Neural Process Lett (2018) 47:697–716
https://doi.org/10.1007/s11063-017-9676-5

Nonlinear Dimensionality Reduction for Data with Disconnected Neighborhood Graph

Jicong Fan[1] · Tommy W. S. Chow[1] · Mingbo Zhao[2] · John K. L. Ho[1]

Published online: 3 August 2017

Abstract Neighborhood graph based nonlinear dimensionality reduction algorithms, such as Isomap and LLE, perform well under an assumption that the neighborhood graph is connected. However, for datasets consisting of multiple clusters or lying on multiple manifolds, the neighborhood graphs are often disconnected, or in other words, have multiple connected components. Neighborhood graph based dimensionality reduction techniques cannot recognize both the local and global properties of such datasets. In this paper, a new method, called enhanced neighborhood graph, is proposed to solve the problem. The concept is to add edges to the neighborhood graph adaptively and iteratively until it becomes connected. Nonlinear dimensionality reduction can then be performed based on the enhanced neighborhood graph. As a result, both local and global properties of the data can be exactly recognized. In this study, thorough simulations on synthetic datasets and natural datasets are conducted. The experimental results corroborate that the proposed method provides significant improvements on dimensionality reduction for data with disconnected neighborhood graph.

Keywords Nonlinear dimensionality reduction · Isomap · LLE · Disconnected graph · Enhanced neighborhood graph · Multiple manifolds

1 Introduction

Dimensionality reduction [1,2] is an important technique to represent high dimensional data by low-dimensional data and has been widely studied and used in many areas such as data mining and computer vision. Classical linear dimensionality reduction methods such as principal components analysis (PCA) [3] and multidimensional scaling (MDS) [4] are not effective

✉ Tommy W. S. Chow
eetchow@cityu.edu.hk

1 City University of Hong Kong, Kowloon, Hong Kong, China

2 School of Information Science and Technology, Donghua University, Shanghai, P. R. China

in handling nonlinear data. As a result, many nonlinear dimensionality reduction techniques (NLDR) [5–11], such as Sammon's nonlinear mapping [12], curvilinear component analysis [13], self-organizing maps [9], KPCA [7], and auto-encoders [8], were proposed. However, these methods often cannot provide satisfactory solutions for image data especially in 2-D or 3-D visualization tasks because they are global methods which are inapt to preserve local properties [14–16].

To better preserve intrinsic nonlinear structures of data, many neighborhood graph (NG) based techniques [17–24], such as Isomap [18], locally linear embedding (LLE) [17], Hessian LLE [20], Laplacian eigenmaps (LE) [19], local tangent space analysis (LTSA) [22], and maximum variance unfolding (MVU) [25], were proposed. Generally, these NG-NLDR techniques start from the construction of a neighborhood graph on which each data point is connected with its nearest neighbors. Then using the neighborhood graph, the local property or structure is defined. At last, low-dimensional embedding is implemented to preserve the global and/or local properties of the data. Many extended or modified methods have also been proposed to overcome different limitations of NG-NLDR techniques. In [26–29], several methods were proposed to determine the number of nearest neighbors, because different numbers of neighbors often have significant effect on the dimensionality reduction results. In [30–33], supervised and semi-supervised modifications of NG-NLDR techniques were proposed to take advantages of label information.

The concept of current NG-NLDR techniques are based on an assumption that the neighborhood graph is connected, which means there exists a path between every pair of vertices in the graph. In most practical cases, it is rather common that a neighborhood graph is disconnected, i.e., consists of multiple connected components. In graph theory, a subgraph H is said to be a connected component of graph G if it meets the following conditions: H is connected and is not contained in any connected subgraph of G which has more vertices or edges than H has. Hence, all connected components are disjoint mutually. A disconnected neighborhood graph is usually caused by 1) dataset consisting of multiple classes/clusters or lying on multiple manifolds; 2) missing data; 3) using too small a value of k. For such data, we may apply NG-NLDR methods to each connected component individually but the inter-component information and global information cannot be captured. If we disregard the disconnection and direct apply the NG-NLDR methods to the whole graph, dimensionality reduction and visualization will fail. The reason is that the multiple components are characterized with no relationship at all and there is lack of unfolding among different components. In some cases, subsets of the data points are even stretched into a line, or compressed into a point, which results in losing both global and local properties. For example, performances of NG-NLDR techniques on Broken Swiss Roll data and COIL-20 [34] data are found largely unsatisfactory. As a result, the generalization errors of 1-NN classifiers in the low-dimensional space are significantly increased [5,6]. In [25], the supervised LLE method compressed certain classes of data into small points. In [26], the M-Isomap method failed to handle disconnected graphs.

A few attempts have been done to handle the problem of disconnected graph or disjoint manifolds [35–38]. For example, in [38], a robust LLE was proposed for data with noises, outliers, or disjoint manifolds. In [38], if disjoint manifolds are caused by noises or outliers, the corresponding points will be added into the largest connected component to form a connected graph on which LLE is performed. Otherwise, LLE is performed on each connected component individually. As can be seen, the method is still unable to capture the inter-manifold information and global information when disjoint manifolds are not caused by noises or outliers. Recently, in [36], an extension for Laplacian Eigenmaps was proposed to analyze multiple manifolds. In [36], the inter-manifold information was captured by a soft correspondence matrix between pair-wise manifolds constructed over intra-manifold weight matrices. More recently, in [39], similar to [36], a joint embedding method was proposed to handle multiple manifolds but each inter-manifold affinity matrix was constructed by computing a permutation matrix for the Gaussian kernel matrix between pair-wise manifolds. It is a more reliable approach than the method described in [36]. It is worth noting that in [36] and [39], the inter-manifold affinity matrices are constructed by global techniques but not local techniques. From their experimental results of [36,39], the data belonging to different classes were not cleanly separated in the 2-D space. Therefore, the general problem of NG-NLDR (e.g., LLE and Isomap) for data with disconnected neighborhood graph or multiple manifolds have not been solved completely.

In this paper, we proposed a technique called enhanced neighborhood graph (ENG) to solve the problems of the disconnected graph and multiple manifolds. For a disconnected neighborhood graph, ENG connects each graph component to its nearest neighbor. Then pair-wise components are combined into a new connected component. The number of connected components in the neighborhood graph will decrease as a result. By repeating the above procedures, all components will be eventually integrated into a single component and the neighborhood graph is then connected. It is important to determine which and how many pair-wise points from two components should be connected. Our preliminary study shows that this issue often has noticeable effect on preserving local properties. It is known that local intrinsic dimensionalities are the same as the global intrinsic dimensionality. Therefore, the new edges between two components (to be linked) can be constructed by making the intrinsic dimensionality of the data points corresponding to the new edges be the global intrinsic dimensionality. Hence, we evaluate the local tangent space [26,28,29] by computing the contribution ratio of the d largest singular values of the space spanned by each data point and its nearest neighbors. The contribution ratio is utilized to construct connections between pair-wise components adaptively.

The main contributions of this paper are as followings. First, we have thoroughly analyzed and illustrated the shortcomings of existing NG-NLDR methods when handling datasets with disconnected neighborhood graphs. Second, we have proposed an approach, ENG, to solve the problem of disconnected neighborhood graph for NG-NLDR. ENG is able to preserve both inter-component properties and intra-component properties from high-dimensional space to low-dimensional space. Third, we apply the proposed method to Isomap, LLE, LE and HLLE, and the experimental results on synthetic and real-world datasets show that the performance of NG-NLDR techniques for datasets with disconnected neighborhood graphs can be significantly improved by using ENG. In addition, our method can also outperform state-of-the-art method of multiple manifold learning.

The rest of this paper is organized as follows. In Sect. 2, a brief overview of NG-NLDR techniques is given, and then the problem of disconnected neighborhood graph is discussed. In Sect. 3, the proposed method, enhanced neighborhood graph is illustrated. Section 4 gives the experimental results on synthetic datasets and natural datasets to show the effectiveness and superiority of the proposed method. The conclusion of this paper is drawn in Sect. 5.

2 Related Works and Problem Formulation

2.1 Neighborhood Graph Based Nonlinear Dimension Reduction (NG-NLDR)

A common framework of NG-NLDR consists of the following steps [40]. The first step is to determine the neighbors of each data point and construct a connected neighborhood graph $G(V, E)$, where Vdenotes a set of vertices and Edenotes a set of edges. In $G(V, E)$, each data point (vertices) is connected with its neighbors by edges and every pair-wise data points is connected by a path. The neighbors of each data point can be chosen as its k nearest neighbors or all points in some fixed radius. The method of k nearest neighbors is more robust than the method of fixed radius especially when the data points are not uniformly distributed. However, a too large or too small k may fail to recognize the local property or/and global property of the data, which results in further impeding the performance of dimensionality reduction and 2-D or 3-D visualization. In [26–29], several methods were proposed to determine the neighbors of each data point automatically and adaptively, which have an effect of improving the performances of NG-NLDR techniques in certain cases. Based on the neighborhood graph constructed previously, the second step of NG-NLDR is to compute an affinity matrix $\boldsymbol{A}(n \times n)$, which could be any metrics such as distance and similarity, for pair-wise data points. The element a_{ij} in $\boldsymbol{A}$ defines the relationship between ith data point and jth data point. For example, in Isomap, $\boldsymbol{A}$ is a geodesic distance matrix; in LLE, $\boldsymbol{A}$ is a variation of reconstruction weights matrix. Finally, with the affinity matrix, low-dimensional representation is obtained through general eigenvalue-decomposition or other approaches.

