Minimal Dirichlet energy partitions for graphs

Braxton Osting[*], Chris D. White[†], and Édouard Oudet[‡]

August 22, 2013

Abstract

Motivated by a geometric problem, we introduce a new non-convex graph partitioning objective where the optimality criterion is given by the sum of the Dirichlet eigenvalues of the partition components. A relaxed formulation is identified and a novel rearrangement algorithm is proposed, which we show is strictly decreasing and converges in a finite number of iterations to a local minimum of the relaxed objective function. Our method is applied to several clustering problems on graphs constructed from synthetic data, MNIST handwritten digits, and manifold discretizations. The model has a semi-supervised extension and provides a natural representative for the clusters as well.

Keywords: graph partition, clustering, graph Laplacian, Dirichlet eigenvalues, rearrangement algorithm, nonnegative matrix factorization, semi-supervised algorithm

1 Introduction

Given a graph $G = (V, E)$ with non-negative edge weights $\{w_e\}_{e\in E}$, we consider the problem of "optimally" partitioning the vertex set, V, into k subsets. This *graph partitioning problem* frequently arises in the machine learning community, where the vertices represent observed data points and the goal is to identify meaningful groups (*i.e.* "clusters") within the data. One difficulty arises in choosing a measure of optimality which is computable and satisfies certain desirable properties which can be application dependent.

In this paper, we introduce a new measure of optimality for graph partitions, based on the sum of the Dirichlet eigenvalues of the partition components,

$$\min_{V=\amalg_{i=1}^k V_i} \sum_{i=1}^k \lambda(V_i).$$

Here $\lambda(V_i)$ denotes the first Dirichlet eigenvalue of the partition component V_i; a precise definition is given in §2. One benefit of the proposed objective is that it is not based on minimizing perimeter, and can thus take the interior of the partition components into account.

[*]Department of Mathematics, University of California, Los Angeles, CA 90095 USA (braxton@math.ucla.edu).
[†]Department of Mathematics, University of Texas at Austin, Austin, TX 78712 USA (cwhite@math.utexas.edu)
[‡]LJK, Université Joseph Fourier, Grenoble, France (edouard.oudet@imag.fr).

The eigenvalue partitioning problem as stated is combinatorial. To solve the problem, in §3 we introduce a relaxation and an algorithm to solve the relaxed problem. Our algorithm was motivated by a geometric interpretation of the problem and is novel in that it does *not* rely on a convex approximation or gradient descent method. We show that the algorithm is strictly decreasing and converges in a finite number of iterations to a local minimum of the relaxed objective function. In §4, we demonstrate with numerical examples some properties of the proposed model. For example, it can handle arbitrary cluster shapes, which is important in most applications. Moreover, our algorithm naturally provides confidences for label assignments. Consequently, the vertex with the highest confidence can be interpreted as a representative for each cluster; the problem of representative-based clustering has many variants [ESV12, FD07], and is especially informative in mining image and video data. In §3.2 we also describe a semi-supervised extension of our algorithm and apply this to the MNIST handwritten digit dataset in §4. Another property of our algorithm is that it can produce geometrically meaningful partitions on discretizations of manifolds; we demonstrate this ability in §4, where we apply it to the torus and the sphere. In both cases, there are open questions concerning the nature of these manifold partitions.

Our model was motivated by an analogous geometric problem phrased for domains $\Omega \subset \mathbb{R}^n$. Indeed, a recent trend in the applied math community is to utilize models and tools derived from PDEs and geometry to motivate analogous data-driven formulations and methods. This results in models which are highly interpretable and have well understood properties. Perhaps the most relevant examples of this are the use of Cheeger Cuts for graph partitioning [ST96, SM00, vL07] and the use of PDE models in image processing [CS05]. Diffusion maps, which are used to embed a dataset into a relatively low-dimensional Euclidean space via the eigenpairs of a diffusion operator, are another PDE-based approach and have become a powerful tool for dimensionality reduction and multi-scale analysis of data sets [CL06, NLCK06]. Functionals and flows coming from materials science (*e.g.*, motion by mean curvature) have been successfully applied to image segmentation and inpainting [MBO92, MBO93, MKB12, GCMB+13, EO13] and have inspired research into curvature flows on graphs [vGGOB13]. We believe our current model fits into this story; the analogous geometric model which inspired this work will be discussed properly in §2.

Outline. In §2, we introduce the eigenvalue partitioning problem for a graph. In §3, we introduce a relaxed formulation of the eigenvalue partitioning problem and propose a rearrangement algorithm for its solution. In §4, we apply our proposed partitioning algorithm to a number of examples, both data driven and purely geometric. We conclude in §5 with a discussion.

2 The graph partitioning model

Define the *Dirichlet energy* of a subset $S \subset V$,

$$\lambda(S) := \inf_{\substack{\|\psi\|_V=1 \\ \psi|_{S^c}=0}} \|\nabla\psi\|_{w,E}^2, \tag{1}$$

where $S^c := V \setminus S$ denotes the complement of S,

$$\|\nabla\psi\|_{w,E}^2 := \sum_{(i,j)\in E} w_{ij}(\psi_i - \psi_j)^2, \quad \|\psi\|_S^2 := \sum_{i\in S} d_i^r \psi_i^2, \quad \text{and} \quad d_i := \sum_j w_{ij}.$$

Here, we take the parameter $r \in [0,1]$. The value $\lambda = \lambda(S)$ in (1) satisfies the following Dirichlet eigenvalue problem in S, for some corresponding eigenvector, $\psi = \psi(S)$,

$$\begin{aligned} \Delta_r \psi &= \lambda \psi \qquad \text{on } S \subset V \\ \psi &= 0 \qquad\quad \text{on } S^c. \end{aligned}$$

The *graph Laplaican*[1], Δ_r, can be written $\Delta_r := D^{1-r} - D^{-r/2} W D^{-r/2}$ where D is the degree matrix and W is the weight matrix for the graph. When $r = 0$, this is the *unnormalized symmetric graph Laplacian* and when $r = 1$ this is the *normalized symmetric graph Laplacian*. In this work, we will take $r \in [0,1]$ to be a fixed parameter, the choice dependent on the application. As such, we suppress the dependency of Δ_r on r. A more complete discussion of these graph objects can be found in [Chu97].

We define the *energy of a k-partition*, $V = \amalg_{i=1}^k V_i$ to be $\Lambda_k(\amalg_{i=1}^k V_i) = \sum_{i=1}^k \lambda(V_i)$ where λ is defined in (1). The graph partitioning problem is then formulated as the following optimization problem

$$\Lambda_k^* := \min_{V = \amalg_{i=1}^k V_i} \sum_{i=1}^k \lambda(V_i). \tag{2}$$

A minimizer always exists, as there are only finitely many partitions of the graph.

The proposed model (2) is closely connected to several other problems and methods, which we outline now.

