

# Improved Spectral Clustering via Embedded Label Propagation

Xiaojun Chang<sup>1</sup>, Feiping Nie<sup>2</sup>, Yi Yang<sup>1</sup> and Heng Huang<sup>2</sup>

<sup>1</sup>School of Information Technology & Electrical Engineering, The University of Queensland.

<sup>2</sup>Department of Computer Science and Engineering, University of Texas at Arlington.  
cxj273@gmail.com, feipingnie@gmail.com, yi.yang@uq.edu.au, heng@uta.edu.

## Abstract

Spectral clustering is a key research topic in the field of machine learning and data mining. Most of the existing spectral clustering algorithms are built upon Gaussian Laplacian matrices, which are sensitive to parameters. We propose a novel parameter-free, distance-consistent Locally Linear Embedding (LLE). The proposed distance-consistent LLE promises that edges between closer data points have greater weight. Furthermore, we propose a novel improved spectral clustering via embedded label propagation. Our algorithm is built upon two advancements of the state of the art: 1) label propagation, which propagates a node's labels to neighboring nodes according to their proximity; and 2) manifold learning, which has been widely used in its capacity to leverage the manifold structure of data points. First we perform standard spectral clustering on original data and assign each cluster to  $k$ -nearest data points. Next, we propagate labels through dense, unlabeled data regions. Extensive experiments with various datasets validate the superiority of the proposed algorithm compared to current state-of-the-art spectral algorithms.

## Introduction

Data clustering is a fundamental research topic that is widely used for many applications in the fields of artificial intelligence, statistics and the social sciences (Jain, Murty, and Flynn 1999)(Jain and Dubes 1988a)(Girolami 2002)(Ye, Zhao, and Liu 2007a). The purpose of clustering is to partition the original data points into various groups so that data points within the same cluster are dense while those in different clusters are far apart (Jain and Dubes 1988b)(Filippone et al. 2008a)(Chang et al. 2015).

Among various implementations of clustering,  $k$ -means is one of the most popular in reality because of its simplicity and effectiveness (Wu et al. 2012). The general procedure of traditional  $k$ -means (TKM) is to randomly initialize  $c$  clustering centers, assign each data point to its nearest cluster and compute a new clustering center. Current researchers claim the curse of dimensionality may deteriorate the performance of TKM (Ding and Li 2007). A straight-forward solution to this problem is to project original datasets to

a low-dimensional subspace by dimensionality reduction, *i.e.*, PCA, before performing TKM. Discriminative analysis has proven effective in enhancing clustering performance (Ding and Li 2007) (la Torre, Fernando, and Kanade 2006) (Ye, Zhao, and Liu 2007b). Thus, discriminative  $k$ -means (DKM) (Ye, Zhao, and Wu 2007) is proposed as a way to incorporate discriminative analysis and clustering into a single framework to formalize clustering as a trace maximization problem. However, TKM and DKM fail to take low-dimensional manifold structure of data into consideration.

On a further note, spectral clustering (SC) (Yu and Shi 2003) (Filippone et al. 2008b) (Shi and Malik 2000) has gradually attracted increasing research attention for its capacity to mine intrinsic data geometric structures, which facilitates partitioning data with more complicated structures (Belkin and Niyogi 2003) (Yang et al. 2011) (Nie et al. 2009) (Wu and Schlkopf 2006) (Yang et al. 2010). The basic idea of SC is to find a cluster assignment of data points by adopting a spectrum of the similarity matrix that leverages the nonlinear and low-dimensional manifold structure of original data. Inspired by the benefits of spectral clustering, different variants of the SC method have been proposed to demonstrate its effectiveness. For example, local learning-based clustering (LLC) (Wu and Schlkopf 2006) utilizes a kernel regression model for label prediction based on the assumption that the class label of a data point can be determined by its neighbors. Similarly, self-tuning SC (Zelnik-Manor and Perona 2004) is able to tune parameters automatically in an unsupervised scenario. Normalized cuts are capable of balancing the volume of clusters for the usage of data density information (Shi and Malik 2000).

Label propagation has been shown effective in propagating labels through the dataset along high density areas defined by unlabeled data in (Zhu and Ghahramani 2002) (Wang and Zhang 2008). Central to label propagation is the following cluster assumption (Chapelle, Weston, and Schlkopf 2002): (1) nearby data points are likely to belong to the same cluster; and (2) data points on the same structures are likely to have the same label. Motivated by the benefits inherent to label propagation, we intend to introduce label propagation into the field of spectral clustering (Kang, Jin, and Sukthankar 2006)

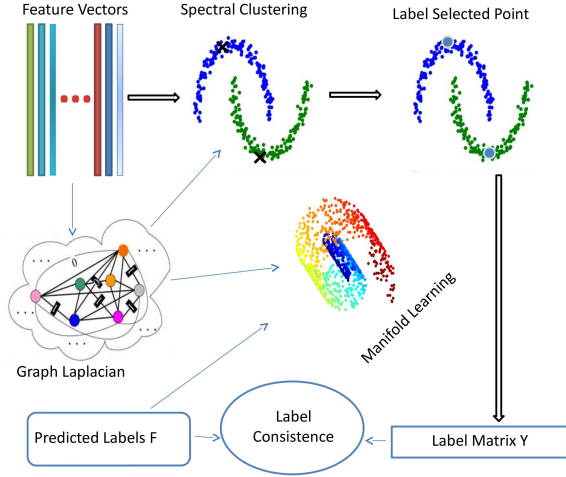


Figure 1: Framework of the proposed algorithm

(Cao, Luo, and Huang 2008) (Cheng, Liu, and Yang 2009).

