Spectral Partitioning with Blends of Eigenvectors

James P Fairbanks   Geoffrey D Sanders   David A Bader

August 26, 2018

Abstract

Many common methods for data analysis rely on linear algebra. We provide new results connecting data analysis error to numerical accuracy in the context of spectral graph partitioning. We provide pointwise convergence guarantees so that spectral blends (linear combinations of eigenvectors) can be employed to solve data analysis problems with confidence in their accuracy. We apply this theory to an accessible model problem, the ring of cliques, by deriving the relevant eigenpairs and finding necessary and sufficient solver tolerances. Analysis of the ring of cliques provides an upper bound on eigensolver tolerances for graph partitioning problems. These results bridge the gap between linear algebra based data analysis methods and the convergence theory of iterative approximation methods. These results explain how the combinatorial structure of a problem can be recovered much faster than numerically accurate solutions to the associated linear algebra problem.

1   Introduction

Spectral methods are a valuable tool for finding cluster structure in data. While all spectral methods rely on approximating the eigenvectors of a matrix, the impact of numerical accuracy on the quality of the partitions is not fully understood. Spectral partitioning methods proceed in two steps, first one or more vectors approximating eigenvectors of a graph matrix are computed, and then a partitioning scheme is applied to those vectors. While many theoretical results quantify the relationship between the exact solution to the numerical problem and the solution to the original data mining problem, few address data analysis errors introduced by error in the numerical solution. For instance [9] studies the runtime and quality (in terms of conductance) of partitioning algorithms including spectral methods. Often the eigenvector computation is used as a primitive operation without accounting for the trade-off between run time and numerical accuracy. Guattery and Miller [6] studies various methods of applying exact eigenvectors to partition graphs by producing examples where each variation does not find the optimal cut. Our paper addresses the effect of numerical error in the eigenvector computation on the quality of sweep cuts which reveal graph structure.

1

In order to understand the impact of numerical error on spectral partitioning, we study both general matrices and a specific family of graphs. Finding error and residual tolerances for general graphs is a difficult problem. Section 2 provides tools for deriving a residual tolerance for arbitrary graphs. Section 3 analyzes a model problem with clear cluster structure, where linear combinations of eigenvectors represent a space of multiple good partitions, and applies Section 2 results to derive a residual tolerance sufficient for solving this model problem. This use of a model problem is well established in the linear algebra literature where the Laplace equation on a regular grid is common in papers and software regarding the solution of systems of equations. Analysis of this model problem allows us to derive a solver tolerance for correctly recovering the clusters with a sweep cut scheme. This analysis illustrates the difference between accurately solving the equation and correctly recovering the combinatorial structure.

This approach to approximate eigenvectors can be applied to other applications where a numerical method solves a data mining problem, such as solving personalized Pagerank as a linear system [3] to rank vertices in a graph, or evaluating commute times [4] to produce a metric distance on the vertices. These methods also apply numerical solvers to infer a combinatorial or data analysis structure from the graph. A similar treatment, in terms of a model problem, of these methods would benefit our understanding of the relationship between numerical accuracy and data analysis accuracy.

Here we introduce the necessary concepts of data analysis quality and eigensolver accuracy. For this work we focus on partitioning graphs to minimize conductance as defined below. For any $S \subset V$, $S \cup \bar{S} = V$ represents a cut of the graph. Define $\text{Vol}(S) = \sum_{i,j \in S} a_{i,j}$ as the total weight of the edges within S. And define $E(S, \bar{S}) = \sum_{i \in S, j \notin S} a_{i,j}$ as the total weight of edges with one vertex in S and one vertex in the complement of S. The conductance of a cut S is thus given by the formula [9]:

$$\phi(S) = \frac{E(S, \bar{S})}{\min(\text{Vol}(S), \text{Vol}(\bar{S}))}.$$

For any vector $\mathbf{x}$, represent the sweep cut of $\mathbf{x}$ at t as in Equation (1).

$$S = S_{\mathbf{x}}(t) = \{i \mid x_i > t\} \tag{1}$$

We denote by $\phi(\mathbf{x})$ the minimal conductance of a sweep cut of $\mathbf{x}$, that is $\min_t \phi(S_x(t))$. The conductance of the graph is defined as $\phi_G = \min_S \phi(S)$. If the graph has multiple partitions with conductance less than a value ψ, then the application might accept any of them.

