samples are contaminated by noise or the data is corrupted. Thus, these corrupted or noisy samples may be close to the response image, whose coefficients are big (Fig. 8).

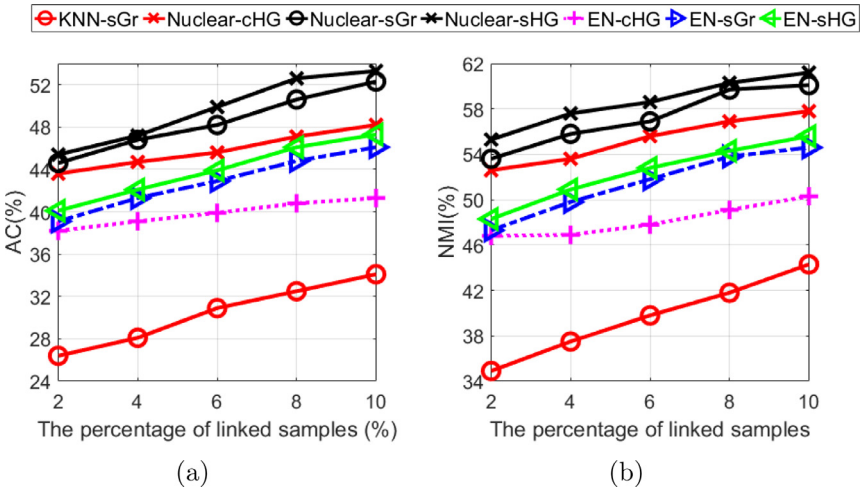


Fig. 9. The clustering results with varied percentages of linked samples. (a) The AC results vs the percentage of link samples. (b) The NMI results vs the percentage of the linked samples.

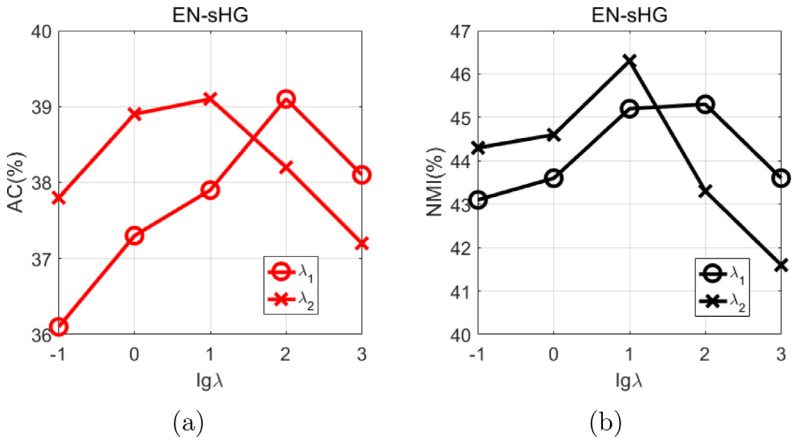


Fig. 10. Parameter tunning experimental results of EN-sHG. (a) The AC results vs the different values of λ_1 and λ_2. (b) The NMI results vs the different values of λ_1 and λ_2.

The experimental results demonstrate that the link constraints can enhance the representation ability, which is robust to noise and data corruption. For the exploitation of link information, directly adopting the link constraints have the limited learning effects, whereas, our proposed two methods adopt the propagated constraints scheme, which achieves the better learning performance.

5.7. On the link effects and parameter tuning

To evaluate how link information of the data affects the clustering results, we further conducted the clustering experiments using the AR dataset (containing 2600 images) with varying percentages of linked samples. Fig. 9 shows the experimental results.

As shown in Fig. 9, as the percentages of the linked samples increases, performance of these methods firstly increases accordingly and then becomes fairly stable. For the different percentages of linked samples, our proposed Nuclear-sHG is superior to Nuclear-cHG and Nuclear-HG; our proposed EN-sHG is superior to EN-cHG and EN-HG. The experimental results demonstrate that our methods leverage the link information to enhance performance.

Furthermore, we present the parameter tuning results to check how different parameter settings affect the clustering results. (1) For our proposed EN-sHG method, it has two key parameters λ_1 and λ_2, where λ_1 is the ℓ_1-norm regularization parameter, and λ_2 is the Frobinous-norm regularization parameter. (2) For our proposed Nuclear-sHG, we tune parameter λ_3, which is the ℓ_2-norm regularization parameter. Figs. 10 and 11 shows the clustering results with different parameter settings on the AR dataset, where 1% samples are imposed on the link constraints.

