

# Unsupervised and Semi-Supervised Learning via $\ell_1$ -Norm Graph

Feiping Nie, Hua Wang, Heng Huang, Chris Ding  
Department of Computer Science and Engineering  
University of Texas, Arlington, TX 76019, USA

{feipingnie, huawangcs}@gmail.com, {heng, chqding}@uta.edu

## Abstract

*In this paper, we propose a novel  $\ell_1$ -norm graph model to perform unsupervised and semi-supervised learning methods. Instead of minimizing the  $\ell_2$ -norm of spectral embedding as traditional graph based learning methods, our new graph learning model minimizes the  $\ell_1$ -norm of spectral embedding with well motivation. The sparsity produced by the  $\ell_1$ -norm minimization results in the solutions with much clearer cluster structures, which are suitable for both image clustering and classification tasks. We introduce a new efficient iterative algorithm to solve the  $\ell_1$ -norm of spectral embedding minimization problem, and prove the convergence of the algorithm. More specifically, our algorithm adaptively re-weight the original weights of graph to discover clearer cluster structure. Experimental results on both toy data and real image data sets show the effectiveness and advantages of our proposed method.*

## 1. Introduction

Graph-based learning provides an efficient approach for modeling data in clustering or classification problems. An important advantage of working with a graph structure is its ability to naturally incorporate diverse types of information and measurements, such as the relationship between unlabeled data (clustering) or both labeled and unlabeled data (semi-supervised classifications).

As an important task in machine learning and computer vision, the clustering analysis has been well studied and solved from different perspectives such as  $K$ -means clustering, spectral clustering, support vector clustering, and maximum margin clustering. Among them, the use of manifold information in spectral clustering has shown the state-of-the-art clustering performance. Laplacian embedding provides an approximation solution of the ratio cut clustering [1], and the generalized eigenvectors of the Laplace matrix provides an approximation solution of the normalized cut clustering [7]. The main drawback of graph-based clustering methods is their solutions can not be directly used as

clustering results because the solutions don't have the clear cluster structure, hence further clustering algorithm such as  $K$ -means need to be applied to obtain the final clustering result. However, the  $K$ -means algorithm converges to local optimum and causes non-unique clustering results.

On the other hand, the graph-based learning models have been used to develop the main algorithms for semi-supervised classifications. Given a data set with pairwise similarities ( $W$ ), the semi-supervised learning can be viewed as the label propagation from labeled data to unlabeled data. Using the diffusion kernel, the semi-supervised learning is like a diffusive process of the labeled information. The harmonic function approach [9] emphasizes the harmonic nature of the diffusive function and consistency labeling approach [8] considers the spread of label information in an iterative way. All existing graph-based semi-supervised learning methods used the quadratic form of graph embedding, thus the results are sensitive to noise and outliers. A more robust graph-based learning model is desired in real world applications to handle the noisy data.

In this paper, to solve the above problems, instead of using the quadratic form of graph embedding, we propose a novel  $\ell_1$ -norm graph method to learn manifold information via the  $\ell_1$ -norm of spectral embedding. The new  $\ell_1$ -norm based objective are applied to both unsupervised clustering and semi-supervised classification problems. The efficient optimization algorithms are introduced to solve both sparse learning objectives. In our methods, the  $\ell_1$ -norm of spectral embedding formulation leads to sparse and direct clustering results. Both unsupervised and semi-supervised computer vision tasks performing on synthetic and real-world image benchmark data sets are used demonstrate the superior performance of our methods.

## 2. Unsupervised Clustering by $\ell_1$ -Norm Graph

### 2.1. Problem Formulation and Motivation

Suppose we have  $n$  data points  $\{x_1, \dots, x_n\} \in \mathbb{R}^{d \times 1}$ , and construct a graph using the data with weight matrix  $W \in \mathbb{R}^{n \times n}$ . Let the cluster indicator matrix  $Q =$

$[q_1, q_2, \dots, q_c] \in \mathbb{R}^{n \times c}$ , where  $q_k \in \mathbb{R}^{n \times 1}$  is the  $k$ -th column of  $Q$ . Without loss of generality, suppose the data points within each cluster are adjacent.

