

# A Polynomial Time Algorithm for Rayleigh Ratio on Discrete Variables: Replacing Spectral Techniques for Expander Ratio, Normalized Cut, and Cheeger Constant

Dorit S. Hochbaum

Department of Industrial Engineering and Operations Research, University of California, Berkeley, Berkeley, California 94720,  
hochbaum@ieor.berkeley.edu

A general form of minimizing the Rayleigh ratio on discrete variables is shown here, for the first time, to be polynomial time solvable. This is significant because major problems in clustering, partitioning, and imaging can be presented as the Rayleigh ratio minimization on discrete variables and an *orthogonality* constraint. These challenging problems are modeled as the *normalized cut* problem, the graph expander ratio problem, the Cheeger constant problem, or the *conductance* problem, all of which are NP-hard. These problems have traditionally been solved, heuristically, using the “spectral technique.” A unified framework is provided here whereby all these problems are formulated as a constrained minimization form of a quadratic ratio, referred to here as the *Rayleigh ratio*. The quadratic ratio is to be minimized on discrete variables and a single sum constraint that we call the *balance* or *orthogonality* constraint. When the discreteness constraints on the variables are disregarded, the resulting continuous relaxation is solved by the spectral method. It is shown here that the Rayleigh ratio minimization subject to the discreteness constraints requiring each variable to assume one of two values in  $\{-b, 1\}$  is solvable in strongly polynomial time, equivalent to a single minimum  $s, t$  cut algorithm on a graph of same size as the input graph, for any nonnegative value of  $b$ . This discrete form for the Rayleigh ratio problem was often assumed to be NP-hard. Not only is it shown here that the discrete Rayleigh ratio problem is polynomial time solvable, but also the algorithm is more efficient than the spectral algorithm. Furthermore, an experimental study demonstrates that the new algorithm provides in practice an improvement, often dramatic, on the quality of the results of the spectral method, both in terms of approximating the true optimum of the Rayleigh ratio problem on both the discrete variables and the balance constraint, and in terms of the subjective partition quality.

A further contribution here is the introduction of a problem, the *quantity-normalized cut*, generalizing all the Rayleigh ratio problems. The discrete version of that problem is also solved with the efficient algorithm presented. This problem is shown, in a companion paper, to enable the modeling of features essential to clustering that are valuable in practical applications.

*Subject classifications:* Cheeger constant; parametric cut algorithm; Fiedler eigenvector; quantity-normalized cut.

*Area of review:* Optimization.

*History:* Received July 2010; revision received May 2012; accepted September 2012. Published online in *Articles in Advance* February 8, 2013.

## 1. Introduction

A general form of minimizing the Rayleigh ratio on discrete variables is shown here, for the first time, to be polynomial time solvable. This is significant because major problems in clustering, partitioning, and imaging can be presented as the Rayleigh ratio minimization on discrete variables and an *orthogonality* constraint. The clustering problems addressed here include the well-known partitioning problems of *normalized cut* (Shi and Malik 2000), the graph *expander ratio* (Hoory et al. 2006), the *Cheeger constant* (Cheeger 1970), the *uniform sparsest cut* (Leighton and Rao 1999), and the *conductance* problem (Jerrum and Sinclair 1997). All these problems are NP-hard and there is no known approximation algorithm for any of them that

guarantees to generate a solution that is at most within a bounded factor of the optimum.

The dominant, and most commonly used, solution method for these problems is the “spectral” method based on finding the eigenvector(s) of a related matrix. We devise here a new combinatorial algorithm alternative to the spectral method that delivers solutions to the problems that are often of better quality and with more efficient run times.

The theoretical underpinnings for the spectral method were motivated by Perron-Frobenius theorem (Perron 1907, Frobenius 1912). The spectral method has been used in ranking forms of clustering since the 1950s (see, e.g., Keener 1993). The use of the spectral technique for graph partitioning is based on links between the values of the eigenvalues of the Laplacian matrix of a graph and assorted

graph properties. Such properties are investigated within the framework of “spectral graph theory.” A good exposition on this subject is present in a classic book by Chung (1997). Although eigenvectors and their respective eigenvalues can be found in polynomial time, for an undirected graph on  $n$  nodes and  $m$  edges the Laplacian matrix is of size  $n \times n$ . For images of size  $1,000 \times 1,000$ , for instance, the respective Laplacian matrix contains  $10^{12}$  elements, presenting a formidable computational challenge. This challenge is magnified by the fact that eigenvector algorithms are “algebraic” and as such are prone to round-off errors and run times affected by the size of the numbers in the matrix. Because in Laplacian matrices only entries corresponding to the  $m$  edges of the graph are nonzero, these are typically sparse. A great deal of recent research is focused on algorithms that can find *approximate* eigenvectors for Laplacian matrices, that are symmetric and diagonally dominant, and with high probability, in exchange for improved run times. The recent significant works in this area vary from the Spielman and Teng (2004) probabilistic algorithm that solves a system of linear equations for symmetric and diagonally dominant matrices with relative error  $\epsilon$  with complexity  $O(m \log^{32} n \log(1/\epsilon))$ , to a recent improvement by Koutis et al. (2011) with complexity  $O(m \log n (\log \log n)^2 \log(1/\epsilon))$ . These recent research efforts attempt to address the computational shortcomings of the spectral method.

Yet, even when the eigenvector is found, there is no obvious way of translating it to a discrete bipartition. The most common approach for doing that is to test all possible *threshold* values, and, for each, partition the  $n$  entries of the vector to the set of those of value exceeding the threshold and those below the threshold. The threshold that gives the best (lowest) value of the objective function is then selected. This technique is called *sweep* and its use is justified by theoretical results in the form of approximability bounds, given in (4).

In contrast, the technique proposed here solves optimally the discrete relaxation of the problem, with a strongly polynomial combinatorial algorithm with complexity  $O(mn \log(n^2/m))$  for a graph on  $n$  nodes and  $m$  edges. The algorithm is flow based and combinatorial, using HPF (Hochbaum’s Pseudo Flow) as a subroutine, which is efficient enough to be used to solve these problems on millions of elements and more than 300 million edges within less than 10 minutes—instances that are way beyond the limits of the spectral technique. Moreover, it is demonstrated, in an extensive experimental study (Hochbaum et al. 2012) that the results of the combinatorial algorithm here improve, often dramatically, on the quality of the results of the spectral method, both in terms of providing better value to the NP-hard objective, and in terms of better visual quality segmentation.

The problems are presented here within a unifying framework whereby a form of the problems is formulated as a *Rayleigh ratio*, generalizing the standard Rayleigh

ratio. The spectral technique can be viewed as solving a continuous relaxation of the Rayleigh ratio problem that disregards the discrete nature of the variables and their values. Our main result is to show that the discrete Rayleigh ratio problem (that excludes the balance constraint) is solvable in strongly polynomial time, and often provides better quality solutions than the spectral approach. Further results include the introduction of the *quantity-normalized cut* that generalizes these problems and experimental evidence to the superiority of the combinatorial algorithm in terms of run time, substantially better approximations, and better subjective quality of the delivered clustering.

### 1.1. Notation

Vectors are denoted in boldface. Let  $G = (V, E)$  be an undirected graph with edge weights  $w_{ij}$  for  $[i, j] \in E$ . For  $A, B \subseteq V$ , let  $C(A, B) = \sum_{[i, j] \in E, i \in A, j \in B} w_{ij}$ . A bipartition of the set of nodes in a graph  $\{S, \bar{S}\}$  is identified with a *cut*,  $(S, \bar{S}) = \{[i, j] \mid i \in S, j \in \bar{S}\}$ , where  $\bar{S} = V \setminus S$ . The capacity of a cut is  $C(S, \bar{S})$ . For inputs with two sets of weights associated with each edge,  $w_{ij}^{(1)}, w_{ij}^{(2)}$  we let  $C_1(A, B) = \sum_{i \in A, j \in B} w_{ij}^{(1)}$  and  $C_2(A, B) = \sum_{i \in A, j \in B} w_{ij}^{(2)}$ .

All the Rayleigh problems, except conductance, are defined for undirected graphs. For a directed graph,  $G = (V, A)$ , an arc in  $A$  is denoted by an ordered pair  $(i, j)$ . This is to differentiate the notation from that of the undirected edge that is an unordered pair,  $[i, j]$ .

Let  $d_i = \sum_{[i, j] \in E} w_{ij}$  be the weighted degree of node  $i$ . For a subset of nodes  $V' \subseteq V$  we let  $d(V') = \sum_{i \in V'} d_i = C(V', V)$ . For arbitrary scalars  $q_i$  for all  $i \in V$  we define  $q(V') = \sum_{i \in V'} q_i$  and a diagonal matrix  $Q$  with  $Q_{ii} = q_i$ .

We use the notation of  $n = |V|$  for the number of nodes, and  $m = |E|$  for the number of edges in the graph  $G$ . Let  $\mathbf{1}$  be the vector of all ones.

Let  $D$  be a diagonal matrix  $n \times n$ , with  $D_{ii} = d_i = \sum_{[i, j] \in E} w_{ij}$ . Let  $W$  be the weighted node–node adjacency matrix of the graph, where  $W_{ij} = W_{ji} = w_{ij}$ . The matrix  $\mathcal{L} = D - W$  is called the Laplacian of the graph and is known to be positive semi-definite (Hall 1970). For a directed graph  $G = (V, A)$  with arc weights  $w_{ij}$ , the matrix  $W$  is not symmetric, and  $d_i = \sum_{(i, j) \in A} w_{ij}$ . Otherwise, all derivations here apply to the nonsymmetric, or directed graph, cases as well.

Let  $n = |V|$  and  $m = |E|$ . Denote by  $T(n, m)$  the complexity of the minimum  $s, t$ -cut problem. This complexity is, for example,  $O(mn \log(n^2/m))$  using the push-relabel algorithm (Goldberg and Tarjan 1988), or using HPF algorithm; see Hochbaum (2008) and Hochbaum and Orlin (2013).

### 1.2. The Problems and Applications

We sketch here the major Rayleigh ratio problems and their applications.

The *expansion ratio* of a graph (also known as the *isoperimetric value*) is the value of  $\min_{S \subseteq V, |S| \leq |V|/2} (C(S, \bar{S})/|S|)$ .