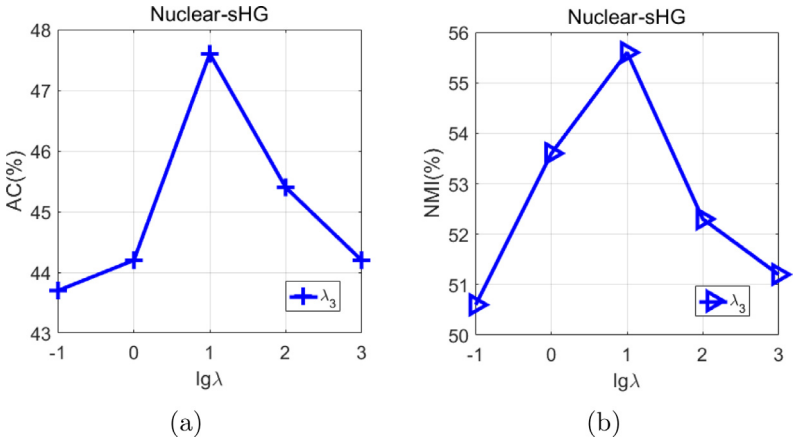


Fig. 11. The parameter tuning experimental results of Nuclear-sHG. (a) The AC results vs the different values of λ_3. (b) The NMI results vs the different values of λ_3.

As shown in Figs. 10 and 11, our proposed EN-sHG method achieves good clustering performance with different values of parameter λ_1. When the parameter value is larger than 100, clustering performance begins to degenerate. The above experimental results demonstrate that the ℓ_1-norm regularization is crucial to discover the discriminative structures of the data.

As the values of parameters λ_2 and λ_3 increase, our proposed EN-sHG and Nuclear-sHG methods obtains good clustering performance. When the parameter values are large ($\lambda_2 > 10$ and $\lambda_3 > 10$), the corresponding performances begin to be degraded a little. The experimental results demonstrate that the ℓ_2-norm regularization is crucial to discover the grouping structures of the data for generating the informative hyperedges.

5.8. Further discussions

Our proposed link-aware hypergraph model and its two variants outperform the compared methods on the noisy data with different illuminations, contiguous occlusions, real-world occlusions and the uncontrolled setting. We consider superiority of performance is attributed to the following two reasons:

On the one hand, the superiority of our model lies in that our proposed two methods belong to the representation-based hypergraph model, which generates a set of hyperedges by solving a linear regression problem. Since the samples with the large regression coefficients are more likely to reflect the affinity relations of the data, the generated hyperedges are robust to noise by linking each vertex to the samples with the large coefficients.

On the other hand, our model further enhances the robustness to noise by exploiting the link information, which is cheaply available. It is worth emphasizing that our model doesn't directly exploit the link information, which has limited effects on the linked samples. Instead, our model propagates the link information of a small number of samples to the entire dataset, where the learned regression coefficients are affected by both the data reconstruction and the link constraints imposed on a small number of samples. Thus, the link-aware and noise resistant hyperedges are generated, which results in the better performance on noisy datasets.

6. Conclusion

In this article, we proposed a novel hypergraph model, which belongs to the representation-based hypergraph construction scope. Compared with the existing methods, our proposed hypergraph model has the following advantages: our methods construct a hypergraph in a semi-supervised manner, where the link information of the data are leveraged to generate a link-aware hyperedge set. Specifically, the adjusted coefficients matrix obtained by link constraint propagation is used for hyperedge generation. Currently, the real-world data is usually contaminated by different types of noise and outliers, which are one key issue for image processing. Thus, it is crucial to develop the novel hypergraph models to be robust against various data noise. In the future, we plan to extend our method to handle more complex noise.

Declaration of Competing Interest

The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affi liations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (no. 2017YFC0113000, no. 2016YFB1001503, no. 2018YFC0830105), by the Natural Science Foundation of China (no. U1705262, no. 61802324, no. 61732005, no. 61772443, no. 61672361, no. 61702136, no. 61572410) and by the Beijing Municipal Education Commission-Beijing Natural Fund Joint Funding Project (KZ201910028039).