The multi-way graph Ratio Cut clustering is to minimize  $Tr(Q^T L Q)$  under the constraint that  $q_k =$

$(0, \dots, 0, \overbrace{\frac{1}{\sqrt{n_k}}, \dots, \frac{1}{\sqrt{n_k}}}^{n_k}, 0, \dots, 0)^T$ , while the multi-way graph Normalized Cut clustering is to minimize the same term  $Tr(Q^T L Q)$  under a different constraint that

$$q_k = (0, \dots, 0, \overbrace{\frac{1}{\sqrt{q_k^T D q_k}}, \dots, \frac{1}{\sqrt{q_k^T D q_k}}}^{n_k}, 0, \dots, 0)^T.$$

It is known that minimizing  $Tr(Q^T L Q)$  under such constraints of  $Q$  is NP hard [1, 7]. Traditional method solve the following relaxed problem for the Ratio Cut clustering:

$$\min_{Q^T Q = I} Tr(Q^T L Q), \quad (1)$$

and solve the following relaxed problem for the Normalized Cut clustering:

$$\min_{Q^T D Q = I} Tr(Q^T L Q). \quad (2)$$

The solution  $Q$  of the relaxed problems can not be directly used as clustering results, further clustering algorithm such as  $K$ -means has to be used on  $Q$  to obtain the final clustering results.

## 2.2. A New $\ell_1$ -Norm Graph Model

The Normalized Cut clustering has been shown the state-of-the-art performance, thus we focus on the Normalized Cut clustering in this paper. The problem (2) in Normalized Cut clustering can be rewritten as:

$$\min_{Q^T D Q = I} \sum_{i,j=1}^n W_{ij} \|q^i - q^j\|_2^2. \quad (3)$$

An ideal solution  $Q$  for clustering is that  $q^i = q^j$  if  $x_i$  and  $x_j$  belong to the same cluster, which means many rows of  $Q$  are equal and thus has strong clustering structure. Therefore, it is desired that  $\|q^i - q^j\|_2 = 0$  for many pairs of  $(i, j)$ . To this end, we propose to solve the following  $\ell_1$ -norm spectral embedding problem for clustering:

$$\min_{Q^T D Q = I} \sum_{i,j=1}^n W_{ij} \|q^i - q^j\|_1. \quad (4)$$

Denote a  $n^2$ -dimensional vector by  $p$ , where the  $((i-1) * n + j)$ -th element of  $p$  is  $W_{ij} \|q^i - q^j\|_2$ , we can re-write the above problem as:

$$\min_{Q^T D Q = I} \|p\|_1, \quad (5)$$

where  $\|p\|_1$  is the  $\ell_1$ -norm of  $p$ . It is widely recognized in the compressed sensing community [3] that minimizing an  $\ell_1$ -norm of  $p$  usually produces the sparse solutions, *i.e.*, many elements of  $p$  are zeros (for matrix data, either  $\ell_1$ -norm or  $R_1$ -norm [2] can be applied for different purposes). Thus, the proposed problem (4) will provide a more ideal solution of  $Q$  as clustering results.

Although the motivation of Eq. (4) is clear and consistent with the ideal clustering intuition, it is a non-smooth objective and difficult to be solved efficiently. Thus, in the next subsection, we will introduce an iterative algorithm to solve the problem (4). We will show that the original weight matrix  $W$  would be adaptively re-weighted to capture clearer cluster structures after each iteration.

## 2.3. Proposed Algorithm

The Lagrangian function of the problem (4) is

$$\mathcal{L}(Q) = \sum_{i,j=1}^n W_{ij} \|q^i - q^j\|_2 - Tr(\Lambda(Q^T D Q - I)). \quad (6)$$

