LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-CONF-513768

Locally-Scaled Spectral Clustering using Empty Region Graphs

C. D. Correa, P. Lindstrom

November 14, 2011

ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2012
Beijing, China
August 12, 2012 through August 16, 2012

Disclaimer

Improving Spectral Clustering with Empty Region Graphs

Carlos D. Correa
Center for Advanced Scientific Computing
Lawrence Livermore National Laboratory
correac@llnl.gov

Peter Lindstrom
Center for Advanced Scientific Computing
Lawrence Livermore National Laboratory
pl@llnl.gov

ABSTRACT

This paper introduces a new method for estimating the local neighborhood and density of samples to improve the robustness of spectral clustering algorithms. We employ empty region graphs – geometric neighborhoods that connect two points as neighbors based on the inclusion or exclusion of points from a geometric region – to estimate the similarity or affinity between points. We show that averages of the distance to the neighbors of a point, such as the mean and the median, are good estimates of the local scale of points at the interior of clusters, and alleviates the need to manually select an appropriate scale parameter or an optimal number of neighbors. We also introduce a diffusion approach that improves this estimate for points at the boundaries of clusters. Our approach is validated on a number of synthetic data sets in low dimensions and on a few real classification examples, including image segmentation.

General Terms

Theory

Keywords

Spectral Clustering, Proximity Graphs

1. INTRODUCTION

Clustering is at the core of modern data mining tools. Common techniques, such as those based on K-means or explicit density models are being replaced by *spectral methods* for clustering, where points are clustered based on a spectral analysis of a matrix of pairwise similarities or affinities, instead of relying on a particular cluster model.

Spectral clustering has been applied successfully in a number of fields, including image segmentation, text mining, and data analysis in general. However, there remain a number of open questions: (1) How to define the neighborhood around data points to estimate a "good" affinity matrix, (2) how to accurately adapt the algorithm to the local scale or density of the data, and (3) how to automatically select the number of clusters. This paper concerns the former two questions.

KDD '12

The most common approaches to date rely on simple neighborhood techniques, such as k-nearest neighbor (kNN) graphs or ε-regions. However, clustering results may change dramatically for different values of k or ε. Alternatively, one can simply connect all points in a fully connected graph and rely on a scaling parameter σ to define the affinity between two points, similar to choosing an ε- graph. In either case, the technique relies on setting a specific parameter that varies from data set to data set.

In this paper, we exploit geometric graphs, namely *empty region graphs*, to construct neighborhood graphs without requiring a particular choice of the neighborhood extents. We show that these graphs improve the accuracy of spectral clustering algorithms. In many cases, the local density of data points changes along the domain, and strategies such as kNN fail to capture that change in density. We also introduce a diffusion-based mechanism that estimates the density based on the average neighborhood size around a point. We show that this local scaling algorithm, when combined with empty region neighborhoods, results in a better classification that is robust to noise and geometric (both linear and non-linear) transformations of the data points.

We show results on a number of synthetic benchmark data sets, as well as real multi-dimensional classification problems, including image segmentation.

2. RELATED WORK

Spectral Clustering. Spectral clustering is becoming a successful alternative to techniques based on k-means [21] or density models [11], and dates back to Donath and Hoffman [9] and Fiedler [12]. Recently, spectral clustering has found a niche in image segmentation [27], text mining [8] and as a data mining tool in general [25, 17]. Since then, there has been a trend in improving spectral clustering through a detailed analysis of the underlying graph structure [22], the scale and density parameters [31, 1], and the stability [15] and consistency [30] of the algorithm. Most related to our work are the techniques that attempt to estimate the local scale or density to improve spectral clustering of data with varying densities, shapes and levels of noise. One of the first to address this problem for data mining were Zelnik-Manor and Perona [31], who improve the general algorithm by Ng et al. [25] with local scaling. This approach, although effective even in high dimensions, was shown to be suboptimal for noisy data sets, or data with clusters of different densities [24]. To alleviate this problem, Nadler and Galun [24] introduce a coherence measure of set of points of belonging to the same cluster. Although not exclusive of spectral methods, the authors show that it alleviates some of the intrinsic limitations of spectral clustering. To deal with noise, Li et al. [19] propose a warping model that maps the data into a new space more suitable for clustering and more resilient to noise. Different alter-

natives are able to cluster data consisting of regions of arbitrary shapes, such as density based clustering [26] and, in a similar spirit to Zelnik-Manor and Perona's method, locally scaled density based clustering [1].

In this paper, we address the problem of locally-scaled and noise robust spectral clustering. We take a different approach and identify the problem as early as the selection of the neighborhood graph. Maier et al. suggest that the construction of the graph has a measurable effect on the results of spectral clustering [22]. Inspired by this paper, we turned to alternative neighborhood graphs, namely empty region graphs, in an effort to obtain better neighborhoods. Combined with a new locally-scaled affinity matrix and a diffusion approach, these graphs produce clusterings as good as or better than those in current alternatives.

Empty Region Graphs. Neighborhood or proximity graphs create a geometric structure that connects two points if they are close in some sense. These graphs have been well studied and include the relative neighborhood graph [16], the Gabriel graph [14], β-skeletons [18], σ-local graphs [2] and Delaunay triangulations [13]. A subset of these, called the empty region graphs, define a neighborhood graph where two points are connected if their neighborhood, defined by a geometric region, does not contain any other point [3]. These graphs have been well studied in terms of their geometric properties [6, 3], and have been applied in geographic analysis, pattern recognition and machine learning. Proximity graphs have been applied to clustering as well. Urquhart et al. [29] use the Gabriel graph and the Relative Neighbor graph to improve hierarchical clustering, noting that these graphs result in natural clusters that can be separated depending on the local graph density [29]. Carreira and Zemel apply an ensemble of minimum spanning trees to form neighborhood graphs that are more resilient to noise and varying densities [4]. Choo et al. propose an agglomerate method for hierarchical clustering, where an algorithm merges candidate clusters that belong to the same connected component in the Gabriel graph [5]. In this paper, we propose the use of empty region graphs, mainly the Gabriel graph and the Relative Neighbor graph, to construct more accurate and locally-scaled affinity matrices. We show that this approach is more effective when combined with a diffusion step that improves the separability of clusters and subsequently the block structure of the affinity matrix.

3. BACKGROUND

Our approach combines geometric neighborhoods, such as empty region graphs, with density estimation techniques and spectral clustering.

3.1 Spectral Clustering

Spectral clustering refers to a general algorithm where data are clustered into groups based on spectral analysis of a matrix of pairwise affinities or similarities between data points.