Specifically, we first take Isomap [18] as an example to illustrate the framework of NG-NLDR. In this paper, we denote a given dataset as $\boldsymbol{Y}(n \times m)$, where m is the number of attributes and n is the number of samples. Isomap firstly determines the neighbors of each data point and construct a neighborhood graph $G(V, E)$. Then, the geodesic distance $d_G(i, j)$ between point i and point j in G is approximated by the shortest path that can be computed by Dijkstra's or Floyd's shortest-path algorithm. Finally, an affinity matrix $\boldsymbol{D}_G$ consisting of the geodesic distances can be obtained and classical scaling is implemented to obtain the low-dimensional representation:

$$\boldsymbol{X} = \arg\min_{X} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\| - d_G(i, j)\right)^2, \tag{1}$$

where $\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|$ denotes the Euclidean distance between the i-th sample and j-th sample of $\boldsymbol{X}$. The fundamental concept of Isomap is to convert geodesic distances in high-dimension space to Euclidean distances in low-dimensional space. Another example is LLE [17], which is to find a low-dimensional embedding that preserves the local properties formulated as the reconstruction weights of the neighborhoods, by minimizing

$$\phi(\boldsymbol{X}) = \sum_{i=1}^{n} \left\| \boldsymbol{x}_i - \sum_{j=1}^{k} w_{i,j} \boldsymbol{x}_{i_j} \right\|^2 \text{subject to } \|\boldsymbol{x}^{(k)}\|^2 = 1, \tag{2}$$

where $w_{i,j}$is the reconstruction weight of $\boldsymbol{y}_{i_j}$ from its j th neighbor $\boldsymbol{y}_{i_j}$, and $w_{i,j}$ can be obtained by minimizing $||\boldsymbol{y}_i - \sum_{j=1}^{k} w_{i,j}\boldsymbol{y}_{i_j}||^2$. The solution of LLE is the eigen-decomposition problem of $\boldsymbol{M} = (\boldsymbol{I} - \boldsymbol{W})^T(\boldsymbol{I} - \boldsymbol{W})$, where $\boldsymbol{W}$is a sparse matrix with entries of $w_{i,j}$. Then $\boldsymbol{X}$ is computed as the d eigenvectors corresponding to the d smallest non-zero eigenvalues of $\boldsymbol{M}$.

2.2 Problem of Disconnected Neighborhood Graph

The major assumption [6,35–39] for NG-NLDR techniques is that the neighborhood graph should be connected. In practice, neighborhood graphs of many datasets are often disconnected [35–39]. The disconnected neighborhood graphs are usually caused by dataset containing multiple classes/clusters or lying on multiple manifolds, dataset with missing data, or scarce nearest neighbors. For example, the COIL-20 [34] dataset consists of images of 20 different subjects, and its neighborhood graph with $k = 8$ has 8 connected components that are mutually disjoint. It should be noted that though larger k may result in fewer connected components (for COIL-20, when $k = 50$, we get 2 connected components), larger k will have a negative effect on preserving the intrinsic properties. In a disconnected neighborhood graph, there is no path between any pair-wise points $\{i, j\}$ belonging to two different components because the two components are disjoint mutually. Physically, it means the relationship between the corresponding data pair $\{x_i, x_j\}$ is never defined directly or indirectly in an obtained affinity matrix of NG-NLDR. This is the fundamental reason why NG-NLDR techniques give unsatisfactory performance on data with disconnected neighborhood graph.

As discussed in Introduction, there exist a few attempts to solve the problem of disconnected neighborhood graph or multiple manifolds. For example, [39] proposed to add inter-manifold correspondence structures to make the multiple manifolds integrated into one global manifold and then low-dimensional embedding was performed. The method was called joint manifold embedding (JME). Specifically, the weight matrix is replaced as $\boldsymbol{W} \leftarrow \boldsymbol{W}+\boldsymbol{C}$, where $\boldsymbol{C}$ is a soft correspondence matrix [39] to incorporate the inter-manifold information. However, JME sometimes failed completely because the value of $\boldsymbol{W}$ may be too small compared with the value of $\boldsymbol{C}$. In this paper, to improve the performance of JME, we reformulate the weight matrix as $\boldsymbol{W} \leftarrow \boldsymbol{W} + \lambda\boldsymbol{C}$ and adjust the value of λ within [0.01, 1] to give the best performances in different cases. It is worth noting that JME makes all graph components connected with each other and every vertex of one component directly connected with every vertex of another component. Therefore, the local structures and global structures cannot be well preserved in low-dimensional space. In Fig. 1a we use a Broken Swiss Roll data (3-D) as an example. The neighborhood graph of the data has two connected components. Figure 1b shows the ground truth of the 2-D embedding for the Broken Swiss Roll. The 2-D

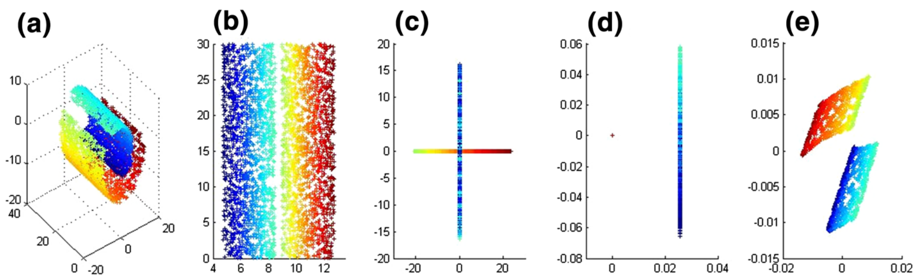

Fig. 1 **a** Broken Swiss Roll data; **b** ground truth of 2-D embedding; **c** 2-D embedding of Isomap; **d** 2-D embedding of LLE; **e** 2-D embedding of JME

embeddings given by Isomap, LLE and JME are shown by Fig. 1c–e respectively. As can be seen, Isomap stretched the two subsets to two lines; LLE stretched one subset into a line and compressed another subset into a point; JME outperformed Isomap and LLE but the result was still far away from the ground truth because both the local property and global structure were distorted.

In sum, to solve the problem of disconnected neighborhood graph, we need to make the graph become connected. There are two existing methods of connecting neighborhood graph components. The first one is to connect smaller components with the largest component [38], which is denoted by Largest-Connect in this paper. The second one is to connect each component with all the others [39], which is denoted by All-Connect in this paper. These two methods could be problematic in practice because the pair-wise components may not be necessarily direct connected in the intrinsic structure. Pair-wise components may be indirectly connected by a media or path formed by other components. In addition, the two methods often entail considerable arbitrariness in determining the edges to be added between two components, which cannot ensure exact local or/and global property preservation. In this paper, we propose a novel method to solve the problem of NG-NLDR on data with disconnected neighborhood graph.

3 Enhanced Neighborhood Graph for NLDR

Given a dataset $\boldsymbol{Y}(n \times m)$, we first construct a neighborhood graph $G(V, E)$. The connected components of the undirected graph can be found by breadth-first search (BFS) [41] or depth-first search (DFS) [42]. BFS and DFS are two algorithms for traversing or searching tree or graph data structures. In the case of a graph, BFS starts at an arbitrary node of a graph and explores the neighbor nodes first, before moving to the next level neighbors. DFS also starts at some arbitrary node of a graph but explores as far as possible along each branch before backtracking. When using BFS or DFS to compute the connected components, a search beginning at an arbitrary vertex v will find the entire connected component containing v (and no more). To find all the connected components of a graph, loop through its vertices, starting a new BFS or DFS whenever the loop reaches a vertex that has not already been included in a previously found connected component. Both BFS and DFS can compute the connected components of a graph in line time. Compared with BFS, DFS usually consumes less memory. We therefore propose to use DFS to compute the connected components of disconnected neighborhood graphs. More details about the algorithm of computing connected components by DFS can be found in [43].