Denote a Laplacian matrix  $\tilde{L} = \tilde{D} - \tilde{W}$ , where  $\tilde{W}$  is a re-weighted weight matrix defined by

$$\tilde{W}_{ij} = \frac{W_{ij}}{2\|q^i - q^j\|_2}, \quad (7)$$

$\tilde{D}$  is a diagonal matrix with the  $i$ -th diagonal element as  $\sum_j \tilde{W}_{ij}$ . Taking the derivative of  $\mathcal{L}(Q)$  w.r.t  $Q$ , and setting the derivative to zero, we have:

$$\frac{\partial \mathcal{L}(Q)}{\partial Q} = \tilde{L}Q - DQ\Lambda = 0, \quad (8)$$

which indicates that the solution  $Q$  is the eigenvectors of  $D^{-1}\tilde{L}$ . Note that  $D^{-1}\tilde{L}$  is dependent on  $Q$ , we propose an iterative algorithm to obtain the solution  $Q$  such that Eq. (8) is satisfied. The algorithm is guaranteed to converge to a local optimum, which will be proved in the next subsection.

The algorithm is described in Algorithm 1. In each iteration,  $\tilde{L}$  is calculated with the current solution  $Q$ , then the solution  $Q$  is updated according to the current calculated  $\tilde{L}$ . The iteration procedure is repeated until converges. From the algorithm we can see, the original weight matrix  $W$  is adaptively re-weighted to minimize the objective in Eq. (4) during the iteration. As can be seen in the experiments, the converged re-weighted matrix  $\tilde{W}$  will demonstrate more clear cluster structure than the original weight matrix  $W$ .

## 2.4. Convergence Analysis

To prove the convergence of the Algorithm 1, we need the following lemma [5]:

**Lemma 1** For any nonzero vectors  $q, q_t \in \mathbb{R}^c$ , the following inequality holds:

$$\|q\|_2 - \|q\|_2^2/2\|q_t\|_2 \leq \|q_t\|_2 - \|q_t\|_2^2/2\|q_t\|_2. \quad (9)$$

**Input:** The original weight matrix  $W \in \mathbb{R}^{n \times n}$ .  $D$  is a diagonal matrix with the  $i$ -th diagonal element as  $\sum_j W_{ij}$ .

$t = 1$ . Initialize  $Q_t \in \mathbb{R}^{n \times c}$  such that  $Q_t^T D Q_t = I$ ;

**while** not converge **do**

1. Calculate  $\tilde{L}_t = \tilde{D}_t - \tilde{W}_t$ , where  $(\tilde{W}_t)_{ij} = \frac{W_{ij}}{2\|q_t^i - q_t^j\|_2}$ ,  $\tilde{D}_t$  is a diagonal matrix with the  $i$ -th diagonal element as  $\sum_j (\tilde{W}_t)_{ij}$ ;
2. Calculate  $Q_{t+1} = [(q_t^1)^T, (q_t^2)^T, \dots, (q_t^n)^T]^T$ , where the columns of  $Q_{t+1}$  are the first  $c$  eigenvectors of  $D^{-1}\tilde{L}_t$  corresponding to the first  $c$  smallest eigenvalues;
3.  $t = t + 1$ ;

**end**

**Output:**  $\tilde{W}_t \in \mathbb{R}^{n \times n}$ ,  $Q_t \in \mathbb{R}^{n \times c}$ .

**Algorithm 1:** The algorithm to solve the problem (4).

As a result, we have the following theorem:

**Theorem 1** *The Algorithm 1 will monotonically decrease the objective of the problem (4) in each iteration, and converge to a local optimum of the problem.*

**Proof:** According to the step 2 in the Algorithm 1, we know that

$$Q_{t+1} = \arg \min_{Q^T D Q = I} \sum_{i,j=1}^n (\tilde{W}_t)_{ij} \|q^i - q^j\|_2^2. \quad (10)$$

Note that  $(\tilde{W}_t)_{ij} = \frac{W_{ij}}{2\|q_t^i - q_t^j\|_2}$ , so we have

$$\sum_{i,j=1}^n \frac{W_{ij} \|q_{t+1}^i - q_{t+1}^j\|_2^2}{2\|q_t^i - q_t^j\|_2} \leq \sum_{i,j=1}^n \frac{W_{ij} \|q_t^i - q_t^j\|_2^2}{2\|q_t^i - q_t^j\|_2}. \quad (11)$$

