Graph Partitioning by Spectral Rounding: Applications in Image Segmentation and Clustering *

Gary L. Miller[1] David Tolliver[2]

[1]Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15213, glmiller@cs.cmu.edu

[2]Robotics Institute Carnegie Mellon University, Pittsburgh, PA 15213, tolliver@cs.cmu.edu

December 11, 2007

1 Abstract

We introduce a new family of spectral partitioning methods. Edge separators of a graph are produced by iteratively reweighting the edges until the graph disconnects into the prescribed number of components. At each iteration a small number of eigenvectors with small eigenvalue are computed and used to determine the reweighting. In this way spectral rounding directly produces discrete solutions where as current spectral algorithms must map the continuous eigenvectors to discrete solutions by employing a heuristic geometric separator (*e.g.* k-means). We show that spectral rounding compares favorably to current spectral approximations on the Normalized Cut criterion (NCut). Results are given in the natural image segmentation, medical image segmentation, and clustering domains. A practical version is shown to converge.

*This work was supported in part by the National Science Foundation under grants CCR-9902091, CCR-9706572, ACI 0086093, CCR-0085982 and CCR-0122581. CMU disclosure submitted.

2 Introduction

Several problems in machine perception and pattern recognition can be formulated as partitioning a graph. In general, most formulations of partition quality yield NP-hard optimization problems. This raises two important questions. First, does a particular optimization problem capture good partitions for the image segmentation domain, especially in light of the optimization being NP-hard and thus we may never know the true optimum anyway. Second, given that the optimal value is a good characterization are approximations quickly constructible and do they return good partitions?

One popular formulation, used in image processing and clustering, is the normalized cut (NCut) of a graph introduced by Shi and Malik [14]. The ideas contained therein were further explored by Ng *et al*. [12] and Yu and Shi [17] both of whom motivated multi-way partitioning algorithms. In part, our method was motivated by observations made in [12, 17]. Now, how does the NCut optimization problem fare against our two questions?

It is not difficult to construct image examples for which common image percepts do not correspond to the optimal NCut of the image (*e.g.* see Shental *et al*.'s example [13] and see Figure 4 for similar sub-optima). This is unsurprising, and an acknowledged attribute of all objective measures of cluster or partition quality (see Kleinberg [5] and Meilă [10] for treatment of this issue). But, for many images, as we shall show, there are segmentations with a smaller normalized cut value than in those generated by earlier methods that are at the same time more pleasing. For example, one of the main empirical advantages of spectral rounding technique seems to be that it is less likely to split the image in homogeneous regions, see Figure 3, while returning smaller NCut values. Thus good image segmentations are generated as graph partitions without reformulating the underlying combinatorial problem.

The two common paradigms for approximating such objective functions are 1) linear or semidefinite programming [8, 16, 1]. and 2) spectral methods [3, 14]. In this paper we introduce a spectral technique that empirically improves upon existing spectral algorithms for quotient cuts. Earlier spectral methods consisted of a two stage algorithm. In the first stage a small collection of, say k, eigenvectors with small eigenvalues are computed. While in the second stage these vectors are used to map the graph vertices into $\mathbb{R}^k$ and a geometric separator is then applied [2]. Recently, Lang and Rao have proposed a compelling alternative, a flow-based technique [6] that bests geometric rounding for quotient 2-cuts of a graph.

Our approach is to skip the geometric separator step by iteratively reweighting the graph in such a fashion that it eventually disconnects. At each iteration we will use the eigenvalues and eigenvectors of the reweighted graph to determine new edge weights. At first hand this may seem very inefficient, since the most expensive step in the two stage method is the eigenvector calculation. By using the eigenvector from the prior step as a starting point, for finding the new eigenvector, simple powering methods seem to work in only a small number of steps.

2.1 Problems and Results

Let $G = (V, E, w)$ be a graph on n vertices with connectivity E; we wish to find a partitioning of vertices such that some cut metric is minimized. An apparently equitable measure of the quality of a cut is the **normalized cut** (NCut). As NCut minimizes the ratio cut cost over the balance in total degree of the partitions. The normalized cut criterion is:

$$nc(G) = \underset{V_1,..,V_p}{argmin} \; : \; \frac{1}{p} \sum_{i=1}^{p} \frac{vol(V_i, V \setminus V_i)}{vol(V_i)} \tag{2.1}$$

where $vol(V_i)$ is the sum of edge weights associated with the vertices in V_i, $vol(V_i, V \setminus V_i)$ is the sum of the edge weights connecting V_i to remainder of the graph, and $V_i \cap V_j = \emptyset$. The combinatorial problem in Equation 2.1 can be stated as a Quadratically Constrained Quadratic Program (QCQP)[17]. This QCQP admits a straightforward eigenvector relaxation, stated as minimization of the Rayleigh Quotient over orthogonal functions.

2.2 Quality of the Spectral Bound

Spectral methods are so named because the second smallest eigenvalue λ_2 of the **normalized Laplacian** bounds the best cut obtained from a continuous vector. The associated eigenvectors provide a means of obtaining a discrete solution that satisfies the eigenvalue bound. The eigenvalue bound on the **isoperimetric number** of a graph $\Phi(G)$, see [11, 3], can be ported to the normalized 2-cut as $\frac{1}{2}\lambda_2 \leq nc(G) \leq \Phi(G) \leq \sqrt{2\lambda_2}$, as $\Phi(G)$ is an upper bound on $nc(G)$. The upper bound on $\Phi(G)$ is loose in general, as demonstrated by the pathological graphs constructed by Guattery and Miller in [4]. While, guaranteed $O(\frac{1}{\sqrt{n}})$ cut bounds were exhibited for planar graphs in [15].

Equation 2.1 has been studied in the context of image segmentation in the vision community [14, 17] and clustering in the learning community [12, 16]. In all cases a standard spectral algorithm is used. The methods [14, 17, 12] differ primarily in how the eigenvectors are used to find a feasible solution satisfying the constraints in Equation 2.1.

3 Preliminaries

Throughout this paper we let $G = (V, E, w)$ denote an edge weighted undirected graph without multiple edges or selfloops, where V is a set of n vertices numbered from 1 to n, E is a set of m edges, and $w : E \to [0, 1]$ is the edge weighting.

We associate four matrices with the graph: First, W the weighted adjacency matrix,

$$W_{ij} = \begin{cases} w_{ij} = w(i,j) & \text{if } (i,j) \in E \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

The **weighted degree** of vertex i is $d_i = \sum_{j=1}^{n} w_{ij}$. We assume that no vertex has zero degree. Second, the **weighted degree matrix** D is

$$D_{ij} = \begin{cases} d_i & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} . \tag{3.2}$$

Third, the **generalized Laplacian** or simply the **Laplacian** of G is $L = D - W$. Finally, the **normalized Laplacian** of G is $\mathcal{L} = D^{-1/2} L D^{-1/2}$.

Rather than working directly with the normalized Laplacian we shall work with a similar system. If $D^{1/2} f = g$ and g is an eigenvector of $\mathcal{L}$ with eigenvalue λ, i.e., $\mathcal{L} g = \lambda g$ then it is an easy calculation to see that f is a generalized eigenvector for the pair (L, D) with eigenvalue λ. That is $Lf = \lambda D f$. It will be convenient to work with the generalized eigenvalues and vectors of (L, D). In this case the normalized Rayleigh quotient is $f^T L f / f^T D f$ of the valuation f.

We make a simple, but important, observation about these Rayleigh quotients:

Lemma 1. *Given a weighted symmetric graph $G = (V, E, w)$ then the normalized Rayleigh quotient can be written as*

$$\frac{f^T L f}{f^T D f} = \frac{\sum_{(i,j)\in E, i<j} (f_i - f_j)^2 w_{ij}}{\sum_{(i,j)\in E, i<j} ((f_i)^2 + (f_j)^2) w_{ij}} \tag{3.3}$$

where $f_i = f(v_i)$

The main importance of Lemma 1 is that for each valuation f and each edge e_{ij} we get the fraction $\frac{(f_i-f_j)^2}{(f_i)^2+(f_j)^2}$. We will use this fraction to reweight the edge e_{ij}. The simplest reweighting scheme would be to replace the edge weight w_{ij} with the weight $\frac{(f_i)^2+(f_j)^2}{(f_i-f_j)^2}w_{ij}$. There are several issues with this scheme that will be address in the next section.