We denote the g disjoint subsets of $\boldsymbol{Y}$ as $\{\boldsymbol{Y}^{\{1\}}, \boldsymbol{Y}^{\{2\}}, \cdots, \boldsymbol{Y}^{\{g\}}\}$ corresponding to the g individual connected components $\{G_1,\ G_2,\ \cdots\ ,\ G_g\}$ of $G(V, E)$. If g=1, $G(V, E)$ is connected. Otherwise, the graph is not connected but has multiple connected components. As described in Sect. 2.2, NG-NLDR techniques experience difficulties when handling data with disconnected neighborhood graph. A direct approach is to add edges among the components of $G(V, E)$ to obtain a connected graph. Essentially, to construct a reliable connected graph, we need to solve the following two problems: (1) which and how many pairs of components should be linked? (2) which and how many pairs of data points from the two components should be connected?

We first consider the second problem, i.e., constructing edges between two components. It is assumed that a data point in one component should be connected to only one

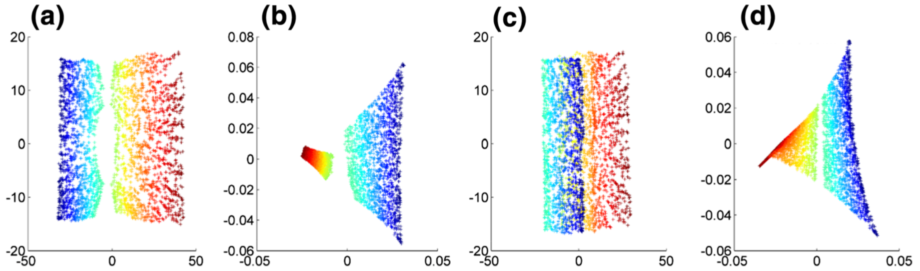

Fig. 2 Two-dimensional embeddings of ENG-Isomap and ENG-LLE: **a** ENG-Isomap with 3 connections; **b** ENG-LLE with 3 connections; **c** ENG-Isomap with 70 connections; **d** ENG-LLE with 70 connections

data point of another component. Then edges can be constructed between certain nearest pair-wise data points from the two components. The number of the constructed edges is essential to the performance of dimensionality reduction. Taking the Broken Swiss Roll data as an example, Fig. 2 illustrates the 2D-embeddings of Isomap and LLE based on ENG when 3 edges or 70 edges are constructed between the two components of the neighborhood graph. It is found that adding insufficient edges and superfluous edges cannot provide satisfactory performances in NLDR. To solve the problem, we propose an adaptive method to construct connections between two components, which will be detailed later.

In order to preserve both inter-component information and intra-component information from high-dimensional space to low-dimensional space, the intrinsic dimensionality of the connection vectors constructed between the two components is assumed to be the same as the intrinsic dimensionality of the data. This assumption is always true when the all the subsets of the data points have a common intrinsic dimensionality. The intrinsic dimensionality can be evaluated by the contribution ratio of the d dominant singular values of the local tangent space [26,28,29]. Denoting the singular values of $\boldsymbol{Y}_i^{(k)} - \boldsymbol{y}_i\boldsymbol{e}^T$ as $\sigma_{i1} > \sigma_{i2} > \cdots > \sigma_{i\,\min(m,k)}$, where $\boldsymbol{Y}_i^{(k)}$ is a matrix form by $\boldsymbol{y}_i$'s k nearest neighbors, $\boldsymbol{e}$ is a column vector of 1s, the contribution ratio of the d dominant singular values of the local tangent space at $\boldsymbol{y}_i$ is given as

$$\eta_i^{(d)} = \sum_{j=1}^{d} \sigma_{ij} \Big/ \sum_{j=1}^{\min(m,k)} \sigma_{ij}. \tag{3}$$

Then the average contribution ratio is

$$\eta^{(d)} = \frac{1}{n} \sum_{i=1}^{n} \eta_i^{(d)}. \tag{4}$$

It is worth noting that larger $\eta^{(d)}$ indicates less loss in dimensionality reduction. The vector space defined by l connections between component p and component q is

$$\boldsymbol{\Delta}^{\{p,q\}(l)} = \boldsymbol{Y}^{\{p\}(l)} - \boldsymbol{Y}^{\{q\}(l)}, \tag{5}$$

where $\boldsymbol{Y}^{\{p\}(l)}$ is a matrix formed by $l$data points in component p, $\boldsymbol{Y}^{\{q\}(l)}$ is a matrix formed by l data points in component q, and the data pairs are sorted by their distances in ascending order, i.e., $D(\boldsymbol{y}^{\{p\}(1)}, \boldsymbol{y}^{\{q\}(1)}) < D(\boldsymbol{y}^{\{p\}(2)}, \boldsymbol{y}^{\{q\}(2)}) < \cdots < D(\boldsymbol{y}^{\{p\}(l)}, \boldsymbol{y}^{\{q\}(l)})$. Denoting the singular values of $\boldsymbol{\Delta}^{\{p,q\}(l)}$ as $\sigma_1^{(l)} > \sigma_2^{(l)} > \cdots > \sigma_{l\,\min(m,l)}^{(l)}$, the contribution ratio of the d dominant singular values of $\boldsymbol{\Delta}^{\{p,q\}(l)}$ is computed as

$$\eta^{(d,l)} = \sum_{j=1}^{d} \sigma_j^{(l)} \bigg/ \sum_{j=1}^{\min(m,l)} \sigma_j^{(l)}. \tag{6}$$

We need to find as many as connections, i.e., the largest value of l, that satisfies$\eta^{(d,l)} \geq \eta^{(d)}$ for any $l \leq l_{\max}$. In practice, $\eta^{(d,l)}$ is more sensitive to noise compared with $\eta^{(d)}$ because l is much smaller than n, where $\eta^{(d)}$ is the average of n values. To increase the robustness, we set $\eta^{(d,l)} \geq \xi\eta^{(d)}$, where ξ is set to be 0.95 or 0.99 empirically. The detail of this strategy is shown in Algorithm 1.

Algorithm 1 Construct edges between two components to be linked

Input: data subsets $\boldsymbol{Y}^{\{p\}}$ and $\boldsymbol{Y}^{\{q\}}$ corresponding to the two components to be linked, d, ξ, and $s = \min(|\boldsymbol{Y}^{\{p\}}|, |\boldsymbol{Y}^{\{q\}}|)$.

1. Form data pairs $\{\boldsymbol{y}^{\{p\}(u)}, \boldsymbol{y}^{\{q\}(u)}\}_{u=1}^{s}$ in ascending order according to their distances.
2. For l from $d+1$ to s, do
   $\boldsymbol{\Delta}^{\{p,q\}(l)} = \boldsymbol{Y}^{\{p\}(l)} - \boldsymbol{Y}^{\{q\}(l)}$
   apply SVD on $\boldsymbol{\Delta}^{\{p,q\}(l)}$ and compute $\eta^{(d,l)}$ by using equation (6)
   if $\eta^{(d,l)} < \xi\eta^{(d)}$, stop.
3. $l_{\max} = l - 1$.

Output: the edges constructed between $\{\boldsymbol{y}^{\{p\}(u)}, \boldsymbol{y}^{\{q\}(u)}\}_{u=1}^{l_{\max}}$.

The step 2 of Algorithm 1 constructs a hyperplane or hypersurface H defined by the edges added between the two components. The intrinsic dimensionality of H is $d_H \leq d$. If the spatial relationship or distance between the two components cannot be represented on a hyperplane or hypersurface with an intrinsic dimensionality $d_H \leq d$, Algorithm 1 will construct very few edges. The relationship between the two components is defined by the few edges as a result. Otherwise, Algorithm 1 will construct many edges to define the relationship between the two components.

Then we consider the second problem, i.e., determining the pairs of components to be linked. In this study, we propose to connect each component to its nearest neighbor. The distance between pair-wise components can be evaluated by the shortest Euclidean distances between point pairs belonging to two components of the graph. Through connecting each component with its nearest neighbor, pair-wise components are fused into a new component and the number of the components in the reconstructed graph will decrease but may not be one because certain pair-wise components maybe nearest neighbors mutually. We then iteratively construct connections between two nearest components until the neighborhood graph is connected. It is worth noting that if the vector space of the edges added between two components has the same dimensionality as the spaces of the two components have, the relative spatial positions of the two components will be definite and permanent. Similar to a jigsaw puzzle, connecting each component with its nearest neighbor iteratively will implicitly determine the relative spatial positions of all components and they will finally be fused into one component consisting of the whole graph. We call this method enhanced neighborhood graph (ENG) and its general framework is illustrated in Algorithm 2.

Algorithm 2 Enhanced Neighborhood Graph

Input: dataset and k.