An analogous geometric problem. The eigenvalue partitioning problem has an analogous geometric formulation. Namely, given a bounded open set $\Omega \subset \mathbb{R}^2$, or more generally a compact manifold, find the partition $\Omega = \amalg_{i=1}^k \Omega_i$ which attains

$$\inf_{\Omega = \amalg_{i=1}^k \Omega_i} \sum_{i=1}^k \lambda(\Omega_i), \tag{3}$$

where $\lambda(\Omega_i)$ denotes the first Dirichlet-Laplace eigenavalue of Ω_i. Existence of optimal partitions for (3) in the class of quasi-open sets was proved in [BBH98]. Subsequently, several papers have investigated (3) and similar problems, focusing on the regularity of partitions, properties of optimal partitions, the asymptotic behavior of optimal partitions as $k \to \infty$, and computational methods [CL07, BNHV07, BBO10, HHO09, Hel10, HHOT10, Oud11, BV13]. In particular, the relaxation of the eigenvalue partitioning problem proposed here is analogous to a relaxation of (3) proposed in [BBO10]. The loss of infinitesimal scale on a graph has many consequences for the Dirichlet spectrum and partitioning problem. For example, the following statement is true in the continuum but fails on a graph: Any eigenvalue of the Laplace-Dirichlet operator is also the first eigenvalue for each of the nodal domains of the eigenfunction.

Cheeger cut partitioning. The Cheeger cut [Chu97] (or balanced cut) for a graph, G, is defined

$$B^*(G) := \min_{S \subset V} B(S) \quad \text{where} \quad B(S) := \frac{|\partial S|}{\min\left(\text{vol}(S), \text{vol}(S^c)\right)}, \tag{4}$$

[1] We comment that the sign convention for the graph Laplacian is opposite to that used for the continuum Laplacian in most of the PDE literature (in particular, Δ_r is a positive semidefinite operator).

where $\partial S := \{(i,j) \in E \colon i \in S \text{ and } j \in S^c\}$ is the edge boundary of S. The vertex subset $S^* \subset V$ attaining the minimum in (4) is a 2-class vertex partition that is often used for the bipartitioning problem [ST96, SM00, vL07]. This approach can be used recursively for the k-class problem [BL12]. Alternatively, [BLUvB13] mathematically formulates the general k-class vertex partitioning problem by generalizing the Cheeger cut (4) and solving

$$B_k^*(G) := \min_{V = \amalg_{\ell=1}^k V_\ell} \sum_{\ell=1}^k B(V_\ell).$$

By considering the test function $\psi = \chi_S$ for any $S \subset V$, it follows that $\lambda(S) \leq \frac{|\partial S|}{\mathrm{vol}(S)}$. Further relationships between the Dirichlet eigenvalues and "local" Cheeger cuts are studied in [Chu07]. The Cheeger partitioning problem shares the attribute of having a geometric analogue; the goal consists of partitioning a manifold into k submanifolds, and methods for solving the Cheeger partitioning problem can be interpreted in terms of a mean curvature flow on a graph [vGGOB13].

Nonnegative matrix factorization (NMF). Nonnegative matrix factorization is the general algebraic problem of finding a factorization of a matrix $A = \prod_i^K N_i$ where some, or all of the N_i are constrained to be nonnegative. This type of problem naturally arises in variable selection [LS99, Hof99] and clustering [YHD+12]. A popular approach for clustering applications is to solve

$$\min_{V \in \mathcal{X}} \|W - VV^T\|_F^2, \qquad \text{where } \mathcal{X} := \{V \in \mathbb{R}^{n \times k} \colon V^T V = \mathrm{Id}, V_{ij} \geq 0\}.$$

Here W is a similarity matrix constructed from the data, and k is the desired number of clusters. The following proposition shows that in certain instances our proposed objective is equivalent to an NMF objective, where the matrix to be factorized is the transpose of the asymmetric graph Laplacian.

Proposition 2.1. *Let $\Psi^* := \left[\psi_1^D | \cdots | \psi_k^D\right]$ be the matrix where the columns are Dirichlet eigenvectors corresponding to the optimal partition for $r = 1$. Then*

$$D^{-1/2}\Psi^* = \arg\min_{U \in \mathcal{M}} \|WD^{-1} - UU^T\|_F^2, \quad \textit{where} \quad \mathcal{M} := \{U \in \mathbb{R}^{n \times k} \colon U^T U = \mathrm{Id}, U_{ij} \geq 0\},$$

$D = \mathrm{diag}(d)$ is the degree matrix and W is the similarity weight matrix.

Proof. Let V be a collection of Dirichlet eigenvectors corresponding to some partition. Then, by definition, we have $\Delta V = V \mathrm{diag}(\vec{\lambda})$ where $\mathrm{diag}(\vec{\lambda})$ is a $k \times k$ diagonal matrix, with the Dirichlet eigenvalues along the diagonal. Moreover, V satisfies $V_{ij} \geq 0$ and $V^T DV = \mathrm{Id}$. Thus the partitioning problem (2) is equivalent to

$$\Lambda_k^* = \min_{V \in \mathbb{R}^{n \times k}} \mathrm{tr}(V^T D \Delta V) \tag{5a}$$

$$\text{s.t. } V_{ij} \geq 0, \quad V^T DV = \mathrm{Id}. \tag{5b}$$

Using the definition of the graph Laplacian, the objective function can be expanded to

$$\mathrm{tr}(V^T DV) - \mathrm{tr}(V^T D^{1/2} W D^{-1/2} V).$$

Thus, (5) is equivalent to

$$\min_{V\in\mathbb{R}^{n\times k}} \|WD^{-1} - D^{1/2}VV^TD^{1/2}\|_F^2$$
$$\text{s.t. } V_{ij}\geq 0, \quad V^TDV = \mathrm{Id}.$$

After the change of variables $U := D^{1/2}V$, we arrive at the stated proposition. □

Remark 2.2. A variant of Proposition 2.1 can be shown for the $r = 0$ Laplacian whenever the graph is regular, as is the case for an unweighted k-nearest neighbor graph.

We are not the first to connect NMF with spectral-based methods; [DHS05] describes a connection between various spectral clustering objectives and NMF. The algorithm proposed in §3 for solving (2) is new for this NMF objective; typical approaches to NMF problems are algebraic and involve finding good convex approximations. Our relaxation is not convex, and our algorithm is novel in that it does *not* rely on a convex approximation or gradient descent method. An interesting future direction of research might be to extend and analyze our geometric algorithm for other NMF objectives.

3 Relaxation and a rearrangement algorithm

In this section, we find a relaxation of the graph partitioning problem (2) and introduce an efficient algorithm for solving the relaxed problem.

For a vertex function, $\phi\colon V\to[0,1]$ and $\alpha>0$, consider the energy

$$\lambda^\alpha(\phi) := \inf_{\|\psi\|=1} \|\nabla\psi\|_{w,E}^2 + \alpha\|\psi\|_{(1-\phi)}^2, \quad \text{where} \quad \|\psi\|_f^2 := \sum_{i\in V} d_i^r f_i\psi_i^2. \tag{6}$$

Observe that $\lambda^\alpha(\phi)$ in (6) is the first eigenvalue of the Schrödinger operator $\Delta+\alpha(1-\phi)$, and the minimizer ψ^α is the corresponding eigenfunction, satisfying

$$[\Delta+\alpha(1-\phi)]\,\psi = \lambda\psi \qquad \text{in } V.$$

The minimizer ψ^α is unique up to a scaling and can be chosen to be strictly positive, *i.e.*, for all $i\in V$, $\psi_i^\alpha>0$.[2] Throughout, we will take ψ^α to be positive with $\|\psi^\alpha\|_V=1$.

If $\phi=\chi_S$ is the indicator function for the set $S\subset V$, then we intuitively think of $\lambda^\alpha(\chi_S)$ as an approximation to $\lambda(S)$. The following lemma shows that this approximation is exact in the limit that $\alpha\to\infty$; moreover, as α becomes large the eigenfunction corresponding to $\lambda^\alpha(\chi_S)$ becomes strongly localized on S. In the continuous case, on may interpret this relaxation as a "fictitious domain method" [Oud04, BBO10].