4 Spectral Rounding

In this section spectral rounding is introduced as a procedure for obtaining graph cuts. At each iteration a small number of eigenvectors with small eigenvalue are computed and used to determine the reweighting w' for the graph $G = (V, E, w)$. We show this process induces a k-way multiplicity in the k smallest eigenvalues of L (*i.e.* $\lambda_i(L) = 0$ for $1 \leq i \leq k$). By obtaining a Laplacian with this nullspace property we guarantee that the matrix represents k disconnected subgraphs, whose vertex membership can be read off directly from the first k eigenvectors.

§4.1 defines the spectral rounding algorithm. §4.2 connects decreasing a Rayleigh quotient to reweighting the graph. §4.6 the spectral rounding algorithm is shown to converge for a **reweighting scheme** proposed in §4.1.

4.1 The `SR-Algorithm`

For a graph $G = (V, E, w^0)$ prescribe the number of partitions k that the edge cut is to yield. Given a valid **reweighting scheme**, iteration of the `SR-Step` produces a sequence of N weightings $\{w^{(N)}\}$ such that the graph $G^N = (V, E, w^N)$ is disconnected into k components by the weighting w^N.

Algorithm 1 `SR-Step(w:` $||w||_k > 0$`)`
Let $F_k = [f_1 \ldots f_k]$ *denote the* k *generalized eigenvectors of* $L(G; w), D(G; w)$ *associated with the* k *smallest eigenvalues* $\Lambda_k = diag([\lambda_1 \ldots \lambda_k])$

1. `compute` $w_r = \mathcal{R}(F_k, \Lambda_k)$, `set` $\alpha = 1$ & $w' = w_r$
2. `while` $||w'||_k \geq ||w||_k$
   $\alpha \leftarrow \frac{1}{2}\alpha$, $w' = (1 - \alpha)w + \alpha w_r$
3. `return` w'

The function $\mathcal{R}$ computes a new weighting of the graph given the first k eigenpairs of L, D. The norm $|| \cdot ||_k$ is taken over weightings of the graph, such that

$||w||_k = 0$ *iff* the weighting w disconnects the graph into at least k pieces. A pair $\mathcal{R},\ ||\cdot||_k$ is called a **reweighting scheme** if the `SR-Step` converges in a finite number of iterations. We define Algorithm 2, the `SR-Algorithm`, as the iteration of Algorithm 1 until $||w^{(N)}||_k \approxeq 0$. In the following sections we propose $\mathcal{R}$s and corresponding norms $||\cdot||_k$ such that the `SR-Step` and `SR-Algorithm` converge in the desired fashion.

In §4.6 a simplified version of `SR-Algorithm` is shown to converge on graphs with $||\cdot||_k < 1$. In the case of a 2-cut this reduces to $\lambda_2(\mathcal{L}) < 1$. The class of graphs satisfying this spectral constraint is very general, excluding an uninteresting collection of graphs for our purposes. In particular, if $\lambda_2 \geq 1$ then no subset of the vertices exists with more than half of its edge volume contained within it (entailed by the Cheeger bound $2\Phi(G) \geq \lambda_2$ [3]). Such graphs are often called expander graphs.

4.2 Fractional Averages: a reweighting function

By Lemma 1 we saw that the Rayleigh quotient could be written as a sum of formal fractions where the numerators are added separately from the denominators. Define a **formal fraction** as a pair of real numbers $\frac{a}{b}$ and its **value** as the real number a/b. We call the average of a set of **formal fractions** the **fractional average**. We now prove a few simple but important facts about fractional averages.

Definition 1. *Given formal fractions*

$$\frac{a_1}{b_1}, \cdots, \frac{a_n}{b_n}$$

the **fractional average** *is the formal fraction*

$$\frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i}$$

where the a_i's and b_i's are reals.

We will simply call formal fractions fractions and only make a distinction between the formal fraction and its value when needed. In the case when the a_i's and b_i's are nonnegative we first observer that the fractional average is a convex combination of the fractions. That is we can rewrite the sum as

$$\sum_{i=1}^{n} \frac{b_i}{\overline{b}} \cdot \frac{a_i}{b_i}$$

where $\bar{b} = \sum_{i=1}^{n} b_i$. Thus fractional average lies between the largest and smallest fraction.

Possibly a more important interpretation is by viewing each fraction $\frac{a_i}{b_i}$ as the point $P_i = (b_i, a_i)$ in the plane and the value of the fraction is just its slope. The fractional average is just the vector sum of the points. Since we are only interested in the value of the fraction, the slope, we will think of the fractional average as the centroid of the points. If we multiply the numerator and denominator by a scalar w we shall say we **reweighted the fraction** by w. Geometrically, we are scaling the vectors or points P_i and then computing the centroid.

In the next lemma we show that we can control the slope of the fractional average by reweighting.

Lemma 2. *If* $\frac{a_1}{b_1} \leq \cdots \leq \frac{a_n}{b_n}$ *and* $w_1 \geq \cdots \geq w_n$ *then*

$$\frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} \geq \frac{\sum_{i=1}^{n} a_i w_i}{\sum_{i=1}^{n} b_i w_i}$$

The inequality is strict if for some pair $1 \geq i < j \leq n$ *we have that* $\frac{a_i}{b_i} < \frac{a_j}{b_j}$ *and* $w_i > w_j$.

Proof. It will suffice to show that

$$\frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} - \frac{\sum_{i=1}^{n} a_i w_i}{\sum_{i=1}^{n} b_i w_i} \geq 0 \tag{4.1}$$

Multiplying the left hand side through by its denominators we get

$$\sum_{i,j} a_j b_i w_i - \sum_{i,j} a_j b_i w_j = \sum_{i,j} a_j b_i w_i - a_j b_i w_j \tag{4.2}$$

Observe that term where $i = j$ are zero. Thus we can write the sum as:

$$\sum_{i<j} a_j b_i (w_i - w_j) + a_i b_j (w_j - w_i) \tag{4.3}$$

Rearranging the last term in the sum gives:

$$\sum_{i<j} a_j b_i (w_i - w_j) - a_i b_j (w_i - w_j) \tag{4.4}$$

Finally we get:

$$\sum_{i<j}(a_j b_i - a_i b_j)(w_j - w_i) \tag{4.5}$$

By the hypothesis each term in the sum above is nonnegative which proves the inequality. The strict inequality follows when one of the pair of terms in the sum are both positive as prescribed in the hypothesis. □

We get an exact expression by observing that the only time we got an inequality was when we cleared the denominators. Thus we have the following equation.

$$\sum_{i=1}^{n} a_i \sum_{i=1}^{n} b_i w_i - \sum_{i=1}^{n} a_i w_i \sum_{i=1}^{n} b_i = \sum_{i<j}(a_j b_i - a_i b_j)(w_j - w_i) \tag{4.6}$$

The Rayleigh quotient in Lemma 1 associates a formal fraction with each edge in G. One of the simplest ways to get weights satisfying the hypothesis of Lemma 2, for such a system, is to pick $w_i = \frac{b_i}{a_i} = \frac{f(u_i)^2+f(v_i)^2}{(f(u_i)-f(v_i))^2}$, if a_i is not zero. We shall call this **inverse fractional reweighting**. This reweighting scheme gives very large values for small values of a_i. We have found that using the stereographic map to normalized the inverse fractions between zero and one works well.

Observation 1. *The stereographic projection $\Psi : \mathbb{R}^d \to \mathcal{S}^d$ preserves the order of points on the real line, mapping points at ∞ to 1 and points at 0 to 0. Thus the* **inverse weight ordering** *of the edge update values is preserved by the stereographic map.*

If we think of the Φ as mapping points in $\mathbb{R}^d$ to $\mathbb{R}^{d+1}$, where we are only interested in the value in the $d+1$ dimension, then the images of $v \in \mathbb{R}^d$ is $\frac{v^T v}{v^T v+1} \geq 0$. We use Ψ_h to denote the map which returns the value in this dimension (*i.e.* the "height" on the sphere).

4.3 Reweighting for Multiple Valuations (`step 1`)

In the discussion so far we have assumed only one valuation is being used. To produce cuts of higher degree (*i.e.* $k > 2$) it is crucial that we simultaneously handle multiple valuations in computing the reweighting. Given two valuations, say f and g, we need to pick a single new weight per edge, *i.e.* a reweighting function $\mathcal{R}$. This can be thought of as finding a reweighting of the graph, w', such that

$$\frac{f^T L' f + g^T L' g}{f^T D' f + g^T D' g} < \frac{f^T L f + g^T L g}{f^T D f + g^T D g} \tag{4.7}$$

we begin be proving that such a weighting exists. We state a corollary of lemma 2 for the case of two fractional averages, but extending it to a larger number of terms may be done with the same argument.