1. Connect each data point with its k nearest neighbors to form a neighborhood graph G.
2. Use DFS to compute the connected components of G, i.e., $G = \{G_1, G_2, \cdots, G_g\}$.
3. Set $G^{old} = G$, $g^{old} = g$,

   While $g^{old} > 1$ do

   For i from 1 to g^{old} do

   Construct edges from G_i^{old} to its nearest neighbor by Algorithm 1;

   G^{new} is obtained;

   Use DFS to compute the connected components of G^{new}, i.e., $G^{new} = \{G_1^{new}, G_2^{new}, \cdots, G_{g^{new}}^{new}\}$;

   Set $G = G^{new}$, $g^{old} = g^{new}$.

Output: Enhanced neighborhood graph $\bar{G} = G^{new}$, on which NG-NLDR is performed.

The second step of Algorithm 2 utilizes DFS to get g components

$$G = \bigcup_{i=1}^{g} G_i. \tag{7}$$

There is no edge between any two vertices from G_p and G_q where $p \neq q$. In the third step of Algorithm 2, the graph components in each iteration are also acquired by applying DFS to G^{new}. In each iteration, edges are added between each component and its nearest neighbor. If $g = 1$, G^{new} is a connected graph, on which NG-NLDR can be performed. As $G = (V, E)$, the enhanced neighborhood graph is given as $\bar{G} = (V, E \cup \hat{E})$, where $\hat{E}$ denotes all the edges constructed by Algorithm 2.

In Algorithm 2, the time complexity of computing connected components by DFS is $\mathrm{O}(\max(|V|, |E|))$ [43], where $|E|$denotes the number of edges and $|V|$denotes the number of vertices (or the number of data points in a given dataset). The major computational cost of Algorithm 1 is the SVD on $\boldsymbol{\Delta}^{\{p,q\}(l)}$which has a time complexity of $\mathrm{O}(l^3)$but $d + 1 < l << |V|/c = n/c$, where $c$denotes the number of disconnected components. With Algorithm 1 and Algorithm 2, our method is scalable to large datasets. It worth noting that in the JME method proposed in [39], the major computational cost was the SVD on the $n_i \times n_j$weight matrices between pair-wise components or manifolds, where $\sum_i^c n_i = n$and $\sum_j^c n_j = n$. The time complexity of JME is about $\mathrm{O}((n/c)^3)$. As $l << n/c$, the computational complexity our method is much smaller than that of JME.

4 Experiments

4.1 NLDR for Synthetic Datasets

As shown in Figs. 1a and 3, five 3-D synthetic datasets including broken Swiss Roll, two parallel Swiss Rolls, broken S-curve, Four Moons, and Two Swiss Rolls of arbitrary positions

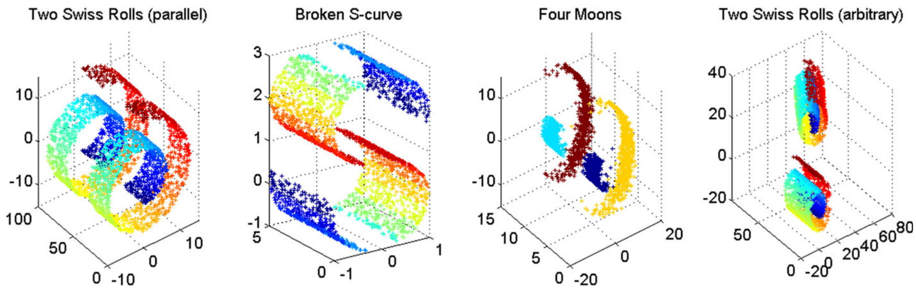

Fig. 3 Four synthetic datasets

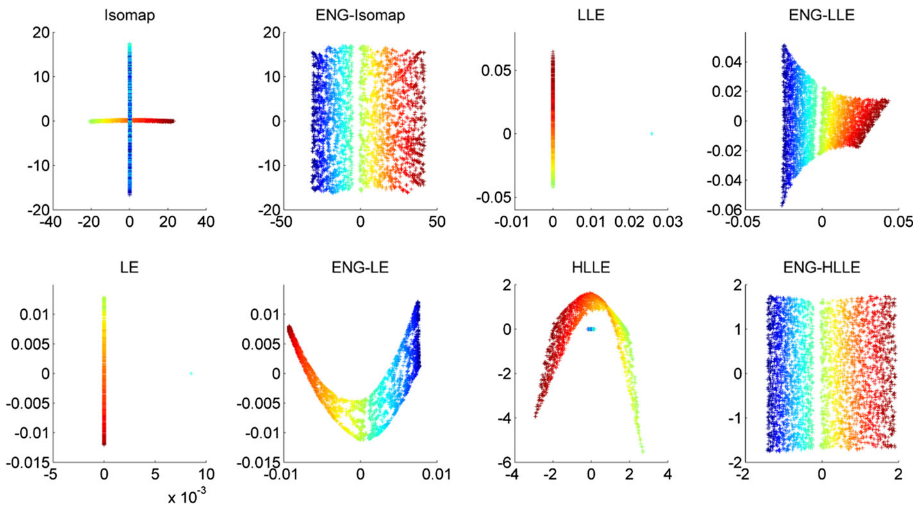

Fig. 4 Dimensionality reduction for Broken Swiss Roll

are used in this study. Each dataset has 3000 data points. We first compare JME, Isomap, LLE, LE, and HLLE with the ENG-based extensions, i.e., ENG-Isomap, ENG-LLE, ENG-LE, and ENG-HLLE on the five datasets. The number of nearest neighbors, i.e., the parameter k, affects the results of dimensionality reduction because all methods considered in this paper are local techniques and based on neighborhood graph. In addition, k has comparable effect to each classical method and its ENG-based extension because the construction of ENG is independent of k. We study their performance on whether they are able to capture the local properties accurately, provide satisfactory 2-D visualizations results and satisfy different metric evaluations. The performance of using different k was also evaluated. The results show that best performance can often be obtained when $k = 8$.

Figure 4 shows the results of 2-D embedding for Broken Swiss Roll data. Obviously, Isomap, LLE, LE, and HLLE fail to reduce the dimensionality. On the contrary, ENG-Isomap, ENG-LLE, ENG-LE, and ENG-HLLE show significant improvements, where both local properties and global structure can be well preserved, especially in the results of ENG-Isomap and ENG-HLLE. Comparing Fig. 4 with Fig. 2, we can notice that suitable number of connections constructed between two components can significantly improve the performance.

The results for Two Swiss Rolls (parallel), Broken S-curve, Four Moons, and Two Swiss Rolls (arbitrary) are given in Figs. 5, 6, 7, and 8 respectively. It can be seen that ENG-Isomap,

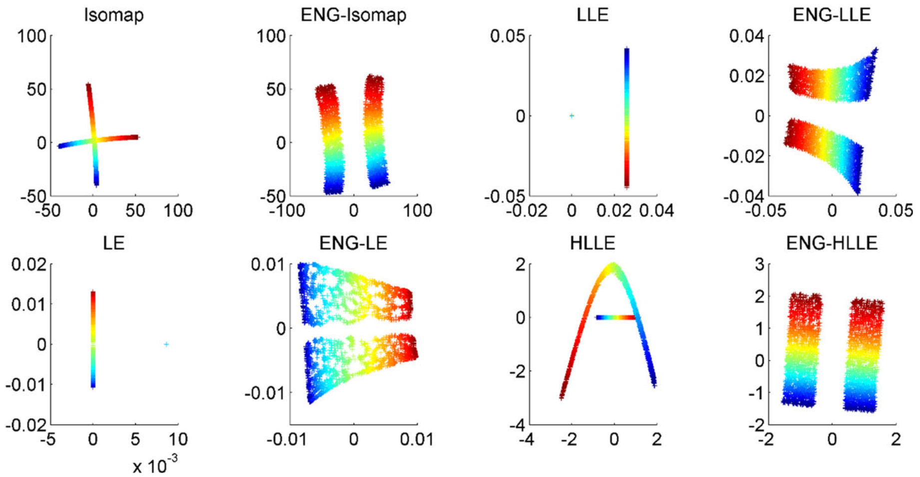

Fig. 5 Dimensionality reduction for Two Swiss Rolls (*parallel*)