References

[1] L. Cao, X. Liu, W. Liu, R. Ji, T.S. Huang, Localizing web videos using social images, Inf. Sci. 302 (2015) 122–131.
[2] G. Chen, J. Zhang, F. Wang, C. Zhang, Y. Gao, Efficient multi-label classification with hypergraph regularization, in: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE, 2009, pp. 1658–1665.
[3] R. Datta, D. Joshi, J. Li, J.Z. Wang, Image retrieval: ideas, influences, and trends of the new age, ACM Comput. Surv. (Csur) 40 (2) (2008) 1–66.
[4] W. Ding, C. Lin, M. Prasad, Hierarchical co-evolutionary clustering tree-based rough feature game equilibrium selection and its application in neonatal cerebral cortex MRI, Expert Syst. Appl. 101 (2018) 243–257, doi:10.1016/j.eswa.2018.01.053.
[5] D. Du, H. Qi, L. Wen, Q. Tian, Q. Huang, S. Lyu, Geometric hypergraph learning for visual tracking, IEEE Trans. Cybern. 47 (12) (2017) 4182–4195.
[6] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, et al., Least angle regression, Ann. Stat. 32 (2) (2004) 407–499.
[7] L. Gao, J. Song, F. Nie, Y. Yan, N. Sebe, H.T. Shen, Optimal graph learning with partial tags and multiple features for image and video annotation, IEEE Trans. Cybern. 44 (7) (2014) 1225–1236.
[8] Y. Gao, M. Wang, R. Ji, Z. Zha, J. Shen, K-partite graph reinforcement and its application in multimedia information retrieval, Inf. Sci. 194 (2012) 224–239.
[9] T. Goldstein, B. O'Donoghue, S. Setzer, R. Baraniuk, Fast alternating direction optimization methods, SIAM J. Imaging Sci. 7 (3) (2014) 1588–1623.
[10] C. Hong, J. Yu, J. Wan, D. Tao, M. Wang, Multimodal deep autoencoder for human pose recovery, IEEE Trans. Image Process. 24 (12) (2015) 5659–5670.
[11] G.B. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled faces in the wild: a database for studying face recognition in unconstrained environments, Technical Report, Technical Report 07–49, University of Massachusetts, Amherst, 2007.
[12] S. Huang, D. Yang, B. Liu, X. Zhang, Regression-based hypergraph learning for image clustering and classification, arXiv: Comput. Vis. Pattern Recognit. (2016).
[13] Y. Huang, Q. Liu, F. Lv, Y. Gong, D.N. Metaxas, Unsupervised image categorization by hypergraph partition, IEEE Trans. Pattern Anal. Mach. Intell. 33 (6) (2011) 1266–1273.
[14] Y. Huang, Q. Liu, S. Zhang, D.N. Metaxas, Image retrieval via probabilistic hypergraph ranking, in: Computer Vision and Pattern Recognition, 2010 IEEE Conference on, IEEE, 2010, pp. 3376–3383.
[15] M. Jian, C. Jung, Interactive image segmentation using adaptive constraint propagation, IEEE Trans. Image Process. 25 (3) (2016) 1301–1311.
[16] T. Jin, J. Yu, J. You, K. Zeng, C. Li, Z. Yu, Low-rank matrix factorization with multiple hypergraph regularizer, Pattern Recognit. 48 (3) (2015) 1011–1022.
[17] T. Jin, Z. Yu, L. Li, C. Li, Multiple graph regularized sparse coding and multiple hypergraph regularized sparse coding for image representation, Neurocomputing 154 (2015) 245–256.
[18] A. Kapoor, H. Ahn, Y. Qi, R.W. Picard, Hyperparameter and kernel learning for graph based semi-supervised classification, in: Advances in Neural Information Processing Systems, 2006, pp. 627–634.
[19] D. Klein, S.D. Kamvar, C.D. Manning, From instance-level constraints to space-level constraints: making the most of prior knowledge in data clustering, Technical Report, Stanford, 2002.
[20] B. Kulis, S. Basu, I. Dhillon, R. Mooney, Semi-supervised graph clustering: a kernel approach, Mach. Learn. 74 (1) (2009) 1–22.
[21] K.-C. Lee, J. Ho, D.J. Kriegman, Acquiring linear subspaces for face recognition under variable lighting, IEEE Trans. Pattern Anal. Mach. Intell. 27 (5) (2005) 684–698.
[22] H. Lee-Kwang, C.H. Cho, Hierarchical reduction and partition of hypergraph, IEEE Trans. Syst. Man Cybern.Part B 26 (2) (1996) 340–344.
[23] X. Li, W. Hu, C. Shen, A.R. Dick, Z.M. Zhang, Context-aware hypergraph construction for robust spectral clustering.
[24] Q. Liu, Y. Sun, C. Wang, T. Liu, D. Tao, Elastic net hypergraph learning for image clustering and semi-supervised classification, IEEE Trans. Image Process. 26 (1) (2017) 452–463.