Finding the isoperimetric value of a graph is an NP-hard problem. Computing this value is of importance because graphs with large isoperimetric value are desirable in the construction of good error-correcting codes. In such graphs messages can be decoded correctly, even if multiple bits are corrupted in the transmission (Sipser and Spielman 1996; Spielman 1996, 1999; Hoory et al. 2006). Codes corresponding to graphs with large isoperimetric values feature large Hamming distances between code words. Consequently, a message arriving with few erroneous bits can likely be decoded correctly by mapping it to the nearest code word. Graphs with small isoperimetric values are called *concentrators*.

The *Cheeger constant* is the quantity

$$\min_{S \subset V, |S| \leq |V|/2} \frac{C(S, \bar{S})}{d(S)}.$$

An associated quantity is the *normalized cut* value

$$\min_{S \subset V} \left\{ \frac{C(S, \bar{S})}{d(S)} + \frac{C(\bar{S}, S)}{d(\bar{S})} \right\}$$

discussed in detail in in §2.

The Cheeger constant of graphs has been associated with the static/dynamic load-balancing problem in parallel computing (Van Driessche and Roose 1995). The set up for this application is grid computing, where each element is a grid point and the goal is to assign grid points to each processor, so as to minimize the amount of required communication between the processors, while increasing the intergrid similarity within each subgrid. To partition the grid points into more than two processors, one iteratively repeats the partition generated by the optimal solution to Cheeger constant.

The Cheeger constant partition and the related normalized cut's partition produce a cluster that is dissimilar from its complement whereas each subset of the bipartition retains as much similarity as possible between the elements of the subset (see Lemma 2). These models have been used extensively for *image segmentation*. Normalized cut has also been used in numerous other application areas including gene clustering, e.g., Narayanan et al. (2010). In other contexts, Hu et al. (2005) used normalized cut as part of a scalable algorithm for mining dense and large subgraphs, and Filippone et al. (2008) provide an excellent survey on the use of kernel, spectral, and normalized cut methods for clustering.

A directed version of Cheeger constant is of importance for assessing the *mixing rate* of Markov chains—that is, the rate of convergence to the stationary probability of the states. In that context, we refer to the analog of Cheeger constant as the *conductance ratio*. Directed graphs with nodes representing the states, and arcs with transition probability associated with them, that have good conductance ratio have rapidly mixing Markov chains. The conductance ratio of a directed graph  $G = (V, A)$  is defined as

$\min_{S \subset V, \pi(S) \leq 1/2} (C(S, \bar{S})/\pi(S))$ . Here the weights assigned to the arcs,  $w_{ij}$ , are the transition probabilities from  $i$  to  $j$ , and the node weights  $\pi_i$  are set equal to  $\sum_{(i,j) \in A} w_{ij}$ ,  $\pi(S) = \sum_{i \in S} \pi_i$ , and  $\pi(V) = 1$ . Therefore this is a directed version of Cheeger constant with the scaling  $\pi_i = d_i/d(V)$ .

The mixing rate was shown to be related to the conductance ratio of the graph in Jerrum and Sinclair (1997). The conductance ratio is also useful in an algorithmic design toolbox that generates *fully polynomial time randomized approximation scheme* (Jerrum and Sinclair 1997). That is, in combination with commonly used statistical techniques (Jerrum and Sinclair 1997) or reduction techniques from sampling to set counting (Sinclair and Jerrum 1989), one can find, for graphs with good conductance ratio, efficient probabilistic algorithms addressing problems that include approximate counting, combinatorial optimization by stochastic heuristic search, and a variety of problems in theoretical computer science and statistical physics (Sinclair 1992, Jerrum and Sinclair 1997). This algorithmic framework is called *Markov chain Monte Carlo* method.

Because in most clustering applications of interest the weights on the edges are (symmetric) similarity weights, we focus the presentation here on undirected graphs. Nevertheless all the results presented carry to directed graphs as well, and the only change for the spectral method is that the Laplacian matrix becomes asymmetric.

One of the contributions here is a generalization of Cheeger constant, which we call *half-quantity-normalized-cut problem*, or half- $q$ -NC:

$$\text{(half-}q\text{-NC)} \quad \phi_G = \min_{S \subset V, q(S) \leq q(V)/2} \frac{C(S, \bar{S})}{q(S)}. \quad (1)$$

The half-quantity-normalized-cut problem has not been discussed previously in the literature. Yet, this problem, and its discrete relaxation, shown here to be polynomial time solvable, are both good models for clustering. We demonstrate its use in image segmentation, where each node (or pixel in the image) has a  $q_i$  value that is higher the likelier is the pixel to belong to the segmented feature. In the experiments presented in §6 the pixel weights are the respective entropy values of the nodes in the associated graph. This ability to assign node weights different from the sum of weights of adjacent edges has applications in, e.g., medical imaging where the goal is to delineate certain tissue area such as knee cartilage. The value of  $q_i$  in that case increases with growing similarity of the texture of the pixel, for instance, to that of the segmented feature. This is unlike existing models such as the normalized cut or Cheeger where the node weights depend only on the similarity to adjacent pixels or nodes.

We present here a unifying framework for all these problems, where the optimal solution is shown to be the solution to a quadratic discrete optimization problem, which we call the *Rayleigh problem* because it is a variant of the well-known Rayleigh ratio (or quotient):

$$\text{(Rayleigh problem)} \quad \min_{\mathbf{y}^T \mathbf{Q} \mathbf{1} = 0, \mathbf{y}_i \in \{-b, 1\}} \frac{\mathbf{y}^T \mathcal{L} \mathbf{y}}{\mathbf{y}^T \mathbf{Q} \mathbf{y}}. \quad (2)$$

For a problem on  $n$  variables the vector  $y$  is  $n$ -dimensional, the matrix  $Q$  is diagonal and nonnegative, and  $\mathcal{L} = D - W$  is the *Laplacian* of the graph. For different settings of  $Q$  we get the formulations of these problems, as shown in §2. We refer to the constraint  $y^T Q \mathbf{1} = 0$  as the *orthogonality* or *balance* constraint.

The Rayleigh problem (2) is NP-hard, even for  $Q = I$  (where it is a “double form” of the expander ratio and NP-hard, discussed in detail in §2). A common approach for generating solutions and bounds for NP-hard problems is to relax the constraints that make the respective problem hard. The Rayleigh problem is a useful formulation because it generalizes a large number of bipartitioning models and because the spectral method can be viewed as relaxing the discretizing constraints  $y_i \in \{-b, 1\}$  in this formulation. We refer to this relaxation here as the *continuous Rayleigh* problem. Notice that this relaxation relaxes more than the discreteness requirement that omits the discrete choice between  $-b$  and one and replaces it by the constraints  $y_i \in [-b, 1]$ . The continuous Rayleigh relaxation further omits the constraints  $y_i \in [-b, 1]$ .

The spectral method is the dominant technique used to date for solving the problems of normalized cut, expander ratio and Cheeger constant (Shi and Malik 2000, Sharon et al. 2006, Tolliver and Miller 2006). One justification for the use of the spectral method is the approximation bound known for the optimal objective value of Cheeger constant,  $h_G$ , based on the second smallest eigenvalue, the Fiedler value  $\lambda_1$  (Fiedler 1973, 1975; Cheeger 1970; Chung 1997):

$$\frac{\lambda_1}{2} \leq h_G \leq \sqrt{2\lambda_1}. \quad (3)$$

The constructive proof of the second inequality of (3) utilizes the bipartition generated by applying the sweep technique. Denoting Cheeger constant’s objective value for this bipartition by  $h_{s\text{-spec}}$ , (“ $s$ ” stands for sweep and “spec” stands for the spectral method) the following inequalities are known to be satisfied (Chung 1997):

$$h_G \leq h_{s\text{-spec}} \leq 2\sqrt{h_G}. \quad (4)$$

Although the theoretical bounds for the spectral solution appear promising, the spectral technique’s solutions are often poor approximations compared to the combinatorial technique. The spectral method is also computationally demanding, noncombinatorial, and imposes severe memory requirement because of the challenge of dealing with the very large (though sparse) Laplacian. For an image of size  $n \times n$ , the Laplacian matrix has  $n^4$  entries, which makes applying the spectral method to images of sizes exceeding  $100 \times 100$  very challenging. The running time results reported in Hochbaum et al. (2012) indeed confirm that in practice the running times of the spectral method are substantially higher than those of the combinatorial method, and furthermore, this gap grows with the size of

the input, indicating that the combinatorial method scales well, whereas the spectral method scales poorly.

The combinatorial algorithm establishes, for the first time, that the binary optimization problem, the discrete Rayleigh ratio (DRR),

$$(b\text{-DRR}) \quad \alpha(b) = \min_{y_i \in \{-b, 1\}} \frac{\mathbf{y}^T \mathcal{L} \mathbf{y}}{\mathbf{y}^T Q \mathbf{y}} \quad (5)$$

is polynomial time solvable and for any value of  $b \geq 0$ . We refer henceforth to this algorithm as the DRR algorithm

### 1.3. Previous Work on the Discrete Rayleigh Ratio Problems

DRR problems have been investigated previously, often assuming implicitly that they are NP-hard, or even stating so explicitly.

Sharon et al. (2006) defined the following problem, calling it “normalized cut,” and stating it is NP-hard:

$$NC'_G = \min_{S \subseteq V} \frac{C(S, \bar{S})}{C(S, S)}. \quad (6)$$

The solution to this problem, referred to here as  $NC'$ , is a subset  $S$  that is as dissimilar as possible to  $\bar{S}$  and that also has similarity within, measured by  $C(S, S)$ , that is as large as possible. This problem was shown to be polynomial time solvable in Hochbaum (2010). However, that algorithm devised in Hochbaum (2010) is different from the DRR algorithm here for the discrete Rayleigh ratio problem as explained next.

In matrix form the problem that corresponds to  $NC'$  can be written as

$$NC'_G = \min_{x_i \in \{0, 1\}} \frac{\sum w_{ij}(x_i - x_j)^2}{\sum w_{ij}x_i \cdot x_j} = \min_{x_i \in \{0, 1\}} \frac{\mathbf{x}^T \mathcal{L} \mathbf{x}}{\mathbf{x}^T W \mathbf{x}}. \quad (7)$$

(See proof of Lemma 3 to verify this.)

The algorithm of Sharon et al. (2006) relaxed the discrete constraints  $x_i \in \{0, 1\}$  and the resulting continuous problem was solved heuristically with an approximate procedure to find the Fiedler eigenvector. However, as noted in Hochbaum (2010), this problem is polynomial time solvable. The algorithm used in Hochbaum (2010) is different from the DRR algorithm and it does not apply directly to Rayleigh ratio problems.