The accuracy of a solution to the eigenvector problem can be measured in three quantities: Rayleigh quotient, error, and residual. Spectral methods for finding low-conductance partitions rely on computing vectors $\mathbf{x}$ and corresponding scalars λ that solve the equations $M\mathbf{x} = \lambda \mathbf{x}$ for some graph-associated matrix M. The Rayleigh quotient, $\mu = \mathbf{x}^t M \mathbf{x}$ is an approximation to the eigenvalue λ. The error $\|\mathbf{v} - \mathbf{x}\|$ where $\mathbf{v}$ is the closest exact solution

is not accessible to a solver in general. The solver can use the norm of the eigenresidual, $\|M\mathbf{x} - \mu\mathbf{x}\|$, to determine when to stop iterations. Throughout this paper $\|\cdot\|$ will be taken to mean the 2-norm with subscripts used to clarify when necessary. In order to practically use an eigensolver, one must choose a residual tolerance $\|M\mathbf{x} - \mu\mathbf{x}\| < \epsilon$ sufficient to ensure that the computed eigenvector is accurate enough to solve the application problem. This paper provides concrete residual tolerances for a specific model problem and provides tools for finding such tolerances for more general graphs.

We briefly summarize notation for various graph matrices. Let $\mathbf{1}$ be the all ones vector. If A is the adjacency matrix of a graph and D is the diagonal matrix whose entries are $d_{i,i} = (A\mathbf{1})_i$, then $L = D - A$ is the combinatorial Laplacian, and $\hat{L} = I - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ is the normalized Laplacian. Solutions of the generalized eigenvector problem $L\mathbf{y} = D\mathbf{y}$ are scaled solutions to the eigenvector problem $D^{-\frac{1}{2}}AD^{-\frac{1}{2}}\mathbf{x} = \mathbf{x}$ where the scaling is $\mathbf{x} = D^{\frac{1}{2}}\mathbf{y}$. We refer to $D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ as $\hat{A}$, and use the identity $\lambda_k(\hat{L}) = 1 - \lambda_{n-k}(\hat{A})$ to replace computations involving small eigenvalues of the normalized Laplacian matrix with computations involving large eigenvalues of the adjacency matrix.

Conductance is an appropriate measure of partition quality for spectral partitioning because of Cheeger's inequality which bounds the conductance of the graph in terms of the eigenvalues of the Laplacian matrix.

Theorem 1. *General Cheeger Inequality[11] If $\mathbf{x}$ is a unit vector orthogonal to $D^{\frac{1}{2}}\mathbf{1}$ such that $\mathbf{x}^T\hat{L}\mathbf{x} = \mu$ then $D^{-\frac{1}{2}}\mathbf{x}$ has a sweep cut S such that $\phi(S) = \phi(\mathbf{x}) \leq \sqrt{2\mu}$.*

When $\mathbf{x}$ satisfies $\hat{L}\mathbf{x} = \lambda_2\mathbf{x}$, $\phi_G \leq \phi\left(D^{-\frac{1}{2}}\mathbf{x}\right) \leq \sqrt{2\lambda_2(L)}$. This general form of Cheeger's inequality indicates that finding low-energy Laplacian eigenvectors is sufficient for constructing low-conductance partitions of the graph.

In graph partitioning, the goal is to compute a partition of the graph that optimizes the chosen objective. When applying spectral methods to graph partitioning, our goal is not to compute very accurate eigenpairs, but instead to partition the vertex set of a graph correctly. Because spectral partitioning can be used recursively to find small clusters, we focus on splitting a graph into two parts. Our results on the model problem indicate that approximate eigenvectors are sufficient to solve the data analysis problem and are much faster to compute if the graph has the right structure.