Our proposed spectral clustering algorithm therefore combines the strengths of spectral clustering and label propagation. The main process of our algorithm is shown in Fig. 1. We first perform standard spectral clustering on our original dataset and obtain  $c$  clusters. Then, we select  $k$  data points that are respectively close to each cluster center and form label matrix  $Y$ . By way of manifold learning, we propagate labels through dense unlabeled data regions. We call the proposed method improved Spectral Clustering via embedded Label Propagation (SCLP).

The main contributions of this paper can be summarized as follows:

1. To the best of our knowledge, this is the first time spectral clustering and embedded label propagation have been incorporated into a single framework. We propagate the labels obtained by spectral clustering to other unlabeled data points.
2. We integrate the advantage of manifold learning, which is capable of leveraging manifold structure among data points, into the proposed framework.
3. A novel distance-consistent Locally Linear Embedding is proposed herein as well. Unlike a traditional Gaussian graph approach, the proposed graph is parameter-free.
4. Extensive experiments on seven real-world datasets demonstrate that the proposed spectral clustering framework (SCLP) outperforms state-of-the-art clustering algorithms.

After revisiting related work on Locally Linear Embedding and spectral clustering in Section 2, we detail our SCLP algorithm in Section 3. Extensive experiments are given in Section 4 and Section 5 concludes this paper.

## Related Work

### Locally Linear Embedding

Locally Linear Embedding (LLE) (Roweis and Saul 2000) aims to identify low-dimensional global coordinates that lie on, or very near to, a manifold embedded in a high-dimensional space. The purpose is to combine the data points with minimal discrepancy after completing a different linear dimensionality reduction at each point.

The main procedure of LLE can be summarized in three steps: (1) build a neighborhood for each data point; (2) find the weights in order to linearly approximate the data in said neighborhood; and (3) find the low-dimensional coordinates best reconstructed by those weights.

By way of example, given a dataset matrix  $X = \{x_1, x_2, \dots, x_n\}$ , the main steps of LLE are as follows:

1. For each data point  $x_i$ , find its  $k$  nearest neighbors.
2. Compute the weight matrix  $A$  by minimizing the residual sum of squares to reconstruct each  $x_i$  from its neighbours as such:

$$\min_A \sum_{i=1}^n \left\| x_i - \sum_{j=1}^k x_j a_{ij} \right\|_2^2, \quad (1)$$

where  $a_{ij} = 0$  if  $x_j$  is not one of  $x_i$ 's  $k$ -nearest neighbours, and for each data point  $x_i$ ,  $\sum_j a_{ij} = 1$ .

3. Obtain the coordinates  $S$  by minimizing the following reconstruction error using the weights:

$$\arg \min_S \sum_{i=1}^n \left\| s_i - \sum_{j=1}^k y_j a_{ij} \right\|_2^2,$$

where  $y_j$  is the cluster indicator vector for the datum  $x_j$ ,  $\sum_i S_{ij} = 0$  for each  $j$  and  $S^T S = I$ .

### Spectral Clustering

Consider a dataset  $X = \{x_1, x_2, \dots, x_n\} \in \mathbb{R}^{d \times n}$ , where  $d$  is the dimension of the data point and  $n$  is the total number of data points. The objective of clustering is to partition  $X$  into  $c$  clusters  $C_i |_{i=1}^c$  so as to keep data points within the same cluster close to one another, while data points from different clusters remain apart. Let us denote  $Y = [y_1, y_2, \dots, y_n]^T \in \mathbb{R}^{n \times c}$  as the cluster indicator matrix, where  $y_i$  is the cluster indicator vector for the datum  $x_i$ . The  $j$ -th element of  $y_i$  is 1 if  $x_i$  belongs to the  $j$ -th cluster and 0 otherwise. Following the work in (Ye, Zhao, and Wu 2007),

we denote the scaled cluster indicator matrix  $F$  as follows:

$$F = [F_1, F_2, \dots, F_n]^T = Y(Y^T Y)^{-1/2}, \quad (2)$$

where  $F_i$  is the scaled cluster indicator of  $x_i$ . The  $j$ -th column of  $F$  is defined as follows by (Ye, Zhao, and Wu 2007):

$$f_j = \left[ \underbrace{0, \dots, 0}_{\sum_{i=1}^{j-1} n_i}, \underbrace{\frac{1}{\sqrt{n_j}}, \dots, \frac{1}{\sqrt{n_j}}}_{n_j}, \underbrace{0, \dots, 0}_{\sum_{i=j+1}^c n_i} \right], \quad (3)$$

and indicates which data points are partitioned into the  $j$ -th cluster  $C_j$ . Meanwhile,  $n_j$  is the number of data points in cluster  $C_j$ .