Corollary 1. *Given two sets of n fractions $\{\frac{a_i}{b_i}\}$ and $\{\frac{c_i}{d_i}\}$ and weights $\{w_i\}$ and $\{\omega_i\}$ such that*

$$\frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i} > \frac{\sum_{i=1}^n a_i w_i}{\sum_{i=1}^n b_i w_i} \quad , \quad \frac{\sum_{i=1}^n c_i}{\sum_{i=1}^n d_i} > \frac{\sum_{i=1}^n c_i \omega_i}{\sum_{i=1}^n d_i \omega_i}$$

there exists a single reweighting $\{\gamma_i\}$ satisfying

$$\frac{\sum_{i=1}^n a_i + \sum_{i=1} c_i}{\sum_{i=1}^n b_i + \sum_{i=1}^n d_i} > \frac{\sum_{i=1}^n a_i \gamma_i + \sum_{i=1}^n c_i \gamma_i}{\sum_{i=1}^n b_i \gamma_i + \sum_{i=1}^n d_i \gamma_i}. \tag{4.8}$$

Proof. First observe that the *l.h.s.* of equation 4.8 is a fractional average. Thus by Lemma 2 there exists a weighting of the formal fractions $\frac{\bar{a}}{\bar{b}} = \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i}$ and $\frac{\bar{c}}{\bar{d}} = \frac{\sum_{i=1}^n c_i}{\sum_{i=1}^n d_i}$ such that $\frac{\bar{a}+\bar{c}}{\bar{b}+\bar{d}} < \frac{\alpha_1 \bar{a}+\alpha_2 \bar{c}}{\alpha_1 \bar{b}+\alpha_2 \bar{d}}$. This condition insures the existence of a single reweighting $\{\alpha_k\}$ where $\gamma_i = \alpha_1 w_i + \alpha_2 \omega_i$ satisfying the inequality in equation 4.8. □

An example weighting $\{\alpha_k\}$ for the fractional average of fractional averages is given by $\alpha_k = \frac{\bar{b}_k}{\bar{a}_k}$. In §? we show how this reweighting of the fractional average of two values will connect the inequality in equation 4.7.

The fractional average over the Rayleigh quotients $\mathcal{R}_{fa}$, takes the fractional sum of f and g per edge, yielding one fraction per edge. Formally, for an edge (uv), let $a_i(u,v) = \frac{1}{\lambda_i}(f_i(u) - f_i(v))^2$, and $b_i(u,v) = f_i(u)^2 + f_i(v)^2$ for k valuations. Define the reweighting function as

$$\mathcal{R}_{fa}(F, \Lambda, w_{uv}) \quad = \quad \Psi_h \left(\frac{\sum_{i=1}^k b_i(u,v)}{\sum_{i=1}^k a_i(u,v)} \right) w_{uv} \tag{4.9}$$

To successfully link the reweighting functions $\mathcal{R}$ to eigenvalue optimization we must specify a norm function $||\cdot||_k$ on weightings of G. For example, given the function $\mathcal{R}_{fa}$ we specify the norm $||w'||_k = \frac{\sum_{i=1}^k \sum_{(uv)\in E} w_{uv}(f_i'(u)-f'(v))^2}{\sum_{i=1}^k \sum_{(uv)\in E} w_{uv}(f_i'(u)^2+f_i'(v)^2)}$, *i.e.* the fractional average of the updated eigenvectors $f_i' \mid L' f_i' = \lambda_i' D' f_i'$. Given the reweighting function $\mathcal{R}_{sa}$ the norm $||w'||_k$ is the sum of k smallest eigenvalues of L', D'. We now connect linear combinations of weightings to their eigenvalues.

4.4 From Rayleigh Quotients to Eigenvalues

In §4.2 we showed how to, given a valuation or set of valuations of a graph, reweight the edges so as to reduce the Rayleigh quotient. In general this does not mean that if the valuation f is an eigenvector with eigenvalue λ of the old graph that the corresponding eigenpair f' and λ' of the new graph will have the property that $\lambda' \leq \lambda$.

Given a new edge weighting w' such that the fractional average of an eigenvector is decreased, we show that there is a linear combination of the weights of the form $w + t \cdot w'$ for $t > 0$ such that the associated eigenvalue is also decreased. This yields an algorithm which forces the target eigenvalue to zero. And motivates a matrix where the entries are linear functions of a scalar t and the following lemma.

Definition 2. *Given two weightings w and w' of G we define the* **matrix curve***, a $1-$dimensional family of matrices, as:*

$$W(t) = W + tW'$$

for $t \geq 0$.

A direct consequence of the scale invariance of the Rayleigh quotient $\frac{f^T L f}{f^T D f}$ is that any linear combination $W(t) = W + tW'$ may be expressed as a convex combination $W(\alpha) = (1-\alpha)W + \alpha W'$ on $0 \leq \alpha \leq 1$ (*i.e.* $\alpha = \frac{t}{t+1}$). The eigenstructure of normalized laplacians defined on $W(\alpha)$ and $W(t)$ are identical by the scale invariance of the Rayleigh quotient.

Lemma 3. *Given a weighted graph $G = (V, E, w)$, matrices L and D, the simple eigenpair $(f, \lambda) \mid Lf = \lambda Df$, and a new weighting w' such that $\frac{f^T L' f}{f^T D' f} < \frac{f^T L f}{f D f} = \lambda$ then the derivative of the eigenvalue function*[1] *$\lambda(t)$ of the matrix curve $W(t) = W + tW'$ is well defined for small t and*

$$\frac{d\lambda(t)}{dt} < 0$$

at $t = 0$.

[1]The proof that $d\lambda$ exists follows from properties of the characteristic polynomial of L, D and relies on the implicit function theorem. Details can be found on the differentiability of λ and f in Lax [7] chapter 9.

Proof. For a simple eigenpair (f, λ), recall that $\frac{f^T L f}{f^T D f} = \lambda$, as $W(0) = W$ and thus $L(0) = L,\ D(0) = D$ by definition. We deduce the bound on $\frac{d\lambda}{dt}$ with a simple derivation.

$$\begin{aligned}
\frac{d\lambda(t)}{dt} &= \frac{f^T \left(\frac{dL_{(t)}}{dt} - \lambda \frac{dD_{(t)}}{dt} \right) f}{f^T D f^T} && (4.10)\\
&= \frac{f^T (L' - \lambda D') f}{f D f^T} \quad \left\{ \frac{dL(t)}{dt} = L' \right\} && (4.11)\\
&= f^T (L' - \lambda D') f && (4.12)\\
&= f^T L' f - \lambda f^T D' f && (4.13)\\
&= \left(\frac{f^T L' f}{f^T D' f} - \lambda \frac{f^T D' f}{f^T D' f} \right) f^T D' f && (4.14)\\
&= \left(\frac{f^T L' f}{f^T D' f} - \lambda \right) f^T D' f && (4.15)\\
&= \left(\frac{f^T L' f}{f^T D' f} - \frac{f^T L f}{f^T D f} \right) f^T D' f && (4.16)
\end{aligned}$$

The bound is entailed by observing that 1) the term $f^T D' f = \sum_{v \in V} f^2(v) d'(v)$ is positive, and 2) thus the sign of $\frac{d\lambda(t)}{dt}$ is strictly negative due to the inequality $\frac{f^T L' f}{f^T D' f} < \frac{f^T L f}{f D f}$. The above derivation demonstrates that the slope $\frac{d\lambda(t)}{dt}$ is negative at $t = 0$. By the continuity of λ there exists a $t > 0$ such that $\lambda(t) < \lambda(0)$. This insures that the eigenvalue λ associated with f can be decreased algebraically using the procedure in §4.2 on fractional averages. □

To handle multiple eigenvectors one hope might be to simultaneously bound the derivatives of the target eigenvalues $\{\lambda_2, ..., \lambda_k\}$ of $L(t), D(t)$. To do this one arrives at the update criteria for the re-weighting w'; select a w' such that $\frac{f_i^T L' f_i}{f_i^T D' f_i} < \frac{f_i^T L f_i}{f_i^T D f_i}$ insuring that $\lambda_i(L(t)) < \lambda_i(L)$ for $2 \leq i \leq k$. But in general this may be unrealistic since we must use the same weight per edge for all the eigenvectors f_i. In the case where the aforementioned inequalities do not hold, we determine w' so as to decrease the fractional average of the Rayleigh quotients (see $\mathcal{R}_{fa}$ in §4.3). The average of the target eigenvalues tends to zero as a consequence of the decrease in the fractional average. We now state a Lemma connecting the derivative of two eigenvalues λ_f and λ_g with the reweighting scheme suggested in the previous section.