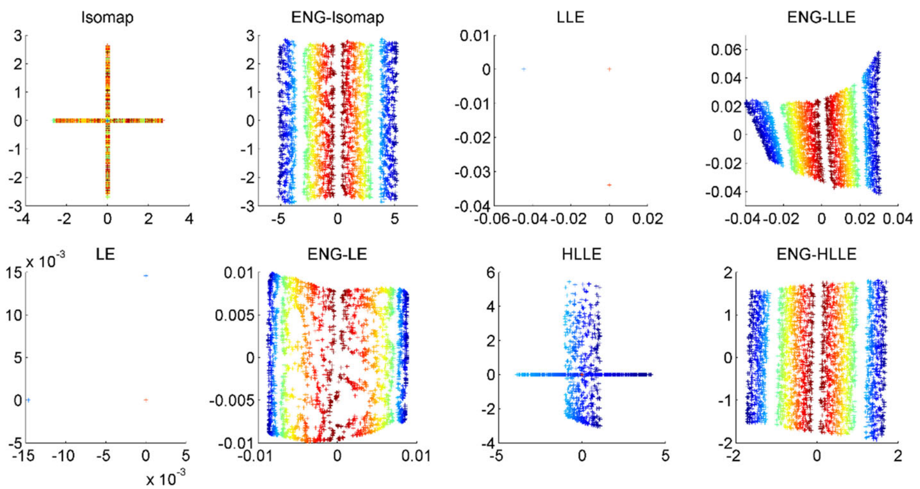

Fig. 6 Dimensionality reduction for Broken S-curve

ENG-LLE, ENG-LE, and ENG-HLLE all outperformed Isomap, LLE, LE, and HLLE significantly. It is worth noting that, for Broken S-curve, dimensionality reduction failed if we connected each sub-NG with the rest three components. This is because certain pair-wise components should be indirectly connected by a path formed by other components. Generally, for these datasets, Isomap, LLE, LE, and HLLE have the following drawbacks. They tend to stretch subsets of the data points into lines, or compress them into dots; they make the components overlapped or intersected; they underperform when the number of the graph components increase. In contrast, the ENG improves the performances of Isomap, LLE, LE, and HLLE in the following ways: they exactly connect pair-wise components; they find the suitable number of connections between the two components; they deliver remarkable dimensionality reduction results by preserving intra-component properties and inter-component properties. In other words, the presented results demonstrate that the enhanced neighborhood

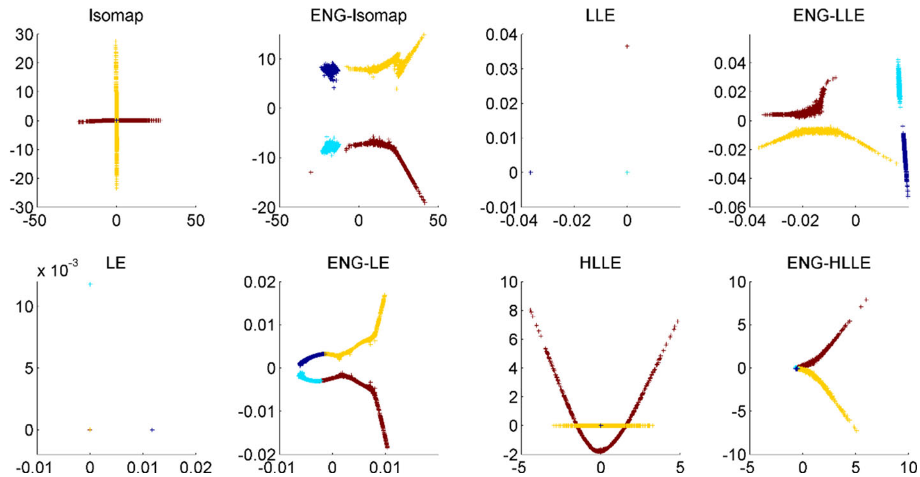

Fig. 7 Dimensionality reduction for Four Moons

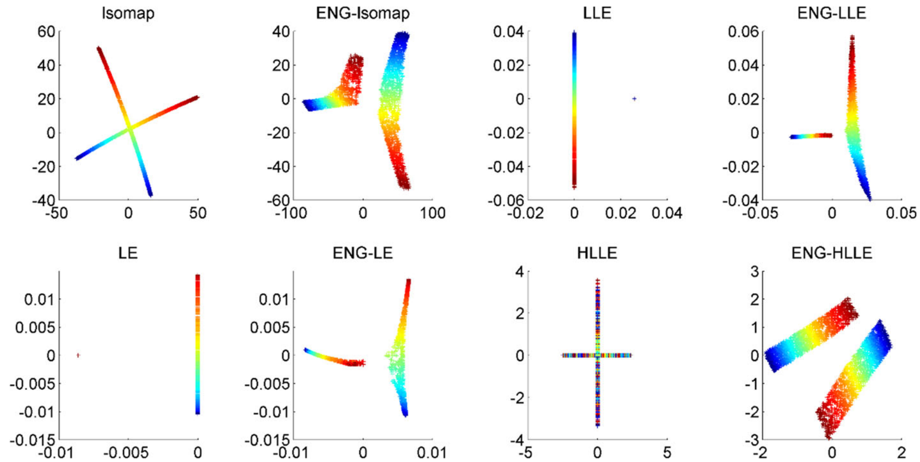

Fig. 8 Dimensionality reduction for Two Swiss Rolls (arbitrary)

graph can significantly improve the abilities of NLDR techniques for data with disconnected neighborhood graph. The 2-D embeddings given by JME for the five synthetic datasets are shown in Fig. 9. As can be seen, JME failed to recognize the intra-manifold structures and inter-manifold structures of Broken S-curve and Four Moons. Comparing Fig. 9 with Figures 4, 5, 6, 7 and 8, it can be seen that ENG-Isomap, ENG-LLE, ENG-LE, and ENG-HLLE outperformed JME significantly.

To compare ENG with existing methods of connecting neighborhood graph components, we show the results of Isomap and LLE with disconnected neighborhood graph, Max-Connect neighborhood graph, All-Connect neighborhood graph, and ENG on the S-curve dataset. In Max-Connect and All-Connect, we tried different numbers of edges added between two graph components. As shown in Figs. 10 and 11, all methods except ENG-Isomap and ENG-LLE, failed to visualize the data. The results indicate that the existing methods Max-Connect

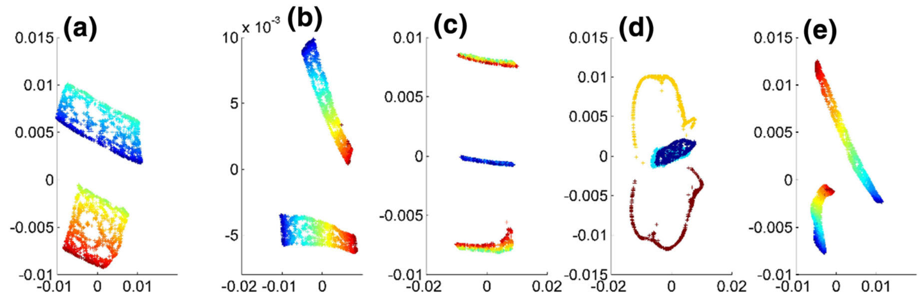

Fig. 9 JME dimensionality reduction for the five synthetic datasets: **a** Broken Swiss Roll, **b** Two Swiss Rolls (*parallel*), **c** Broken S-curve, **d** Four Moons, **e** Two Swiss Rolls (*arbitrary*)

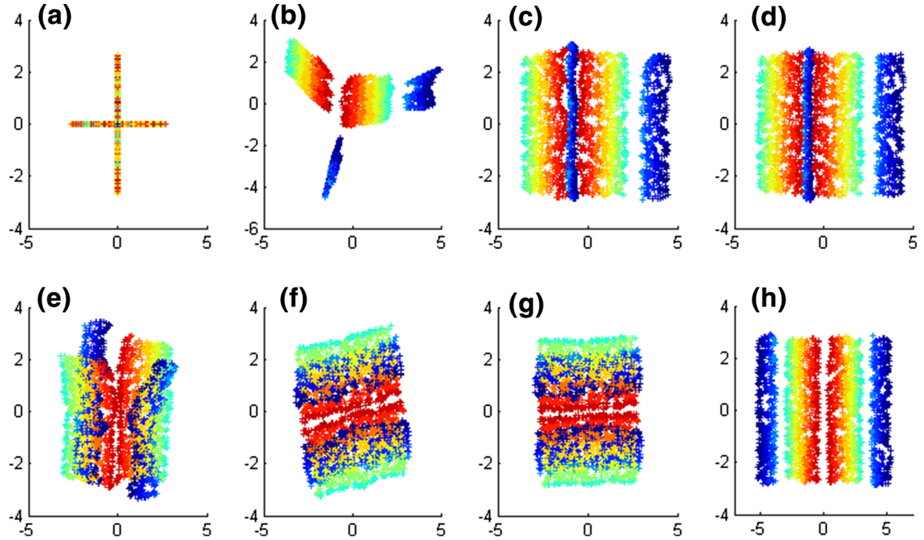

Fig. 10 Dimensionality reduction by Isomap with different methods of connecting graph components on the S-curve dataset: **a** Isomap with disconnected graph, **b–d** Isomap with Max-Connect graph of different number of added edges between two components, **e–g** Isomap with All-Connect graph of different number of added edges between two components, **h** ENG-Isomap

and All-Connect cannot provide exact connected neighborhood graph for dimensionality reduction and visualization.