According to (Dhillon, Guan, and Kulis 2004), the overall function of spectral clustering can be defined as follows:

$$\begin{aligned} \min_F \text{Tr}(F^T L F) \\ \text{s.t. } F = Y(Y^T Y)^{-1/2}, \end{aligned} \quad (4)$$

where  $\text{Tr}(\cdot)$  denotes the trace operator and  $L$  is a graph Laplacian matrix computed in accordance with local data structure. Among different strategies, a common way to compute the weight matrix is thus:

$$a_{ij} = \begin{cases} \exp(-\frac{\|x_i - x_j\|^2}{\delta^2}), & \text{if } x_i \text{ and } x_j \text{ are } k \text{ nearest} \\ & \text{neighbours.} \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

where  $\mathcal{N}_k(x_j)$  denotes  $k$  nearest neighbors of  $x_j$  and  $\delta$  is utilized to control the spread of neighbors. The Laplacian graph  $L$  is then computed by  $L = D - W$ , where  $D$  is a diagonal matrix with its diagonal elements as  $D_{ii} = \sum_j A_{ij}$ .

By replacing  $L$  in Eq. (4) by the normalized Laplacian matrix  $L_n$ ,

$$L_n = D^{-1/2} L D^{-1/2} = I - D^{-1/2} A D^{-1/2} \quad (6)$$

the objective function becomes the well-known SC algorithm normalized cut (Shi and Malik 2000). In the same manner, if we replace  $L$  in Eq. (4) with  $L_l$ , which is a Laplacian matrix obtain(Yang et al. 2010)ed by local learning (Wu and Scholkopf 2006), the objective function is then modified to Local Learning Clustering (LLC).

## The Proposed Framework

In this section, we illustrate the detailed framework of our algorithm.

We aim to cluster the dataset into  $c$  clusters. Suppose  $X \in \mathbb{R}^{d \times n}$  indicates the dataset;  $d$  is the dimension of data points and  $n$  is the total number of data points.

### Distance-Consistent Similarity Learning

Following the work in (Karasuyama and Mamitsuka 2013), we propose leveraging manifold regularization built upon the Laplacian graph for label propagation. To begin, we first present a novel distance-consistent Local Linear Embedding (LLE).

Intuitively, we expect close data points to have similar labels. We create a graph in which all data points are considered nodes. If  $x_i$  ( $x_j$ ) is in  $k$ -Nearest-Neighbor of  $x_j$  ( $x_i$ ), then the two nodes are connected. The edge between them is weighted so that the closer the nodes are in Euclidean distance, the larger the weight  $a_{ij}$ . Because we have  $\sum_j a_{ij} = 1$ , the objective function of LLE in Eq. (1) can be safely rewritten as follows:

$$\min_A \sum_{i=1}^n \left\| [x_i, \dots, x_i] a_{ij} - \sum_{j=1}^k x_{ij} a_{ij} \right\|_2^2, \quad (7)$$

By way of simple mathematical deduction, we can rewrite the above objective function in this manner:

$$\min_A \sum_{i=1}^n \left\| [x_i - x_{i_1}, \dots, x_i - x_{i_k}] a_i \right\|_2^2 \quad (8)$$

$$\Rightarrow \min_A \sum_{i=1}^n \text{Tr}(a_i^T [x_i - x_{i_1}, \dots, x_i - x_{i_k}]^T [x_i - x_{i_1}, \dots, x_i - x_{i_k}] a_i) \quad (9)$$

For simplicity's sake, we assign all the non-diagonal elements of  $[x_i - x_{i_1}, \dots, x_i - x_{i_k}]^T [x_i - x_{i_1}, \dots, x_i - x_{i_k}]$  to zero. The above objective function is equivalent to the following:

$$\min_A \sum_{i=1}^n \sum_{j=1}^k \|x_i - x_{i_j}\|_2^2 a_{ij}^2 \quad (10)$$

$$\text{s.t. } \sum_{j=1}^k a_{ij} = 1, a_{ij} \geq 0$$

From the above function, we can observe that the proposed distance-consistent LLE suggests that the edge between closer nodes has a greater weight.

The Lagrangian function of problem of Eq. (10) can be written as

$$\sum_{j=1}^k \|x_i - x_{i_j}\|_2^2 a_{ij}^2 - \gamma \left( \sum_{j=1}^k a_{ij} - 1 \right), \quad (11)$$

where  $\gamma$  is a Lagrange multiplier. By setting the derivative of Eq. (11) w.r.t.  $a_{ij}$  to zero, we have

$$a_{ij} = \frac{\gamma}{2 \|x_i - x_{i_j}\|_2^2}. \quad (12)$$

By substituting the resultant  $a_{ij}$  in Eq. (12) into the constraint  $\sum_{j=1}^k a_{ij} = 1$ , we arrive at

$$\gamma = \frac{1}{\sum_{j=1}^k \frac{1}{2 \|x_i - x_{i_j}\|_2^2}}. \quad (13)$$

By integrating Eq. (12) and Eq. (13), we obtain the final solution for  $a_{ij}$ .