Lemma 3.1. *For $S\subset V$, $\lim_{\alpha\to\infty}\lambda^\alpha(\chi_S)=\lambda(S)$ and $\lim_{\alpha\to\infty}\psi^\alpha(\chi_S)=\psi^D(S)$.*

Proof. A simple computation shows that $\frac{d\lambda}{d\alpha} = \|\psi\|_{S^c}^2 > 0$, where ψ is the corresponding normalized eigenvector. Moreover, it is clear that $\lambda^\alpha(\chi_S)\leq\lambda(S)$. Consequently $\lim_{\alpha\to\infty}\lambda^\alpha(\chi_S)$ exists and satisfies $\lim_{\alpha\to\infty}\lambda^\alpha(\chi_S)\leq\lambda(S)$.

[2] These facts can be obtained by applying the Perron-Frobenius theorem to the matrix $\beta\mathrm{Id}-[\Delta+\alpha(1-\phi)]$ for sufficiently large β.

For the reverse inequality, observe that if we normalize all eigenvectors, then after possibly passing to a subsequence, there exists a $\tilde{\psi}$ such that $\psi^\alpha \to \tilde{\psi}$, and $\|\tilde{\psi}\|^2_{S^c} = 0$. Thus $\tilde{\psi}$ is admissible for the Dirichlet eigenvalue problem in S, giving us that $\lambda(S) \leq \lim_{\alpha\to\infty} \lambda^\alpha(\chi_S)$.

Since the minimizer of the Dirichlet problem is unique, $\tilde{\psi} = \psi^D$. Thus, the previous argument shows that the only limit point of $\{\psi^\alpha\}_\alpha$ is ψ^D and so $\lim_{\alpha\to\infty} \psi^\alpha = \psi^D$. □

Define the admissible class

$$\mathcal{A}_k = \{\{\phi_i\}_{i=1}^k \colon \phi_i \colon V \to [0,1] \text{ and } \sum_{i=1}^k \phi_i = 1\}.$$

Observe that the set of indicator functions for any k-partition of the vertices is a member of $\mathcal{A}_k$. For $\{\phi_i\}_{i=1}^k \in \mathcal{A}_k$ and $\alpha > 0$, we define the *relaxed energy*, $\Lambda_k^\alpha(\{\phi_\ell\}_{\ell=1}^k) = \sum_{i=1}^k \lambda^\alpha(\phi_i)$, where λ^α is defined in (6). Thus, a relaxed version of the graph partitioning problem (2) can be formulated

$$\Lambda_k^{\alpha,*} := \min_{\{\phi_i\}_{i=1}^k \in \mathcal{A}_k} \sum_{i=1}^k \lambda^\alpha(\phi_i). \tag{7}$$

It is a consequence of Lemma 3.1 that for any $\{\phi_i\}_{i=1}^k \in \mathcal{A}_k$, $\Lambda_k^\alpha(\{\phi_i\}_{i=1}^k)$ is monotonically increasing in α and for any partition $V = \amalg_{i=1}^k V_i$, $\lim_{\alpha\to\infty} \Lambda_k^\alpha(\{\chi_{V_i}\}_{i=1}^k) = \Lambda_k(\amalg_{i=1}^k V_i)$. However, in practice, we desire a solution to (7) for finite $\alpha > 0$. We observe that Λ_k^α is bounded below by zero, and is being minimized over the compact set $\mathcal{A}_k$. Thus a minimizer always exists. Supposing momentarily that we are able to find it, it is not yet clear how to interpret the collection $\{\phi_i^*\}$, which attains the minimum, as a vertex partition, as sought in (2). The following theorem, which is analogous to a continuous version in [BBO10, Thm. 2.3], tells us how this is accomplished.

Theorem 3.2. *Let $k \in \mathbb{Z}^+$ and $\alpha > 0$ be fixed. Every (local) minimizer of Λ_k^α over $\mathcal{A}_k$ is a collection of indicator functions.*

To prove Theorem 3.2, we first prove the following lemma.

Lemma 3.3. *For $\alpha > 0$ fixed, $\lambda^\alpha(\phi)$ is a concave function of ϕ.*

Proof. Let $t \in (0,1)$ and $\phi_i \colon V \to \mathbb{R}$ for $i = 1, 2$. Using (6) and the fact that the minimum is achieved for some normalized ψ, we compute

$$\begin{aligned}
\lambda^\alpha(t\phi_1 + (1-t)\phi_2) &= \|\nabla\psi\|^2_{w,E} + \alpha\|\psi\|^2_{(1-t\phi_1-(1-t)\phi_2)} \\
&= \|\nabla\psi\|^2_{w,E} + t\alpha\|\psi\|^2_{(1-\phi_1)} + (1-t)\alpha\|\psi\|^2_{(1-\phi_2)} \\
&\geq t\lambda^\alpha(\phi_1) + (1-t)\lambda^\alpha(\phi_2).
\end{aligned}$$

□

Proof of Theorem 3.2. Our proof closely follows the proof of [BBO10, Thm. 2.3]. The set $\mathcal{A}_k$ is the probability simplex in $\mathbb{R}^k$, and its extreme points are clearly given by the indicator functions. As Lemma 3.3 shows, Λ^α is a concave function on $\mathcal{A}_k$, so has a minimum and at least one minimizer is an extreme point of $\mathcal{A}_k$.

Now suppose that there exists some $\{\phi_i\}_{i=1}^k \in \mathcal{A}_k$ that achieves the minimum and is not an extreme point. Since $\sum_{i=1}^k \phi_i = 1$ there exist at least two ϕ's which are not $\{0,1\}$-valued at a vertex $v \in V$. After re-indexing, suppose these are given by ϕ_1 and ϕ_2. Thus, there exists $\epsilon > 0$ such that $\epsilon < \phi_i(v) < 1 - \epsilon \quad i = 1, 2$. By concavity of λ^α, we have

$$\lambda^\alpha(\phi_1) \geq \frac{1}{2}\lambda^\alpha(\phi_1 + \epsilon 1_v) + \frac{1}{2}\lambda^\alpha(\phi_1 - \epsilon 1_v) \tag{8a}$$

$$\lambda^\alpha(\phi_2) \geq \frac{1}{2}\lambda^\alpha(\phi_2 + \epsilon 1_v) + \frac{1}{2}\lambda^\alpha(\phi_2 - \epsilon 1_v) \tag{8b}$$

Adding these, and recognizing the right-hand side as an average, we must have

$$\lambda^\alpha(\phi_1) + \lambda^\alpha(\phi_2) \geq \min\{\lambda^\alpha(\phi_1 + \epsilon 1_v) + \lambda^\alpha(\phi_2 - \epsilon 1_v), \lambda^\alpha(\phi_1 - \epsilon 1_v) + \lambda^\alpha(\phi_2 + \epsilon 1_v)\}$$

But both terms in the minimum are feasible perturbations, thus by optimality of $\{\phi_i\}$, we must have equality:

$$\lambda^\alpha(\phi_1) + \lambda^\alpha(\phi_2) = \lambda^\alpha(\phi_1 + \epsilon 1_v) + \lambda^\alpha(\phi_2 - \epsilon 1_v) = \lambda^\alpha(\phi_1 - \epsilon 1_v) + \lambda^\alpha(\phi_2 + \epsilon 1_v)$$