According to Lemma 1, we have

$$\begin{aligned} & \sum_{i,j=1}^n W_{ij} \left( \|q_{t+1}^i - q_{t+1}^j\|_2 - \frac{\|q_{t+1}^i - q_{t+1}^j\|_2^2}{2\|q_t^i - q_t^j\|_2} \right) \\ & \leq \sum_{i,j=1}^n W_{ij} \left( \|q_t^i - q_t^j\|_2 - \frac{\|q_t^i - q_t^j\|_2^2}{2\|q_t^i - q_t^j\|_2} \right) \end{aligned} \quad (12)$$

Summing Eq. (11) and Eq. (12) in the two sides, we arrive at

$$\sum_{i,j=1}^n W_{ij} \|q_{t+1}^i - q_{t+1}^j\|_2 \leq \sum_{i,j=1}^n W_{ij} \|q_t^i - q_t^j\|_2. \quad (13)$$

Thus the Algorithm 1 will monotonically decrease the objective of the problem (4) in each iteration  $t$  until the algorithm converges. In the convergence, the equality in Eq. (13) holds, thus  $Q_t$  and  $\tilde{L}_t$  will satisfy Eq. (8), the KKT condition of problem (4). Therefore, the Algorithm 1 will converge to a local optimum of the problem (4).  $\square$

### 3. Semi-Supervised Classification Using $\ell_1$ -Norm Graph

#### 3.1. Problem Formulation and Motivation

Denote  $Y = [(y^1)^T, (y^2)^T, \dots, (y^n)^T] \in \mathbb{R}^{n \times c}$  as the initial label matrix. If  $x_i$  is the unlabeled data, then  $y^i = \mathbf{0}$ . If  $x_i$  is labeled as class  $k$ , then the  $k$ -th element of  $y^i$  is 1 and the other elements of  $y^i$  is 0.

Traditional graph based semi-supervised learning usually solve the following problem [9, 8]:

$$\min_Q Tr(Q^T L Q) + Tr(Q - Y)^T U (Q - Y), \quad (14)$$

where  $L$  is the Laplacian matrix defined as before,  $U$  is a diagonal matrix with the  $i$ -th diagonal element to control the impact of the initial label  $y^i$  of  $x_i$ ,  $Q \in \mathbb{R}^{n \times c}$  is the label matrix to be solved.

Similarly, problem (14) can be rewritten as:

$$\min_Q \sum_{i,j=1}^n W_{ij} \|q^i - q^j\|_2^2 + Tr(Q - Y)^T U (Q - Y). \quad (15)$$

To obtain a more ideal solution  $Q$ , we propose to solve the following problem for semi-supervised classification:

$$\min_Q \sum_{i,j=1}^n W_{ij} \|q^i - q^j\|_2 + Tr(Q - Y)^T U (Q - Y) \quad (16)$$

#### 3.2. Proposed Algorithm

Taking the derivative of Eq. (16) w.r.t  $Q$ , and setting the derivative to zero, we have:

$$\tilde{L}Q + U(Q - Y) = \mathbf{0} \Rightarrow Q = (\tilde{L} + U)^{-1}UY. \quad (17)$$

Note that  $\tilde{L}$  is dependent on  $Q$ , we propose an iterative algorithm to obtain the solution  $Q$  such that Eq. (17) is satisfied. The algorithm is also guaranteed to converge to a local optimum, which will be proved in the next subsection.

The algorithm is described in Algorithm 2. In each iteration,  $\tilde{L}$  is calculated with the current solution  $Q$ , then the solution  $Q$  is updated according to the current calculated  $\tilde{L}$ . The iteration procedure is repeated until converges. From the algorithm we can see, the original weight matrix  $W$  is adaptively re-weighted to minimize the objective in Eq. (16) during the iteration. Similarly, in the convergence, the re-weighted matrix  $\tilde{W}$  will demonstrate more clear cluster structure than the original weight matrix  $W$ . It is worth noting that the step 2 in Algorithm 2 can be efficiently solve by a sparse linear equation system instead of computing the inverse.

#### 3.3. Convergence Analysis

The following theorem guarantees that the Algorithm 2 will converge to the global optimum of the problem (16):

**Input:** The original weight matrix  $W \in \mathbb{R}^{n \times n}$ .  $D$  is a diagonal matrix with the  $i$ -th diagonal element as  $\sum_j W_{ij}$ . The initial label matrix  $Y \in \mathbb{R}^{n \times c}$ .