Lemma 4. *For a weighted graph $G = (V, E, w)$ with matrices L and D and simple eigenpairs $(f, \lambda_f) \mid Lf = \lambda_f Df$ and $(g, \lambda_g) \mid Lg = \lambda_g Dg$, given a reweighting w' such that*

$$\frac{\alpha f^T L' f + \beta g^T L' g}{\alpha \lambda_f f^T D' f + \beta \lambda_g g^T D' g} < \frac{\frac{1}{\lambda_f} f^T L f + \frac{1}{\lambda_g} g^T L g}{f^T D f + g^T D g} = 1 \tag{4.17}$$

then

$$\frac{d\lambda_f(t)}{dt} + \frac{d\lambda_g(t)}{dt} < 0$$

at $t = 0$.

Proof. We begin by stating a related quantity of interest, the derivative of the fractional average of Rayleigh quotients on f and g for the matrix curve $w = w + t \cdot w'$ as:

$$\frac{d}{dt}\left[\frac{\frac{1}{\lambda} f^T(t) L(t) f(t) + \frac{1}{\lambda} g^T(t) L(t) g(t)}{f^T(t) D(t) f(t) + g^T(t) D(t) g(t)}\right] \tag{4.18}$$

and examine its derivative centered at $t = 0$. First we must fix the scale of the eigenvectors $f(t)$ and $g(t)$, we choose $f(t)^T D(t) f(t) = g(t)^T D(t) g(t) = 1$ *w.l.o.g.* Thus equation 4.18 simplifies to

$$\begin{aligned} \frac{d}{dt}\left[\frac{\frac{1}{\lambda_f} f^T(t) L(t) f(t) + \frac{1}{\lambda_g} g^T(t) L(t) g(t)}{1+1}\right] &= \\ \frac{1}{2}\left(\frac{1}{\lambda_f}\frac{d}{dt}\lambda_f(t) + \frac{1}{\lambda_g}\frac{d}{dt}\lambda_g(t)\right) & \end{aligned} \tag{4.19}$$

by the linearity of the derivative. We may now substitute the functional form of $\frac{d\lambda(t)}{dt}$ in

$$\begin{aligned} \frac{1}{2}\left(\frac{1}{\lambda_f}\frac{d}{dt}\lambda_f(t) + \frac{1}{\lambda_g}\frac{d}{dt}\lambda_g(t)\right) &= \\ \frac{1}{2}\left(\frac{1}{\lambda_f} f^T (L' - \lambda_f D') f + \frac{1}{\lambda_g} g^T (L' - \lambda_g D') g\right) & \end{aligned} \tag{4.20}$$

assume the bound holds

$$\begin{aligned} \frac{1}{2}\left(\frac{1}{\lambda_f} f^T L' f - \frac{\lambda_f}{\lambda_f} f^T D' f + \frac{1}{\lambda_g} g^T L' g - \frac{\lambda_g}{\lambda_g} g^T D' g\right) &< 0 \\ \left(\frac{1}{\lambda_f} f^T L' f - f^T D' f + \frac{1}{\lambda_g} g^T L' g - g^T D' g\right) & \\ < 0 & \end{aligned} \tag{4.21}$$

arriving at

$$\frac{1}{\lambda_f} f^T L' f + \frac{1}{\lambda_g} g^T L' g \quad < \quad f^T D' f + g^T D' g$$

which is equivalent to the hypothesis in equation 4.17. The remainder of the proof follows the continuity argument in Lemma 3. □

4.5 Termination of SR-Step

In order to prove that the algorithm SR-Algorithm converges to a k-way partition we first need to show that each step of SR-step terminates. In the next subsection we use this termination to show convergence. To simplify the discussion we only consider using one eigenvector. The SR-Algorithm has two majors steps. In the first step given a valuation f it computes a reweighting w_r. We claim with a reasonable assumption about the Rayleigh quotient that

$$\frac{f^T L f}{f^T D f} > \frac{f^T L_r f}{f^T D_r f}. \tag{4.22}$$

By lemma 2 we know that equation 4.22 is true as long as not all the fractions $\delta_{ij} = (f_i - f_j)^2/(f_i^2 + f_j^2)$ are equal. We show that if the fraction are all equal the Rayleigh quotient is at least one, contradicting how we construct affinity graphs. We take

$$\mathcal{R}(f)_{ij} = \Phi_h(\delta_{ij}^{-1}) \cdot w_{ij} = \Phi_h\left(\frac{f_i^2 + f_j^2}{(f_i - f_j)^2}\right) \cdot w_{ij} \tag{4.23}$$

as the reweighting function $\mathcal{R}$ used in `SR-step`.

Lemma 5. *Let $G = (V, E, w)$ be a connected graph and f a valuation such that all the fractions $\delta_{ij} = (f_i - f_j)^2/(f_i^2 + f_j^2)$ are equal then $\frac{f^T L f}{f^T D f} \geq 1$*

Proof. Let f be a valuation of G with Rayleigh quotient λ. Suppose all $\delta_{ij} = (f_i - f_j)^2/(f_i^2 + f_j^2)$ over all edges of G are the same value. Observe that: 1) $\delta_{ij} > 1$ if the sign of f_i and f_j differ, 2) $\delta_{ij} = 1$ if $f_i f_j = 0$, 3) $\delta_{ij} < 1$ if $f_i f_j > 0$. If we are not case 3) then the value of each fraction is at least one. Thus by Lemma 1 the Rayleigh quotient is at least one. □

Lemma 6 can be easily generalized to the case of multiple valuations under the update rule $\mathcal{R}_{fa}$ in §4.3. As the fractional average of the generalized Rayleigh quotients is a convex sum it lies between the largest λ_k and smallest λ_2 eigenvalues. If $\lambda_k < 1$ holds the method will terminate as stated above with the update rule $\mathcal{R}_{fa}$.

Lemma 6. *Algorithm SR-step terminates for graphs such that* $\lambda_k < 1$

Proof. Follows from Lemma 3 and Lemma 5. □

4.6 Convergence of SR-Algorithm

By Lemma 6 we know that each step of procedure SR-step produces a new weighted graph such that the norm $||\ ||_k$ is strictly decreasing. We show that in fact the limit norm must be zero, i.e., SR-Algorithm converges.

Again for simplicity of the exposition we only consider the case of a single eigenvector.

Let G^i, D^i, λ^i, and x^i be the ith graph, its degree matrix, its second eigenvalue, and eigenvector. That is $G^i x^i = \lambda^i D^i x^i$. We also assume that each x^i is a unit length vector. Thus all the eigenvectors all belong to a compact region in $\mathbb{R}^n$, the unit sphere. The Bolzano-Weierstrass theorem dictates such a space contains an accumulation point, say, x. Let $\{x^{(i)}\}$ be a subsequence of eigenvectors that converge to x, and let G', D', and λ' be the corresponding graph, degree matrix and eigenvalue. The eigenvectors converge to x and the eigenvalues converge to some value λ. To insure that the graph also converges to a unique graph we apply the Bolzano-Weierstrass theorem again to the weighted graphs which belong to a compact set in $\mathbb{R}^m$. Thus, we may also assume the graphs and degree matrix converge to G', D', and λ' such that $G'x' = \lambda' D' x'$.

Lemma 7. *The limit of the sequence of* λ_i*, as defined above, converges to* $\lambda = 0$

Proof. Suppose that $\lambda' > 0$. We know by Lemma 6 that if we run SR-step on G' we will get a new graph with $\lambda'' < \lambda'$. Let $\epsilon = \lambda' - \lambda''$. We can also run SR-Step on each G'_i getting a new eigenvalue λ''_i. Let $\epsilon_i = \lambda'_i - \lambda''_i$. Since SR-step is a continuous function in G'_i and x'_i we get that the ϵ_i converge to ϵ. For sufficiently large i it follows that $\epsilon_i \geq \epsilon/2$. But this implies that λ'_i goes to $-\infty$ which contradicts the fact the they a bounded below by zero. □

4.7 Convergence & Global Convergence for multi-valuation cases

In the previous section a proof of convergence was given for spectral rounding where one eigenvector, say the k^{th}, used to drive the first k eigenvalues to zero. In this section we provide an analogous proof for the multi-valuation case. Unlike the previous proof, the following depends directly on a generalization of Fiedler's

theorem constraining the number of **sign connected components** induced by the k^{th} eigenvector f_k.

Need some defs – **sign connected components**, the Lovasz statement of Fiedler's theorem, and the joint nodal domain lemma.