But this implies equality in (8) as well:

$$\lambda^\alpha(\phi_1) = \tfrac{1}{2}\lambda^\alpha(\phi_1 + \epsilon 1_v) + \tfrac{1}{2}\lambda^\alpha(\phi_1 - \epsilon 1_v)$$

$$\lambda^\alpha(\phi_2) = \tfrac{1}{2}\lambda^\alpha(\phi_2 + \epsilon 1_v) + \tfrac{1}{2}\lambda^\alpha(\phi_2 - \epsilon 1_v)$$

From the proof of Lemma 3.3, we conclude that the eigenvector ψ corresponding to ϕ_1 is also an eigenvector for $\phi_1 + \epsilon 1_v$ and $\phi_1 - \epsilon 1_v$. We can subtract the following equations

$$\begin{cases} \Delta_G\psi + \alpha(1 - \phi_1 - \epsilon 1_v)\psi = \lambda^\alpha(\phi_1 + \epsilon 1_v)\psi \\ \Delta_G\psi + \alpha(1 - \phi_1 + \epsilon 1_v)\psi = \lambda^\alpha(\phi_1 - \epsilon 1_v)\psi \end{cases}$$

and using $\psi > 0$, simplify to yield

$$\phi_1 + \epsilon 1_v - (\phi_1 - \epsilon 1_v) = 2\epsilon 1_v \equiv C > 0$$

for some constant C, which is clearly a contradiction. □

For fixed $\alpha > 0$, we now consider the problem of solving the relaxed partitioning problem (7). Since $\Lambda_k^\alpha \colon \mathcal{A}_k \to \mathbb{R}$ is Fréchet differentiable, we could apply a gradient descent algorithm analogous to the continuous method proposed in [BBO10]. Instead, we propose a *rearrangement algorithm* (Algorithm 1). This is illustrated in Fig. 1. In Lemma 3.4, we prove that Algorithm 1 strictly decreases Λ_k^α at each iteration. This result is then strengthened in Theorem 3.5, to show that not only do the iterates decrease the objective function, but the iterates terminate in a finite number of steps to a local minimum.

Lemma 3.4. *Assume $\{\phi_i\}_{i=1}^k \in \mathcal{A}_k$ is not fixed by the rearrangement algorithm (Algorithm 1). Then one iteration of the rearrangement algorithm results in a strict decrease in Λ_k^α.*

Algorithm 1 A rearrangement algorithm for (7).

Input: An initial $\{\phi_i\}_{i=1}^k \in \mathcal{A}_k$.

while not converged, **do**
For $i = 1, \dots, k$, compute the (positive and normalized) eigenfunction ψ_i corresponding to $\lambda^\alpha(\phi_i)$ in (6).
Assign each node $v \in V$ the label $i = \arg\max_j \ \psi_j(v)$.
Let $\{\phi_i\}_{i=1}^k$ be the indicator functions for the labels.
end while

Proof. Suppose $\{\phi_i\}_{i=1}^k \in \mathcal{A}_k$ is not fixed by the rearrangement algorithm and let $\{\phi_i^+\}_{i=1}^k \in \mathcal{A}_k$ be the next iterate. Let ψ_i denote the first (normalized, positive) eigenvector of the operator $\Delta + \alpha(1 - \phi_i)$. We compute

$$\Lambda^\alpha(\{\phi_i\}_{i=1}^k) = \sum \lambda^\alpha(\phi_i) = \sum \|\nabla\psi_i\|_2^2 + \alpha\|\psi_i\|_{(1-\phi_i)}^2 \tag{9a}$$

$$\geq \sum \|\nabla\psi_i\|_2^2 + \alpha\|\psi_i\|_{(1-\phi_i^+)}^2 \tag{9b}$$

$$\geq \sum \lambda^\alpha(\phi_i^+). \tag{9c}$$

The inequality in (9b) follows from the construction of the algorithm. The inequality in (9c) follows from (6). Moreover, equality in (9c) holds if and only if ψ_i is also an eigenvector for the updated Schrödinger operator $\Delta + \alpha(1 - \phi_i^+)$ for all $i = 1, \dots, k$. From the proof of Theorem 3.2, we find that $\phi_i - \phi_i^+$ is a constant function for all $i = 1, \dots, k$. This contradicts the assumption that $\{\phi_i\}_{i=1}^k$ is not fixed by the rearrangement algorithm. □

Theorem 3.5. *Let $\alpha > 0$. For any initialization, the rearrangement algorithm 1 terminates in a finite number of steps at a local minimum of Λ_k^α, as defined in (7).*

Proof. It follows from Lemma 3.4 and the finiteness of V that for any initialization, the rearrangement algorithm 1 converges to a fixed point in a finite number of iterations. Thus, it suffices to show that every fixed point of the algorithm is locally optimal. Let $\{\phi_i\}_{i=1}^k$ be a fixed point of the rearrangement algorithm and let ψ_i denote the first (normalized, positive) eigenvector of the operator $\Delta + \alpha(1 - \phi_i)$. The Fréchet derivative of $\Lambda_k^\alpha \colon \mathcal{A}^k \to \mathbb{R}$ in the direction $\{\delta\phi_i\}_{i=1}^k$ is written

$$\left\langle \frac{\delta\Lambda_k^\alpha}{\delta\{\phi\}}, \{\delta\phi_i\} \right\rangle = -\alpha \sum_i \langle \psi_i^2, \delta\phi_i \rangle. \tag{10}$$

For $\{\phi_i\}_{i=1}^k \in \mathcal{A}_k$, any admissible perturbation can be written

$$\tilde{\phi}_i = \phi_i + \sum_{v \in V} t_{i,v} \chi_{\{v\}}, \qquad i = 1, \dots, k$$

for constants $t_{i,v}$, such that for every $v \in V$, $\sum_{i=1}^k t_{i,v} = 0$ and

$$t_{i,v} \begin{cases} \geq 0 & \text{if } \phi_i(v) = 0 \\ \leq 0 & \text{if } \phi_i(v) = 1. \end{cases}$$

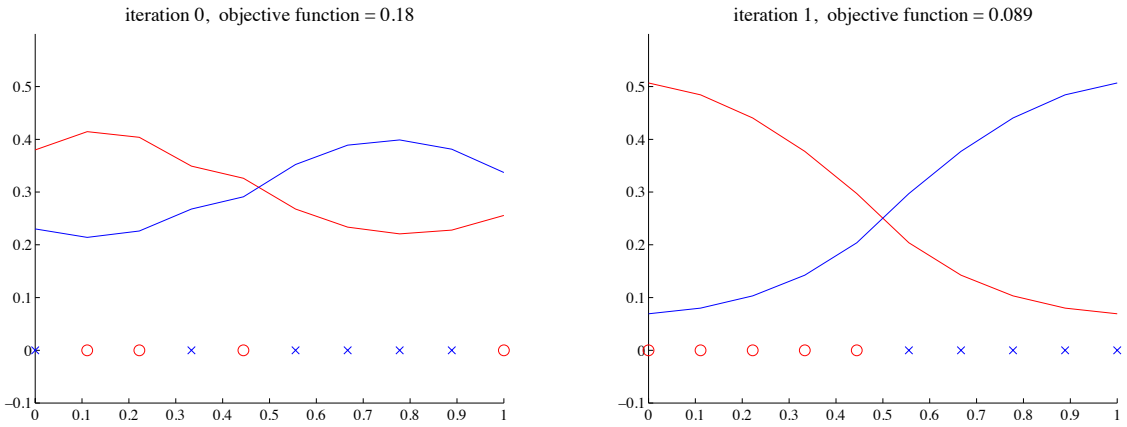

Figure 1: An illustration of the rearrangement algorithm 1 for $k = 2$ on a line graph with $n = 10$ vertices, an equally weighted nearest-neighbor similarity matrix, and graph Laplacian with $r = 0$. **(left)** The algorithm is randomly initialized as indicated by the blue `x` and red `o`. The eigenfunctions corresponding to the potential given by the indicator function on each partition is also indicated in the same color. **(right)** The algorithm converges to the globally optimal stationary state on the right in one iteration. See §3.