$t = 1$ . Initialize  $Q_t \in \mathbb{R}^{n \times c}$  such that  $Q_t^T D Q_t = I$ ;

**while** not converge **do**

1. Calculate  $\tilde{L}_t = \tilde{D}_t - \tilde{W}_t$ , where  $(\tilde{W}_t)_{ij} = \frac{W_{ij}}{2\|q_t^i - q_t^j\|_2}$ ,  $\tilde{D}_t$  is a diagonal matrix with the  $i$ -th diagonal element as  $\sum_j (\tilde{W}_t)_{ij}$ ;
2. Calculate  $Q_{t+1} = (\tilde{L}_t + U)^{-1} U Y$ ;
3.  $t = t + 1$ ;

**end**

**Output:**  $\tilde{W}_t \in \mathbb{R}^{n \times n}$ ,  $Q_t \in \mathbb{R}^{n \times c}$ .

**Algorithm 2:** The algorithm to solve the problem (16).

**Theorem 2** *The Algorithm 2 will monotonically decrease the objective of the problem (16) in each iteration, and converge to the global optimum of the problem.*

**Proof:** Denote  $f(Q) = \text{Tr}(Q - Y)^T U (Q - Y)$ . According to the step 2 in the algorithm 2, we know that

$$Q_{t+1} = \arg \min_Q \sum_{i,j=1}^n (\tilde{W}_t)_{ij} \|q^i - q^j\|_2^2 + f(Q). \quad (18)$$

Note that  $(\tilde{W}_t)_{ij} = \frac{W_{ij}}{2\|q_t^i - q_t^j\|_2}$ , so we have

$$\begin{aligned} & \sum_{i,j=1}^n \frac{W_{ij} \|q_{t+1}^i - q_{t+1}^j\|_2^2}{2\|q_t^i - q_t^j\|_2} + f(Q_{t+1}) \\ & \leq \sum_{i,j=1}^n \frac{W_{ij} \|q_t^i - q_t^j\|_2^2}{2\|q_t^i - q_t^j\|_2} + f(Q_t) \end{aligned} \quad (19)$$

Summing Eq. (19) and Eq. (12) on both sides, we have

$$\begin{aligned} & \sum_{i,j=1}^n W_{ij} \|q_{t+1}^i - q_{t+1}^j\|_2 + f(Q_{t+1}) \\ & \leq \sum_{i,j=1}^n W_{ij} \|q_t^i - q_t^j\|_2 + f(Q_t). \end{aligned} \quad (20)$$

Thus the Algorithm 2 will monotonically decrease the objective of the problem (16) in each iteration  $t$ . In the convergence,  $Q_t$  and  $\tilde{L}_t$  will satisfy the Eq. (17). As the problem (16) is a convex problem, satisfying the Eq. (17) indicates that  $Q_t$  is the global optimum solution to the problem (16). Therefore, the Algorithm 2 will converge to the global optimum of the problem (16).  $\square$

## 4. Experiments

We present experiments on synthetic data and real data to validate the effectiveness of the proposed method, and compare the performance with the traditional spectral clustering (*i.e.*, Normalized Cut) method [4] and the commonly used label propagation method [9].

We use Gaussian function to construct the original weight matrix  $W$ . The weight  $W_{ij}$  is defined as  $\exp(-\frac{\|x_i - x_j\|_2^2}{\sigma^2})$ ,  $x_i$  and  $x_j$  are neighbors; 0, otherwise. The number of neighbors and the parameter  $\sigma$  should be predefined by user. In the experiments, we set the number of neighbors to be 4 in all data sets, and set the  $\sigma$  according to the distances of neighbors as suggested in [6].

### 4.1. Synthetic Data

In this experiment, two synthetic data are used for evaluation. The first synthetic data include data points distributed on two half-moon shapes with noise, and the second synthetic data include data points distributed on three ring shapes with noise.