4.7.1 termination

Lemma 8. *Let $G = (V, E, w)$ be a connected graph with n vertices and m edges. Further let f and g be valuations, with generalized Rayleigh quotients $\lambda = \frac{f^T L f}{f^T D f}$ and $\lambda' = \frac{g^T L g}{g^T D g}$ respectively. If all fractions of the form*

$$\delta_{ij} = \frac{\frac{1}{\lambda}(f_i - f_j)^2 + \frac{1}{\lambda'}(g_i - g_j)^2}{f_i^2 + f_j^2 + g_i^2 + g_j^2}$$

are equal then f and g must define n joint nodal domains (sign connected components).

Proof. W.l.o.g. assume that all the fractions are equal to 1. The following calculation entails the theorem statement.

$$\frac{\frac{1}{\lambda}(f_i - f_j)^2 + \frac{1}{\lambda'}(g_i - g_j)^2}{f_i^2 + f_j^2 + g_i^2 + g_j^2} = 1$$

$$\begin{aligned}
\frac{1}{\lambda}(f_i - f_j)^2 + \frac{1}{\lambda'}(g_i - g_j)^2 &= f_i^2 + f_j^2 + g_i^2 + g_j^2 \\
\frac{1}{\lambda}(f_i^2 + f_j^2 - 2f_i f_j) + \frac{1}{\lambda'}(g_i^2 + g_j^2 - 2g_i g_j) &= f_i^2 + f_j^2 + g_i^2 + g_j^2 \\
(\frac{1}{\lambda} - 1)(f_i^2 + f_j^2) + (\frac{1}{\lambda'} - 1)(g_i^2 + g_j^2) &= \frac{2}{\lambda} f_i f_j + \frac{2}{\lambda'} g_i g_j \qquad (4.24)
\end{aligned}$$

this is wrong it seems. notes:

$$\begin{aligned}
\frac{1}{\lambda}(f_i - f_j)^2 &= f_i^2 + f_j^2 \qquad (4.25) \\
(f_i - f_j)^2 &= \lambda(f_i^2 + f_j^2) \qquad (4.26) \\
f_i^2 f_j^2 - 2f_i f_j &= \lambda(f_i^2 + f_j^2) \qquad (4.27)
\end{aligned}$$

4.8 Offset Inverse Fractional Reweighting

There are an inexhaustible number of reweighting functions and norms that satisfy the conditions in §4.5. In the following subsection we introduce a reweighting function with improved performance over the simple reweighting rule specified above. The reweighting in the `SR-step` convergences for a large class of graphs (those satisfying a modest eigenvalue constraint). Recall that if we wish to break the graph into k pieces we must drive the k^{th} eigenvalue to zero. We begin by fixing the following **reweighting scheme**, define $f \mid Lf = \lambda_k Df$ where λ_k is the k^{th} smallest generalized, let

$$\mathcal{R}(f_k, \lambda_k) = \Phi_h \left(\frac{f_i^2 + f_j^2 + 2\alpha^2}{(f_i - f_j)^2} \right) \cdot w_{ij} \tag{4.28}$$

$$||\cdot||_k = \lambda_k. \tag{4.29}$$

The term α is taken to be $1/\sqrt{vol(G)}$, which can thought of as an entry in the normalized $\mathbf{1}$s vector (associated with $\lambda = 0$). We now show that this scheme converges, in the desired sense, on a large set of graphs (in essence we must exclude a larger class of expanders than in §4.5).

At a high level it's worth motivating the use of Equation 4.28 over 4.23. In particular what deficit in Equation 4.23 does it overcome? It is clear that reweighting according to Equation 4.23 will most aggressively down-weight edges that span 0 in the valuation f (*i.e.* $f_i f_j < 0$). Immediately a number of graphs come to mind where this property is undesirable. For example, a symmetrically weighted odd-length line graph with a heavy-weight edge at the center will yield a suboptimal 2-cut under Equation 4.23 but not Equation 4.28. This is because the Fiedler vector of the line crosses zero at the center edge for such a weighting. Thus the update rule in 4.28 biases toward cutting edges with large gap $(f_i - f_j)^2$ and high magnitude $f_i^2 + f_j^2$.

The following Lemmas and definitions will be combined to supplant Lemma 5, and allow the remaining technology in §4.6 to be employed to demonstrate convergence. Their purpose is to establish the greatest lower bound of SR-Algorithm as $\lambda_k = 0$. To do so we establish that update rule only admits fixed offset fractions, of $a_i,\ b_i$ as show in Figure 1, for graphs of large eigenvalue $\lambda > \frac{1}{2}$. We will call valuations that produce a constant update **iso**$-\delta$ **valuations**.

Lemma 9. *The Rayleigh quotient $\frac{f^T L f}{f^T D f}$ is invariant to a uniform rescaling of the weighting.*

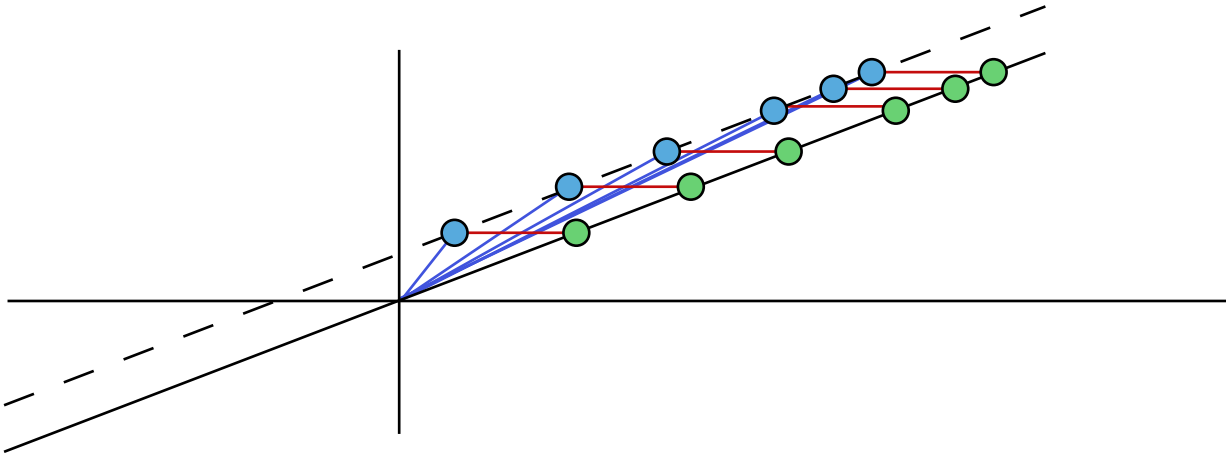


Figure 1: An illustration of shifted formal fractions producing a constant slope (update value). The offset $2\alpha^2$ is given in red and maps the slopes corresponding to the blue points to the green co-linear points (*i.e.* the green points produce a constant update fraction δ_{ij} for all edges.)

Proof. A direct consequence of the rewriting Lemma in Chapter 3 □

Given Lemma 9 we may choose a global scale for w such that $\sum_{i=1}^{n} d_i = 2$, and by a simple calculation verify that $2\alpha^2 = 1$, simplifying the reweighting scheme. The second degree of freedom that we fix in Equation 4.28 is the norm of $f \mid f^T D f = 1$. While this has no effect on the Rayleigh quotient $\frac{f^T L f}{f^T D f}$, this scaling has a dramatic effect on Equation 4.28.

Having fixed α and the scale of f it is possible to derive an **iso**$-\delta$ **valuation** $f : V \to \mathbb{R}$, such as that shown in Figure 1, for the weighted line graph. Given such an f, a weighting of the line graph G may be constructed such that $L(G)f = \lambda_2 D(G)f$ is satisfied. We will show that the weights drop off exponentially as one walks from the center to end of the line, indicating that the weighted line is an expander graph. Further, from the properties of f we can bound the eigenvalue λ_2 for all lines or arbitrary length.