Using (10), we compute

$$\left\langle \frac{\delta \Lambda_k^\alpha}{\delta\{\phi\}}, \{\delta\phi_i\} \right\rangle = -\alpha \sum_v \sum_i t_{i,v} \psi_i^2(v) \geq -\alpha \sum_v \sum_i t_{i,v} \psi_{i^*(v)}^2(v) = 0$$

where $i^*(v) = \arg\max_i \psi_i(v)$. This proves local optimality. □

Remark 3.6. We refer to algorithm 1 as a rearrangement algorithm since at each iteration, the vertex functions $\{\phi_i\}$ are rearranged to decrease (7). These types of methods were introduced by Schwarz and Steiner and have wide applications in variational problems [Kaw00]. For example, Steiner rearrangement can be used to prove the isoperimetric inequality that the ball is the minimal perimeter domain amongst all regions of equal measure. More recently, rearrangement algorithms have been used in eigenvalue optimization problems including Krein's problem: Given an open, bounded connected domain $\Omega \subset \mathbb{R}^2$ and a prescribed amount of two materials of different density, find the distribution which minimizes the smallest frequency of the clamped drum [Cox91, CGI+00, KLY08, KS12].

Algorithm 1 also shares many attributes with the Merriman, Bence, and Osher (MBO) algorithm for approximating the motion by mean curvature [MBO92, MBO93, MKB12, GCMB+13, EO13, vGGOB13].

3.1 Advice on α

All machine learning algorithms require user input in various forms. For the rearrangement algorithm to find a meaningful optimum, the parameter α must be chosen carefully. Lemma 3.1 might

suggest to the reader that taking α large is a good idea; however, this is not the case. Large values of α force the associated eigenvectors to localize immediately, and the rearrangement algorithm terminates. We make this precise with Lemma 3.7 below. Moreover, comparing partition quality for different values of α is not possible; as was noted earlier, a consequence of the proof of Lemma 3.1 is that the relaxed energy is monotonic in α.

Lemma 3.7. *Consider any partition of V consisting entirely of connected sets, $V = \amalg_{i=1}^{k} V_i$. Then there exists α sufficiently large so that this partition is locally optimal for $\Lambda_k^{\alpha,*}$.*

We first require the following lemma, which has a well-known analogue in the continuous setting.

Lemma 3.8. *For any connected $S \subseteq V$, the first Dirichlet eigenvector, $\psi^D(S)$, attaining the minimum in (1), is strictly positive on S.*

Proof. We assume $r = 0$ so that $\Delta = D - W$ in this calculation, but the result holds for all $r \in [0, 1]$ by applying a transformation to the eigenvector which preserves positivity. By Lemma 3.1, $\psi^\alpha \to \psi^D$. Moreover, using the Perron-Frobenius Theorem we see that $\psi^D \geq 0$. Assume $\psi^D(i) = 0$ for some $i \in S$. Then by definition we have

$$\Delta\psi^D(i) = \lambda^D(S)\psi^D(i) = 0 \qquad \Longrightarrow \qquad \sum_j w_{ij}(-\psi^D(j)) = 0.$$

Since the left-hand side summation consists entirely of non-positive terms and using the connectedness of S, we must have at least one $w_{ij} > 0$ implying $\psi^D(j) = 0$; iterating this argument implies $\psi^D \equiv 0$, which is a contradiction. □

Proof of Lemma 3.7. Now, consider any partition of V consisting entirely of connected sets, $V = \amalg_{i=1}^{k} V_i$. After possibly passing to a subsequence we know $\psi_i^\alpha \to \psi_i^D$ pointwise, and thus we can choose α large enough so that $\psi_i^\alpha(v) = \max_j\{\psi_j^\alpha(v)\}$ for all $v \in V_i$ and for all i. This implies that this partition results in a stationary point of the rearrangement algorithm for this choice of α large, which proves local optimality by Theorem 3.5. □

Conversely, taking α too small also has undesirable consequences. We find experimentally that for α chosen too small, partitions can be lost. By considering the decomposition

$$\Delta + \alpha(1 - \chi_S) = \sum_i \lambda_i \psi_i \psi_i^t + \alpha \sum_{i \in S} \delta_i \delta_i^t,$$

where δ_i indicates the Kronecker-delta function on vertex i, a convincing heuristic is to pick α on the scale of the eigenvalues of the graph Laplacian. Following this line of thought, in most of our numerical examples (see §4) we pick α to be approximately $k\lambda_2$, where λ_2 is the second smallest eigenvalue of the graph Laplacian. We hope to make this heuristic more precise in future work.

3.2 A semi-supervised extension

In many clustering applications, a small percentage of the data labels are known and thus it is desirable for a clustering algorithm to have a *semi-supervised extension* that allows for the incorporation of such information. The rearrangement algorithm 1 has a natural semi-supervised variant. The label membership of a subset of the points can be fixed in the algorithm and the reader may check that all proofs of convergence remain valid. Moreover, fixing these points will force the eigenvectors to 'spread out' accordingly. We apply this variant to the MNIST handwritten digit data in Section 4.3.

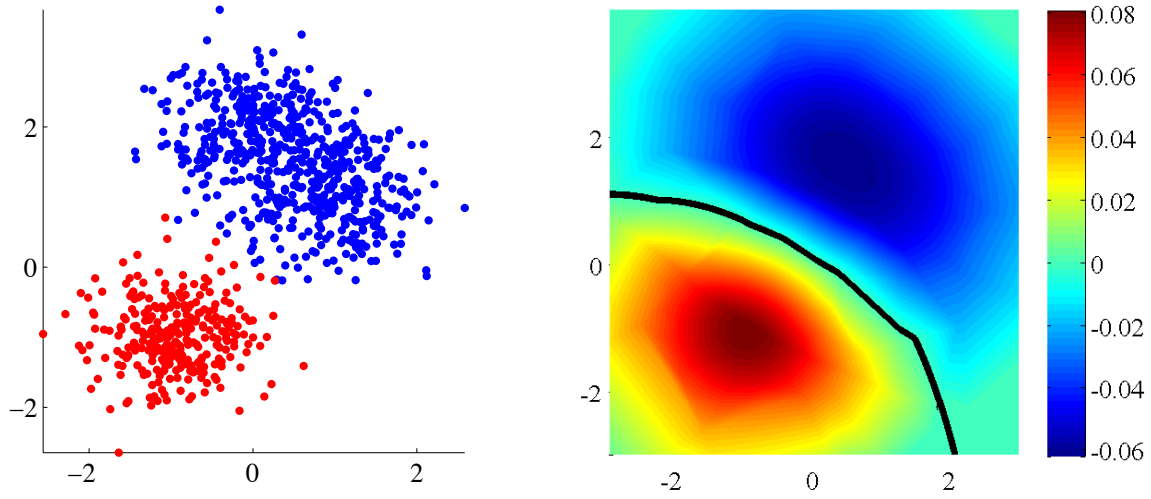

Figure 2: **(left)** The outputted partition of the rearrangement algorithm 1. **(right)** An interpolated contour plot of $\psi_1 - \psi_2$ in the feature space. The black line represents the decision boundary for clustering a new datapoint. See §4.1.