[25] W. Liu, D. Tao, Multiview hessian regularization for image annotation, IEEE Trans. Image Process. 22 (7) (2013) 2676–2687.
[26] X. Liu, M. Wang, B.-C. Yin, B. Huet, X. Li, Event-based media enrichment using an adaptive probabilistic hypergraph model, IEEE Trans. Cybern. 45 (11) (2015) 2461–2471.
[27] R. Lu, W. Xu, Y. Zheng, X. Huang, Visual tracking via probabilistic hypergraph ranking, IEEE Trans. Circuits Syst. Video Technol. 27 (4) (2017) 866–879.
[28] Z. Lu, Y. Peng, Exhaustive and efficient constraint propagation: a graph-based learning approach and its applications, Int. J. Comput. Vis. 103 (3) (2013) 306–325.
[29] M. Mao, J. Lu, J. Han, G. Zhang, Multiobjective E-commerce recommendations based on hypergraph ranking, Inf. Sci. 471 (2019) 269–287.
[30] A. Martínez, R. Benavente, The ar face database, 1998, Computer Vision Center, Technical Report 3 (2007) 5.
[31] A.M. Martinez, The ar face database, CVC Technical Report 24 (1998).
[32] S.A. Nene, S.K. Nayar, H. Murase, et al., Columbia object image library (1996).
[33] J. Qian, L. Luo, J. Yang, F. Zhang, Z. Lin, Robust nuclear norm regularized regression for face recognition with occlusion, Pattern Recognit. 48 (10) (2015) 3145–3159.
[34] M.H. Rohban, H.R. Rabiee, Supervised neighborhood graph construction for semi-supervised classification, Pattern Recognit. 45 (4) (2012) 1363–1372.
[35] D. Wang, X. Gao, X. Wang, Semi-supervised nonnegative matrix factorization via constraint propagation, IEEE Trans. Cybern. 46 (1) (2016) 233–244.
[36] M. Wang, H. Li, D. Tao, K. Lu, X. Wu, Multimodal graph-based reranking for web image search, IEEE Trans. Image Process. 21 (11) (2012) 4649–4661.
[37] M. Wang, X. Liu, X. Wu, Visual classification by ℓ_1-hypergraph modeling, IEEE Trans. Knowl. Data Eng. 27 (9) (2015) 2564–2574.
[38] Q. Wang, Z. Gong, An application of fuzzy hypergraphs and hypergraphs in granular computing, Inf. Sci. 429 (2018) 296–314.
[39] J. Yang, L. Luo, J. Qian, Y. Tai, F. Zhang, Y. Xu, Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes, IEEE Trans. Pattern Anal. Mach. Intell. 39 (1) (2017) 156–171.
[40] J. Yu, Y. Rui, Y.Y. Tang, D. Tao, High-order distance-based multiview stochastic learning in image classification, IEEE Trans. Cybern. 44 (12) (2014) 2431–2442.
[41] J. Yu, D. Tao, J. Li, J. Cheng, Semantic preserving distance metric learning and applications, Inf. Sci. 281 (2014) 674–686.
[42] J. Yu, D. Tao, M. Wang, Adaptive hypergraph learning and its application in image classification, IEEE Trans. Image Process. 21 (7) (2012) 3262–3272.
[43] J. Yu, X. Yang, F. Gao, D. Tao, Deep multimodal distance metric learning using click constraints for image ranking, IEEE Trans. Cybern. 47 (12) (2017) 4014–4024.
[44] R. Zass, A. Shashua, Probabilistic graph and hypergraph matching, in: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, IEEE, 2008, pp. 1–8.
[45] L. Zhang, Y. Gao, C. Hong, Y. Feng, J. Zhu, D. Cai, Feature correlation hypergraph: exploiting high-order potentials for multimodal recognition, IEEE Trans Cybern 44 (8) (2014) 1408–1419.
[46] Z. Zhang, L. Bai, Y. Liang, E. Hancock, Joint hypergraph learning and sparse regression for feature selection, Pattern Recognit. 63 (2017) 291–309.

[47] D. Zhou, O. Bousquet, T.N. Lal, J. Weston, B. Schölkopf, Learning with local and global consistency, in: Advances in Neural Information Processing Systems 16 [Neural Information Processing Systems, NIPS 2003, December 8–13, 2003, Vancouver and Whistler, British Columbia, Canada], 2003, pp. 321–328.
[48] D. Zhou, J. Huang, B. Schölkopf, Learning with hypergraphs: clustering, classification, and embedding, in: Advances in neural information processing systems, 2007, pp. 1601–1608.
[49] L. Zhuang, Z. Zhou, S. Gao, J. Yin, Z. Lin, Y. Ma, Label information guided graph construction for semi-supervised learning, IEEE Trans. Image Process. 26 (9) (2017) 4182–4192.
[50] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, J. R. Stat. Soc. Ser.B-stat. Methodol. 67 (2) (2005) 301–320.