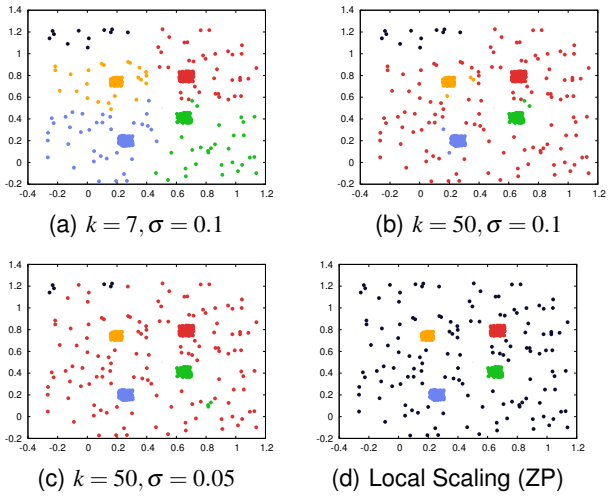
The general algorithm to produce a spectral clustering, as described by Ng et al. [25] , is as follows: Given a set of n points $S = \{s_1, s_2, \ldots, s_n\}$ in $\mathbb{R}^d$ that we want to cluster into K clusters,

- Construct an affinity matrix $A \in \mathbb{R}^{n \times n}$, where

$$A_{ij} = \begin{cases} e^{\frac{-d(s_i,s_j)^2}{\sigma^2}} & i \neq j \\ 0 & i = j \end{cases} \quad (1)$$

  for $d(s_i, s_j)$ a distance function, commonly Euclidean, and σ a scale parameter.

- Define D as a diagonal matrix $D_{ii} = \sum_{j=1}^{n} A_{ij}$.



(a) $k = 7, \sigma = 0.1$ (b) $k = 50, \sigma = 0.1$ (c) $k = 50, \sigma = 0.05$ (d) Local Scaling (ZP)

Figure 1: A single scale parameter and/or neighborhood parameter k are insufficient to correctly cluster data with varying densities. In all these cases, at least one of the small clusters is connected to the background noise. Local scaling (d) alleviates the need for tuning these parameters.

- Define the normalized Laplacian matrix $L = I - D^{-1/2} A D^{-1/2}$.

- Find the K smallest eigenvectors of L, and form the matrix $X \in \mathbb{R}^{n \times K}$ with these eigenvectors as columns.

- Form the matrix Y after normalizing the rows of X, so that $Y_{ij} = X_{ij} / \sqrt{\sum_j X_{ij}^2}$.

- Treat each row of Y as a point in $\mathbb{R}^K$ and cluster via k-means [21].

- Each point s_i is assigned to a given cluster c if the corresponding row i in Y is assigned to cluster c.

Clearly, the accuracy of the clustering depends, among other factors, on the selection of the scale parameter σ. Fig. 1 shows an example where global neighborhood parameters (k nearest neighbors and scaling σ) are used to cluster five groups of 2D points, four of which define small dense clusters, while another forms a sparse and noisy background. We see that exploring the different values of k and σ helps refining the shape of some of the small clusters, but it fails to separate one of them from the background, or creates spurious clusters. In general, it may occur that no combination of k and σ yields the correct clustering.

To deal with disparate densities, Zelnik-Manor and Perona define a more general affinity that incorporates local scaling [31]. Instead of a single scale parameter, they define the affinity between two points as:

$$A_{ij} = e^{\frac{-d(s_i,s_j)^2}{\sigma_i \sigma_j}} \quad (2)$$

where σ_i and σ_j are the local scale parameters estimated for points s_i and s_j, respectively. In the original paper, this parameter is defined as $\sigma_i = d(s_i, s_J^{(i)})$, where $s_J^{(i)}$ is the J'th neighbor of s_i. In practice, it was found that a single parameter $J = 7$ gave acceptable results. An example is shown in Fig. 1(d), where the algorithm is able to cluster the data correctly.

However, we found that the quality of the results using this approach still depends on a single parameter. In fact, as we will discuss in Section 5, a single value of J may not cluster data in the presence of noise or under nonlinear geometric transformations.

3.2 Empty Region Graphs

As an alternative to kNN graphs, a number of simpler, less costly neighborhood graphs have been proposed, such as the relative neighbor graph (RNG) and the Gabriel graph (GG), as surveyed by Jaromczyk et al. [16]. A family of these, known collectively as the *empty region graphs*, are more representative of the neighborhood of a point and less redundant than kNN, and more efficient to compute than simplicial tessellations such as the Delaunay triangulation.

Definition 1. A graph $G(V,R) = (V,E)$ is an empty region graph if for every edge $(p,q) \in E$, a canonical region $R(p,q)$ does not contain any other point in V:

$$pq \in E \iff R(p,q) \cap V = \varnothing \tag{3}$$

where the region R defines the neighborhood and is called the *empty region*.

Some common ERGs are:

Nearest Neighbor Graph (NNG). This is the directed graph that results from the empty region $R(p,q)$ formed by the open d-ball centered on p with radius $d(p,q)$.

$$pq \in E \iff \forall r \in V,\ d(p,r) \geq d(p,q) \tag{4}$$

Relative Neighborhood Graph (RNG). This graph is defined by a lune-shaped region consisting of the intersection of two d-balls of radius $d(p,q)$, one centered on p and the other centered on q, i.e.,

$$pq \in E \iff \forall r \in V,\ \max\{d(p,r), d(q,r)\} \geq d(p,q) \tag{5}$$

Gabriel Graph (GG). This is the graph defined by a d-ball centered at $\frac{1}{2}(p+q)$ with diameter $d(p,q)$, i.e.,

$$pq \in E \iff \forall r \in V,\ 2d(\tfrac{1}{2}(p+q), r) \geq d(p,q) \tag{6}$$

β-Skeleton. The so-called lune-based β-skeleton is a one-parameter generalization of the RNG and GG, defined as follows:

- For $0 < \beta < 1$, the empty region is the intersection of all d-balls with diameter $d(p,q)/\beta$ that have p and q on the boundary.
- For $\beta \geq 1$, the empty region is the intersection of two d-balls with diameter $\beta d(p,q)$ centered at $(1-\frac{\beta}{2})p + \frac{\beta}{2}q$ and $\frac{\beta}{2}p + (1-\frac{\beta}{2})q$.

It follows that $\beta = 2$ gives the RNG, while $\beta = 1$ is the GG. Finally, we note that geometric inclusion of one region within another also implies a partial order of the resulting neighborhood graphs (in terms of their edges), so that:

$$RNG \subseteq GG \subseteq (\beta \leq \tfrac{1}{\sqrt{2}})\text{-skeleton} \tag{7}$$

Fig. 2 shows the empty regions R associated with the RNG, GG and β-skeleton. The smaller the region, the denser the resulting graph becomes.