By denoting  $D$  as a diagonal matrix with its diagonal  $d_i = \sum_j a_{ij}$ , the graph Laplacian can be calculated as

$$L = D - A. \quad (14)$$

### Refined Spectral Clustering

After initially clustering the dataset into  $c$  clusters through traditional spectral clustering, we select  $k$  data points per cluster, which are nearest to each clustering center, and mark them as labeled data points. The remaining points are marked as unlabeled data points. Note that we assume these  $k$  data points are grouped into the proper clusters. Hence, we obtain the label matrix  $Y \in \mathbb{R}^{n \times c}$  and diagonal selection matrix  $U \in \mathbb{R}^{n \times n}$ , where  $Y_{i,j} = 1$  if  $X_i$  is labeled, and  $X_i$  belongs to the  $j$ -th cluster.  $Y_{i,j} = 0$  otherwise.  $U_{ii} = \infty$  if

$X_i$  is labeled and  $U_{ii} = 0$  otherwise. In the experiment, we use  $10^{10}$  to approximate  $\infty$ . We propose propagating labels of labeled data points to unlabeled data points.

Moreover, we denote a predicted label matrix  $F \in \mathbb{R}^{n \times c}$  for the data points in  $X$ . According to (Nie et al. 2010),

$F$  should satisfy the smoothness on both the obtained label matrix  $Y$  and the manifold structure. Hence,  $F$  can be obtained as follows (Zhu 2006):

$$\min_F Tr(F^T L F) + Tr((F - Y)^T U (F - Y)), \quad (15)$$

where  $Tr(\cdot)$  denotes the trace operator. The purpose of this definition is to keep the predicted labels  $F$  consistent with the ground truth labels  $Y$ .

We further incorporate a regularization term into the objective function to correlate the features with the predicted labels. Consequently, the objective function arrives at

$$\min_F Tr(F^T L F) + Tr((F - Y)^T U (F - Y)) + \alpha \|X^T W - F\|_F^2 + \beta \|W\|_F^2 \quad (16)$$

where  $\|\cdot\|_F$  denotes the Frobenius norm of a matrix.

Since the least square loss function is very sensitive to outliers, we employ  $l_{2,1}$ -norm on the regularization term to handle this issue. Hence, we can rewrite the objective function as follows:

$$\min_F Tr(F^T L F) + Tr((F - Y)^T U (F - Y)) + \alpha \|X^T W - F\|_F^2 + \beta \|W\|_{2,1} \quad (17)$$

It is worth noting that the proposed framework can be readily applied to out-of-sample clustering. By calculating  $X^T W$ , we obtain the label predictor matrix for outside samples.

### Optimization

The proposed function involves the  $l_{2,1}$ -norm, which is difficult to solve in a closed form. We propose to solve this problem in the following steps. By setting the derivative of Eq. (17) *w.r.t.*  $W$  to zero, we have

$$W = \alpha(\alpha X X^T + \beta D)^{-1} X F, \quad (18)$$

where  $I$  is an identity matrix and  $D$  is a diagonal matrix which is defined as:

$$D = \begin{bmatrix} \frac{1}{2\|w^1\|_2} & & \\ & \ddots & \\ & & \frac{1}{2\|w^d\|_2} \end{bmatrix}. \quad (19)$$

Letting  $H$  represent  $(\alpha X X^T + \beta D)^{-1}$ , the objective becomes

$$\min_F Tr(F^T L F) + Tr((F - Y)^T U (F - Y)) + \alpha \|\alpha X^T H X F - F\|_F^2 + \beta \|\alpha H X F\|_F^2 \quad (20)$$

By denoting  $P$  as  $P = \alpha X^T H X - I$ , the objective becomes

$$\min_F Tr(F^T L F) + Tr((F - Y)^T U (F - Y)) + \alpha \|P F\|_F^2 + \beta \|\alpha H X F\|_F^2. \quad (21)$$

By setting the derivative of Eq. (21) *w.r.t.*  $F$  to zero, we have

$$F = (L + U + \alpha P P + \alpha^2 \beta X^T H H X)^{-1} U Y \quad (22)$$

Based on the above mathematical deduction, we propose an efficient iterative algorithm to optimize the objective function Eq. (17), which is summarized in Algorithm 1.

---

#### Algorithm 1: Optimization Algorithm for SCLP

---

**Data:** Data  $X \in \mathbb{R}^{d \times n}$

The number of clusters  $c$

Parameters  $\alpha$  and  $\beta$

**Result:** The cluster indicator matrix  $F$

- 1 Construct the Laplacian matrix  $L$  according to Eq. (14) ;
  - 2 Compute the selecting matrix  $U \in \mathbb{R}^{n \times n}$  ;
  - 3 Cluster data  $X$  into  $c$  clusters through Spectral Clustering ;
  - 4 Pick out  $n_L$  data points that are close to each cluster center and construct  $Y$  ;
  - 5 **repeat**
  - 6     Compute  $D$  according to
 
$$D = \begin{bmatrix} \frac{1}{2\|w^1\|_2} & & \\ & \ddots & \\ & & \frac{1}{2\|w^d\|_2} \end{bmatrix}.$$
  - 7     Compute  $H$  according to  $H = (\alpha X X^T + \beta D)^{-1}$  ;
  - 8     Compute  $P$  according to  $P = \alpha X^T H X - I$  ;
  - 9     Compute  $F$  according to
 
$$F = (L + U + \alpha P P + \alpha^2 \beta X^T H H X)^{-1} (U * Y) ;$$
  - 10 **until** *Convergence* ;
  - 11 **Return**  $F$
- 

## Experiments

In this section, we conduct extensive experiments to validate the proposed SCLP's performance and compare it with related state-of-the-art spectral clustering algorithms, following a study of parameter sensitivity.