Lemma 10. *Given an* **iso**$-\delta$ **valuation** f *for a connected graph* G *the following properties hold*

I. for all vertices $(u, v) \in E(G)$ *where* $f(u) > f(v)$ *then* $f(u) = f(v) + \epsilon(f(v))$,

II . *for two vertices v and u where $(u, v) \in E(G)$ then $f(u) \neq f(v)$,*

III . *graphs with cliques of size 3 or greater do not admit iso$-\delta$ valuations.*

Proof. We prove the properties for the $i^{th} + 1$ term, and assume that $f_{i+1} \geq f_i$ is satisfied. *Property I* follows from the observation that equation 4.31 admits a unique positive solution for ϵ_{i+1} in terms of the known quantities f_i and δ. This dictates that every edge joins vertices spanning exactly one threshold in f. *Property II* is simple to observe, as all the fractions would equal 0 and as G is connected this would contradict the orthogonality conditions $fD\mathbf{1} = 0$. *Property III* depends on *I* and *II*. Label the vertices of a 3-clique as a, b, and c using P*I* and P*II* we may assume that $f(a) < f(b) < f(c)$. The unique monotonicity of the edge equations contradicts this ordering relationship, demonstrating *III* for the minimal case. □

4.8.1 Constructing an iso$-\delta$ valuation for the line graph

For the weighted line graph on n vertices it is possible to construct an **iso$-\delta$ valuation** f insuring that for all edges (i, j) the update $\delta_{ij} = c$. For simplicity we assume that n is odd and number the vertices, over the integers, as $-\lfloor \frac{n}{2} \rfloor$ to $\lfloor \frac{n}{2} \rfloor$, as the line is odd length $f_0 = 0$. Given f_0 there must be an offset ϵ so that $f_1 = f_0 + \epsilon$. Thus we can write down the first update value, to the right, as

$$\delta_{01} = \frac{\epsilon_1^2}{\epsilon_1^2 + 1} \tag{4.30}$$

or more generally

$$\delta_{i,i+1} = \frac{(f_i - f_{i+1})^2}{f_i^2 + f_{i+1}^2 + 1} = \frac{\epsilon_{i+1}^2}{f_i^2 + (f_i + \epsilon_{i+1})^2 + 1} \tag{4.31}$$

the valuation f_{i+1} can be written as

$$f_{i+1} = f_i + \epsilon_{i+1} = f_i + \frac{\delta f_i + \sqrt{\delta\left(1 + 2f_i^2 - (1 + f_i^2)\delta\right)}}{|\delta - 1|} \tag{4.32}$$

an **iso$-\delta$ valuation** can be constructed by applying this recurrence, starting at the center of the line and moving toward end points.

Claim. *Given the update equation δ_{ij} with $2\alpha^2 = 1$ a vertex valuation f can be constructed such that the fraction δ_{ij} is constant for all edges in the graph.*

Claim. *A weighting of the line graph, with Laplacian L and degree matrix D, can be derived from a vector f such that $f \mid Lf = \lambda_2 Df$*

The construction of a such a beast – give the recurrence relation.

Lemma 11. *The constant update vector associated with a line graph f grows exponentially with n, the length of the graph.*

Proof. The recurrence relationship in equation 4.31 can be expressed purely in terms of the first offset, ϵ, between valuations of vertices v_1 and v_2. Thus we can express f_{i+1} terms of f_1, and ϵ by recursively back substituting ϵ_i for ϵ_{i+1}. Routine calculation reveils that the resulting recurrence grows more quickly, in magnitude, than the simplified polynomial below

$$f_i = \epsilon(i + 2\epsilon^2)(1 + 2\epsilon^2)^{i-1} \tag{4.33}$$

which grows exponentially in magnitude as n, the limit of i, tends toward infinity. □

Lemma 12. *The eigenvalue λ is fixed for a line graph weighted by a valuation f.*

Lemma 13. *A $\delta-$valuation weighted line graph has minimum λ_2 of $\frac{1}{2}$.*

Proof. currently a numerical argument, clean version forthcoming □

4.8.2 All graphs with iso$-\delta$ valuations reduce to the line

Lemma 14. *Given an iso$-\delta$ valuation f of a graph G such that $Lf = \lambda Df$ then a line graph G' may be constructed such that $L'f' = \lambda D'f'$.*

Proof. We begin by demonstrating that the thesis holds for a contraction of two vertices with the same value in f. *w.l.o.g.* assume that these vertices are associated with the least valued vertices of f and order the graph accordingly, generality holds as this is accomplished by permuting the associated matrices L and D. We expand $Lf = \lambda Df$ to clarify our assumptions

$$\begin{pmatrix} d_1 & 0 & -w_{1,i} & . & . & . \\ 0 & d_2 & -w_{2,j} & . & . & . \\ . & & . & & & \\ . & & & . & & \\ . & & & & . & \\ . & & & & & \end{pmatrix} \begin{pmatrix} a \\ a \\ b_1 \\ . \\ . \\ . \end{pmatrix} = \lambda \begin{pmatrix} d_1 \cdot a \\ d_2 \cdot a \\ d_3 \cdot b_1 \\ . \\ . \\ . \end{pmatrix}. \tag{4.34}$$

For the first vertex, rewrite the constraint equation $Lf = \lambda Df$ as

$$\begin{aligned} d_1 \cdot a - \lambda d_1 - \sum_{v_i \sim v_1} w_{1i} b_i &= 0 && (4.35) \\ (a - \lambda a) \sum_{v_i \sim v_1} w_{1i} - \sum_{v_i \sim v_1} w_{1i} b_i &= 0 \ \text{(def. of } d_1) && (4.36) \\ (a - \lambda a) \sum_{v_i \sim v_1} w_{1i} - b \sum_{v_i \sim v_1} w_{1i} &= 0 \ \text{(iso}-\delta \text{ hyp.)} && (4.37) \\ (a - \lambda a - b) \sum_{v_i \sim v_1} w_{1i} &= 0 && (4.38) \end{aligned}$$

analogously the linear constraint for v_2 can be written as $(a-\lambda a-b) \sum_{v_j \sim v_2} w_{2j} = 0$. The replacement of the b_is with b is a direct consequence of *P I* in Lemma 10. Thus we see that the constraint applies only to the weighted degrees, *i.e.* $\sum_{v_i \sim v_1} w_{1i} = \sum_{v_j \sim v_2} w_{2j}$. Define L' and D' as the graph with v_1 and v_2 contracted to v_1', removing double edges. The contraction reduces to adding the first two rows and dividing by two, which preserves the total weight constraint and thus the eigenvector equation $L'f' = \lambda D'f'$ preserves λ.

We now handle the next contraction, along the order induced by f. The only novel term introduced is the backward edge to the previously contracted layer. We drop the $'$ notation to simplify the expressions. We have

$$(b - \lambda b - c) \sum_{v_i \sim v_2} w_{2i} - a \cdot w_{21} = (b - \lambda b - c) \sum_{v_j \sim v_3} w_{3j} - a \cdot w_{31} \qquad (4.39)$$

as the weights w_{21} and w_{31} may be taken to be equal the lower diagonal terms cancel and thus the previous argument applies. This operation is repeated along the ordering induced by f, producing a line graph with matrices L' and D' such that if $Lf = \lambda Df$ then $L'f' = \lambda D'f'$. □

4.9 Multi-valuation Updating

In this section we describe two methods for integrating multiple valuations into a single update. The first treats the δ_{ij} update associated the valuations $f_1, ..., f_k$ as a coordinate and maps the resulting coordinates onto the $k-$sphere using a stereographic map. The second takes the fractional average of the valuations and maps the resulting combined δ_{ij} update value to the circle under a stereographic projection.

In the factored form the update function $\mathcal{R}$ is taken to be

$$\mathcal{R}(F_k, \Lambda_k)_{ij} = \Phi_h \left(\left[\frac{F_1^2(i) + F_1^2(j)}{(F_1(i) - F_1(j))^2}, \ \ldots \ , \frac{F_k^2(i) + F_k^2(j)}{(F_k(i) - F_k(j))^2} \right] \right) \cdot w_{ij} \quad (4.40)$$

4.9.1 A fixed update condition

Given that the $\delta_{ij}^{(k)}$ points are mapped onto the sphere a fixed update is only possible when updates for all edges (ij) are co-circular prior to the mapping. The

5 Empirical Evaluation

We show that spectral rounding compares favorably to recent spectral approximations for the Normalized Cut criterion (NCut). For the purpose of our evaluation we compared our method, denoted *SR*, to that of Yu and Shi [17], denoted *EIG*, as it reliably returns superior cuts relative to methods which use K-means and less of the eigenvectors, those based on [12]. In the following sections results are given in geometric clustering, natural image segmentation, and medical image segmentation. In all cases, the same graph $G = (V, E, w)$, is passed to both algorithms. To compare the partitioning we report the expected improvement in NCut value, on the initial graph $G = (V, E, w)$, and the **variation of information** between clusterings.

The first comparison, the expected improvement in value, can be expressed as a positive scaler $\mathbf{c}$ such that $nc(SR) \approxeq \mathbf{c} \cdot nc(EIG)$ on average. In the case where $\mathbf{c} < 1$ *spectral rounding* finds an improved cut on average.