4.2 NLDR for Natural Datasets

In this subsection, Isomap with different neighborhood graph (i.e., disconnected, Max-Connect, and All-Connect), JME, and ENG-Isomap are tested on two natural image datasets. The first dataset is COIL-20 [34], which contains images of 20 objects and each object has 72 images of different poses and the size of all images is 32×32. The parameter k for all methods was set as 8 that can outperform other values in terms of preserving local and global property. We perform dimensionality reduction on the whole dataset. The 2-D visualization results are shown in Fig. 12, where dots with different colors and markers indicate images of different subjects. It is worth mentioning that in Fig. 12, we only showed 33% of each

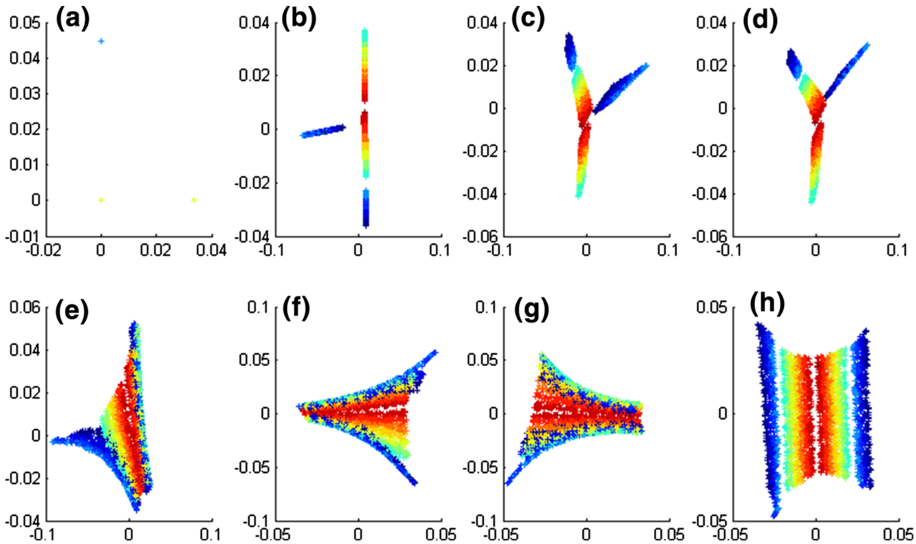

Fig. 11 Dimensionality reduction by LLE with different methods of connecting graph components on the S-curve dataset: **a** LLE with disconnected graph, **b**–**d** LLE with Max-Connect graph of different number of added edges between two components, **e**–**g** LLE with All-Connect graph of different number of added edges between two components, **h** ENG-LLE

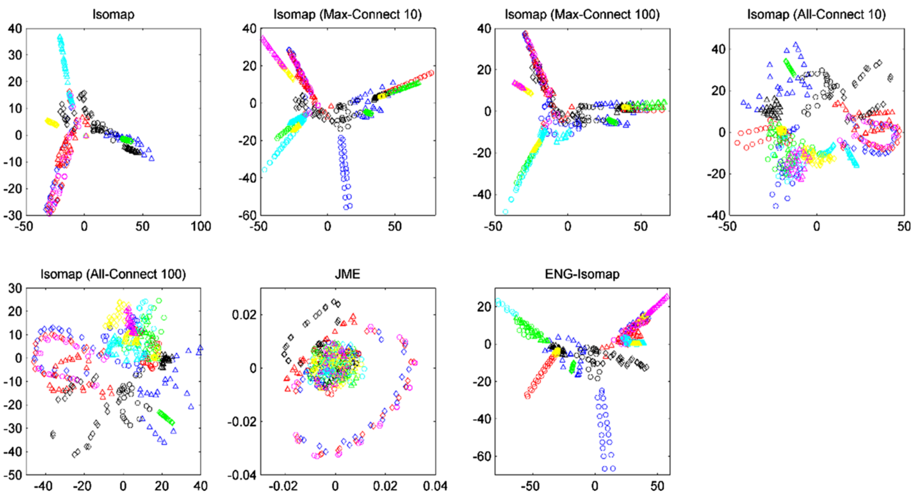

Fig. 12 Dimensionality reduction for COIL-20 data

object's images for convenience; otherwise the dots will be too dense and the markers cannot be clearly distinguished. We see that in all methods except Isomap with Max-Connect 10 and ENG-Isomap, there are significant overlaps among different objects' images. In addition, ENG-Isomap slightly outperformed Isomap with Max-Connect 10.

The MIT-CBCL face database [44] is the second studied dataset. The dataset contains face images of 10 persons and the number of images for each person is 324. We resized all images to 40×40 and visualize the data in 2-D space. In Fig. 13, images of different subjects

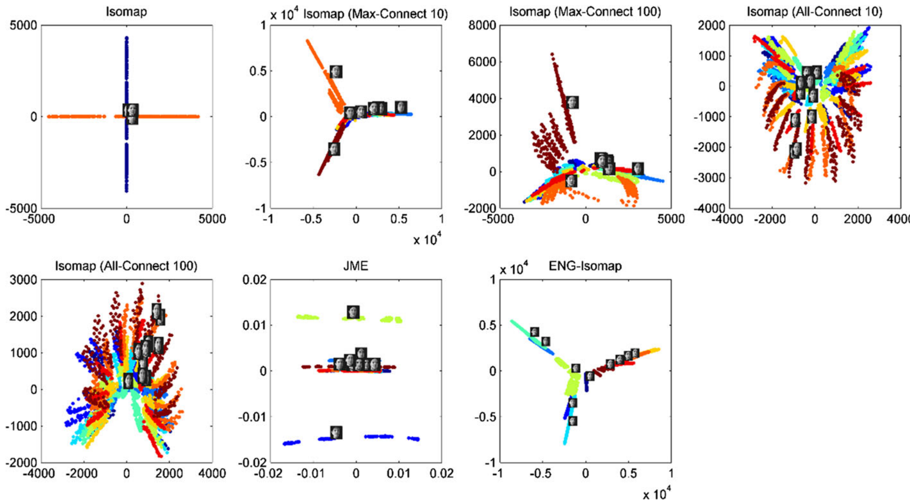

Fig. 13 Dimensionality reduction for MIT-CBCL face data

are shown by different color dots. It is clear that Isomap with disconnected neighborhood graph compressed the data of 8 subjects into a small cluster and the relative distributions of different subjects' images are completely lost. Isomap with connected graph given by Max-Connect or All-Connect also result in significant overlaps among different subjects' images. JME outperformed Isomap related methods but the images of 6 subjects are not cleanly separated. In contrast, ENG-Isomap placed different subjects' images in different regions of the 2-D space, where data of similar faces are relatively close. In general, compared with other methods, ENG-Isomap shows significant improvements for 2-D visualization.

4.3 Numerical Evaluation for NLDR

To numerically evaluate the effectiveness of the proposed method, we first refer to three classical metrics that are the generalization error of 1-NN classifier, trustworthiness, and continuity [6]. The generalization error of 1-NN classifier is to evaluate whether or not the nearest neighbor of each data point in high-dimensional space can be preserved in a low-dimensional space; the classifier is trained in a high-dimensional space but tested in a low-dimensional space. The trustworthiness measures the proportion of data points that are still very close together in the low-dimensional space. The continuity is to measure whether or not data points adjacent in low-dimensional space are originally neighboring in high-dimensional space. The parameter K for the trustworthiness, and continuity was set as 8 because the NLDR methods considered in this study is based on 8-nearest neighborhood graph. The results for the synthetic datasets are the averages of 20 repeated experiments. As shown in Table 1, generalization error rates of 1-NN classification, given by ENG-Isomap, ENG-LLE, ENG-LE, and ENG-HLLE, are significantly lower than those given by Isomap, LLE, LE, HLLE and JME. More specifically, in terms of the Broken S-curve, the effectiveness of ENG is the most significant, i.e., error rates of all ENG-based methods are lower than 20%, whereas error rates of all other studied methods are higher than 50%. The values of trustworthiness are reported in Table 2. As can be seen, for all datasets, the trustworthiness given by all methods based on ENG are larger than 0.9 and most of the values are close

Table 1 Generalization error rates of 1-NN classifiers (smaller values are desirable)

	Broken Swiss Roll	Two Swiss Rolls (parallel)	Broken S-curve	Four Moons	Two Swiss Rolls (arbitrary)	COIL-20	MIT -CBCL
None	7.23	9.17	9.49	0.1	11.63	0	0
Isomap	23.03	15.13	65.64	19.8	23.93	22.08	30.25
ENG-Isomap	7.7	9.83	10.33	0.33	15.97	15.56	3.27
LLE	31.46	22.4	66.11	0.1	30.37	19.03	0.03
ENG-LLE	8.8	9.73	12.65	0.1	11.81	12.8	1.75
LE	30.43	27.76	67.13	0.1	27.17	56.67	9.75
ENG-LE	12.27	11.81	18.49	0.1	16.17	12.36	0
HLLE	26.77	21.6	56.25	5.03	35.73	13.33	4.11
ENG-HLLE	11.5	10.3	13.02	0.2	12.8	17.35	0.09
JME	12.56	15.33	31.28	2.62	18.65	38.19	16.54