To see why, notice that problem (7) is *not* a Rayleigh ratio problem because the matrix  $W$  in the denominator is not a diagonal matrix. The algorithm in Hochbaum (2010) formulated this problem directly, as an integer program on monotone inequalities. This formulation of this problem includes a variable for each node and a variable for each edge (or arc). The problem is then shown to be equivalent to a parametric cut problem on a graph on  $n + m$  nodes and  $n + m$  edges. Consequently the running time is  $T(m, m)$ . The DRR algorithm here based on the Rayleigh ratio formulation is more efficient, of complexity  $T(n, m)$ .

To see that problem (7) is in fact equivalent to the discrete Rayleigh problem, observe that

$$\frac{\mathbf{x}^T \mathcal{L} \mathbf{x}}{\mathbf{x}^T \mathbf{W} \mathbf{x}} = \frac{\mathbf{x}^T \mathcal{L} \mathbf{x}}{\mathbf{x}^T \mathbf{D} \mathbf{x} - \mathbf{x}^T \mathcal{L} \mathbf{x}} = 1 / \left( \frac{\mathbf{x}^T \mathbf{D} \mathbf{x}}{\mathbf{x}^T \mathcal{L} \mathbf{x}} - 1 \right).$$

Therefore, minimizing this ratio for  $x_i \in \{0, 1\}$  is equivalent to maximizing  $(\mathbf{x}^T \mathbf{D} \mathbf{x}) / (\mathbf{x}^T \mathcal{L} \mathbf{x})$  on binary variables, which in turn is equivalent to minimizing the reciprocal quantity  $(\mathbf{x}^T \mathcal{L} \mathbf{x}) / (\mathbf{x}^T \mathbf{D} \mathbf{x})$ , which is the discrete Rayleigh problem. Therefore the problem is solved more efficiently than in Hochbaum (2010) with the DRR algorithm given here.

A variant of the NP-hard problem of the expander ratio,  $\min_{|S| \leq n/2} (C(S, \bar{S}) / |S|)$ , was addressed under the name *ratio regions* by Cox et al. (1996). The ratio region problem is motivated by seeking a segment, or region, where the boundary is of low cost and the segment itself has high node weight:

$$\min_{S \subset V} \frac{C(S, \bar{S})}{|S|}. \quad (8)$$

Note that this formulation does not contain the constraint  $|S| \leq n/2$  that is present in the expander ratio problem. Hence, the segments' sizes are not necessarily balanced. The ratio regions problem studied by Cox et al. (1996) is restricted to planar graphs and thus, in the context of images, to planar grid images with four neighbors only. In case of planar graphs, the length of the path along the boundary of the region is the same as the capacity of a cut in the dual graph. This observation is key to the algorithm in Cox et al. (1996). For graph nodes of weight  $q_i$  the problem is generalized to,  $\min_{S \subset V} (C(S, \bar{S}) / \sum_{i \in S} q_i)$ . Cox et al. (1996) showed how to solve the weighted problem on planar graphs where all node weights are positive.

This weighted problem, for *any* general graph and for arbitrary weights, was shown in Hochbaum (2010) to be polynomial time solvable. That algorithm is a special case of the DRR algorithm here solving the 0-DRR  $\min_{x_i \in \{0, 1\}} (\mathbf{x}^T \mathcal{L} \mathbf{x}) / (\mathbf{x}^T \mathbf{Q} \mathbf{x})$ .

Ding et al. (2001) suggested an approach of utilizing the Fiedler eigenvector to better balance the partition. The idea is to check all thresholds with respect to a different objective function than normalized cut and find the best solution for this alternative objective. Such an approach certainly provides a more balanced partition than the sweep technique, but a worse solution (larger value) to the normalized cut problem than the solution provided by the sweep method. In their presentation they use the discrete Rayleigh ratio in which the discreteness constraints are relaxed. That formulation, as shown here, could be solved in polynomial time optimally.

#### 1.4. Summary of Main Contributions

Our main contributions here include the following:

1. Devising a unifying framework that presents all the problems discussed here as a Rayleigh ratio problem. This

has not been shown explicitly up until now for all these problems and the proof offered here is the simplest compared to proofs that exist for special cases, e.g., Chung (1997) and Shi and Malik (2000).

2. Providing the first known polynomial time algorithm for DRR—the Rayleigh ratio minimization subject to the discrete two-valued (binary) constraints.

3. Introducing the half-quantity-normalized-cut problem as a generalization of the expander, Cheeger constant, and conductance problems.

4. Showing that the quantity-normalized cut' problem is solved in polynomial time with arbitrary node weights.

5. Identifying the orthogonality constraint as the “cause” of the NP-hardness of the problem and showing its equivalent to the “balance” constraint  $q(S) \leq q(V)/2$ .

6. Providing a simple and easy proof that for  $\alpha(b)$  (the optimal solution value to  $b$ -DRR for a fixed value  $b$ ),  $h_G$  and the half- $q$ -normalized cut  $\phi_G$  both satisfy  $h_G, \phi_G \geq \alpha(b)/2$  and  $h_G, \phi_G \geq \lambda_1/2$ .

7. Demonstrating that the proposed polynomial time algorithm is also efficient in practice and should thus be the methodology of choice for deriving good solutions for these problems. An expanded experimental study is provided in Hochbaum et al. (2012).

8. Introducing the directed version of normalized cut and quantity normalized cut and proving that the discrete Rayleigh ratio of these problems is solved in polynomial time.

#### 1.5. Overview and Organization of Paper

In §2 we explore the relationships between the different bipartition problems and introduce the double form of the problems, which are called here normalized cut problems, as they are variant of the problem known in the literature as normalized cut (Shi and Malik 2000). The algebraic and graph representations of these problems are linked by showing in §3 that all these normalized cut problems can be formulated as Rayleigh problems. In §4 we describe the spectral method that solves the continuous relaxation of the Rayleigh problem. Section 5 presents the new combinatorial algorithms for solving the discrete Rayleigh problem. Section 6 contains a summary of an experimental study on the performance of both the spectral approach and the combinatorial approaches for all the problems studied for image-based data sets. Section 7 contains a brief summary of the results presented.

## 2. The Problems and Their Relationships

In the problems addressed the objective is to identify a nonempty subset  $S \subset V$  so that a certain criterion will be optimized for the set and its complement  $\bar{S}$ . A feasible solution to any of these problems forms a bipartition  $(S, \bar{S})$ . In general, the edge weights given are *similarity* weights. That is, the magnitude of the similarity weight between two nodes is growing as the two nodes are more “similar.”

A desirable bipartition therefore corresponds to minimizing the cut  $C(S, \bar{S})$  so the two sets are as dissimilar from each other as possible. Indeed, for all problems here, part of the objective includes a term for the cut  $C(S, \bar{S})$ . The optimization criteria are in the form of *ratio* for all the problems discussed. These ratio problems are defined as follows:

- *Expander ratio of a graph*. This is also known as the isoperimetric problem. It appears as either one of the two formulations:

$$\min_{S \subset V} \frac{C(S, \bar{S})}{\min\{|S|, |\bar{S}|\}} = \min_{S \subset V, |S| \leq n/2} \frac{C(S, \bar{S})}{|S|}.$$

Here the goal is to achieve a bipartition that is as balanced as possible in terms of the size of the two sets  $S$  and  $\bar{S}$ .

- *Cheeger constant* (Cheeger 1970, Chung 1997).

$$h_G = \min_{S \subset V} \frac{C(S, \bar{S})}{\min\{d(S), d(\bar{S})\}}.$$

The problem can be written equivalently as,

$$h_G = \min_{S \subset V, d(S) \leq d(V)/2} \frac{C(S, \bar{S})}{d(S)}.$$

The goal here is to get both sets to have large similarity within. Lemma 2 demonstrates why this is the case for this objective function.

- *Normalized cut* (Shi and Malik 2000). This problem has one term equal to the objective of Cheeger constant, and a second term

$$NC_G = \min_{S \subset V} \frac{C(S, \bar{S})}{d(S)} + \frac{C(S, \bar{S})}{d(\bar{S})}. \quad (9)$$

The normalized cut problem is considered as a “double form” of Cheeger constant problem, or Cheeger constant problem is viewed as “half-normalized cut.” That is because Cheeger constant is the larger of the two terms in the objective of normalized cut. These bounds, attained for  $h_G$  in terms of the second smallest eigenvalue  $\lambda_1$ , were established via the relationship to  $NC_G$ , which was then related to the eigenvalue, (see Theorem 1 for proof).

- *Conductance problem*. This problem is a directed version of Cheeger constant. It is defined on a directed graph  $G = (V, A)$ , where the weight of arc  $(i, j)$  is the transition probability from  $i$  to  $j$ ,  $p_{ij}$ , and the sum of all arc weights is one. Here the node weights  $\pi_i$  are equal to  $\sum_{(i, j) \in A} p_{ij}$ , the weight of outgoing arcs from node  $i$ :

$$\phi_G = \min_{S \subset V, \pi(S) \leq 1/2} \frac{C(S, \bar{S})}{\pi(S)}. \quad (10)$$

As mentioned in the introduction, conductance ratio arises in the context of Markov chains where an important goal is to establish the “mixing rate” of Markov chains. For a conductance ratio equal to  $\phi$ , the upper bound on the mixing rate is a function of  $1/\phi^2$ . For a discussion, proof of this result, and additional details on the problem the reader is referred to Sinclair (1992) and to Jerrum and Sinclair (1997).

- *Quantity-normalized cut*. The quantity normalized cut is a strict generalization of normalized cut. Instead of restricting the denominators to be the sum of the total weighted degrees of the nodes in the set, here each node has a quantity  $q_i$  associated with it, and the sum is of these quantities are not necessarily equal to the total weighted degrees:

$$q\text{-}NC_G = \min_{S \subset V} \frac{C(S, \bar{S})}{q(S)} + \frac{C(S, \bar{S})}{q(\bar{S})}.$$

- *Half-quantity-normalized cut*. This problem is the one term form of the quantity-normalized cut. Its relationship to the quantity-normalized cut is analogous to the relation of Cheeger to normalized cut:

$$\text{half-}q\text{-}NC_G = \min_{S \subset V, q(S) \leq q(V)/2} \frac{C(S, \bar{S})}{q(S)}. \quad (11)$$

- *Size-normalized cut*. This quantity is the “double form” of the expander problem and is used to bound the expander ratio of a graph. It can be viewed as an unweighted normalization version of the normalized cut problem where all node weights are equal to one. We thus call it “size normalized”:

$$\min_{S \subset V} \frac{C(S, S)}{|S|} + \frac{C(\bar{S}, \bar{S})}{|\bar{S}|}.$$

- *The uniform sparsest cut*.  $\min_{S \subset V} (C(S, \bar{S}) / (|S| \cdot |\bar{S}|))$ .