4 Numerical experiments

In this section we apply our partitioning method to several examples.

4.1 Gaussian mixture model

As a simple illustration, we first consider a Gaussian Mixture Model with two Gaussian clouds of differing sizes as illustrated in the left panel of Figure 2. To construct the similarity matrix, we used a standard Gaussian kernel, $e^{-d^2(x,y)/\sigma^2}$, with $\sigma = 1$. We choose the graph Laplacian with $r = 1$ for this experiment and the algorithm is randomly initialized. We choose $\alpha = 2\lambda_2$, where λ_2 is the second eigenvalue for the graph Laplacian. The algorithm converges to the global optimum in 5 iterations. In the left panel of Figure 2 we plot the optimal partition.

In the feature space for the model, we interpolate the values of the eigenvectors ψ_i corresponding to each partition and plot the difference $\psi_1 - \psi_2$ in the right panel. The eigenvectors concentrate where the clusters are most dense; this is in contrast to the $r = 0$ case, where the eigenvectors tend to be flat. The values of the eigenvectors can be interpreted as a confidence for the labeling of a particular vertex. The two points given by $x^* = \arg\max \psi_1$ and $z^* = \arg\max \psi_2$ are thus "good representatives" for the two clusters in that they have the highest confidence. Moreover, the contour plot in the right panel of Figure 2 illustrates the supervised extension of the algorithm discussed in §3.2. Namely, if we wish to cluster a new datapoint $z \in \mathbb{R}^2$, we can use interpolated values for the eigenvectors to assign it to the cluster $i = \arg\max_j \psi_j(z)$. The decision boundary for this assignment is given by the black line in Figure 2.

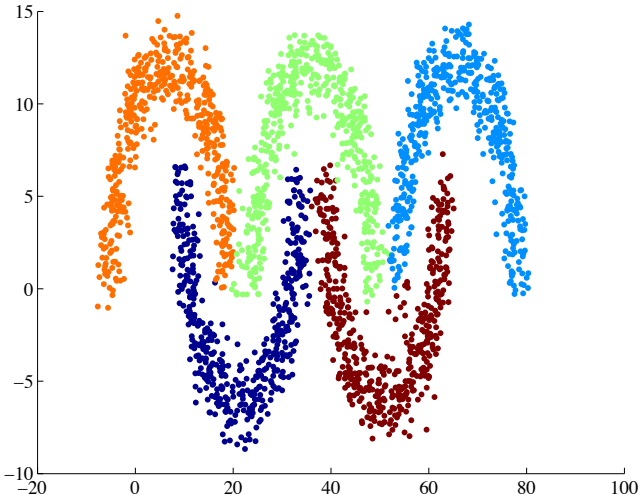

Figure 3: A partitioning of a five moons dataset after ten iterations of the rearrangement algorithm. See §4.2.

4.2 Five moons

Next we consider the five moons partitioning problem. We construct a similarity matrix using a Gaussian kernel, $e^{-d^2(x,y)/\sigma^2}$, with $\sigma = 1$, and the Laplacian is constructed using $r = 1$. We choose $\alpha = 5\lambda_2$, where λ_2 is the second eigenvalue for the graph Laplacian. For ten different random initializations, we run the algorithm until convergence and choose the lowest energy partition. In Fig. 3, we give a scatter plot of the points where colors represent the labels assigned to each point by the algorithm. For eight of the ten initializations, the algorithm generated a partition very similar to that shown in Fig. 3. For each initialization, the algorithm converges in approximately 10 iterations.

4.3 MNIST handwritten digits

The MNIST handwritten digit dataset consists of 70,000 28×28 greyscale images of handwritten digits 0 to 9. As input we used the similarity matrix for the MNIST dataset obtained from the website of Zhirong Yang[3] [YHD+12]. We symmetrize this matrix via $\tilde{W}_{ij} = \max\{W_{ij}, W_{ji}\}$, take $r = 0$, and set $\alpha = 10\lambda_2$. Moreover, we randomly sampled 3% of the data and used the semi-supervised variant of our algorithm (see §3.2). The remaining initialized labels were assigned randomly.

For ten different random initializations, we run the algorithm until convergence and choose the lowest energy partition. In each case, the algorithm converges in approximately 20 iterations. The purity obtained, as defined in [YHD+12], is 0.961 which is comparable to the performance of state-of-the-art clustering algorithms. We note that the partitions identified for other initial configurations had similar energy and purity values. Figure 4 is a graphical display of the quality

[3] http://users.ics.aalto.fi/rozyang/nmfr/NIPS2012_experiments.zip

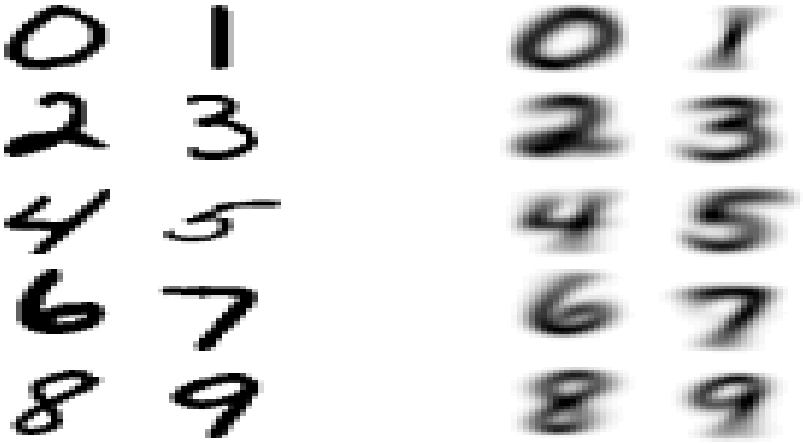

Figure 4: MNIST handwritten digits; each image is 28×28 pixels. **(left)** Representative images for each of the $k = 10$ clusters. **(right)** The cluster means. See §4.3.

	0	1	2	3	4	5	6	7	8	9
0	**0.9864**	0.0147	0.0021	0.0008	0.0002	0.0006	0.0005	0.0009	0.0040	0.0000
1	0.0012	**0.9717**	0.0023	0.0006	0.0011	0.0012	0.0047	0.0046	0.0013	0.0012
2	0.0012	0.0047	**0.9537**	0.0000	0.0079	0.0001	0.0041	0.0076	0.0054	0.0003
3	0.0005	0.0004	0.0000	**0.9894**	0.0002	0.0045	0.0009	0.0001	0.0332	0.0001
4	0.0000	0.0003	0.0213	0.0002	**0.9736**	0.0060	0.0003	0.0076	0.0045	0.0010
5	0.0018	0.0001	0.0000	0.0005	0.0032	**0.9811**	0.0000	0.0029	0.0000	0.0009
6	0.0038	0.0065	0.0003	0.0027	0.0006	0.0000	**0.9792**	0.0009	0.0992	0.0001
7	0.0033	0.0008	0.0033	0.0029	0.0112	0.0035	0.0011	**0.9675**	0.0080	0.0014
8	0.0013	0.0003	0.0168	0.0027	0.0011	0.0004	0.0090	0.0054	**0.8440**	0.0017
9	0.0005	0.0004	0.0001	0.0003	0.0008	0.0026	0.0002	0.0024	0.0004	**0.9932**

Table 1: **MNIST confusion matrix.** The columns represent the ground truth labels and the rows represent the labels assigned by the proposed algorithm. Each column sums to one and represents the distribution of the true labels across our partitions. We see that the algorithm does very well, but confused some 8s for 6s. See §4.3.

of the output. On the left-hand side are the representative images for each cluster (where each eigenvector achieves its maximum), and on the right are the averaged images within each cluster. In general, the maximum value of the eigenvectors may be non-unique, but for this dataset, the maximum was unique. In Table 4.3, we display the confusion matrix for the obtained clusters. The columns represent the ground truth labels and the rows represent the labels assigned by the proposed algorithm. Each column sums to one and represents the distribution of the true labels across our partitions.