### Dataset Description

We use seven trademark datasets to validate the performance of the proposed algorithm. The USPS dataset has 9298 gray-scale handwritten digital images with an image size of 256 scanned from envelopes with the U.S. Postal Service. The Yale-B dataset (Georghiadis, Belhumeur, and Kriegman 2001)

Table 1: DATASET DETAILS

Dataset	Matrix Size	Dataset Size	Class #
LUNG	3312	203	4
PALM	256	2000	100
MSRA50	1024	1799	12
FRGC	1296	5658	275
AR	768	840	120
YaleB	1024	2414	38
USPS	256	9298	10

consists of 2414 near frontal images from 38 persons under different illuminations. The AR dataset (Martinez and Benavente 1998) has 840 images with a dimension of 768. The FRGC dataset (Phillips et al. 2005) was collected at the University of Notre Dame and contains 50,000 images. All the images were taken across 13 different poses, under 43 different illumination conditions and with four different expressions per person. The MSRA50 dataset (He et al. 2004) consists of 1799 images and 12 classes. The PALM dataset consists 700 right-hand images, seven samples per person across 100 users, taken via digital camera. The images are resized to the same dimension of  $100 \times 100$ . The human lung carcinomas (LUNG) dataset (Singh et al. 2002) contains 203 samples and 3312 genes.

## Experiment Setup

We compare the proposed SCLP with traditional k-means (TKM) (Wu et al. 2012), discriminative k-means (DKM) (Ye, Zhao, and Wu 2007), Local Learning Clustering (LLC) (Wu and Schlkopf 2006), Non-negative Normalized Cut (NNC) (Shi and Malik 2000), Spectral Clustering (SC), LLC (Wu and Schlkopf 2006), CLGR (Wang, Zhang, and Li 2009) and Spectral Embedding Clustering (SEC) (Nie et al. 2009).

The size of neighborhood  $k$  is set to 5 for all spectral clustering algorithms. For the parameter,  $\delta$ , in NNC, we perform a self-tuning algorithm (Zelnik-Manor and Perona 2004) to determine the best parameter. For parameters in DKM, LLC, CLGR and SEC, we tune them in the range of  $\{10^{-6}, 10^{-4}, 10^{-2}, 10^0, 10^2, 10^4, 10^6\}$  and report the best results. Note that the results of all clustering algorithms vary based on initialization. To reduce the influence of statistical variation, we repeat each clustering 50 times with random initialization and report the results according to the best objective function values. For SCLP, we select 2 data points per cluster nearest to the clustering center.

## Evaluation Metrics

Following related clustering studies, we utilize clustering accuracy (ACC) and normalized mutual information (NMI) as our experiments' evaluation metrics.

Let  $q_i$  represent the clustering label result from a clustering algorithm;  $p_i$  represent the corresponding ground truth

label of arbitrary data point  $x_i$ . From there, ACC is defined as follows:

$$ACC = \frac{\sum_{i=1}^n \delta(p_i, \text{map}(q_i))}{n}, \quad (23)$$

where  $\delta(x, y) = 1$  if  $x = y$  and  $\delta(x, y) = 0$  otherwise.  $\text{map}(q_i)$  is the best mapping function that permutes clustering labels to match the ground truth labels using the Kuhn-Munkres algorithm. A larger ACC indicates a better clustering performance.

For any two arbitrary variables  $P$  and  $Q$ , NMI is defined as follows (Strehl and Ghosh 2003)

$$NMI = \frac{I(P, Q)}{\sqrt{H(P)H(Q)}}, \quad (24)$$

where  $I(P, Q)$  computes mutual information between  $P$  and  $Q$ , and  $H(P)$  and  $H(Q)$  are the entropies of  $P$  and  $Q$ . Let  $t_l$  represent the number of data points in the cluster  $C_l$  ( $1 \leq l \leq c$ ) generated by a clustering algorithm and  $\tilde{t}_h$  represent the number of data points from the  $h$ -th ground truth class. The NMI metric is then computed as follows (Strehl and Ghosh 2003)

$$NMI = \frac{\sum_{l=1}^c \sum_{h=1}^c t_{l,h} \log\left(\frac{n \times t_{l,h}}{t_l \tilde{t}_h}\right)}{\sqrt{(\sum_{l=1}^c t_l \log \frac{t_l}{n}) (\sum_{h=1}^c \tilde{t}_h \log \frac{\tilde{t}_h}{n})}}, \quad (25)$$

where  $t_{l,h}$  is the number of data samples that lies in the intersection between  $C_l$  and  $h$ th ground truth class. Similarly, a larger NMI indicates a better clustering performance.