We claim that this problem is equivalent, and has the same optimal solution, as the size-normalized cut problem.

LEMMA 1. *The uniform sparsest cut has the same set of optimal solutions as the size-normalized cut.*

PROOF. To see the equivalence, note that

$$\begin{aligned} C(S, \bar{S}) \left[ \frac{1}{|S|} + \frac{1}{|\bar{S}|} \right] &= C(S, \bar{S}) \left[ \frac{|S|}{|S||\bar{S}|} + \frac{|\bar{S}|}{|S||\bar{S}|} \right] \\ &= n \cdot \frac{C(S, \bar{S})}{|S| \cdot |\bar{S}|}. \end{aligned}$$

Therefore any optimal solution to the size-normalized cut is an optimal solution to the uniform sparsest cut, and vice versa, the uniform sparsest cut optimal solution value is  $(1/n) \cdot s\text{-}NC_G$ . Q.E.D.

- *Normalized cut'*. Sharon et al. (2006) defined the following problem, calling it “normalized cut,” and stating it is NP-hard:

$$NC'_G = \min_{S \subset V} \frac{C(S, \bar{S})}{C(S, S)}. \quad (12)$$

The solution to this problem is a subset  $S$  that is as dissimilar as possible to  $\bar{S}$  and that also has similarity within, measured by  $C(S, S)$ , as large as possible. This problem was shown to be polynomial time solvable in Hochbaum (2010). This variant of normalized cut is equivalent to the

problem of minimizing one term,  $\min_{S \subset V} (C(S, \bar{S})/d(S))$ , in (9), as proved in the next lemma:

LEMMA 2 (HOCHBAUM 2010). *The sets of optimal solutions to  $\text{NC}'_G$  and  $\min_{S \subset V} (C(S, \bar{S})/d(S))$  are identical.*

PROOF.

$$\frac{C(S, \bar{S})}{C(S, S)} = \frac{C(S, \bar{S})}{d(S) - C(S, \bar{S})} = \frac{1}{(d(S)/C(S, \bar{S})) - 1}.$$

Therefore, minimizing this ratio is equivalent to maximizing  $d(S)/C(S, \bar{S})$ , which in turn is equivalent to minimizing the reciprocal quantity  $C(S, \bar{S})/d(S)$ , which is the first term in (9), as claimed. Q.E.D.

It is important to note that in problem  $\text{NC}'_G$  there is no balance restriction on the size of the set  $S$  or the total weight of  $C(S, S)$  as compared to  $C(V, V)$ . This is in fact the constraint that turns the polynomial problem into an NP-hard problem. With this constraint, the problem  $\min_{d(S) \leq d(V)/2} (C(S, \bar{S})/d(S))$  is the same as the half-normalized-cut problem and is NP-hard.

- *Weighted ratio region.* The unweighted version of this problem was introduced as “ratio region” by Cox et al. (1996). Cox et al. (1996) presented, for the unweighted version on planar graphs, a polynomial time algorithm. Hochbaum (2010) introduced the weighted version of the problem and presented a polynomial time algorithm solving the weighted and unweighted problems, for any graph, without restriction on its topology. For  $q_i$  the weight of node  $i$ ,  $q(S) = \sum_{i \in S} q_i$  the problem is

$$\min_{S \subset V} \frac{C(S, \bar{S})}{q(S)}.$$

This problem, with the balance constraint  $q(S) \leq \frac{1}{2}q(V)$ , added, is the half-quantity-normalized-cut problem.

The double forms of the problems are within a factor of two of the respective single-form problems as shown next. These double forms have been used (at least implicitly) by Cheeger (1970) and Chung (1997) for bounding Cheeger constant and the expander ratio of a graph, and explicitly here for the more general half-normalized cut in Lemma 3:

THEOREM 1. *For a graph  $G$ :*

1.  $\frac{1}{2} \cdot s\text{-NC}_G \leq \text{Exp}_G \leq s\text{-NC}_G$ ;
2.  $\frac{1}{2} \cdot \text{NC}_G \leq h_G \leq \text{NC}_G$ ;
3.  $\frac{1}{2} \cdot q\text{-NC}_G \leq \phi_G \leq q\text{-NC}_G$ .

PROOF. Let  $S^*$  be an optimal solution to  $\min_{S \subset V} (C(S, \bar{S}) \cdot [1/|S| + 1/|\bar{S}|])$ , the size-normalized cut problem. Then,

$$\begin{aligned} \text{Exp}_G &= \min_{S \subset V, |S| \leq |V|/2} \frac{C(S, \bar{S})}{|S|} \leq \frac{C(S^*, \bar{S}^*)}{|S^*|} \\ &\leq C(S^*, \bar{S}^*) \cdot \left[ \frac{1}{|S^*|} + \frac{1}{|\bar{S}^*|} \right] = s\text{-NC}_G. \end{aligned}$$

Now, let  $S^{\text{Exp}}$  be an optimal solution to  $\min_{S \subset V, |S| \leq |V|/2} (C(S, \bar{S})/|S|)$ , the expander ratio. Then,

$$\frac{1}{2}s - \text{NC}_G \leq \frac{1}{2} \cdot C(S^{\text{Exp}}, \bar{S}^{\text{Exp}}) \cdot \left[ \frac{1}{|S^{\text{Exp}}|} + \frac{1}{|\bar{S}^{\text{Exp}}|} \right].$$

Because  $|S^{\text{Exp}}| \leq |V|/2$ , the first term in the latter two terms sum is the larger one. Thus,  $\frac{1}{2}s\text{-NC}_G \leq C(S^{\text{Exp}}, \bar{S}^{\text{Exp}})/|S^{\text{Exp}}| = \text{Exp}_G$ . The proofs for 2 and 3 are similar. Q.E.D.

We conclude that the optimal solutions for the respective problems of size-normalized cut, normalized cut, and quantity-normalized cut, are 2-approximations to expander, Cheeger, and half-quantity-normalized cut, respectively. Because all these problems are NP-hard, any “good” approximation algorithms for the normalized cut problems would translate to a good approximation, with an extra factor of two, for expander, Cheeger, and half-quantity-normalized cut problems. Indeed, the spectral technique and the (equivalent; see Remark 1) Markov chain method used to generate solutions to expander, Cheeger, conductance, and the normalized cut problems, apply the spectral method to the corresponding Rayleigh problem rather than directly for the original formulation. The Rayleigh problem is equivalent to each of the normalized cut problems as shown in Lemma 3, and this accounts for the extra factor of two to the bounds in (3).

REMARK 1. Computing the principal eigenvector of a matrix  $A = (a_{ij})$  can be done by an iterative technique—the power method (Vargas 1962): For a given initial guess of the eigenvector  $\mathbf{x}_0$  (typically equal to  $\mathbf{1}$ ), this is a recursive procedure based on

$$\lim_{k \rightarrow \infty} \frac{A^k \mathbf{x}_0}{|A^k \mathbf{x}_0|} = \mathbf{x}^*. \quad (13)$$

The advantage of the use of this recursion is that it is inherently distributed and localized. Each iteration is implemented by following a walk from each node  $i$  to node  $j$  with probability  $a_{ij}/\sum_{p=1}^n a_{ip}$ . The count of the number of visits to each node after a number of iterations, is an estimate of the relative weight of the respective entry in the eigenvector. The drawbacks include several technical requirements on convergence conditions that often are not satisfied. An adaptation of the same technique is often used to find the Fiedler eigenvector in the context of the conductance problem.

The list of problems addressed here is summarized in Table 1.

### 3. The Normalized Cut Problems as Rayleigh Problems

For a real vector  $\mathbf{y} \in R^n$ , a matrix  $\mathcal{L} = D - W$ , and a diagonal matrix  $Q$ , we introduce the continuous relaxation of the Rayleigh problem as

$$\mathcal{Rr}(Q) = \min_{\mathbf{y}^T Q \mathbf{1} = 0} \frac{\mathbf{y}^T \mathcal{L} \mathbf{y}}{\mathbf{y}^T Q \mathbf{y}}.$$

**Table 1.** List of problems addressed here.

| Problem name                   | Objective  | Complexity  | Notation                                |
|--------------------------------|--|-------------|---|
| Expander                       | $\min_{S \subset V,  S  \leq  V /2} \frac{C(S, \bar{S})}{ S }$                     | NP-hard     | $\text{Exp}_G$ , half- $s\text{-NC}_G$  |
| Cheeger                        | $\min_{S \subset V, d(S) \leq d(V)/2} \frac{C(S, \bar{S})}{d(S)}$                  | NP-hard     | $h_G$ , half- $\text{NC}_G$             |
| Conductance                    | $\min_{S \subset V, \pi(S) \leq 1/2} \frac{C(S, \bar{S})}{\pi(S)}$                 | NP-hard     | $\phi_G$ , directed-half- $\text{NC}_G$ |
| Half-quantity – normalized cut | $\min_{S \subset V, q(S) \leq q(V)/2} \frac{C(S, \bar{S})}{q(S)}$                  | NP-hard     | half- $q\text{-NC}_G$                   |
| Size-normalized cut            | $\min_{S \subset V} \frac{C(S, \bar{S})}{ S } + \frac{C(S, \bar{S})}{ \bar{S} }$   | NP-hard     | $s\text{-NC}_G$                         |
| Normalized cut                 | $\min_{S \subset V} \frac{C(S, \bar{S})}{d(S)} + \frac{C(S, \bar{S})}{d(\bar{S})}$ | NP-hard     | $\text{NC}_G$                           |
| Quantity-normalized cut        | $\min_{S \subset V} \frac{C(S, \bar{S})}{q(S)} + \frac{C(S, \bar{S})}{q(\bar{S})}$ | NP-hard     | $q\text{-NC}_G$                         |
| Uniform-sparsest cut           | $\min_{S \subset V} \frac{C(S, \bar{S})}{ S  \cdot  \bar{S} }$                     | NP-hard     | $= s\text{-NC}_G$                       |
| Normalized cut'                | $\min \frac{C(S, \bar{S})}{C(S, V)}$   | Polynomial* | $\text{NC}'_G$                          |
| Ratio region                   | $\min \frac{C(S, \bar{S})}{ S }$   | Polynomial* | $\text{RR}_G$                           |
| Weighted ratio region          | $\min \frac{C(S, \bar{S})}{q(S)}$  | Polynomial* | $\text{WRR}_G$                          |

\*Polynomial time algorithms are given in Hochbaum (2010).