Figs. 1 and 2 show the re-weighted weights between data points during the iteration by Algorithm 1. In the figures, we use the line width to represent the weight between two data points, bigger width indicates larger weight. Zoom in the figures will give better visualization. The results show that the algorithm converges fast and usually within 10 iterations. During the iterations, the weights between data points with different clusters are gradually suppressed, while the weights between data points within the same cluster are gradually strengthened. Therefore, the cluster structure is more and more clear during the iterations, which validates the effectiveness of the proposed method.

Figs. 3 and 4 show the embedded results by Algorithm 1 (denoted by L1\_un) and by traditional spectral clustering (denoted by SC). From the results we can see, the embedded results by Algorithm 1 demonstrate much more clear cluster structure than that by traditional spectral clustering method. Therefore, our method can perform clustering task directly by using the embedded result, while traditional spectral clustering need additional clustering algorithm on the embedded result to obtain the final clustering result.

We also perform the experiment of the semi-supervised method by Algorithm 2 on two synthetic data, the results are similar to the unsupervised case, thus are not reported here. The label matrix  $Q$  obtained by Algorithm 2 is very close to the ideal label matrix, *i.e.*, one and only one element of each row of  $Q$  is 1 and the other elements are all 0.

### 4.2. Evaluations Using Image Benchmark Data Sets

In this experiment, we present experimental results on five real image data sets, including four face data Jaffe, Yale, Umist and MSRA, and one object data Coil20. Table 1 give a brief description of the five image data sets.

We compare the performance of Algorithm 1 (denoted by L1\_un) with the traditional spectral clustering (*i.e.*, Normalized Cut) method (denoted by SC) and compare the performance of Algorithm 2 (denoted by L1\_semi) with traditional label propagation method [9] (denoted by LP). The

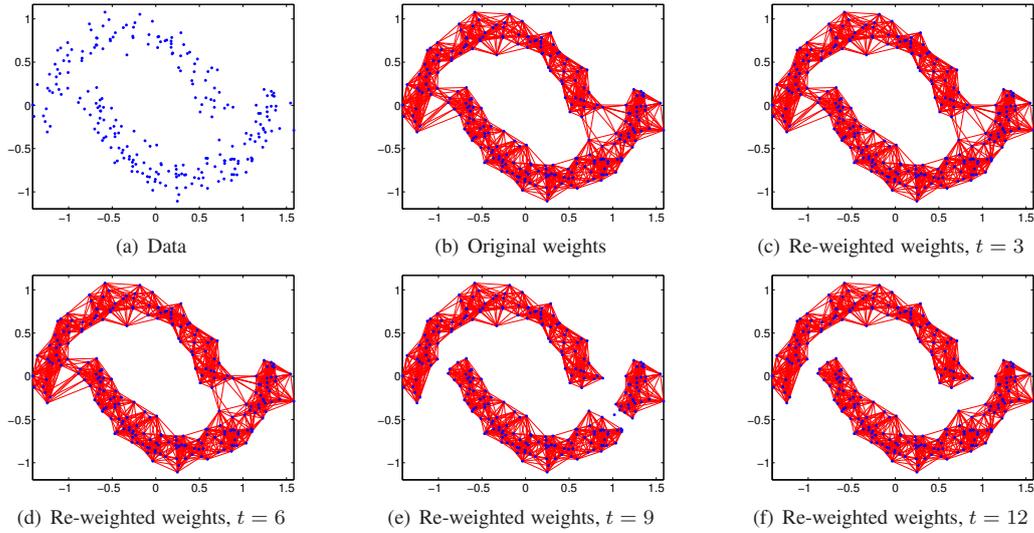


Figure 1. On the two-half-moon synthetic data, the re-weighted weights between data points during the iteration by Algorithm 1, bigger width indicates larger weight. Zoom in the figures will give better visualization.