4.4 Torus

We consider a 120×120 square grid with periodic boundary conditions and construct the nearest-neighbor $r = 0$ graph Laplacian. This is precisely the finite difference approximation of the Laplacian for the flat torus. We initialize the partition with a Voronoi tessellation for a randomly chosen

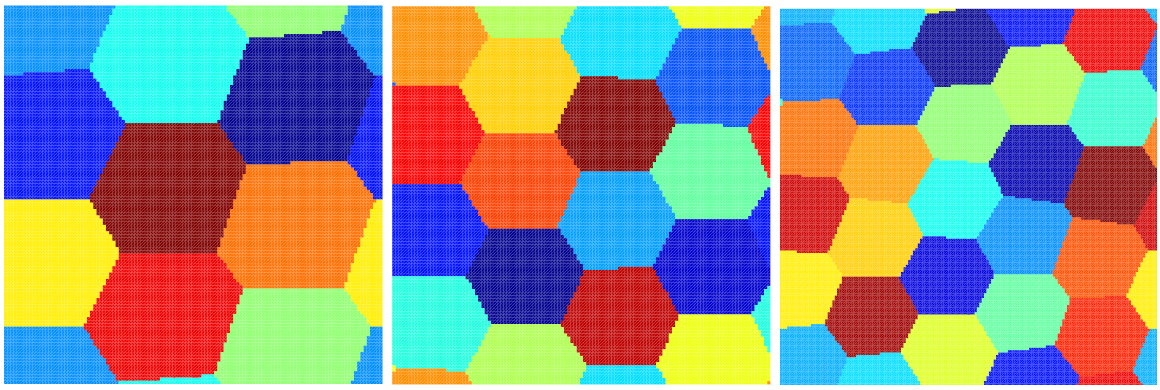

Figure 5: Locally optimal $k = 9$, 25 and 36 partitions of a flat torus. See §4.4.

generators. In Fig. 5, we plot locally optimal partitions with the smallest energy for $k = 9$, 25, and 36. We observe that the partitions are a hexagonal tiling as conjectured by [CL07] and also computationally studied in [BBO10].

4.5 Sphere

An approximately uniform triangular tessellation of the sphere with $n = 4000$ points is generated by minimizing the Reisz s-energy[4]. Using the geodesic distance, we use a standard Gaussian kernel $e^{-d^2(x,y)/\sigma^2}$ with $\sigma^2 = 1/10$ to construct the similarity matrix. The $r = 0$ graph Laplacian is used for partitioning. We choose $k = 3$ and initialize the partition with a Voronoi tessellation for 5 instances of randomly chosen generators. We choose $\alpha = 3\lambda_2$ and the algorithm converges in roughly 13 iterations. In Fig. 6, we plot the lowest energy partition. The conjectured Y-partition is attained; see [HHOT10].

5 Discussion and further directions

In this paper we introduced a new non-convex partitioning objective based on minimizing the sum of the Dirichlet eigenvalues of the partition components and constructed an algorithm for minimizing a relaxed version of this objective. We believe this model is promising for many reasons; in particular, it does not rely on minimizing a functional of perimeter and thus is able to take the interior geometry of the partition components into account. Moreover, our algorithm naturally provides confidences for label assignments and consequently produces a representative for each cluster. Our model was motivated by an analogous continuous problem (3); we have extended many well-established facts from the continuum to an arbitrary graph and, in some cases, proven considerably more. Many interesting research directions remain—perhaps most ambitiously, we wonder if it is possible to prove convergence (in the sense of Gromov-Hausdorff) of the discretized manifold graph partitions to their continuum limit. We would also like to find a tractable relaxation of the similar Neumann boundary conditions problem, and analyze its properties.

[4] http://www.mathworks.com/matlabcentral/fileexchange/37004-uniform-sampling-of-a-sphere

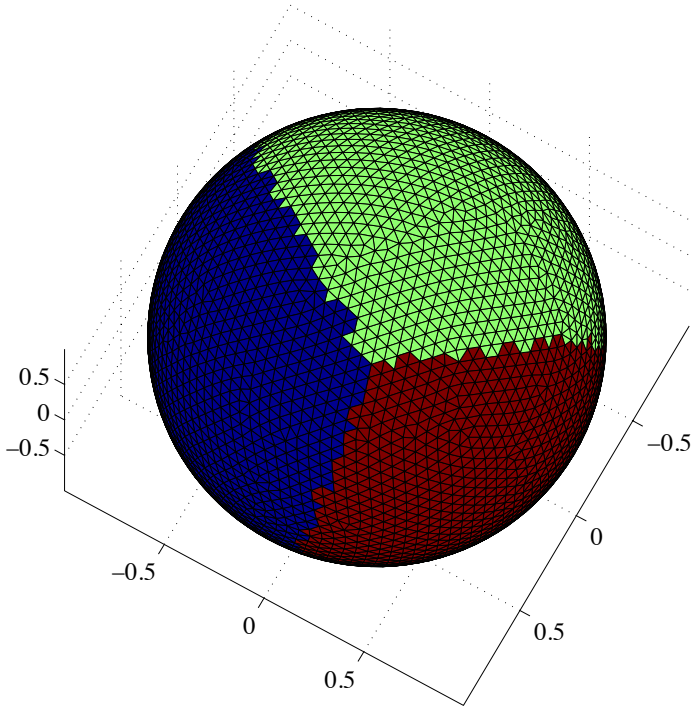

Figure 6: A Y-partition of the sphere obtained via the rearrangement algorithm. See §4.5.

Another more immediate direction of research is in the numerical analysis of the rearrangement algorithm 1. The majority of the computational costs for the rearrangement algorithm are in the eigenvalue computations. Our implementation could possibly be improved via a multi grid approach [BBH98], the Nyström method [FBM01], or the Rayleigh–Chebyshev method [And10]. In §3.1 we discuss optimal choices of the parameter α, and believe optimal choices should be on the scale of the second eigenvalue of the graph Laplacian. Another user-directed choice lies in which graph Laplacian to choose; we have found that optimal choices here are highly application dependent, and also dependent on the sparsity level of the given graph. In a future work, we intend to analyze the numerics further, and will do a more systematic comparison against other algorithms, both unsupervised and semi-supervised.