The discrete Rayleigh ratio problem,  $b\text{-DRR}$ , for a scalar  $b \geq 0$  is

$$R(Q, b) = \min_{y_i \in \{-b, 1\}} \frac{\mathbf{y}^T \mathcal{L} \mathbf{y}}{\mathbf{y}^T \mathbf{Q} \mathbf{y}}.$$

The *Rayleigh problem* is

$$\min_{y_i \in \{-b, 1\}} \mathcal{R}r(Q) = \min_{\mathbf{y}^T \mathbf{Q} \mathbf{1} = 0} R(Q, b) = \min_{\mathbf{y}^T \mathbf{Q} \mathbf{1} = 0, y_i \in \{-b, 1\}} \frac{\mathbf{y}^T \mathcal{L} \mathbf{y}}{\mathbf{y}^T \mathbf{Q} \mathbf{y}}.$$

We show next that for  $Q = I$ ,  $Q = D$ , and general  $Q$ , the Rayleigh problem is equivalent to the size-normalized cut, the normalized cut, and the quantity-normalized cut, respectively.

LEMMA 3. For a diagonal matrix  $Q$ , with  $Q_{ii} = q_i \geq 0$ ,

$$\min_{\mathbf{y}^T \mathbf{Q} \mathbf{1} = 0, y_i \in \{-b, 1\}} \frac{\mathbf{y}^T \mathcal{L} \mathbf{y}}{\mathbf{y}^T \mathbf{Q} \mathbf{y}} = \min_{\emptyset \neq S \subset V} \frac{C(S, S)}{q(S)} + \frac{C(\bar{S}, \bar{S})}{q(\bar{S})}.$$

PROOF. Let the variables  $y_i$  be binary with values  $-b$  or  $1$  defined as follows:

$$y_i = \begin{cases} 1 & \text{if } i \in S \\ -b & \text{if } i \in \bar{S}. \end{cases}$$

We first note that,  $\mathbf{y}^T \mathbf{Q} \mathbf{y} = q(S) + b^2 q(\bar{S})$ . Secondly, the orthogonality constraint,  $\mathbf{y}^T \mathbf{Q} \mathbf{1} = 0$  is equivalent to  $b = q(S)/q(\bar{S})$ . The latter is a form of a *balance* requirement on the two parts of the bipartition, which is why we refer to this constraint as the balance constraint:

$$\begin{aligned} \mathbf{y}^T \mathcal{L} \mathbf{y} &= \mathbf{y}^T \mathbf{D} \mathbf{y} - \mathbf{y}^T \mathbf{W} \mathbf{y} \\ &= \sum_{i \in S} d_i + b^2 \sum_{i \in \bar{S}} d_i \\ &\quad - [C(S, S) - 2bC(S, \bar{S}) + b^2 C(\bar{S}, \bar{S})] \\ &= C(S, S) + C(S, \bar{S}) + b^2 C(S, \bar{S}) + b^2 C(\bar{S}, \bar{S}) \\ &\quad - [C(S, S) - 2bC(S, \bar{S}) + b^2 C(\bar{S}, \bar{S})] \\ &= (1 + b^2 + 2b)C(S, \bar{S}) = (1 + b)^2 C(S, \bar{S}). \end{aligned}$$

Therefore,

$$\min_{\mathbf{y}^T \mathbf{Q} \mathbf{1} = 0, y_i \in \{-b, 1\}} \frac{\mathbf{y}^T \mathcal{L} \mathbf{y}}{\mathbf{y}^T \mathbf{Q} \mathbf{y}} = \min_{\mathbf{y}^T \mathbf{Q} \mathbf{1} = 0, S \subset V} \frac{(1 + b)^2 C(S, \bar{S})}{q(S) + b^2 q(\bar{S})}.$$

We now substitute  $\mathbf{y}^T \mathbf{Q} \mathbf{1} = 0$  by the equivalent expression  $b = q(S)/q(\bar{S})$ ,

$$\frac{(1 + b)^2 C(S, \bar{S})}{q(S) + b^2 q(\bar{S})} = \frac{(1 + q(S)/q(\bar{S}))^2 C(S, \bar{S})}{q(S) + (q(S)/q(\bar{S}))^2 q(\bar{S})}$$



$$\begin{aligned} &= \frac{(1 + q(S)/q(\bar{S}))^2 C(S, \bar{S})}{q(S)(1 + q(S)/q(\bar{S}))} \\ &= \frac{(1 + q(S)/q(\bar{S})) C(S, \bar{S})}{q(S)}. \end{aligned}$$

Hence,

$$\min_{\mathbf{y}^T \mathbf{Q} \mathbf{1} = 0, y_i \in \{-b, 1\}} \frac{\mathbf{y}^T \mathcal{L} \mathbf{y}}{\mathbf{y}^T \mathbf{Q} \mathbf{y}} = \min_{S \subset V} C(S, \bar{S}) \left[ \frac{1}{q(S)} + \frac{1}{q(\bar{S})} \right]$$

as claimed. Q.E.D.

**COROLLARY 1.** *The size-normalized cut, normalized cut, and quantity-normalized cut problems are equivalent to minimizing the discrete Rayleigh ratio on discretely valued variables  $y_i \in \{-b, 1\}$ ,  $b$ -DRR, and for  $Q = I$ ,  $Q = D$  and  $Q$  unrestricted, respectively.*

**COROLLARY 2.** *The normalized cut' and weighted ratio region problems are equivalent to minimizing  $R(Q, 0)$  for  $Q = D$  and  $Q$  unrestricted, respectively.*

**PROOF.** As shown in the proof of Lemma 3,  $\mathbf{y}^T \mathcal{L} \mathbf{y} = (1 + b)^2 C(S, \bar{S})$  which for  $b = 0$  is equal to  $C(S, \bar{S})$ . The denominator  $\mathbf{y}^T \mathbf{Q} \mathbf{y} = q(S) + b^2 q(T)$ , which is  $q(S)$  for  $b = 0$ . When  $Q$  is set equal to  $D$  we get the problem  $\min_{S \subset V} (C(S, \bar{S})/d(S))$ , which is equivalent to the normalized cut' problem,  $\min_{S \subset V} (C(S, \bar{S})/C(S, \bar{S}))$ . When  $Q$  is unrestricted we get the problem  $\min_{S \subset V} (C(S, \bar{S})/q(S))$ , which is the value of  $\text{WRR}_G$ . Q.E.D.

These corollaries then frame the results in Hochbaum (2010) as equivalent to solving the respective Rayleigh ratios for the value of  $b = 0$ . The DRR algorithm introduced here is more general and more efficient than the algorithm in Hochbaum (2010), as noted in §1.3. However, if we wish to minimize the ratio of the cut with respect to one set of weights, divided the similarity within cluster with a second set of weights,  $\min_{S \subset V} (C_1(S, \bar{S})/C_2(S, \bar{S}))$ , the transformation to the Rayleigh ratio problem does *not* apply. This problem is solved by the algorithm in Hochbaum (2010) but not by the more efficient algorithm described here.

## 4. The Spectral Relaxation

As the Rayleigh problem is NP-hard, one possible heuristic approach is to relax the problem by eliminating the requirements on the variables that  $y_i \in \{-b, 1\}$  and permitting each  $y_i$  to assume any real value. The *spectral relaxation* is to minimize the Rayleigh ratio  $\mathcal{R}r(Q)$ .

For  $Q$  nonnegative, the optimal solution to  $\mathcal{R}r(Q)$  is attained by setting the vector  $\mathbf{y}$  equal to the second smallest eigenvector solving  $\mathcal{L} \mathbf{y} = \lambda \mathbf{Q} \mathbf{y}$  for the smallest nonzero eigenvalue  $\lambda$ . This is done by solving  $Q^{-1/2} \mathcal{L} Q^{-1/2} \mathbf{z} = \lambda \mathbf{z}$  for the eigenvector  $\mathbf{z}$  that corresponds to the second smallest eigenvalue (the smallest eigenvalue is zero) and setting  $\mathbf{y} = Q^{-1/2} \mathbf{z}$ . To map this solution to a partition, a feasible

solution, one sets all the positive entries of  $y_i$  to one, and hence  $i$  is in  $S$ , and all the remaining ones are assigned to  $\bar{S}$ . Alternatively, another threshold value is chosen, and all values of  $y_i$  that exceed the threshold value are set to  $S$  and the others to  $\bar{S}$ .

Notice that Theorem 1 together with Lemma 3 imply immediately the lower bound on Cheeger constant,  $\lambda_1/2 \leq h_G$ . That is because the value of the relaxation can only be smaller than the value of the optimum of the Rayleigh problem.

## 5. The DRR Algorithm Solving the Discrete Rayleigh Ratio Problem

The DRR algorithm solves a relaxation of the Rayleigh problem resulting from omitting constraint  $\mathbf{y}^T \mathbf{D} \mathbf{1} = 0$  and specifying the value of  $b$ ,  $R(Q, b)$ . We call this problem *discrete Rayleigh relaxation*, or  $b$ -DRR, and denote its value by  $\alpha_G(b)$ :

$$\begin{aligned} (b\text{-DRR}) \quad \alpha_G(b) &= \min_{y_i \in \{-b, 1\}} \frac{\mathbf{y}^T \mathcal{L} \mathbf{y}}{\mathbf{y}^T \mathbf{Q} \mathbf{y}} \\ &= \min_{\emptyset \subset S \subset V} \frac{(1 + b)^2 C(S, \bar{S})}{q(S) + b^2 q(\bar{S})}. \end{aligned} \quad (14)$$

Because this is a relaxation of the Rayleigh problem we get, as before, from Theorem 1 and Lemma 3 that for  $\psi_G$  the optimal value of the half-quantity normalized cut,  $\alpha_G(b)/2 \leq \psi_G$ . Furthermore,

$$\max_{b \geq 0} \alpha_G(b)/2 \leq \psi_G.$$

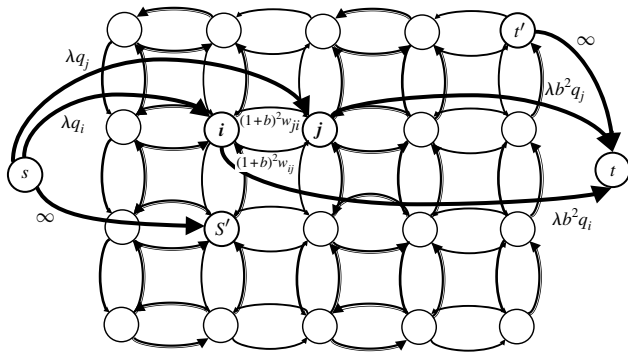
For the special case of Cheeger constant  $\max_{b \geq 0} \alpha_G(b)/2 \leq h_G$ .