3.3 Locally Empty Region Graphs

Empty region graphs are, by definition, constructed in a global sense. For the purposes of clustering, a local definition of neighborhood may also be useful. Locally Gabriel and Locally Delaunay

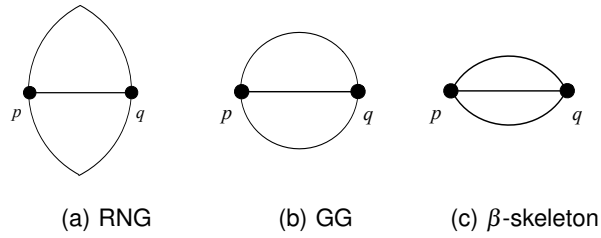


Figure 2: Examples of template regions for defining empty region graphs.

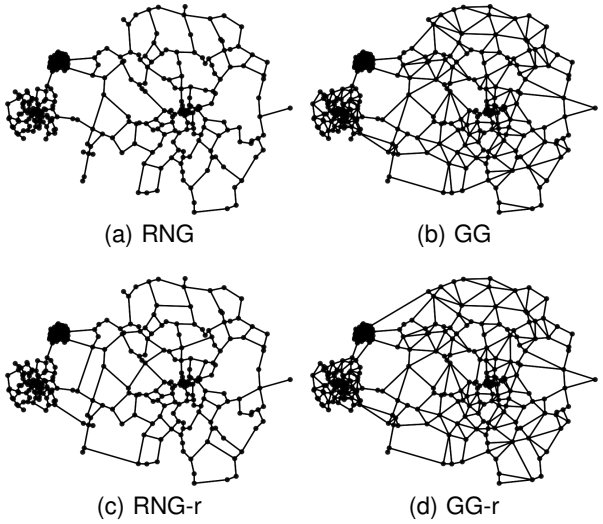


Figure 3: Examples of neighborhood graphs for a data set consisting of three clusters of varying density.

graphs have been studied before [], where the criterion for selecting an edge is applied with respect to a local subset of the points. This definition can be generalized to an empty region graph of arbitrary region template R, as follows:

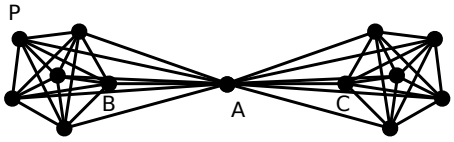
Definition 2. A graph $G = (V,E)$ parameterized by a region template R is a Locally Empty Region Graph (LERG) if for every edge $(p,q) \in E$, the empty region R_{pq} does not contain any neighbor of p or q in G.

Note that, according to this definition, there may be many locally ERGs for a given set of points. In fact XXX et al. [] have shown that finding an optimal LERG is an NP-complete problem.

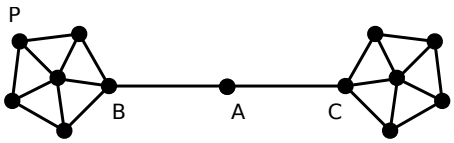
Fig. 3(a-d) shows the RNG, the GG, a local RNG and a local GG, respectively, for a set of 2D points forming three clusters of varying density. Note that the local graphs have more edges than the global graph, and that the local RNG is sparser than the GG. In general, a local ERG is more sparse than an ERG with a smaller region template. We exploit this property to obtain a much richer variety of neighborhood graphs.

4. APPROACH

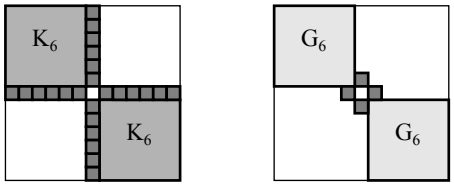
We now motivate the benefits of using empty region graphs for representing the neighborhood around a point, and consequently, for shaping the affinity matrix, and describe a general method for its application in spectral clustering.



(a) 6NN



(b) GG



(c) Affinity Matrices

Point	Mean 6NN	Median 6NN	Mean GG	Median GG
A	3.37	3.91	3	3
B	1.67	1.62	1.67	1
P	1.99	1.31	1	1

(d) Some average distances

Figure 4: Two proximity graphs (a) 6NN, (b) Gabriel graph (GG) and corresponding affinity matrices (c). The presence of a bridge node between clusters destroys the block structure of the matrix for 6NN, while it only adds a small block for GG. (d) The average distance to the neighbors in the 6NN graph is an unreliable measure of the local scale (particularly for a non-boundary point such as B), while it is reliable for the GG

4.1 Constructing the Affinity Matrix

Instead of the prevalent approach of defining the neighborhood of a point based on a fixed number of neighbors k or a fixed radius, we employ empty region graphs. The affinity is defined as:

$$A_{ij} = \begin{cases} e^{\frac{-d(s_i,s_j)^2}{\sigma_i \sigma_j}} & (i,j) \in ERG(V,R) \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

where R is an empty region template. In the examples throughout this paper, we use the Gabriel and Relative Neighbor graphs.

4.1.1 Benefits of ERGs

Common practices, such as the k nearest neighbor graph, are susceptible to short circuiting nearby clusters due to a point in the middle, as shown in Fig. 4(a). In this particular case, the separability of the two clusters is a bit more difficult due to the connecting node. We see in Fig. 4(c) that the corresponding affinity matrix, while we would expect a perfect block diagonal matrix, is not entirely separable into distinct blocks.

Using an empty region graph, however, alleviates this problem, as illustrated in Fig. 4(b). We see that, in this case, the connecting node (which may be due to noise or the presence of a smaller cluster), is only connected via a single edge to each cluster and the block structure of the affinity matrix does not change dramatically. The corresponding affinity matrix 4(c) on the right now can be separated into three blocks. The extra block, formed by the bridge node (A) and the two boundary nodes B and C is increasingly small as the size of the clusters increases.

In fact, it is easy to see that for the Relative Neighbor Graph, there is at most one edge connecting a bridge node and a cluster, regardless of the density of the cluster, when the distance from the bridge node to the boundary of the cluster is larger than the radius of the cluster. Therefore, in those cases we are guaranteed that the affinity matrix will only have two extra entries for each pair of connected block structures.

4.1.2 Computation of Locally ERGs

In practice, the edge complexity of the neighborhood graph helps determine the selection of clusters at multiple scales. The sparser the graph, the easier it is to pick clusters at small scales. As we illustrate in Fig. 2, locally ERGs are slightly denser than their global counterparts, but do not increase the edge complexity in a quadratic or exponential manner. Thus, they provide a good balance between connectivity and sparseness that is useful to effective spectral clustering.