Acknowledgements

We would like to thank Andrea Bertozzi, James von Brecht, Dorin Bucur, Thomas Laurent, Stan Osher, Marc Sigelle, and David Uminsky for helpful conversations. Braxton Osting is supported in part by a National Science Foundation (NSF) Postdoctoral Fellowship DMS-1103959. Chris White was supported by AFOSR MURI grant FA9550-10-1-0569, ONR grant N000141210838 and ONR grant N000141210040; he would like to thank Rachel Ward for her support and Andrea Bertozzi for her hospitality in inviting him to UCLA. Édouard Oudet gratefully acknowledges the support of the ANR, through the projects GEOMETRYA and OPTIFORM.

References

[And10] C. R. Anderson, *A Rayleigh–Chebyshev procedure for finding the smallest eigenvalues and associated eigenvectors of large sparse Hermitian matrices*, Journal of Computational Physics **229** (2010), no. 19, 7477–7487.

[BBH98] D. Bucur, G. Butazzo, and A. Henrot, *Existence results for some optimal partition problems*, Adv. Math. Sci. Appl. **8** (1998), 571–579.

[BBO10] B. Bourdin, D. Bucur, and É. Oudet, *Optimal partitions for eigenvalues*, SIAM Journal on Scientific Computing **31** (2010), no. 6, 4100–4114.

[BL12] X. Bresson and T. Laurent, *Asymmetric Cheeger cut and application to multi-class unsupervised clustering*, UCLA CAM Report 12-27, 2012.

[BLUvB13] X. Bresson, T. Laurent, D. Uminsky, and J. H. von Brecht, *Multiclass total variation clustering*, 2013.

[BNHV07] V. Bonnaillie-Noel, B. Helffer, and G. Vial, *Numerical simulations for nodal domains and spectral minimal partitions*, Preprint Hal 00150455 (2007).

[BV13] D. Bucur and B. Velichkov, *Multiphase shape optimization problems*, submitted (2013).

[CGI+00] S. Chanillo, D. Grieser, M. Imai, K. Kurata, and I. Ohnishi, *Symmetry breaking and other phenomena in the optimization of eigenvalues for composite membranes*, Commun. Math. Phys **214** (2000), 315–337.

[Chu97] F. R. K. Chung, *Spectral graph theory*, AMS, 1997.

[Chu07] ———, *Random walks and local cuts in graphs*, Linear Algebra and its applications **423** (2007), no. 1, 22–32.

[CL06] R. R. Coifman and S. Lafon, *Diffusion maps*, Applied and computational harmonic analysis **21** (2006), no. 1, 5–30.

[CL07] L. A. Cafferelli and F. H. Lin, *An optimal partition pronblem for eigenvalues*, J. Sci. Comp. **31** (2007).

[Cox91] S. J. Cox, *The two phase drum with the deepest bass note*, Japan. J. Indust. Appl. Math **8** (1991), 345–355.

[CS05] T. Chan and J. Shen, *Image processing and analysis: variational, PDE, wavelet, and stochastic methods*, SIAM, 2005.

[DHS05] C. H. Q. Ding, X. He, and H. D. Simon, *On the equivalence of nonnegative matrix factorization and spectral clustering.*, SDM, vol. 5, 2005, pp. 606–610.

[EO13] S. Esedoglu and F. Otto, *Threshold dynamics for networks with arbitrary surface tensions*, accepted to CPAM (2013).

[ESV12] E. Elhamifar, G. Sapiro, and R. Vidal, *Finding exemplars from pairwise dissimilarities via simultaneous sparse recovery*, Advances in Neural Information Processing Systems, 2012, pp. 19–27.

[FBM01] C. Fowlkes, S. Belongie, and J. Malik, *Efficient spatiotemporal grouping using the Nyström method*, Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, vol. 1, 2001, pp. I–231.

[FD07] B. J. Frey and D. Dueck, *Clustering by passing messages between data points*, Science **315** (2007), no. 5814, 972–976.

[GCMB+13] C. Garcia-Cardona, E. Merkurjev, A. L. Bertozzi, A. Flenner, and A. Percus, *Fast multiclass segmentation using diffuse interface methods on graphs*, arXiv preprint arXiv:1302.3913 (2013).

[Hel10] B. Helffer, *On spectral minimal partitions: A survey*, Milan J. Math. **78** (2010), 575–590.

[HHO09] B. Helffer and T. Hoffmann-Ostenhof, *Remarks on two notions of spectral minimal partitions.*

[HHOT10] B. Helffer, T. Hoffmann-Ostenhof, and S. Terracini, *On spectral minimal partitions: the case of the sphere.*

[Hof99] T. Hofmann, *Probabilistic latent semantic indexing*, Proceedings of the 22nd annual international ACM SIGIR conference on research and development in information retrieval, ACM, 1999, pp. 50–57.

[Kaw00] B. Kawohl, *Symmetrization-or how to prove symmetry of solutions to a PDE*, Chapman and Hall CRC Research Notes in Mathematics (2000), 214–229.

[KLY08] C.-Y. Kao, Y. Lou, and E. Yanagida, *Principal eigenvalue for an elliptic problem with indefinite weight on cylindrical domains*, Mathematical Biosciences and Engineering **5** (2008), no. 2, 315–335.

[KS12] C.-Y. Kao and S. Su, *Efficient rearrangement algorithms for shape optimization on elliptic eigenvalue problems*, preprint (2012).

[LS99] D. D. Lee and H. S. Seung, *Learning the parts of objects by non-negative matrix factorization*, Nature **401** (1999), no. 6755, 788–791.

[MBO92] B. Merriman, J. K. Bence, and S. Osher, *Diffusion generated motion by mean curvature*, UCLA CAM Report 06–32 (1992).

[MBO93] _____, *Diffusion generated motion by mean curvature*, AMS Selected Letters, Crystal Grower's Workshop (1993), 73–83.

[MKB12] E. Merkurjev, T. Kostic, and A. Bertozzi, *An MBO scheme on graphs for segmentation and image processing*, UCLA CAM Report 12–46 (2012).

[NLCK06] B. Nadler, S. Lafon, R. R. Coifman, and I. G. Kevrekidis, *Diffusion maps, spectral clustering and reaction coordinates of dynamical systems*, Applied and Computational Harmonic Analysis **21** (2006), no. 1, 113–127.

[Oud04] É. Oudet, *Numerical minimization of eigenmodes of a membrane with respect to the domain*, ESAIM COCV **10** (2004), 315–335.

[Oud11] _____, *Approximation of partitions of least perimeter by γ-convergence*, Experimental Mathematics **20** (2011), no. 3, 260–270.

[SM00] J. Shi and J. Malik, *Normalized cuts and image segmentation*, IEEE Transactions on Pattern Analysis and Machine Intelligence **22** (2000), no. 8, 888–905.

[ST96] D. A. Spielman and S.-H. Teng, *Spectral partitioning works: planar graphs and finite element meshes*, 37th Annual Symposium on Foundations of Computer Science (Burlington, VT, 1996), IEEE Comput. Soc. Press, 1996, pp. 96–105.

[vGGOB13] Y. van Gennip, N. Guillen, B. Osting, and A. Bertozzi, *Mean curvature, threshold dynamics, and phase field theory on finite graphs*, submitted (2013).

[vL07] U. von Luxburg, *A tutorial on spectral clustering*, Statistics and Computing **17** (2007), no. 4, 395–416.

[YHD+12] Z. Yang, T. Hao, O. Dikmen, X. Chen, and E. Oja, *Clustering by nonnegative matrix factorization using graph random walk*, NIPS, 2012.