Throughout this section we report the **variation of information** described in [9] to measure the difference between two partitionings, P_a and P_b of the vertex set V, defined as

$$\mathcal{D}_{vi}(P_a, P_b) = H(P_a) + H(P_b) - 2I(P_a, P_b) \quad (5.1)$$

with $I(P_a, P_b) = -\sum_{i=1}^{p} \sum_{j=1}^{p} \frac{n_{i,j}}{n} \log \frac{n_{i,j}}{n}$, $H(P) = -\sum_{k=1}^{p} \frac{n_k}{n} \log \frac{n_k}{n}$, where n_k is number of vertices in the k^{th} partition, $n_{i,j}$ the size of the overlap between the i^{th} partition in P_a and j^{th} partition in P_b, and $n = |V|$. The entropy term $H(P)$ can be thought of as the number of bits required to represent the distribution of partition sizes. Where as $I(P_a, P_b)$ is the mutual information between the two partitionings P_a and P_b. And so, $\mathcal{D}_{vi}$ can be thought of as the number of bits needed to represent the cumulative disagreement between the overlapping partitions of P_a and P_b. As expected $\mathcal{D}_{vi}(P_a, P_a) = 0$.

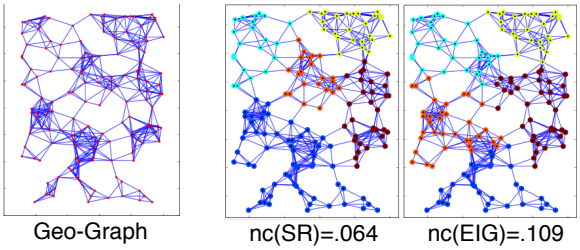


Figure 2: A ($|V| = 300$) geometric graph, and two 5-way cuts.

5.1 Random Geometric Graphs

We compare *SR* and *EIG* in the expected partition cost on a collection of random geometric graphs. The vertices of $G = (V, E, w)$ are associated with uniformly distributed coordinates in $\mathbb{R}^d$. The edge set of $E(G)$ is then constructed using the following rule, for $\{u, v \in V | u \neq v\}$, $(u, v) \in E \iff dist(u, v) < r$. We sampled 10000 graphs with 1000 vertices and chose the radius r such that the expected degree of each vertex was approximately $log(|V|)$. As shown in Figure 2 such graphs afford a large number of relatively small cuts. Table 1 contains the improvement factor, and the cluster divergence. We note that the divergence distance, relative to partition entropy $H(SR)$, highlights that the NCut improvements are not due to a small number of boundary vertex exchanges, but rather that SR and EIG return significantly different subgraphs.

	$\mathcal{D}_{vi}(SR, Eig)$	$nc(SR) = \mathbf{c} \cdot nc(EIG)$
geo-graph	$0.910 \pm .219$	$\mathbf{c} = .690 \pm .113$

Table 1: Comparison between spectral rounding *SR* and the multi-way cut algorithm of Yu and Shi [17] *EIG*. The partition entropy for SR was $H(SR) \approxeq 1.935$.

5.2 Image Segmentation

The parameters used in constructing a weighted graph from an image were fixed for all the results presented in this section. The graph $G = (V, E, w)$ represents an image as follows. For each pixel in the image a vertex in V is assigned. If two pixels are connected in E a weight in w is determined based on the image data. The graph connectivity, E, was generated by connecting pixels to 15% of their neighboring pixels in a 10 pixel radius. The initial weighting w of the graph

$G = (V, E, w)$ was determined using the *Intervening Contour* cue described in [14]. This cue assigns small weights to pixels which lie on opposite sides of a strong image boundary, and large weights otherwise.

5.2.1 Natural Image Segmentation

We compiled a set of a 100 images from Google Images using the keywords *farm, sports, flowers, mountains, & pets*. Examples from this data set, and segmentations can be found in Figure 3. Again, we note that changes in the cut value often correlate with large changes in the co-membership relationships on the image pixels. To quantitatively compare the methods on natural images we report the divergence distance and NCut improvement factor $\mathbf{c}$.

	$\mathcal{D}_{vi}(SR, Eig)$	$nc(SR) = \mathbf{c} \cdot nc(EIG)$
natural	$1.23 \pm .160$	$\mathbf{c} = .536 \pm .201$

Table 2: Comparison between spectral rounding *SR* and the multi-way cut algorithm of Yu and Shi [17] *EIG* on segmentations of natural images. The average cluster entropy over *SR*-segmentations of the image collection is $1.62 \pm .4$.

5.2.2 Medical Image Segmentation

To a degree, clustering methods are only successful in that are useful in servicing a particular task. We have selected a simple medical task, segmenting out the left ventricle (a fluid sack located in the brain), as it is well defined – *i.e.* the boundary of the percept is agreed upon by experts. While this task would seem to be relatively easy, a successful automatic method represents a significant reduction in human effort for a common labeling task.

The test set was constructed from a collection of 200 NMR images containing the left ventricle. The collection was built by taking 5 slices each from scans of 40 individuals. Images were selected randomly from slices containing the left ventricle. As shown in Figure 6 the appearance of the ventricle varies substantially in shape and size.

The comparison of segmentations obtained from spectral rounding and the eigenvector method of Yu and Shi [17] with the human labels is given in Table 3. The divergence distance and expected cut improvement are given in Table 4. The average cluster entropy for *SR* was $0.611 \pm .131$. As this is a two-class problem,

this suggests that one of the segments tends to be much smaller than the other. This is due to the often small size of the ventricle in the image.

	$nc(SR)$	$nc(EIG)$[17]
$Pr(v \in T(Im))$	$.95 \pm .04$	$.37 \pm .12$

Table 3: The value $Pr(v \in T(Im))$ is reported over the population of images, where $T(Im)$ is the expert's hand segmentation and $Pr(v \in T(Im))$ is the probability that a pixel v in a segment is also contained in $T(Im)$ – this statistic was computed for the segment with the largest overlap with $T(Im)$.

	$\mathcal{D}_{vi}(SR, Eig)$	$nc(SR) = \mathbf{c} \cdot nc(EIG)$
medical	$1.856 \pm .192$	$\mathbf{c} = .598 \pm .237$

Table 4: The divergence and expected value improvement for the medical image data set. The average cluster entropy for *SR* segmentations on the medical data set was $0.611 \pm .131$.

References

[1] S. Arora, S. Rao, and U. V. Vazirani. Expander flows, geometric embeddings and graph partitioning. In *STOC*, pages 222–231, 2004. 2

[2] T. F. Chan, J. R. Gilbert, and S.-H. Teng. Geometric spectral bisection. Manuscript, July 1994. 2

[3] F. R. K. Chung. *Spectral graph theory*. Number 92 in Regional Conference Series in Mathematics. Amer. Math. Soc, Providence, 1992. 2, 3, 6

[4] S. Guattery and G. L. Miller. On the quality of spectral separators. *Matrix Analysis & Applications*, 19(3), 1998. 3

[5] J. Kleinberg. An impossibility theorem for clustering. *Advances in Neural Information Processing Systems*, 2003. 2

[6] K. Lang and S. Rao. A flow-based method for improving the expansion or conduectance of graph cuts. *Integer Programming and Combinatorial Optimization*, pages 325–337, June 2004. 2

[7] P. D. Lax. *Linear Algebra*. Pure and Applied Mathematics. Wiley-Interscience, 1st edition, 1997. 10

[8] F. T. Leighton and S. Rao. An approximate max-flow min-cut theorem for uniform multicommodity flow problems with applications to approximation algorithms. In *FOCS*, pages 422–431. IEEE, Oct. 1988. 2

[9] M. Meilă. Comparing clusterings by the variation of information. *COLT*, pages 173–187, 2003. 21

[10] M. Meilă. Comparing clusterings – an axiomatic view. *ICML*, pages 577–584, 2005. 2

[11] B. Mohar. Isoperimetric numbers of graphs. *Journal of Combinatorial Theory, Series B*, 47:274–291, 1989. 3

[12] A. Ng, M. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *NIPS*, 2002. 2, 4, 21

[13] N. Shental, A. Zomet, T. Hertz, and Y. Weiss. Learning and inferring image segmentations with the gbp typical cut algorithm. *ICCV*, 2003. 2

[14] J. Shi and J. Malik. Normalized cuts and image segmentation. In *PAMI*, 2000. 2, 4, 23

[15] D. Spielman and S. Teng. Spectral partitioning works: Planar graphs and finite element meshes. *FOCS*, 1996. 3