However, obtaining an optimal Locally ERG is NP-complete, as demonstrated by XXX et al. []. Here, we derive a greedy approximation of a locally ERG that minimizes the distance of a point to its neighbors.

Let $N(p) = \{q : pq \in E\}$ denote the neighbors of a vertex p, and let q_i denote the i^{th} nearest sample to p, with $q_0 = p$. An rERG with empty region R is defined in terms the following recurrence:

$$N_1(p) = \{q_1\} \quad (9)$$

$$N_i(p) = \begin{cases} N_{i-1}(p) \cup \{q_i\} & \text{if } R(p,q_i) \cap N_{i-1}(p) = \varnothing \\ N_{i-1}(p) & \text{otherwise} \end{cases} \quad (10)$$

$$N(p) = N_{n-1}(p) \quad (11)$$

where n is the number of data points. In other words, we construct $N(p)$ by adding points in order of increasing distance, as long as the respective edges do not contain any point already in $N(p)$. Since we consider only a subset of the points in V in the containment test, it is easy to see that $ERG(R) \subseteq lERG(R)$ for all empty regions R. This approximation has been shown to improve the topological segmentation of sparsely sample points [7].

4.2 Local Scaling

Another issue with clustering is the selection of a local scale, as opposed to a global scale parameter. A natural measure of the local scale around a point include functions of the distance to its neighbors, such as the mean or the median. However, these measures can be brittle in kNN graphs. To illustrate this problem, consider the neighborhood in Fig. 5(a). The average distance of the center point to its neighbors, denoted as a dashed circle, is an accurate estimate of the local scale. In this particular case, the neighborhoods of kNN and the Gabriel graph are the same. Now let us consider a slightly different neighborhood in Fig. 5(b), consisting of a smaller cluster formed by the five points at the center. The average of the 8NN distances overshoots the scale and misrepresents the extents of the cluster. In fact, small perturbations of the points may change this average in both directions. In contrast, the average of the Gabriel graph is an accurate estimate of the small scale of this cluster and is not sensitive to the density of the cluster, as seen in Fig. 5(c).

Based on the above observation, we follow an approach similar to Zelnik and Perona's [31] and estimate the local scale σ_i based on the neighborhood of each point. Unlike kNN graphs, averages of ERGs are good estimates of the local scale. These include measures such as the mean and median distance to the neighbors. For

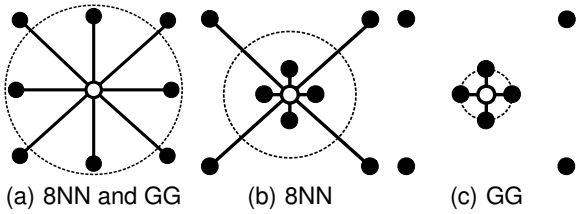


Figure 5: Average distance to neighbors as a measure of local scale (a) Correctly estimates the scale for the GG and 8NN. (b) After moving some of the points, now the average is an unreliable measure of the local scale. (c) The average correctly measures the local scale for the GG.

example, for the mean distance, the local scale at a point s_i is:

$$\sigma_i = \frac{1}{|N(s_i)|} \sum_{j \in N(s_i)} d(s_i, s_j) \tag{12}$$

where $N(s_i)$ is the set of neighbors of s_i.

For kNN neighborhoods, this average changes quickly for small values of k, while it varies slowly for large k, making it an unreliable predictor for the local density of points. On the other hand, neighborhoods in graphs such as the Gabriel graph, where neighbors are picked in sufficiently different directions, the average is a better predictor of density, as depicted in Fig. 5(c).

As an example, consider Fig 4 and the summarized averages for three points A, B and P, where the radius of each cluster is approximately 1 and $|AB| = |AC| = 3$. The averages for the 6NN graph overestimate the local scale, especially for a point P which is not at the boundary with respect to A. On the other hand, the Gabriel graph correctly estimates the local scale of the cluster as 1 for point P. For a boundary point B, both graphs overestimate the local scale when using the average.

The inaccuracies at boundary points is an issue with supersets of the Relative Neighbor graph, including the Gabriel graph, which are connected. This suggests that while most of the edges occur *within* clusters, a number of cluster pairs are still connected via long edges. Therefore, the local scale of boundary points will be larger than those at interior points, as shown for the example in Fig. 4.

To counteract this problem, we observe that, in particular for ERGs, most of the neighbors of a boundary point are points within the same cluster. Therefore, if we average the local scale with that of its neighbors, it will converge to a local scale more representative of the enclosing cluster. This strategy, based on non-linear diffusion, is described below.

4.3 Diffusion-based Scale Refinement

To deal with the local scale of boundary points, we propose a diffusion approach, whereby the local density of a point is smoothly blended with the local densities of its neighboring points. Because shorter edges are likely to correspond to points in the same cluster, diffusion using an inverse distance weighted kernel, such as Gaussian or inverse polynomials is preferred. This diffusion takes place over a number of iterations T, which converges to an estimate of the local estimate of the data points.

We propose to diffuse the local density instead of the local scales, in an attempt to converge to the result of equivalent kernel density estimators [10], using the reciprocal of the local scale as an approximation of density. Therefore, we define the diffusion process as the weighted harmonic mean of the local scales of the neighbors.

We propose an iterative process that progressively smoothes the local scale among the data points. At a given iteration $t+1$, $t \in \{0, \ldots, T\}$,

$$\sigma_i^{(t+1)} = \left(\sum_{j \in N(s_i) \cup \{s_i\}} \hat{w}_{ij} \frac{1}{\sigma_j^{(t)}} \right)^{-1} \tag{13}$$

where $\sigma_i^{(0)}$ is the local scale computed using Eq. (12) and the weights are normalized kernels

$$\begin{aligned} \hat{w}_{ij} &= w_{ij} / \sum_j w_{ij} \\ w_{ij} &= G(d(s_i, s_j), \rho_G) \end{aligned}$$

with G a kernel function. A number of kernel functions can be applied, Gaussians being the most common:

$$G(d, \rho) = e^{-d^2/\rho}$$

with ρ a number controlling how fast the diffusion process works. A smaller ρ diffuses the values slowly, while a larger ρ makes the diffusion converge to the arithmetic mean of the densities. The number of diffusion steps T depends on how noisy the data is. In our experiments, $T = 1$ suffices for data sets with little or no noise. Although this introduces an extra parameter, we observed that this algorithm produces better results even for a single iteration, and exhibits convergence. Therefore, this technique can be self-tuned by picking the first T that produces stable results above a given tolerance level.