To solve (14) we first “linearize” the ratio function: A general approach for minimizing a fractional (or as it is sometimes called, geometric) objective function over a feasible region  $\mathcal{F}$ ,  $\min_{x \in \mathcal{F}} (f(x)/g(x))$ , is to reduce it to a sequence of calls to an oracle that provides the yes/no answer to the following  $\lambda$ -question: Is there a feasible subset  $x \in \mathcal{F}$  such that  $(f(x) - \lambda g(x) < 0)$ ?

If the answer to the  $\lambda$ -question is *yes* then the optimal solution has a value smaller than  $\lambda$ . Otherwise, the optimal value is greater than or equal to  $\lambda$ . A standard approach is then to utilize a binary search procedure that calls for the  $\lambda$ -question  $O(\log(UF))$  times in order to solve the problem, where  $U$  is an upper bound on the value of the numerator and  $F$  an upper bound on the value of the denominator.

Therefore, if the linearized version of the problem, that is the  $\lambda$ -question, is solved in polynomial time, then so is the ratio problem. Note that the number of calls to the linear optimization is not strongly polynomial but rather, if binary search is employed, depends on the logarithm of the magnitude of the numbers in the input. In our case however there is a more efficient procedure.

**Figure 1.** The graph  $G_{st}^b$  illustrated for an image segmentation grid problem with 4-neighborhood set-up and node weights  $q_i$ .



The  $\lambda$ -question of whether the value of  $b$ -DRR is less than  $\lambda$  is equivalent to determining whether

$$\begin{aligned} \min_{y_i \in \{-b, 1\}} \mathbf{y}^T \mathcal{L} \mathbf{y} - \lambda \mathbf{y}^T \mathbf{D} \mathbf{y} \\ = \min_{S \subset V} (1+b)^2 C(S, \bar{S}) - \lambda [q(S) + b^2 q(\bar{S})] \end{aligned} \quad (15)$$

is negative. We next construct an  $s, t$  graph,  $G_{st}^b$ , which corresponds to  $G = (V, E)$  for a fixed value of  $b$  illustrated in Figure 1. We prove that the source set  $S$  of the minimum  $s, t$ -cut in  $G_{st}^b$  is an optimal solution to (15). Furthermore, the solution, for all values of  $\lambda$ , and thus for the optimal ratio is generated in the complexity of a single minimum  $s, t$ -cut.

The graph  $G_{st}^b$  is constructed as follows: We add a source node  $s$  and a sink node  $t$ . For each edge  $[i, j] \in E$  there is a pair of arcs  $(i, j), (j, i) \in A_{st}$  both with capacity  $(1+b)^2 w_{ij}$ . For each node  $i$ , there is an arc  $(s, i)$  of capacity  $\lambda q_i$ , and an arc  $(i, t)$  of capacity  $\lambda b^2 q_i$ . Two nodes  $s', t' \in V$  are designated as seed source and seed sink, respectively. This is done by assigning infinite capacity to  $(s, s')$  and to  $(t', t)$ . This assignment guarantees that in any feasible solution  $s'$  will be part of the source set  $S$  and  $t'$  will be part of the sink set  $\bar{S}$ . This will rule out the trivial solutions of  $S = \emptyset$  or  $S = V$ .

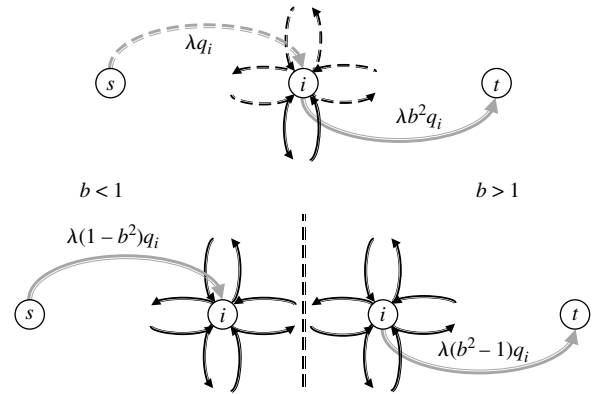
**THEOREM 2.** *The source set of a minimum cut in the graph  $G_{st}^b$  is an optimal solution to the linearized ( $b$ -DRR), (15).*

**PROOF.** Let  $(S \cup \{s\}, T \cup \{t\})$  be a partition of  $V \cup \{s, t\}$  corresponding to a finite capacity  $s, t$ -cut in  $G_{st}^b$ . We compute this cut's capacity:

$$\begin{aligned} C(S \cup \{s\}, T \cup \{t\}) &= \lambda q(T) + \lambda b^2 q(S) + C(S, T) \\ &= \lambda(1+b^2)q(V) - \lambda q(S) - \lambda b^2 q(T) \\ &\quad + C(S, T). \end{aligned}$$

Now the first term,  $\lambda(1+b^2)q(V)$ , is a constant. Thus minimizing  $C(S \cup \{s\}, T \cup \{t\})$  is equivalent to minimizing  $(1+b)^2 C(S, \bar{S}) - \lambda [q(S) + b^2 q(\bar{S})]$ , which is the objective of (15). Q.E.D.

**Figure 2.** Updated source and sink adjacent arcs' capacities.

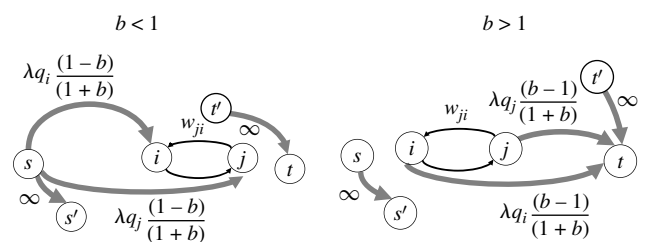


Graph  $G_{st}^b$  can be simplified: We distinguish between the cases where  $b > 1$  or  $b < 1$  or  $b = 1$ . When  $b > 1$  then all source adjacent finite capacity arcs are saturated, as the flow of  $\lambda q_i$  saturates the arc  $(s, i)$  and is below the capacity of  $(i, t)$ ,  $\lambda b^2 q_i$ . We can thus subtract the lower capacity from both resulting in an equivalent graph with sink adjacent capacities equal to  $\lambda(b^2 - 1)q_i$  and source adjacent capacities equal to zero (except for the seed  $s'$ ). Similarly, for  $b < 1$  we subtract  $\lambda b^2 q_i$  from the capacity of  $(s, i)$  and  $(i, t)$ . This results in no node adjacent to sink. We therefore choose a sink “seed” node  $u$  and connect it to  $t$  with infinite capacity arc. The capacity of the arcs from  $s$  to each node  $i$  is then  $\lambda(1 - b^2)q_i$  as illustrated in Figure 2. In the case  $b = 1$  the analogous update results in all source adjacent and sink adjacent arcs having zero capacity.

We next *scale* the graph arc weights by multiplying all arc capacities by a nonnegative scalar  $\alpha$ . For  $\alpha > 0$  a minimum cut of capacity  $\mathcal{C}$  in a graph  $G_{st}$  is also a minimum cut of capacity  $\alpha \mathcal{C}$  in the scaled graph  $\alpha \cdot G_{st}$ . Here we choose  $\alpha = 1/(1+b)^2$  as illustrated in Figure 3.

In the updated graph  $G_{st}$  either all source adjacent nodes are proportional to  $\lambda$  or all sink adjacent nodes are proportional to  $\lambda$ . This is therefore an instance of parametric minimum cut, where all the arcs adjacent to source are monotone nondecreasing with the parameter  $\lambda$  and all arcs adjacent to sink are monotone nonincreasing with the parameter  $\lambda$ . A parametric minimum cut procedure can then be applied to find the largest value of  $\lambda$  so that the  $\lambda$  question is answered in the affirmative. The running time of the procedure is the running time of solving for a single minimum cut, using the push-relabel algorithm as in Gallo

**Figure 3.** Scaling arc weights in  $G_{st}$ .



et al. (1989), or using the HPF algorithm, as in Hochbaum (2008). Within this run time, all *breakpoints* of the parameter  $\lambda$  are generated, where the cut is changing by at least a single node. It is well known that there are no more than  $n$  breakpoints, (see, e.g., Hochbaum 2008).

Because the normalized cut' problem is a special case with  $Q = D$  and  $b = 0$ , we get a polynomial time algorithms for the problem immediately. This running time is  $T(n, m)$  for a graph or digraph on  $n$  nodes and  $m$  edges or arcs. This run time improves on the running time of the polynomial time algorithm in Hochbaum (2010), which is based on a non-Rayleigh formulation and has running time of  $T(m, m)$ . For  $Q = D$  and  $b = 1$  the problem is the simple minimum cut problem (we leave this to the reader to verify).

### 5.1. Solving for All Values of $b$ Efficiently

For each value of  $b$  the optimal solution to the ratio problem may be different. Yet, any optimal solution for  $b$  corresponds to a bipartition from the same universal set of "breakpoint" solutions common to all values of  $b$ , as shown next.

To implement the parametric procedure efficiently, we choose a parameter  $\beta = \lambda \cdot ((1 - b)/(1 + b))$  for  $b < 1$  and  $\beta = \lambda \cdot ((b - 1)/(1 + b))$  for  $b > 1$ . There are no more than  $n$  breakpoints for  $\beta$ . There are  $l \leq n$  nested source sets of minimum cuts each corresponding to one of the  $l$  breakpoints. Given the values of  $\beta$  at the breakpoints,  $\{\beta_1, \dots, \beta_l\}$ , we can generate, for each value of  $b$ , all the breakpoints of  $\lambda^b$ . For  $b < 1$ ,  $\lambda_i^b = \beta_i \cdot ((1 + b)/(1 - b))$ . Consequently, by solving once the parametric problem for the parameter  $\beta$  we obtain simultaneously, all the breakpoint solutions for every value of  $b$ , in the complexity of a single minimum cut,  $T(n, m)$ .