## Experimental Results

We show the clustering results of different algorithms in terms of ACC and NMI over seven benchmark datasets in Tables 2 and 3. Based on the results of our experiment, we can make the following observations:

1. When comparing the k-means based algorithms (*i.e.*, TKM and DKM), DKM generally outperforms TKM because discriminative dimension reduction is integrated into a single framework. Thus each cluster is more identifiable and facilitates clustering performance. We can therefore safely conclude that discriminative information is beneficial for clustering.
2. SC outperforms LLC on the YaleB and USPS datasets, while LLC outperforms SC on all those remaining. That is, CLGR achieves better performance on all datasets than both algorithms combined.
3. SEC obtains the second-best performance over the seven datasets, which indicates that linearity regularization can also facilitate clustering performance. Similar to our algorithm, SEC is capable of dealing with out-of-sample data.
4. The proposed algorithm SCLP generally outperforms the compared clustering algorithms on the seven benchmark datasets, which demonstrates that manifold regularization-based label propagation is beneficial for spectral clustering.

Table 2: Performance Comparison(ACC %) of KM, DKM, NNC, SC, LLC, CLGR, SEC and SCLP. The proposed algorithm, SCLP, generally outperforms the compared algorithms, which indicates that manifold regularization-based label propagation is beneficial for spectral clustering.

	KM	DKM	NNC	SC	LLC	CLGR	SEC	SCLP
LUNG	82.8 ± 1.6	84.1 ± 1.9	83.2 ± 1.5	83.4 ± 2.1	85.2 ± 1.7	85.7 ± 1.8	87.8 ± 1.9	<b>91.6 ± 1.3</b>
PALM	73.4 ± 2.3	73.9 ± 2.1	73.8 ± 1.9	74.3 ± 2.4	74.8 ± 1.8	76.3 ± 2.0	77.8 ± 1.8	<b>79.3 ± 1.6</b>
MSRA50	52.9 ± 1.4	63.4 ± 1.0	55.3 ± 1.3	58.1 ± 1.9	65.8 ± 2.3	67.3 ± 2.1	69.8 ± 1.8	<b>71.7 ± 1.7</b>
FRGC	32.0 ± 1.8	40.3 ± 2.2	34.6 ± 2.1	36.1 ± 1.4	36.6 ± 1.8	38.9 ± 1.6	42.4 ± 1.9	<b>46.1 ± 1.6</b>
AR	37.4 ± 2.1	43.8 ± 1.8	40.5 ± 2.2	42.6 ± 1.9	47.8 ± 1.8	48.2 ± 2.0	60.1 ± 1.8	<b>61.9 ± 1.7</b>
YaleB	13.6 ± 0.9	37.2 ± 1.9	36.5 ± 1.8	35.2 ± 2.1	32.8 ± 1.8	37.3 ± 2.3	46.3 ± 2.1	<b>48.4 ± 1.9</b>
USPS	66.5 ± 3.0	71.2 ± 2.6	67.8 ± 2.8	65.3 ± 3.1	64.6 ± 2.5	65.8 ± 2.8	70.6 ± 2.4	<b>75.1 ± 2.7</b>

Table 3: Performance Comparison(NMI %) of KM, DKM, NNC, SC, LLC, CLGR, SEC and SCLP. The proposed algorithm, SCLP, generally outperforms the compared algorithms, which indicates that manifold regularization-based label propagation is beneficial for spectral clustering.

	KM	DKM	NNC	SC	LLC	CLGR	SEC	SCLP
LUNG	59.1 ± 1.8	60.4 ± 1.6	59.7 ± 1.4	60.0 ± 1.7	63.8 ± 1.9	64.9 ± 2.1	68.2 ± 1.5	<b>73.3 ± 1.5</b>
PALM	81.8 ± 2.3	83.4 ± 1.9	82.1 ± 2.1	85.6 ± 1.8	87.2 ± 2.0	87.5 ± 1.8	89.3 ± 1.6	<b>91.4 ± 1.9</b>
MSRA50	61.7 ± 2.4	63.4 ± 2.2	62.5 ± 1.9	62.8 ± 2.5	69.4 ± 2.1	68.5 ± 1.8	70.9 ± 1.7	<b>73.3 ± 1.6</b>
FRGC	66.3 ± 1.8	68.5 ± 1.4	67.8 ± 1.7	66.7 ± 1.5	68.9 ± 1.9	69.2 ± 2.1	71.2 ± 1.8	<b>73.9 ± 1.4</b>
AR	69.3 ± 2.5	73.8 ± 2.1	72.1 ± 2.4	73.7 ± 1.9	74.2 ± 2.1	74.6 ± 1.7	75.3 ± 2.2	<b>77.4 ± 1.8</b>
YaleB	17.3 ± 1.8	37.2 ± 2.2	36.5 ± 2.5	35.2 ± 1.9	32.8 ± 2.4	37.3 ± 1.8	<b>67.6 ± 1.6</b>	67.1 ± 2.3
USPS	60.2 ± 1.2	62.4 ± 1.7	61.8 ± 3.5	63.2 ± 2.8	63.9 ± 2.4	65.2 ± 1.8	69.6 ± 2.4	<b>74.4 ± 1.9</b>

### Parameter Sensitivity

In this section, we study performance variance *w.r.t.* on the regularization parameters  $\alpha$  and  $\beta$ .