This process ensures that intracluster scales are made more uniform than intercluster scales. However, since this is an isotropic diffusion, eventually the scales are blended together and the solution converges to the arithmetic mean. Ideally, we want to constrain the diffusion process within clusters and prevent it along edges between clusters. To this end, we turn to non-linear diffusion kernels, such as the bilateral filter [28].

The simplest approach is to introduce an additional kernel that penalizes the weight when the difference in scale is high. As we observed in our experiments, nodes between clusters are represented with different scales than those within clusters. Therefore, we weight less those edges where the difference in scale is high. The new weights are defined as:

$$w_{ij} = G(d(s_i, s_j), \rho_D) G(|\sigma_i^{(t)} - \sigma_j^{(t)}|, \rho_S) \tag{14}$$

where ρ_D and ρ_S are parameters controlling the speed at which diffusion propagates the local scales in terms of the distance between points and their difference in scale.

4.3.1 Parameter Selection

Our approach, although designed to reduce the selection of a global parameter k for the number of neighbors or σ for the global scale, requires the selection of a number of parameters, namely the choice of ERG (e.g., Gabriel, RNG) and the diffusion parameters, T, ρ_D and ρ_S. Since there is only a few choices (RNG, GG, β-skeleton, and their corresponding relaxed versions), selecting the ERG is not as tedious as searching for good neighborhood parameters k and σ. Finding the optimal diffusion parameters may require more exploration, but we argue that setting these parameters involve less user intervention than the selection of a global parameter σ. First, the diffusion parameters are similar to those used in kernel density estimation and diffusion in general, and have been well studied [20]. In our experiments, setting these parameters high result in isotropic smoothing of the local densities, which in most cases produce more accurate results than using a global scale. Second, diffusion exhibits convergence. Therefore, one can devise a

mechanism to progressively smooth the local scales in the search of stability regions. Our results seem to confirm that the gain in accuracy is well worth the exploration of these diffusion parameters.

4.4 Algorithm

To summarize, our proposed algorithm for robust spectral clustering is as follows:

input : Data points $S = \{s_1, \ldots, s_N\}$, number of clusters K, number of iterations T, diffusion parameters ρ_D, ρ_S
output: Classification C

$G \leftarrow \text{ERG}(S)$ *Construct ERG*
for $i \leftarrow 1$ *to* $|S|$ **do**
  $N_i \leftarrow \text{NEIGHBORS}(G, s_i)$
  $\sigma_i^{(0)} \leftarrow \text{LOCAL SCALE}(s_i, N_i)$ *Estimate initial local scale*

for $t \leftarrow 1$ **to** T **do**
  for $i \leftarrow 1$ *to* $|S|$ **do**
    $N_i \leftarrow \text{NEIGHBORS}(G, s_i)$
    $\Sigma_i \leftarrow \{\sigma_i^{(t-1)}\}$
    for $j \in N_i$ **do**
      $\Sigma_i \leftarrow \Sigma_i \cup \{\sigma_j^{(t-1)}\}$
    $\sigma_i^{(t)} \leftarrow \text{DIFFUSE}(\Sigma_i, \rho_D, \rho_S)$ *Update local scale*

$A \leftarrow \text{AFFINITY}(\{s_i\}, \{\sigma_i\})$ *Compute affinity matrix*
$C \leftarrow \text{SPECTRALCLUSTER}(A)$ *Cluster using the spectrum... ...of the Laplacian of A [25]*

Algorithm 1: New algorithm for spectral clustering using empty region graphs

4.5 Complexity Analysis

We now describe the computational cost of the key steps in our algorithm.

Construction of ERG Constructing an ERG can be expensive. A brute-force implementation requires

$$O(n^3) \text{ time,}$$

which is prohibitively expensive for most practical applications. Certain ERGs can be computed at a lower cost. For example, Touissant et al. [] provide an algorithm that constructs the Gabriel graph in

$$O(n^2) \text{ time.}$$

We approximate ERGs by restricting the neighbor search to the k_{max} nearest neighbors of a point, where k_{max} is usually larger than the k value selected for k nearest neighbor graphs, but $k_{max} << N$. The overall complexity of computing such a kNN graph is

$$O(n^2 \log k_{max}),$$

and the additional cost of computing an ERG becomes

$$O(n^2 k_{max}).$$

Computing a locally ERG is also faster, and only requires

$$O(nk_{max}^2) \text{ time.}$$

Diffusion The diffusion process involves a sequential walk on the neighborhood graph and requires

$$O(|E|T) \text{ time.}$$

where T is the number of iterations and $|E|$ is the number of edges in the graph, $|E| < nk_{max}$.

Solution Eigenvector Problem In general, computing the eigenvectors of the Laplacian matrix takes $O(n^3)$ time, but, as described by Song et al. [], sparse eigensolvers, such as the variants of Lanczos/Arnoldi factorization, have a cost of

$$O(m^3) + (O(nm) + O(nk_{max}) + O(m-k)) \times (\text{\# restarted Arnoldi}).$$

Clustering the spectrum Finally, using k-means to cluster the eigenvectors has a cost of

$$O(nK) \times \text{\# k-means iterations}$$

Overall, as suggested by Song et al. [] and Ding et al. [], the cost of the clustering depends on the construction of the affinity matrix. Nonetheless, the additional cost of using an ERG instead of a KNN is only linear on the maximum number of neighbors k_{max}.

5. RESULTS

We have validated our algorithm with a number of low dimensional synthetic data sets and a few (higher-dimensional) real classification problems. We compare our results with traditional spectral clustering using a global scale [25], as well as with Zelnik-Manor and Perona's approach (ZP) [31]. To compare the quality of these datasets, we measure the normalized mutual information (NMI):

$$NMI(X;Y) = \frac{2I(X;Y)}{H(X)+H(Y)} \tag{15}$$

where $I(X;Y)$ is the mutual information between sets X and Y and $H(X)$ and $H(Y)$ are the entropies of the clustering and classification sets, respectively. Other metrics, such as Variation of Information [23], can be defined in terms of NMI, e.g. $VI(X;Y) = 1 - NMI(X;Y)/(2 - NMI(X;Y))$.

We have explored our approach using a number of synthetic data sets, also used in [31]. These data sets are characterized for having clusters of varying density, scale and shape, where spectral algorithms using a global scale are known to perform poorly.