Table 2 Trustworthiness (larger values are desirable)

	Broken Swiss Roll	Two Swiss Rolls (parallel)	Broken S-curve	Four Moons	Two Swiss Rolls (arbitrary)	COIL-20	MIT -CBCL
Isomap	0.742	0.876	0.656	0.845	0.891	0.910	0.828
ENG-Isomap	1.000	0.999	1.000	0.998	0.996	0.952	0.925
LLE	0.816	0.858	0.805	0.861	0.839	0.878	0.965
ENG-LLE	0.998	0.999	0.999	0.985	0.988	0.959	0.925
LE	0.787	0.828	0.767	0.826	0.846	0.672	0.795
ENG-LE	0.997	0.998	0.987	0.997	0.950	0.969	0.993
HLLE	0.825	0.832	0.800	0.870	0.830	0.918	0.943
ENG-HLLE	0.999	1.000	1.000	0.983	1.000	0.934	0.969
JME	0.992	0.951	0.933	0.972	0.964	0.863	0.867

to 1. On the contrary, Isomap, LLE, LE, HLLE and JME deliver inferior results and most of the values are found to be smaller than 0.9. In Table 3, the continuity given by methods based on ENG are larger than those given by other methods. Taking the Broken Swill Roll dataset as an example, larger values of trustworthiness and continuity indicate that ENG based methods ensure data points adjacent in 3-D space are also neighboring in 2-D space, and vice versa. For the COIL-20 images and MIT-CBCL face images, the results measured by the three metrics demonstrate that the ENG-based methods are effective in preserving both intra-class properties and inter-class properties. Generally, the obtained results corroborate that our proposed ENG method is able to improve the performances of NLDR considerably under all the studied cases.

We also use the metric $R_{NX}(K)$ [45,46] to verify the superiority of our method. $R_{NX}(K)$ measures the neighborhood agreement from high-dimensional space to low-dimensional space, where K is the number of nearest neighbors. The value of $R_{NX}(K)$is within [0, 1], where 1 indicates a perfect agreement. We report the values of $R_{NX}(K)$at $K = 2^0, 2^1, 2^2, 2^3, \ldots, 2^{11}$. It is worth noting that $R_{NX}(K)$with large K may not be a suitable met-

Table 3 Continuity (larger values are desirable)

	Broken Swiss Roll	Two Swiss Rolls (parallel)	Broken S-curve	Four Moons	Two Swiss Rolls (arbitrary)	COIL-20	MIT -CBCL
Isomap	0.981	0.983	0.978	0.989	0.992	0.986	0.936
ENG-Isomap	1.000	1.000	1.000	0.999	0.998	0.993	0.995
LLE	0.984	0.906	0.927	0.927	0.898	0.960	0.995
ENG-LLE	0.998	0.999	0.999	0.998	0.995	0.982	0.989
LE	0.874	0.873	0.887	0.898	0.908	0.782	0.926
ENG-LE	0.998	0.999	0.993	0.998	0.995	0.994	0.998
HLLE	0.989	0.969	0.986	0.982	0.962	0.983	0.989
ENG-HLLE	0.999	1.000	1.000	0.993	1.000	0.983	0.992
JME	0.981	0.997	0.996	0.991	0.995	0.972	0.986

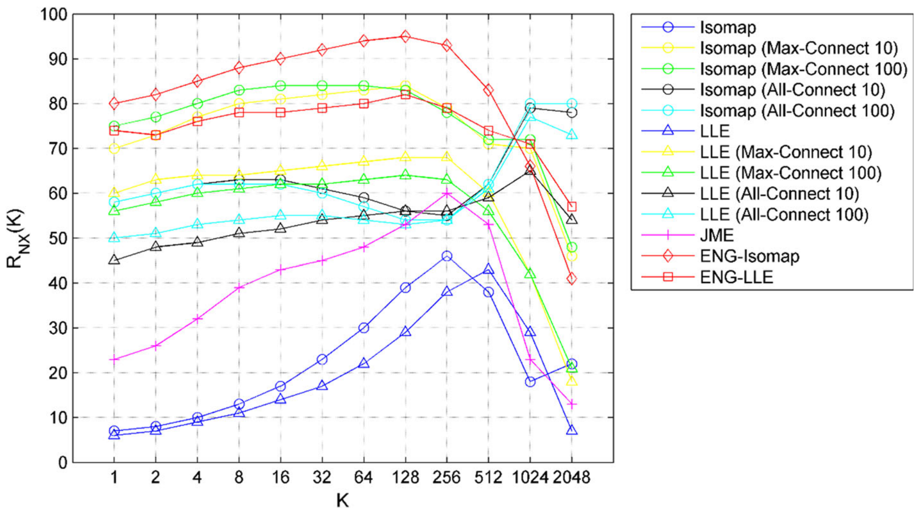

Fig. 14 $R_{NX}(K)$for Broken S-curve

ric of evaluating the performances of local NLDR methods. The reason is that local NLDR methods often do not impose the agreement of distant neighbors from high-dimensional space to low-dimensional space. In this study, the compared methods include ENG-Isomap, ENG-LLE, JME, Isomap/LLE with different methods of connecting graph components (i.e., disconnected, Max-Connect, All-Connect). Figure 14 shows the $R_{NX}(K)$curves for the Broken S-curve. As can be seen, ENG-Isomap/ENG-LLE give significantly higher $R_{NX}(K)$than JME and other Isomap/LLE methods do when $K \leq 512$. Figure 15 shows the $R_{NX}(K)$curves for COIL20 data, in which ENG-Isomap/ENG-LLE outperform JME and other Isomap/LLE methods when $K \leq 256$. Figure 16 reports the $R_{NX}(K)$curves for MIT-CBCL data. It is found that ENG-Isomap/ENG-LLE give higher $R_{NX}(K)$than JME and other Isomap/LLE methods do in most cases. These results confirm that the proposed ENG can significantly improve the performances of Isomap and LLE when the datasets have disconnected neighborhood graphs.