The one breakpoint solution that minimizes the ratio for a given value of  $b$  depends on the value of  $b$ . But any scaling of the numerator or denominator of the ratio may give a different optimal solution that coincides with a different breakpoint. We contend that a ratio is a rather arbitrary form of weighing the two different objectives in the numerator and denominator, and the scalar multiplication of the numerator can change that relative weighing. We use a "good" selection of a breakpoint in the experimental study reported in §6, which is the one that gives the best value of the respective objective function (normalized cut, or  $q$ -normalized cut).

## 6. Experimental Results

An extensive experimentation was conducted comparing the performance of the combinatorial algorithm proposed here, versus that of the spectral technique, on image segmentation instances. A detailed report of the complete experimental study for image segmentation is given in Hochbaum et al. (2012). We sketch here some of the highlights of the experiments that evaluate the performance

of the algorithms in terms of how well they approximate the objective values of normalized cut and  $q$ -normalized cut. Hochbaum et al. (2012) also addressed the subjective visual quality of the segmentation, which we mention briefly below.

Because the spectral method's solution is dependent on the discretizing method applied to the continuous Fiedler eigenvector, we present here results for two different approaches. One is the *sweep technique*, where for each possible threshold cutoff of the eigenvector there is a corresponding bipartition and the objective value of that bipartition is evaluated. The solution of the spectral sweep technique ultimately reports the one bipartition solution with the best (smallest) value of the objective function. Here we consider the objective functions of normalized cut and quantity-normalized cut. A different technique, utilized by Shi, as reported in Yu and Shi (2003) and in what we refer to as Shi's code (2008), generates a bipartition from the Fiedler eigenvector, which is claimed to give a superior approximation to the objective value of the respective normalized cut. We used an implementation of the sweep technique and Shi's code as representing the performance of the spectral method. Because the default size of the images for Shi's code is  $160 \times 160$  the number of solutions compared by the sweep technique is 25,599.

For the DRR algorithm, we evaluate the solutions at all the breakpoints for the objective value of normalized cut and quantity normalized cut, and we select the one that has the smallest value. The number of breakpoints evaluated for all images in this experiment varied from 10 to 20.

The image database consists of 20 benchmark images from the Berkeley Segmentation Data Set and Benchmark (<http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/>). Figure 4 shows the 20 benchmark images used in the experimental study, referred to sequentially from image 1 to image 20.

The benchmark images used here consist of grayscale images. A color intensity value is associated with every pixel, represented as an integer in  $[0, 255]$  in MATLAB. This is normalized and mapped to  $[0, 1]$ . The similarity weight between a pair of pixel nodes is a function of the difference of their color intensities. For  $p_i$  and  $p_j$  the color intensities of two neighboring pixel nodes  $i$  and  $j$ , the *exponential similarity weight* is defined as

$$w_{ij} = e^{-\alpha|p_i - p_j|}, \quad (16)$$

where  $\alpha$  is a parameter that can be adjusted. In the experiments we set  $\alpha = 100$ .

The performance of the algorithms was tested with respect to the normalized cut and the  $q$ -normalized cut problems. For  $q$ -normalized cut the values of the node weights  $q_i$  were set equal to the *entropy* of the respective pixel. The entropy of an image is a measure of randomness in the image that can be used to characterize the texture of an image. In MATLAB, by default the local entropy value

**Figure 4.** The 20 benchmark images.



Source. <http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/>.

of a pixel is the entropy value of the 9-by-9 neighborhood around the pixel. In our experiment, the entropy of a pixel was computed directly via the MATLAB built-in function `entropyfilt`.

In sum, we present here the comparison of the DRR algorithm with the spectral method solution generated by Shi's code and with the sweep method. Both are applied to the Fiedler eigenvector output. The comparison of DRR algorithm with these two spectral methods is evaluated for instances generated by the 20 images in terms of achieving best value for normalized cut and for  $q$ -normalized cut. Altogether there are four sets of experiments applied to the 20 image graphs:

1. DRR algorithm versus Shi's spectral solution for normalized cut;
2. DRR algorithm versus Shi's spectral solution for  $q$ -normalized cut;
3. DRR algorithm versus sweep method solution for normalized cut;
4. DRR algorithm versus sweep method solution for  $q$ -normalized cut.

### 6.1. DRR Algorithm vs. Shi's Spectral Solution for Normalized Cut and for $q$ -Normalized Cut

Each solution generated is a bipartition, which is then plugged into the normalized cut (NC) objective function. The smaller the value, the better the approximation. Because DRR algorithm almost always delivers a better solution, the results are presented here in the form of the ratio of the value of the spectral solution NC value to the value of the DRR-algorithm's solution NC value. The larger

**Table 2.** The ratios of the normalized cut objective value of Shi’s code spectral solution to the normalized cut objective value of DRR algorithm.

|                        |                       |                       |                        |                       |
|------------------------|-----------------------|-----------------------|------------------------|-----------------------|
| Image 1<br>191.52741   | Image 2<br>34.945958  | Image 3<br>228.88261  | Image 4<br>2.3501553   | Image 5<br>522,467.35 |
| Image 6<br>45,242,414  | Image 7<br>757,898.1  | Image 8<br>800.08425  | Image 9<br>11.785929   | Image 10<br>357.12512 |
| Image 11<br>11,514.768 | Image 12<br>125.05640 | Image 13<br>4.8974465 | Image 14<br>1,417.0076 | Image 15<br>233.03002 |
| Image 16<br>13.482913  | Image 17<br>16.897142 | Image 18<br>345.39787 | Image 19<br>471.05938  | Image 20<br>6.5424435 |

the value of the ratio, the better is the improvement of DRR algorithm of the spectral method. Table 2 gives these ratios for the normalized cut objective value. For all images, the DRR algorithm performed better than the spectral method.

Table 3 provides the ratios of the values of the spectral solution divided by the value of the DRR solution for the  $q$ -normalized objective value. Here again, for all images, the DRR algorithm performed better than the spectral method, and the rate of improvement is dramatically higher than that for the respective normalized cut instances.

In Table 4 we summarize the extent of the improvement factors of DRR algorithm over the spectral method by providing the mean and the median values of those ratios.

The improvement ratios of the DRR algorithm over Shi’s code spectral solution are summarized as a bar chart on a logarithmic scale in Figure 5.

**6.2. DRR Algorithm vs. Sweep Method Solution for Normalized Cut and for  $q$ -Normalized Cut**

Recall that the sweep method checks the values, for all possible thresholds, of the bipartitions, and selects the solution that is best with respect to the objective function of normalized cut or  $q$ -normalized cut. Therefore the performance of the sweep method is expected to be better than that of the spectral method that does not choose among all possible threshold partitions. In Table 5 the ratios of the values of the sweep method solution divided by the value of the DRR algorithm solution are indeed smaller than those for the spectral method of Shi’s code, and there are instances where the sweep method improves on the DRR algorithm.

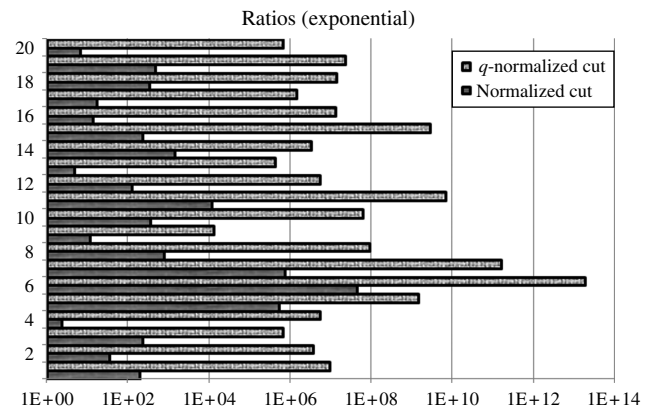
**Table 3.** The ratios of the  $q$ -normalized cut objective value of Shi’s code spectral solution to the  $q$ -normalized cut objective value of the DRR algorithm.

|                                       |                                       |                        |                         |                           |
|---------------------------------------|---------------------------------------|------------------------|-------------------------|---------------------------|
| Image 1<br>9,358,153.9                | Image 2<br>3,599,904.9                | Image 3<br>654,763.23  | Image 4<br>5,346,256.1  | Image 5<br>1,418,996,100  |
| Image 6<br>$1.8295880 \times 10^{13}$ | Image 7<br>$1.6418361 \times 10^{11}$ | Image 8<br>92,852,128  | Image 9<br>13,079.179   | Image 10<br>63,071,852    |
| Image 11<br>6,835,417,700             | Image 12<br>5,388,176                 | Image 13<br>431,894.04 | Image 14<br>3,397,368.9 | Image 15<br>2,755,701,700 |
| Image 16<br>13,524,963                | Image 17<br>1,440,666.1               | Image 18<br>14,317,766 | Image 19<br>22,693,749  | Image 20<br>681,058.90    |

**Table 4.** The means and medians of the improvements of the combinatorial algorithm on Shi’s code with exponential similarity weights.

|                     |                            |                        |
|---------------------|----------------------------|------------------------|
|                     | Mean of improvements       | Median of improvements |
| Normalized cut      | 2,326,927.7                | 230.95632              |
| $q$ -normalized cut | $9.2356550 \times 10^{11}$ | 11,441,558             |

**Figure 5.** Bar chart for the ratios in Table 2 and Table 3.



Note. The darker bars represent ratios for normalized cut (Table 2), and the lighter bars represent ratios for  $q$ -normalized cut (Table 3).

Although the sweep algorithm improves on the spectral method of Shi’s code, and is better in some instances than the DRR algorithm for the normalized cut problem, it is still not in the ballpark of the performance of the DRR algorithm for the  $q$ -normalized cut problem as seen in Table 6.

The means and medians of the improvement of DRR algorithm over the sweep method are given in Table 7. The respective means and medians of the improvement of the sweep method over DRR algorithm are given in Table 8.

Finally, we summarize in the bar chart in Figure 6 the ratios of improvements, as before. Here, the extent of improvement in instances where the sweep method improves on DRR algorithm are marked with the bars that extend to the left, and the bars that extend to the right indicate a better performance of the DRR algorithm.