5.1 Stability of Parameters

To further discuss the effect of our proposed algorithms, we compare the stability of the diffusion mechanism vs. more traditional approaches, which require the search of a scaling parameter σ. **SHOW SURFACE**

5.2 Sensitivity to Transformations

5.3 Validation on UCI Data

5.4 Image Segmentation

EXPLAIN TABLE 1

6. SUMMARY AND CONCLUSION

This paper introduced a method to improve the accuracy of spectral clustering algorithms using empty region graphs and a diffusion-based local scaling approach. We learned that empty region graphs such as the Gabriel graph and the Relative Neighbor graph are especially useful when clusters can be grouped into non overlapping regions in space. Special attention must be given to the cases where clusters enclose each other in extreme shapes, such as concentric rings or spirals. We have also shown that those cases are naturally handled by our diffusion approach, since intercluster edges tend to

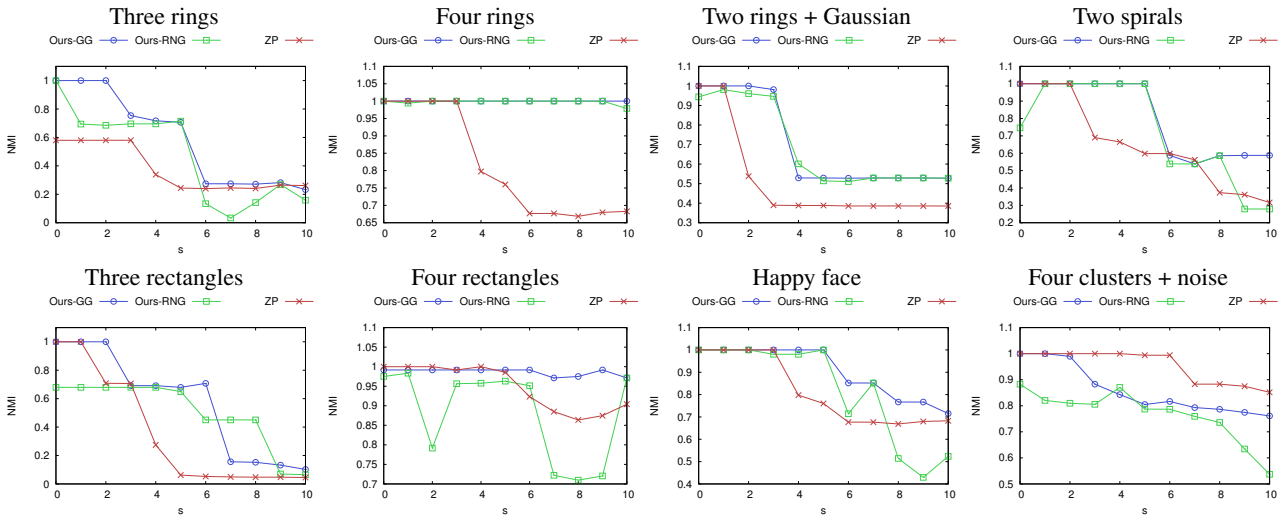
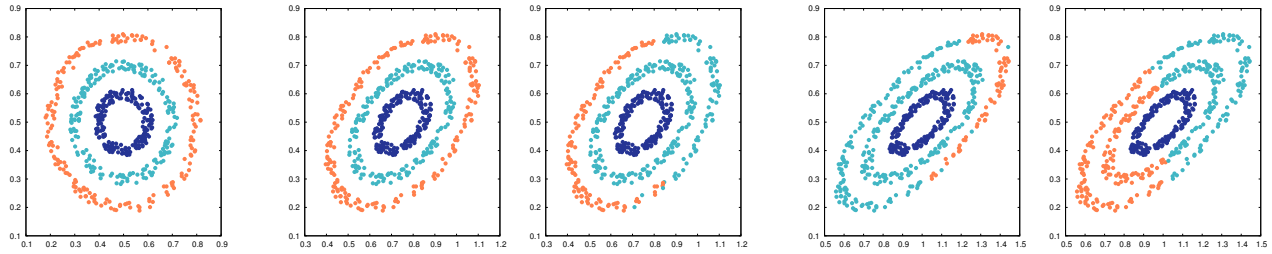


Figure 6: Test of geometric transformations. TBD

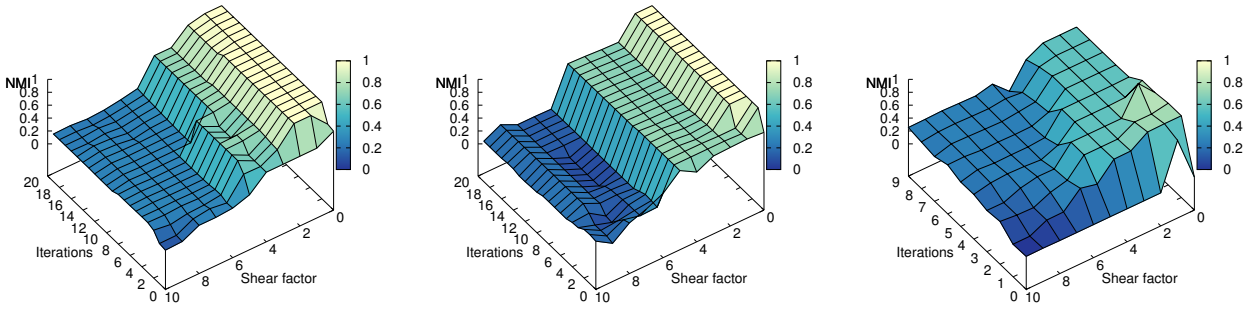
Table 1: Summary of experiments on multi-dimensional data

Data set	Dimensions	Num. Clusters	Ours (Graph, Num.Its)	Global (KNN)	Local (KNN)	DBSCAN
Gaussians	3	3	**0.982283** (GG-r, 1) **0.982283** (RNG-r,1)	0.822146	**0.982283** (K=7)	**0.982283**
Noisy Rings	3	2	0.983877 (GG, 5) **1.000000** (RNG,9)	0.983877	0.948838 (K=10) 0.910986 (K=7)	**0.95720**
Ellipsoids	3	3	**0.979854** (GG, 11) 0.975918 (RNG-r,11)	0.958760	0.964494 (K=5) 0.960750 (K=7)	0.67055
Iris	4	3	**0.847160** (GG-r, 24) 0.810730 (RNG-r,3)	0.833720	0.810730 (K=3) 0.790646 (K=7)	0.73140
Ellipsoids5D	5	5	0.792720 (GG-r, 15) **0.795458** (RNG-r,21)	0.728345	0.771108 (K=5) 0.746859 (K=7)	0.53799
Auto-mpg	5	5	0.658052 (GG-r, 9) 0.701952 (RNG-r,5)	0.648400	**0.734673** (K=1) 0.635056 (K=7)	0.47103
E.coli	7	8	**0.675533** (GG-r) 0.674021 (RNG-r)	0.671797	0.658698 (K=2) 0.620015 (K=7)	0.43751
Breast	9	2	0.750583 (GG-r) **0.756719** (RNG-r)	0.741520	0.752766 (K=1) 0.735777 (K=7)	0.69536
Glass	9	6	**0.447526** (GG-r,5) 0.416217 (RNG,9)	0.355411	0.427606 (K=2) 0.366673 (K=7)	0.45900
Wine	13	3	0.835249 (GG) 0.875893 (RNG)	**0.938507**	0.882215 (K=1) 0.866091 (K=7)	0.52798