[16] E. P. Xing and M. I. Jordan. On semidefinte relaxtion for normalized k-cut and connections to spectral clustering. *TR-CSD-03-1265*, University of California Berkeley, 2003. 2, 4

[17] S. Yu and J. Shi. Multiclass spectral clustering. In *ICCV*, October 2003. 2, 3, 4, 21, 22, 23, 24, 26, 27

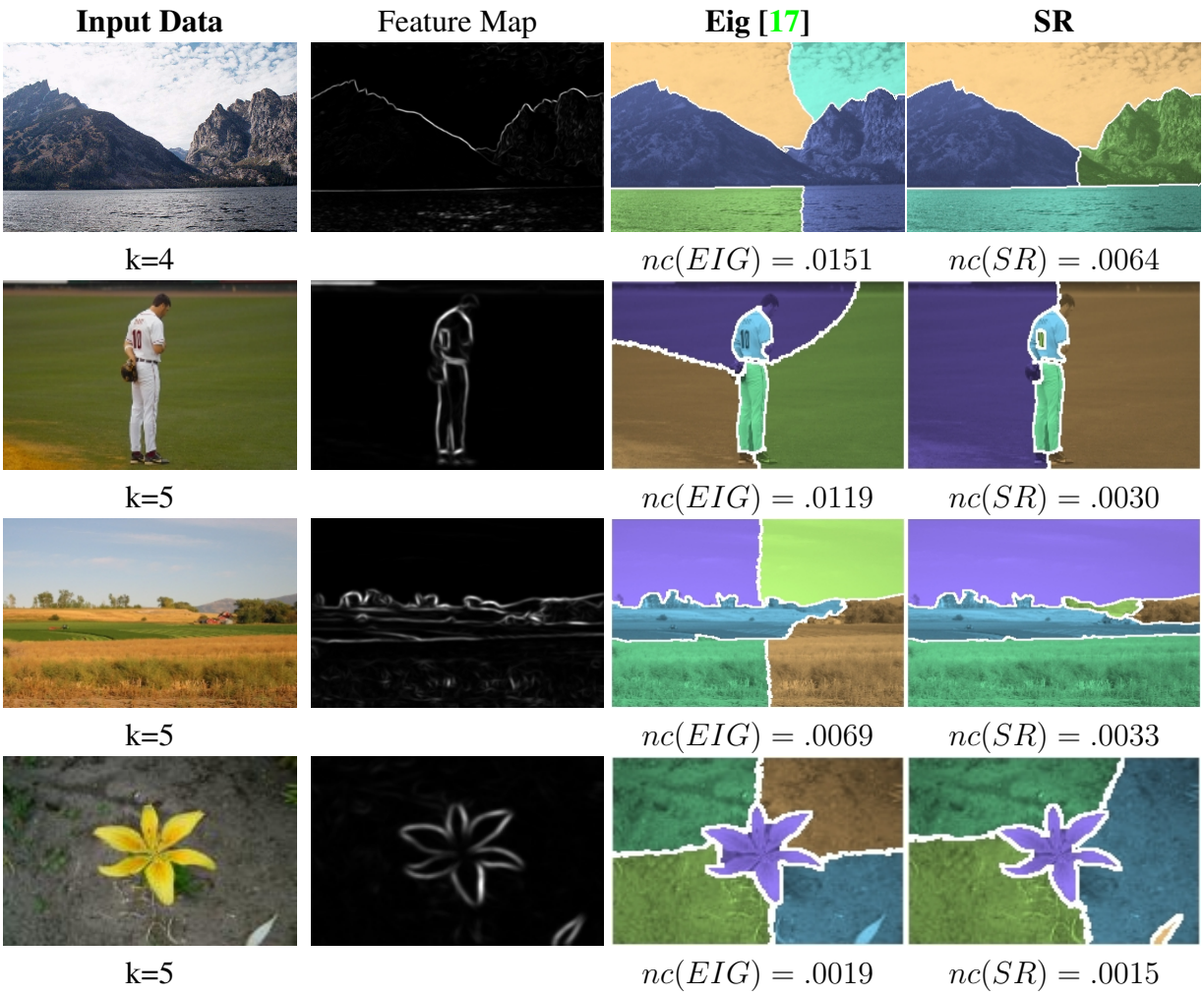


Figure 3: The first four rows illustrate with qualitative examples the improvements in NCut value for natural images. Column three contains segmentations generated by the published code of Yu and Shi [17]. Column four contains results for *Spectral Rounding*. The number of segments k is fixed for each comparison. We emphasize that the cut cost was evaluated on identical combinatorial problems (graphs).

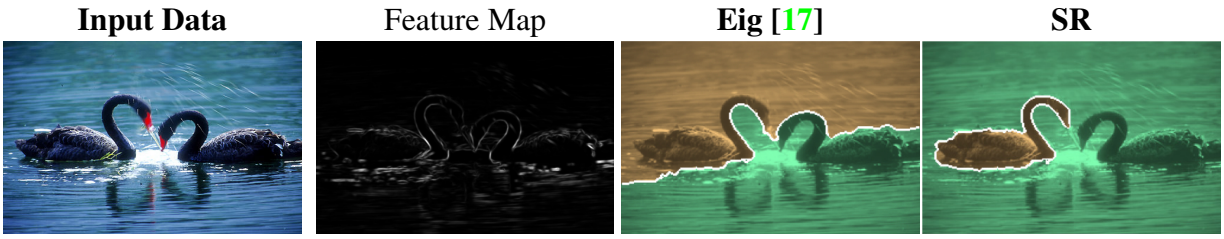


Figure 4: The second to last row illustrates a 2-way cut in which the NCut values are nearly identical, but which support very different percepts.

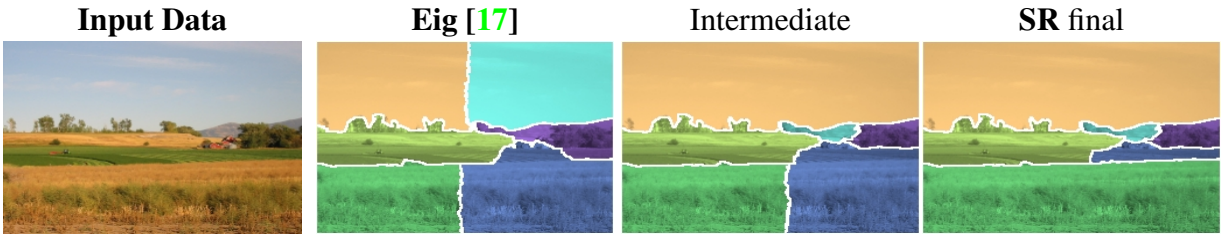


Figure 5: A sequence of iterations projected onto the feasible set, starting left with solution from Yu's method and ending with the fourth and final *SR* iteration on the right. Notice that the large cuts in the sky and field shift to small cuts in the area around the farm.

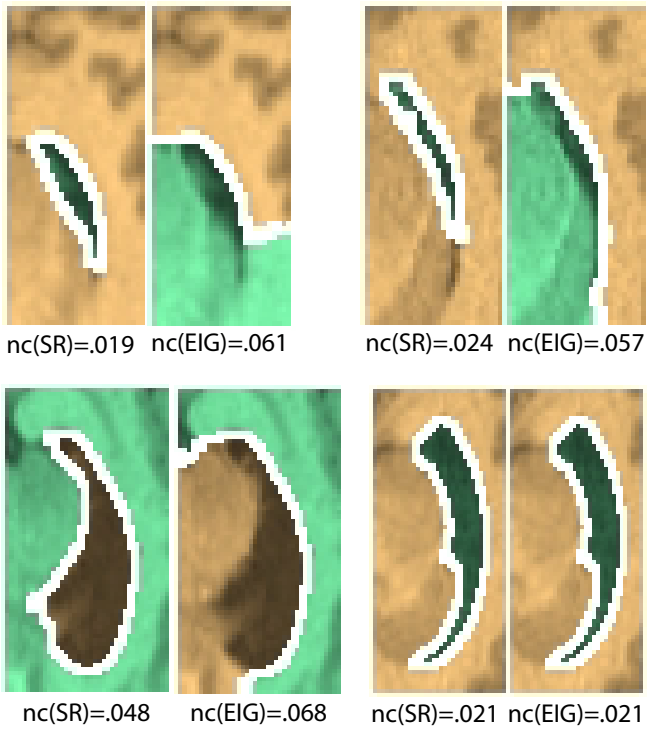


Figure 6: Examples of the left ventricle, and qualitative results for the *SR* and *EIG* algorithms. Segmentations required approximately 1.2 seconds for *EIG* and 1.9 seconds for *SR*.