**Table 5.** The ratios of the normalized cut objective value of the spectral sweep technique to the normalized cut objective value of the DRR algorithm.

|             |             |            |            |            |
|-------------|-------------|------------|------------|------------|
| Image 1     | Image 2     | Image 3    | Image 4    | Image 5    |
| 0.12676509  | 0.030815693 | 1.7318891  | 0.55478004 | 1.021992   |
| Image 6     | Image 7     | Image 8    | Image 9    | Image 10   |
| 32.769017   | 151,053.5   | 1.0124601  | 1.0757747  | 16.970223  |
| Image 11    | Image 12    | Image 13   | Image 14   | Image 15   |
| 3,229.7642  | 1.7051154   | 0.59025832 | 0.07758062 | 13.387584  |
| Image 16    | Image 17    | Image 18   | Image 19   | Image 20   |
| 0.031279526 | 1.2725133   | 1.6439209  | 0.56137852 | 0.22995850 |

**Table 6.** The ratios of the  $q$ -normalized cut objective value of the spectral sweep technique to the  $q$ -normalized cut objective value of the DRR algorithm.

|                |                |             |            |            |
|----------------|----------------|-------------|------------|------------|
| Image 1        | Image 2        | Image 3     | Image 4    | Image 5    |
| 1,612,971.1    | 295,255        | 222,060.29  | 4,904.4776 | 26,524,576 |
| Image 6        | Image 7        | Image 8     | Image 9    | Image 10   |
| 20,416,303,000 | 16,686,194,000 | 248,921.64  | 6,404.4956 | 23,558,163 |
| Image 11       | Image 12       | Image 13    | Image 14   | Image 15   |
| 2,097,520,100  | 1,344,474.1    | 80,934.898  | 921,433.37 | 25,749,626 |
| Image 16       | Image 17       | Image 18    | Image 19   | Image 20   |
| 54,403.132     | 216,141.34     | 1,191,186.4 | 9,983,867  | 70,300.882 |

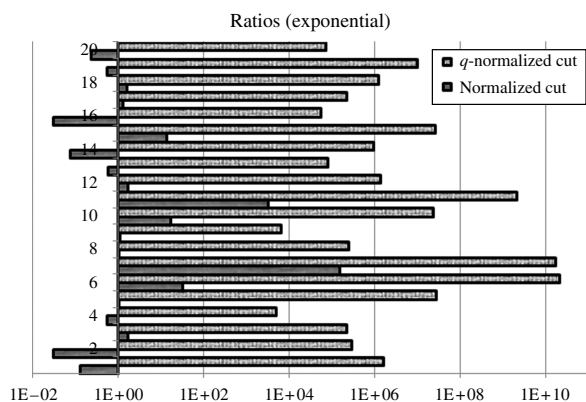
**Table 7.** The means and medians of the improvements of the combinatorial algorithm on the spectral sweep technique with exponential similarity weights.

|                     | Mean of improvements | Median of improvements |
|---------------------|----------------------|------------------------|
| Normalized cut      | 12,862.988           | 1.7185022              |
| $q$ -normalized cut | 1,964,605,100        | 1,056,309.9            |

**Table 8.** The means and medians of the improvements of the spectral sweep technique on DRR algorithm. Notice that for  $q$ -normalized cut, there is no improvement of the sweep method on DRR algorithm.

|                | Mean of improvements | Median of improvements |
|----------------|----------------------|------------------------|
| Normalized cut | 11.853231            | 6.1186089              |

**Figure 6.** Bar chart for the ratios in Table 5 and Table 6.



Note. The darker bars represent ratios for normalized cut (Table 5), and the lighter bars represent ratios for  $q$ -normalized cut (Table 6).

This set of experiments confirms that the DRR algorithm is not only more efficient than the spectral method, it also provides much better solutions than those provided by the spectral method.

In the extensive experimental study described in Hochbaum et al. (2012), the visual quality of the bipartitions, or segmentations, is investigated as well. The conclusion from this study is that the DRR algorithm is far superior to the sweep method in terms of visual quality, and it improves also on variants of Shi’s code that are designed specifically to deliver good visual quality.

## 7. Conclusions

We provide here for the first time an efficient combinatorial algorithm solving a discrete relaxation of a family of

NP-hard Rayleigh problems. Compared to the well-studied spectral approach that solves a continuous relaxation of these problems, the algorithm is efficient, combinatorial, and requires less storage space. Experimental results further demonstrate that in addition to being more efficient in practice, the proposed algorithm delivers solutions that are often much closer to the optimal objective value of the respective NP-hard problem as compared to the spectral technique.

## Acknowledgments

The author gratefully acknowledges the support of the National Science Foundation [Awards CMMI-1200592 and CBET-0736232].

## References

- Cheeger J (1970) A lower bound for the smallest eigenvalue of the Laplacian. Gunning RC, ed. *Problems Analysis* (Princeton University Press, Princeton, NJ), 195–199.
- Chung FRK (1997) *Spectral Graph Theory* (American Mathematical Society, Providence, RI).
- Cox IJ, Rao SB, Zhong Y (1996) “Ratio regions”: A technique for image segmentation. *Proc. Internat. Conf. Pattern Recognition* (IEEE Computer Society, Washington, DC), B:557–B:564.
- Ding C, He X, Zha H, Gu M, Simon H (2001) A min–max cut algorithm for graph partitioning and data clustering. *Proc. First IEEE Internat. Conf. Data Mining* (IEEE Computer Society, Washington, DC), 107–114.
- Fiedler M (1973) Algebraic connectivity of graphs. *Czechoslovak Math. J.* 23(98):298–305.
- Fiedler M (1975) A property of eigenvectors of nonnegative symmetric matrices and its applications to graph theory. *Czechoslovak Math. J.* 25(100):618–633.
- Filippone M, Camastra F, Masulli F, Rovetta S (2008) A survey of kernel and spectral methods for clustering. *Pattern Recognition* 41(1):176–190.
- Frobenius G (1912) Ueber Matrizen aus nicht negativen Elementen. *Sitzungsber. Königl. Preuss. Akad. Wiss.*, 456–477.
- Gallo G, Grigoriadis MD, Tarjan RE (1989) A fast parametric maximum flow algorithm and applications. *SIAM J. Comput.* 18(1):30–55.
- Goldberg AV, Tarjan RE (1988) A new approach to the maximum flow problem. *J. ACM* 35(4):921–940.
- Hall KM (1970) An  $r$ -dimensional quadratic placement algorithm. *Management Sci.* 17(3):219–229.
- Hochbaum DS (2008) The pseudoflow algorithm: A new algorithm for the maximum flow problem. *Oper. Res.* 58(4):992–1009.
- Hochbaum DS (2010) Polynomial time algorithms for ratio regions and a variant of normalized cut. *IEEE Trans. Pattern Anal. Machine Intelligence* 32(5):889–898.
- Hochbaum DS, Orlin JB (2013) Simplifications and speedups of the pseudoflow algorithm. *Networks* 61(1):40–57.
- Hochbaum DS, Lyu C, Bertelli E (2012) Evaluating the performance of image segmentation criteria and techniques. *Eur. J. Comput. Optim.* Forthcoming.
- Hoory S, Linial N, Wigderson A (2006) Expander graphs and their applications. *Bull. (New Series) Amer. Math. Soc.* 43(4):439–561.
- Hu H, Yan X, Huang Y, Han J, Zhou X (2005) Mining coherent dense subgraphs across massive biological networks for functional discovery. *Bioinformatics* 21(Suppl. 1):213–221.
- Jerrum M, Sinclair A (1997) The Markov chain Monte Carlo method: An approach to approximate counting and integration. Hochbaum DS, ed. *Approximation Algorithms for NP-Hard Problems*, Chap. 12 (PWS, Boston), 482–519.
- Keener JP (1993) The Perron-Frobenius theorem and the rating of football teams. *SIAM Rev.* 35(1):80–93.
- Koutis I, Miller GL, Peng R (2011) A nearly  $m \log n$  time solver for SDD linear systems. *Proc. Foundations Comput. Sci. 2011* (IEEE Computer Society, Washington, DC), 590–598.
- Leighton T, Rao S (1999) Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *J. ACM* 46(6):787–832.
- Narayanan M, Vetta A, Schadt E, Zhu J (2010) Simultaneous clustering of multiple gene expression and physical interaction data sets. *PLOS Comput. Biol.* 6(4). <http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1000742>
- Perron O (1907) Zur Theorie der Matrizen. *Mathematische Annalen* 64(2):248–263.
- Sharon E, Galun M, Sharon D, Basri R, Brandt A (2006) Hierarchy and adaptivity in segmenting visual scenes. *Nature* 442(7104):810–813.
- Shi J, Malik J (2000) Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Machine Intelligence* 22(8):888–905.
- Shi’s code (2008) MATLAB normalized cuts segmentation code. <http://www.cis.upenn.edu/~jshi/software/>.
- Sinclair A (1992) Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combinatorics, Probab. Comput.* 1(4): 351–370.
- Sinclair A, Jerrum M (1989) Approximate counting, uniform generation and rapidly mixing Markov chains. *Inform. Comput.* 82(1):93–133.
- Sipser M, Spielman DA (1996) Expander codes. *IEEE Trans. Inform. Theory* 42(6):1710–1722.
- Spielman DA (1996) Linear-time encodable and decodable error-correcting codes. *IEEE Trans. Inform. Theory* 42(6):1723–1731.
- Spielman DA (1999) Constructing error-correcting codes from expander graphs. *IMA Volumes Math. Appl.* (Springer Science + Business Media, LLC, Berlin), 109.
- Spielman DA, Teng SH (2004) Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. *ACM Sympos. Theory Comput.* (ACM, New York), 81–90.
- Tolliver DA, Miller GL (2006) Graph partitioning by spectral rounding: Applications in image segmentation and clustering. *Conf. Comput. Vision and Pattern Recognition* (IEEE Computer Society, Washington, DC), 313–319.
- Van Driessche R, Roose D (1995) An improved spectral bisection algorithm and its application to dynamic load balancing. *Parallel Comput.* 21(1):29–48.
- Vargas RS (1962) *Matrix Iterative Analysis* (Prentice Hall, Englewood Cliffs, NJ).
- Yu SX, Shi J (2003) Multiclass spectral clustering. *Proc. Ninth Internat. Conf. Comput. Vision* (IEEE Computer Society, Washington, DC), 313–319.

---

**Dorit S. Hochbaum** is a full professor and Chancellor Chair in the Department of Industrial Engineering and Operations Research at the University of California, Berkeley. Her research interests are in the areas of approximation algorithms, design and analysis of algorithms, and discrete and continuous optimization. Her recent work focuses on clustering, efficient techniques for network flow related problems, ranking, data mining, and image segmentation problems. In 2004 she received an honorary doctorate of sciences from the University of Copenhagen for her work on approximation algorithms. She is an INFORMS Fellow and the winner of the 2011 INFORMS Computing Society Prize for best paper dealing with the operations research/computer science interface.