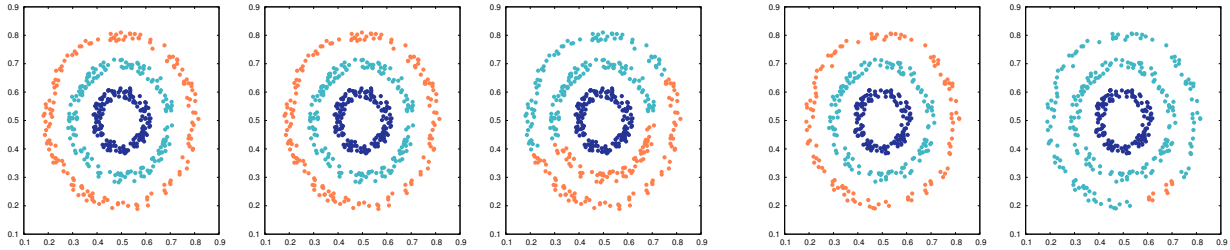
Data GG (NMI=1) ZP (NMI=0.7607) GG (NMI=0.718) ZP (NMI=0.576)

(a) Example data set undergoing a geometric transformation



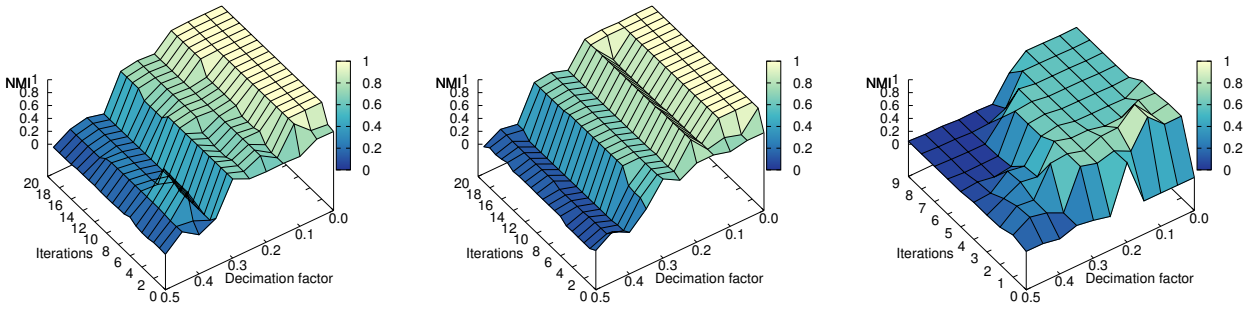
GG RNG ZP

(b) Quality of clustering as we apply a transformation to the data



RG (NMI=1) ZP (NMI=1) ZP (NMI=0.587) RNG (NMI=1) ZP (NMI=0.721)

(c) Example data set undergoing decimation



GG RNG ZP

(d) Quality of clustering as we decimate the data points

Figure 7: Tests

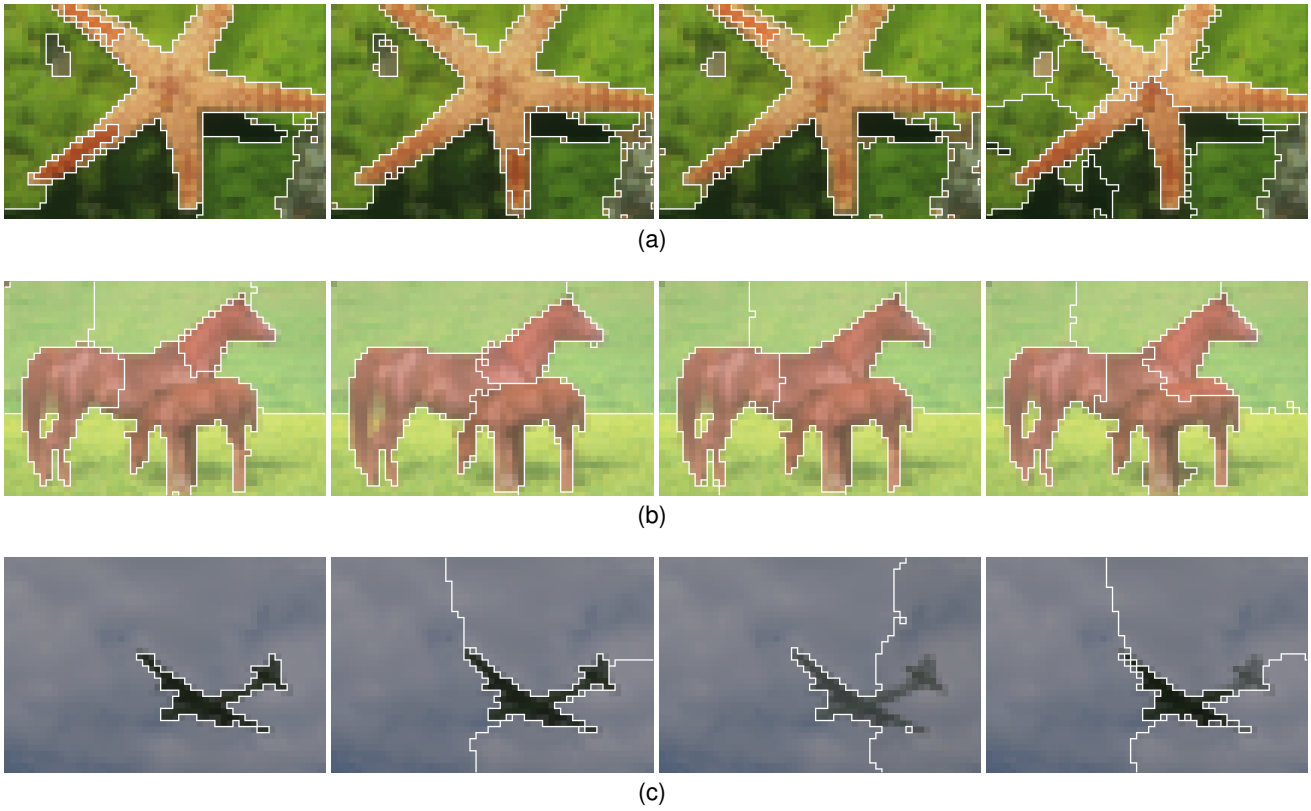


Figure 8: Image Segmentation.