We use the MSRA50 dataset for these experiments. Fig. 2 shows how clustering performance varies *w.r.t.* different combinations of  $\alpha$  and  $\beta$ . We can see that better performance occurs when  $\alpha$  and  $\beta$  are comparable.

### Conclusion

In this paper, we have proposed a novel improved spectral clustering algorithm SCLP. Most of the existing spectral clustering algorithms are based on Gaussian matrices or LLE, each of which are extremely sensitive to parameters. Moreover, the parameters are difficult to tune. We have therefore presented a novel distance-consistent LLE that is parameter-free. The distance-consistent LLE can promise that the edge between closer data points has a greater weight. Utilizing this distance-consistent LLE, we have proposed an improved means of spectral clustering via label propagation. The proposed algorithm takes advantage of label propagation and manifold learning. With label propagation, we can propagate the labels obtained through spectral clustering to other unlabeled data points. By adopting manifold learning, we leverage the manifold structure among data points. Note that our framework can also be readily applied to out-of-sample data. Finally, we have evaluated the clustering per-

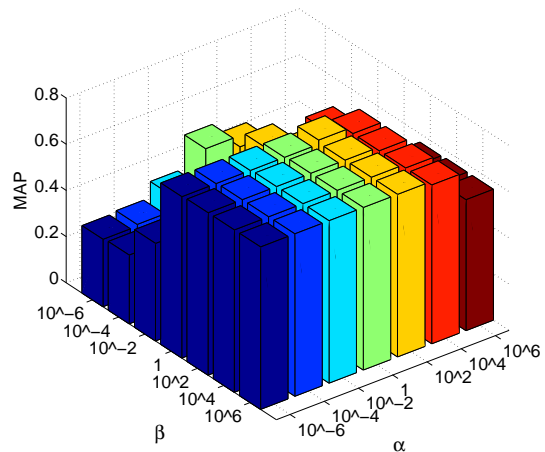


Figure 2: Performance variance *w.r.t.*  $\alpha$  and  $\beta$ .

formance of the proposed algorithm over seven datasets. The experimental results demonstrate that the proposed algorithm consistently outperforms other algorithms to which it is compared.