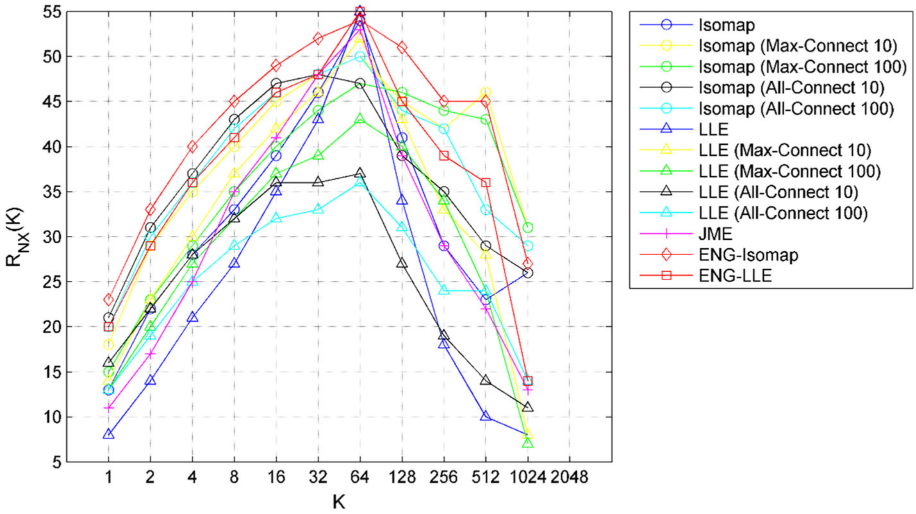

Fig. 15 $R_{NX}(K)$for COIL-20 data

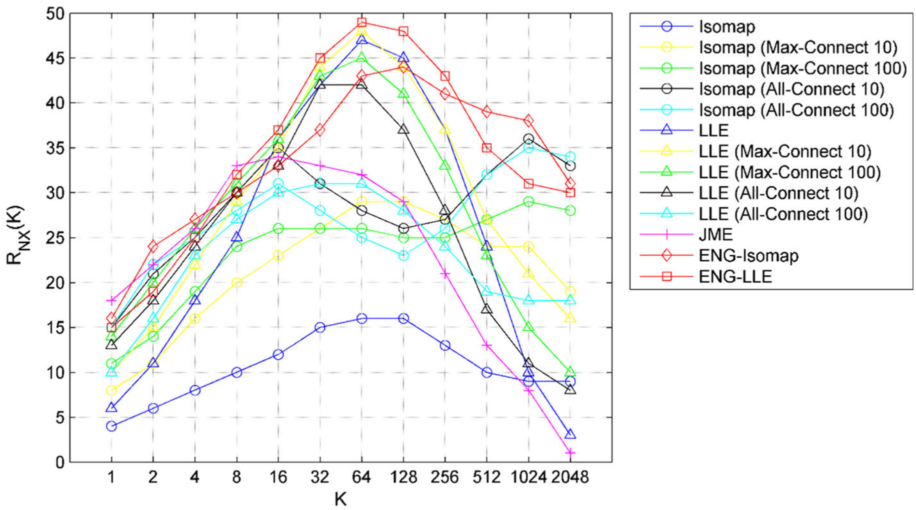

Fig. 16 $R_{NX}(K)$for MIT-CBCL face data

5 Conclusion

This paper showed that neighborhood graph-based nonlinear dimensionality reduction techniques (NG-NLDR) deliver unsatisfactory results when handling data with disconnected neighborhood graphs. We proposed a new approach called enhanced neighborhood graph (ENG) to solve the issue of disconnected neighborhood graph. The approach iteratively and adaptively adds edges among the graph components to construct a connected neighborhood graph. Thus, through combining the proposed ENG approach with conventional techniques like Isomap, LLE, LE, and HLLE, this paper shows that we can preserve the intra-component properties and inter-component properties. We also used four numerical metrics to evaluate the performance of the proposed methods. The thorough analysis together with other visualization results confirmed that ENG is an effective and efficient method to improve the performance of NLDR on datasets with disconnected neighborhood graphs. In addition, ENG is able to outperform existing methods of connecting neighborhood graph components. ENG is scalable to large-scale datasets and its computational complexity is smaller than that of JME.

Acknowledgements This work is partially supported by the National Science Foundation of China under Grant no.61601112, the Fundamental Research Funds for the Central Universities and DHU Distinguished Young Professor Program. This work is also partially supported by the Natural Science Foundation of China under grant no. 61572156.

References

1. Lee JA, Verleysen M (2007) Nonlinear dimensionality reduction. Springer, Berlin
2. Verleysen M, Lee JA (2013) Nonlinear dimensionality reduction for visualization. In: 20th international conference neural information processing, ICONIP 2013. Springer, Berlin, pp 617–622
3. Jolliffe I (2005) Principal component analysis. Encyclopedia of statistics in behavioral science. Wiley, Hoboken
4. Borg I, Groenen PJF (2005) Modern multidimensional scaling: theory and applications. Springer, Berlin
5. DeMers D, Cottrell GW (1993) Non-linear dimensionality reduction. In: Advances in neural information processing systems 5, [NIPS conference]. Morgan Kaufmann Publishers Inc., pp 580–587
6. Van der Maaten LJP, Postma EO, Van den Herik HJ (2009) Dimensionality reduction: a comparative review. J Mach Learn Res 10:66–71
7. Hoffmann H (2007) Kernel PCA for novelty detection. Pattern Recognit 40:863–874
8. Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Science 313:504–507
9. Mohebi E, Bagirov A (2016) Constrained self organizing maps for data clusters visualization. Neural Process Lett 43:849–869
10. Lee JA, Peluffo-Ordóñez DH, Verleysen M (2015) Multi-scale similarities in stochastic neighbour embedding: reducing dimensionality while preserving both local and global structure. Neurocomputing 169:246–261
11. Yang J, Fan L (2014) A novel indefinite kernel dimensionality reduction algorithm: weighted generalized indefinite kernel discriminant analysis. Neural Process Lett 40:301–313
12. Sammon JW (1969) A nonlinear mapping for data structure analysis. IEEE Trans Comput 18:401–409
13. Demartines P, Herault J (1997) Curvilinear component analysis: a self-organizing neural network for nonlinear mapping of data sets. IEEE Trans Neural Netw 8:148–154
14. Wan M, Lai Z, Jin Z (2011) Locally minimizing embedding and globally maximizing variance: unsupervised linear difference projection for dimensionality reduction. Neural Process Lett 33:267–282
15. Wang F, Zhang D (2013) A new locality-preserving canonical correlation analysis algorithm for multi-view dimensionality reduction. Neural Process Lett 37:135–146
16. Zhou Y, Sun S (2016) Local tangent space discriminant analysis. Neural Process Lett 43:727–744
17. Roweis ST, Saul LK (2000) Nonlinear dimensionality reduction by locally linear embedding. Science 290:2323–2326
18. Tenenbaum JB, De Silva V, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. Science 290:2319–2323
19. Belkin M, Niyogi P (2003) Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput 15:1373–1396
20. Donoho DL, Grimes C (2003) Hessian eigenmaps: locally linear embedding techniques for high-dimensional data. Proc Natl Acad Sci 100:5591–5596
21. Saul LK, Roweis ST (2003) Think globally, fit locally: unsupervised learning of low dimensional manifolds. J Mach Learn Res 4:119–155
22. Zhang Z, Zha H (2004) Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. J Shanghai Univ 8:406–424
23. Coifman RR, Lafon S (2006) Diffusion maps. Appl Comput Harmon Anal 21:5–30
24. Lee JA, Verleysen M (2005) Nonlinear dimensionality reduction of data manifolds with essential loops. Neurocomputing 67:29–53

25. Weinberger KQ , Sha F, Saul LK (2004) Learning a kernel matrix for nonlinear dimensionality reduction. In: Proceedings of the twenty-first international conference on Machine learning. ACM, Banff, p 106
26. Mekuz N, Tsotsos J (2006) Parameterless Isomap with adaptive neighborhood selection. Pattern Recognit. Springer, Berlin, pp 364–373
27. Samko O, Marshall AD, Rosin PL (2006) Selection of the optimal parameter value for the Isomap algorithm. Pattern Recognit Lett 27:968–979
28. Zhang Z, Wang J, Zha H (2012) Adaptive manifold learning. IEEE Trans Pattern Anal Mach Intell 34:253–265
29. Jia W et al (2008) Adaptive neighborhood selection for manifold learning. In: International conference on machine learning and cybernetics, 2008
30. Song Y et al (2008) A unified framework for semi-supervised dimensionality reduction. Pattern Recognit 41:2789–2799
31. de Ridder D et al (2003) Supervised locally linear embedding. Artificial neural networks and neural information processing—ICANN/ICONIP 2003. Springer, Berlin, pp 333–341
32. Huang Y, Xu D, Nie F (2012) Semi-supervised dimension reduction using trace ratio criterion. IEEE Trans Neural Netw Learn Syst 23:519–526
33. Zhang Z, Chow TWS, Zhao M (2013) M-Isomap: orthogonal constrained marginal Isomap for nonlinear dimensionality reduction. IEEE Trans Cybern 43:180–191
34. Nene SA, Nayar SK, Murase H (1996) Columbia object image library (COIL-20). Columbia University, New York
35. Liu X, Lu H, Li W (2010) Multi-manifold modeling for head pose estimation. In: 2010 IEEE international conference on image processing
36. Valencia-Aguirre J et al (2011) Multiple manifold learning by nonlinear dimensionality reduction. Springer, Iberoamerican Congress on Pattern Recognition
37. Torki M , Elgammal A, Lee CS (2010) Learning a joint manifold representation from multiple data sets. In: 2010 20th international conference on pattern recognition (ICPR). IEEE
38. Hadid A, Pietikäinen M (2003) Efficient locally linear embeddings of imperfect manifolds. In: Machine learning and data mining in pattern recognition: third international conference, MLDM 2003 Proceedings. Springer, Berlin, pp 188–201
39. Lee C-S, Elgammal A, Torki M (2016) Learning representations from multiple manifolds. Pattern Recognit 50:74–87
40. Yan S et al (2007) Graph embedding and extensions: a general framework for dimensionality reduction. IEEE Trans Pattern Anal Mach Intell 29:40–51
41. Lee CY (1961) An algorithm for path connections and its applications. IRE Trans Electron Comput EC–10:346–365
42. Tarjan R (1972) Depth-first search and linear graph algorithms. SIAM J Comput 1:146–160
43. Hopcroft J, Tarjan R (1973) Algorithm 447: efficient algorithms for graph manipulation. Commun ACM 16:372–378
44. Weyrauch B et al (2004) Component-based face recognition with 3D morphable models. In: Proceedings of the 2004 conference on computer vision and pattern recognition workshop (CVPRW'04), vol 05. IEEE Computer Society, p 85
45. Lee JA et al (2013) Type 1 and 2 mixtures of Kullback–Leibler divergences as cost functions in dimensionality reduction based on similarity preservation. Neurocomputing 112:92–108
46. Lee JA, Verleysen M (2014) Two key properties of dimensionality reduction methods. In: 2014 IEEE symposium on computational intelligence and data mining (CIDM). IEEE