be much longer than intracluster edges. The efficacy of the diffusion approach depends on the interconnectedness of the subgraph formed by the individual clusters. Although sparse graphs (such as the Relative Neighbor Graph) are useful for reducing the number of intercluster edges, clusters with low density may not be connected densely enough through edges that are different enough from the ones between clusters. As a rule of thumb, we apply the Gabriel graph as an initial candidate for generating the affinity matrix. The creation of spurious disconnected clusters may be an indication of a sparse graph, which suggest the exploration of a denser graph, such as the relaxed Gabriel of the β-skeleton with $\beta < 1$. The blending of graphs into larger groups may be an indication of a dense graph, which suggests the exploration of a sparser graph, such as the Relative Neighbor graph. Although we expose a number of diffusion parameters for the user to control, we found that the diffusion mechanism is stable under different degrees of noise and perturbation. In our 2D cases, we noticed that even under sub-optimal diffusion parameters ρ_D and ρ_S, our algorithm produces results comparable or better than previous approaches. Our approach can be extended in a number of ways to retrieve the number of clusters automatically, as suggested by approaches like [31, 1]. We believe our approach, proving to be resilient to noise and other types of data perturbations, is a step forward in the search for robust spectral clustering.

7. REFERENCES

[1] E. Biçici and D. Yuret. Locally scaled density based clustering. In *Proceedings of the 8th international conference on Adaptive and Natural Computing Algorithms, Part I*, ICANNGA '07, pages 739–748, Berlin, Heidelberg, 2007. Springer-Verlag.

[2] P. Bose, S. Collette, S. Langerman, A. Maheshwari, P. Morin, and M. Smid. Sigma-local graphs. *J. of Discrete Algorithms*, 8:15–23, March 2010.

[3] J. Cardinal, S. Collette, and S. Langerman. Empty region graphs. *Comput. Geom. Theory Appl.*, 42:183–195, April 2009.

[4] M. ç. Carreira-perpiŰĞn and R. S. Zemel. Proximity graphs for clustering and manifold learning. In *In*, pages 225–232. MIT Press, 2005.

[5] J. Choo, R. Jiamthapthaksin, C.-S. Chen, O. U. Celepcikay, C. Giusti, and C. F. Eick. Mosaic: A proximity graph approach for agglomerative clustering. In I. Y. Song, J. Eder, and T. M. Nguyen, editors, *DaWaK*, volume 4654 of *Lecture Notes in Computer Science*, pages 231–240. Springer, 2007.

[6] R. J. Cimikowski. Properties of some euclidean proximity graphs. *Pattern Recogn. Lett.*, 13:417–423, June 1992.

[7] C. D. Correa and P. Lindstrom. Towards robust topology of sparsely sampled data. *To appear: IEEE Transactions on Visualization and Computer Graphics (Proceedings Visualization / Information Visualization 2011)*, 17(12), 2011.

[8] I. S. Dhillon. Co-clustering documents and words using bipartite spectral graph partitioning. In *Proceedings of ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 269–274, 2001.

[9] W. Donath and A. Hoffman. Lower bounds for partitioning of graphs. *IBM J. Res. Dev.*, 17:420–425, 1973.

[10] V. A. Epanechnikov. Non-parametric estimation of a multivariate probability density. *Theory of Probability and*

Its Applications, 14, 1969.

[11] M. Ester, H. peter Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Knowledge Discovery and Data Mining*, pages 226–231, 1996.

[12] M. Fiedler. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, 1973.

[13] S. Fortune. Voronoi diagrams and delaunay triangulations. In J. E. Goodman and J. O'Rourke, editors, *Handbook of discrete and computational geometry*, pages 377–388. CRC Press, Inc., Boca Raton, FL, USA, 1997.

[14] R. K. Gabriel and R. R. Sokal. A new statistical approach to geographic variation analysis. *Systematic Zoology*, 18(3):259–278, Sept. 1969.

[15] L. Huang, D. Yan, M. I. Jordan, and N. Taft. Spectral clustering with perturbed data. In *Neural Information Processing Systems*, pages 705–712, 2008.

[16] J. Jaromczyk and G. Toussaint. Relative neighborhood graphs and their relatives. *Proceedings of the IEEE*, 80(9):1502 –1517, Sept. 1992.

[17] R. Kannan, S. Vempala, and A. Vetta. On clusterings: Good, bad and spectral. *Journal of The ACM*, 51:497–515, 2004.

[18] D. Kirkpatrick and J. Radke. A framework for computational morphology. *CG*, 85:217–248, 1985.

[19] Z. Li, J. Liu, S. Chen, and X. Tang. Noise robust spectral clustering. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1 –8, oct. 2007.

[20] T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, Norwell, MA, USA, 1994.

[21] J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In L. M. L. Cam and J. Neyman, editors, *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. University of California Press, 1967.

[22] M. Maier, U. von Luxburg, and M. Hein. How the result of graph clustering methods depends on the construction of the graph. *CoRR*, abs/1102.2075, 2011. informal publication.

[23] M. Meilă. Comparing clusterings—an information based distance. *Journal of Multivariate Analysis*, 98(5):873–895, 2007.

[24] B. Nadler and M. Galun. Fundamental limitations of spectral clustering. In *Neural Information Processing Systems*, pages 1017–1024, 2006.

[25] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS*, pages 849–856. MIT Press, 2001.

[26] J. Sander, M. Ester, H.-P. Kriegel, and X. Xu. Density-based clustering in spatial databases: The algorithm gdbscan and its applications. *Data Min. Knowl. Discov.*, 2:169–194, June 1998.

[27] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888 – 905, 2000.

[28] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proceedings of the Sixth International Conference on Computer Vision*, ICCV '98, pages 839–, Washington, DC, USA, 1998. IEEE Computer Society.

[29] R. Urquhart. Graph theoretical clustering based on limited neighbourhood sets. *Pattern Recognition*, 15(3):173 – 187, 1982.

[30] U. von Luxburg, M. Belkin, and O. Bousquet. Consistency of spectral clustering. *Annals of Statistics*, 36:555–586, 2008.

[31] L. Zelnik-Manor and P. Perona. Self-tuning spectral clustering. In *NIPS*, 2004.

Prepared by LLNL under Contract DE-AC52-07NA27344.