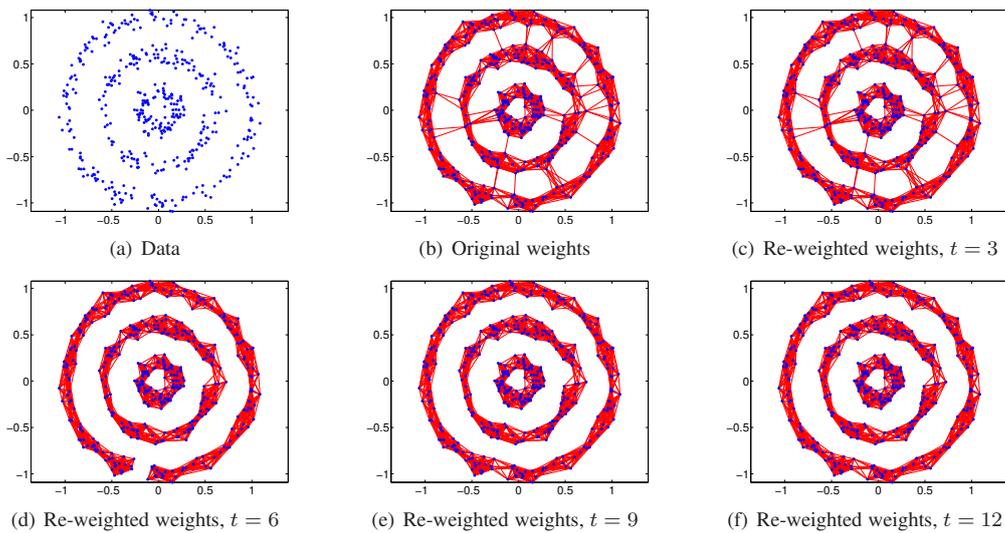


Figure 2. On the three-ring synthetic data, the re-weighted weights between data points during the iteration by Algorithm 1, bigger width indicates larger weight. Zoom in the figures will give better visualization.

Normalized Cut and accuracy are reported in the experiment to measure the performance of compared methods.

In the unsupervised setting, additional clustering algorithm K-means is used to obtain the final clustering results for SC. Since K-means depends on the initialization, we independently repeat K-means for 50 times with random initialization, and then report the results corresponding to the best objective values. In the semi-supervised setting, in each image data set, 10 images in each class are randomly selected as the labeled data, and the rest images are as the unlabeled data. The experiment run 20 times independently and the averaged results are recorded.

The results are shown in Tables 2 and 3. From the results we can see that the proposed methods outperform traditional methods in many real applications. Meanwhile, in all experiments, the Algorithms 1 and 2 converge within 10 iterations.

## 5. Conclusions

We propose a novel unsupervised and semi-supervised learning with  $\ell_1$ -norm graph. Different from minimizing an  $\ell_2$ -norm in traditional graph based learning methods, we propose to minimize the  $\ell_1$ -norm. Minimizing the  $\ell_1$ -norm results in a sparse solution which demonstrates more clear

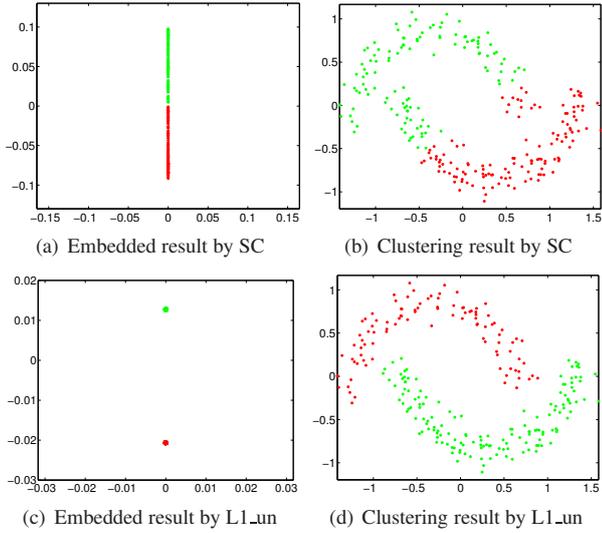


Figure 3. The embedded results and the clustering results by SC and L1\_un, respectively.