## References

- [Belkin and Niyogi 2003] Belkin, M., and Niyogi, P. 2003. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation* 15(6):1373–1396.
- [Cao, Luo, and Huang 2008] Cao, L.; Luo, J.; and Huang, T. S. 2008. Annotating photo collections by label propagation according to multiple similarity cues. In *ACM Multimedia*, 121–130.
- [Chang et al. 2015] Chang, X.; Nie, F.; Ma, Z.; Yang, Y.; and Zhou, X. 2015. A convex formulation for shrunk spectral clustering. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, January 2530, 2015, Austin Texas, USA*.
- [Chapelle, Weston, and Schlkopf 2002] Chapelle, O.; Weston, J.; and Schlkopf, B. 2002. Cluster kernels for semi-supervised learning. In *Proc. NIPS*, 585–592.
- [Cheng, Liu, and Yang 2009] Cheng, H.; Liu, Z.; and Yang, J. 2009. Sparsity induced similarity measure for label propagation. In *Proc. CVPR*, 317–324.
- [Dhillon, Guan, and Kulis 2004] Dhillon, I. S.; Guan, Y.; and Kulis, B. 2004. Kernel k-means: spectral clustering and normalized cuts. In *Proc. ACM SIGKDD*, 551–556.
- [Ding and Li 2007] Ding, C., and Li, T. 2007. Adaptive dimension reduction using discriminant analysis and k-means clustering. In *Proc. ICML*, 521–528.
- [Filippone et al. 2008a] Filippone, M.; Camastra, F.; Masulli, F.; and Rovetta, S. 2008a. A survey of kernel and spectral methods for clustering. *Pattern recognition* 41(1):176–190.
- [Filippone et al. 2008b] Filippone, M.; Camastra, F.; Masulli, F.; and Rovetta, S. 2008b. A survey of kernel and spectral methods for clustering. *Pattern Recognition* 41(1):176–190.
- [Georghiades, Belhumeur, and Kriegman 2001] Georghiades, A. S.; Belhumeur, P. N.; and Kriegman, D. J. 2001. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. PAMI* 23(6):643–660.
- [Girolami 2002] Girolami, M. 2002. Mercer kernel-based clustering in feature space. *IEEE Trans. Neural Networks* 13(3):780–784.
- [He et al. 2004] He, X.; Yan, S.; Hu, Y.; Niyogi, P.; and Zhang, H.-J. 2004. Face recognition using laplacianfaces. *IEEE Trans. PAMI* 27(3):328–340.
- [Jain and Dubes 1988a] Jain, A. K., and Dubes, R. C. 1988a. *Algorithms for clustering data*. Prentice-Hall, Inc.
- [Jain and Dubes 1988b] Jain, A. K., and Dubes, R. C. 1988b. *Algorithms for clustering data*. Englewood Cliffs, NJ: Prentice-Hall.
- [Jain, Murty, and Flynn 1999] Jain, A. K.; Murty, M. N.; and Flynn, P. J. 1999. Data clustering: a review. *ACM computing surveys* 31(3):264–323.
- [Kang, Jin, and Sukthankar 2006] Kang, F.; Jin, R.; and Sukthankar, R. 2006. Correlated label propagation with application to multi-label learning. In *Proc. CVPR*, 1719–1726.
- [Karasuyama and Mamitsuka 2013] Karasuyama, M., and Mamitsuka, H. 2013. Manifold-based similarity adaptation for label propagation. In *Proc. NIPS*, 1547–1555.
- [la Torre, Fernando, and Kanade 2006] la Torre, D.; Fernando; and Kanade, T. 2006. Discriminative cluster analysis. In *Proc. ICML*, 241–248.
- [Martinez and Benavente 1998] Martinez, A. M., and Benavente, R. 1998. The ar face database. Technical report, Centre Vis. Comput., Univ. Autonoma Barcelona, Barcelona, Spain.
- [Nie et al. 2009] Nie, F.; Xu, D.; Tsang, I. W.; and Zhang, C. 2009. Spectral embedded clustering. In *Proc. IJCAI*, 1181–1186.
- [Nie et al. 2010] Nie, F.; Xu, D.; Tsang, I.-H.; and Zhang, C. 2010. Flexible manifold embedding: A framework for semi-supervised and unsupervised dimension reduction. *IEEE Trans. Image Process.* 19(7):1921–1932.
- [Phillips et al. 2005] Phillips, P. J.; Flynn, P. J.; Scruggs, T.; Bowyer, K. W.; Chang, J.; Hoffman, K.; Marques, J.; Min, J.; and Worek, W. 2005. Overview of the face recognition grand challenge. In *Proc. CVPR*, 947–954.
- [Roweis and Saul 2000] Roweis, S. T., and Saul, L. K. 2000. Nonlinear dimensionality reduction by locally linear embedding. *Science* 290(5500):2323–2326.
- [Shi and Malik 2000] Shi, J., and Malik, J. 2000. Normalized cuts and image segmentation. *IEEE Trans. Pattern Analysis and Machine Intelligence* 22(8):888–905.
- [Singh et al. 2002] Singh, D.; Febbo, P. G.; Ross, K.; Jackson, D. G.; Manola, J.; Ladd, C.; and et al., P. T. 2002. Gene expression correlates of clinical prostate cancer behavior. *Cancer cell* 1(2):203–209.
- [Strehl and Ghosh 2003] Strehl, A., and Ghosh, J. 2003. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *Machine Learning Research* 3:583–617.
- [Wang and Zhang 2008] Wang, F., and Zhang, C. 2008. Label propagation through linear neighborhoods. *IEEE Trans. Know. Data Engin.* 20(1):55–67.
- [Wang, Zhang, and Li 2009] Wang, F.; Zhang, C.; and Li, T. 2009. Clustering with local and global regularization. *IEEE Trans. Know. Data Engin.* 21(12):1665–1678.
- [Wu and Schlkopf 2006] Wu, M., and Schlkopf, B. 2006. A local learning approach for clustering. In *Proc. NIPS*, 1529–1536.
- [Wu et al. 2012] Wu, L.; Hoi, S. C.; Jin, R.; Zhu, J.; and Yu, N. 2012. Learning bregman distance functions for semi-supervised clustering. *IEEE Trans. Knowledge and Data Eng.* 24(3):478–491.
- [Yang et al. 2010] Yang, Y.; Xu, D.; Nie, F.; Yan, S.; and Zhuang, Y. 2010. Image clustering using local discriminant models and global integration. *IEEE Trans. Image Process.* 19(10):2761–2773.
- [Yang et al. 2011] Yang, Y.; Shen, H. T.; Nie, F.; Ji, R.; and Zhou, X. 2011. Nonnegative spectral clustering with discriminative regularization. In *Proc. AAAI*.

- [Ye, Zhao, and Liu 2007a] Ye, J.; Zhao, Z.; and Liu, H. 2007a. Adaptive distance metric learning for clustering. In *Proc. CVPR*, 1–7.
- [Ye, Zhao, and Liu 2007b] Ye, J.; Zhao, Z.; and Liu, H. 2007b. Adaptive distance metric learning for clustering. In *Proc. CVPR*, 1–7.
- [Ye, Zhao, and Wu 2007] Ye, J.; Zhao, Z.; and Wu, M. 2007. Discriminative k-means for clustering. In *Proc. NIPS*, 1649–1656.
- [Yu and Shi 2003] Yu, S. X., and Shi, J. 2003. Multiclass spectral clustering. In *Proc. ICCV*, 313–319.
- [Zelnik-Manor and Perona 2004] Zelnik-Manor, L., and Perona, P. 2004. Self-tuning spectral clustering. In *Proc. NIPS*, 1601–1608.
- [Zhu and Ghahramani 2002] Zhu, X., and Ghahramani, Z. 2002. Learning from labeled and unlabeled data with label propagation. Technical report, CMU-CALD-02-107, Carnegie Mellon University.
- [Zhu 2006] Zhu, X. 2006. Semi-supervised learning literature survey. Technical Report 3, Computer Science, University of Wisconsin-Madison 2.