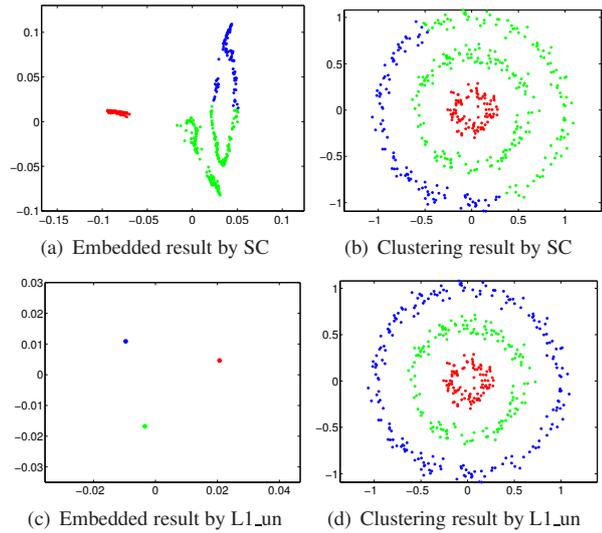


Figure 4. The embedded results and the clustering results by SC and L1\_un, respectively.

Table 1. Dataset Descriptions

Data set	Size	Dimensions	Classes
Jaffe	213	1024	10
Yale	165	3456	15
Umist	575	644	20
MSRA	1799	1024	12
Coil20	1440	1024	20

cluster structure, and thus is more suitable for clustering or classification. Efficient iterative algorithms are proposed to solve the  $\ell_1$ -norm minimization problem both in the unsupervised and in the semi-supervised cases, and the conver-

Table 2. The Normalized Cut and the accuracy results obtained by SC and L1\_un.

Data set	Normalized Cut		Accuracy	
	SC	L1_un	SC	L1_un
Jaffe	0.02	<b>0.00</b>	82.16%	<b>84.04%</b>
Yale	0.54	<b>0.35</b>	61.21%	<b>64.85%</b>
Umist	0.66	<b>0.02</b>	64.35%	<b>72.35%</b>
MSRA	0.17	<b>0.00</b>	47.08%	<b>48.75%</b>
Coil20	0.45	<b>0.01</b>	85.28%	<b>87.29%</b>

Table 3. The Normalized Cut and the accuracy results obtained by LP and L1\_semi, respectively.

Data set	Normalized Cut		Accuracy	
	LP	L1_semi	LP	L1_semi
Jaffe	<b>0.00</b>	<b>0.00</b>	99.38%	<b>99.51%</b>
Yale	2.87	<b>2.85</b>	70.13%	<b>71.33%</b>
Umist	0.15	<b>0.13</b>	98.29%	<b>98.65%</b>
MSRA	0.15	<b>0.09</b>	87.16%	<b>87.83%</b>
Coil20	0.13	<b>0.08</b>	98.29%	<b>99.21%</b>

gence is guaranteed with theoretical analysis. In essence, the iterative algorithms adaptively re-weight the original weights of graph to discover clearer cluster structure. Experimental results on synthetic data and real data validate that the proposed method is effective and attractive.

## Acknowledgement.

This research was supported by NSF-IIS 1117965, NSF-CCF-0830780, NSF-DMS-0915228, NSF-CCF-0917274.

## References

- [1] P. K. Chan, M. D. F. Schlag, and J. Y. Zien. Spectral k-way ratio-cut partitioning and clustering. *IEEE Trans. on CAD of Integrated Circuits and Systems*, 13(9):1088–1096, 1994.
- [2] C. H. Q. Ding, D. Zhou, X. He, and H. Zha. R1-PCA: rotational invariant L1-norm principal component analysis for robust subspace factorization. In *ICML*, pages 281–288, 2006.
- [3] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.
- [4] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *NIPS*, pages 849–856, 2001.
- [5] F. Nie, H. Huang, X. Cai, and C. Ding. Efficient and robust feature selection via joint  $\ell_{2,1}$ -norms minimization. In *NIPS*, 2010.
- [6] F. Nie, S. Xiang, Y. Jia, and C. Zhang. Semi-supervised orthogonal discriminant analysis via label propagation. *Pattern Recognition*, 42(11):2615–2627, 2009.
- [7] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on PAMI*, 22(8):888–905, 2000.
- [8] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf. Learning with local and global consistency. In *NIPS*, 2004.
- [9] X. Zhu, Z. Ghahramani, and J. D. Lafferty. Semi-supervised learning using Gaussian fields and harmonic functions. In *ICML*, pages 912–919, 2003.