1.1 A Model Problem

We use a simple model (the ring of cliques) to study the capabilities of spectral partitioning algorithms, form theory to characterize performance, and potentially enhance these algorithms. Such use of model problems is well-established in the numerical analysis literature regarding iterative solutions to discretized partial differential equations (PDEs). The Dirichlet Laplacian on a unit square discretized on a Cartesian lattice is a simple problem with known eigenpairs and is used to study the properties of various eigensolvers. These

simple model problems do not demonstrate the algorithms perform well on real-world problems, but are incredibly important tools for algorithm development and theoretical analysis. For spectral partitioning, the ring of cliques is one candidate model problem for which we can derive complete knowledge of the eigenpairs. In PDEs, the order of discretization error (difference between continuous solution and discrete solution) provides the solver with a stopping criterion. In spectral graph partitioning, we do not have this luxury, and we must develop theory to understand how perturbations in a spectral embedding impact partitioning quality. Another reason to develop a collection of model problems is to enable careful study of this impact in well-understood situations.

In order to provide a striking example of our improved analysis Section 3 studies our model problem in detail. The goal is to understand when approximate eigenvectors have sweep cuts that correctly identify graph structures. The ring of cliques has been studied as "the most modular network" in order to demonstrate a resolution limit in the modularity maximization procedure for community detection [5]. For this family of highly structured graphs, the correct partition is unambiguous. We use the ring of cliques to investigate how spectral embeddings for partitioning are affected by numerical error. Because of the high degree of symmetry, the ring of cliques allows for a thorough closed form analysis producing formulas for the eigenvectors and eigenvalues. A sweep cut using exact eigenvectors partitions the graph with small conductance and successful recovery of all the clusters. We quantify the effects of approximation error on sweep cut partitions of this graph. Our findings demonstrate that despite a small spectral gap, which implies slow convergence of non-preconditioned eigensolvers, the ring of cliques is well partitioned by low accuracy approximations to the eigenvectors.

Studying the ring of cliques provides guidance for practitioners on useful tolerances for eigensolvers. We are able to construct the smallest perturbation that induces a mistake in the sweep cut partition. This perturbation shows that when looking for clusters of size b in a general graph the eigensolver tolerance must be smaller than $\mathcal{O}\left(b^{-\frac{1}{2}}\right)$. Analysis of the ring of cliques provides an upper bound on the eigensolver accuracy that is sufficient to recover community structure.

1.2   Contributions

This paper provides the following contributions. Section 2 extends a known error bound on the computation of eigenvectors to the computation of linear combinations of eigenvectors. By extending this classical error bound to linear combinations of eigenvectors, we find a condition on the spectrum of where numerically accurate blends are easy to achieve. 4 provides a general condition under which approximate eigenvectors preserve sweep cuts. Section 3 analyzes a model problem and derives necessary and sufficient error tolerances for solving the model problem, which are essentially tight for some parameter regime. We show for the model problem where the number of clusters is polynomial in the size of the clusters, the power method takes $\mathcal{O}(1)$ iterations to identify the clusters.

4

1.3 Related Work

Iterative methods for solving eigenvector problems are well understood. These algorithms are able to generate solutions to arbitrary approximation factors, but have run time which increases in the number of iterations, where more iterations leads to better approximations. Iterative methods [10, 14] have been shown to provide fast approximate solutions for a wide range of problems. Many iterative eigensolvers can be represented as the output $\mathbf{y}$ equals a polynomial p applied to the matrix M times a vector $\mathbf{x}$, $\mathbf{y} = p(M)\mathbf{x}$. The degree of p depends on the number of iterations of the method, which is controlled by the eigenresidual tolerance $\|M\mathbf{y} - \mu\mathbf{y}\| < \epsilon$. The simplest such method is the power method which is easy to analyze because $p(M)$ is always M^k where k is the number of iterations. More sophisticated methods choose $p(M)$ adaptively and typically converge more quickly. A practical implementation of the Arnoldi method can be found in [10], which is commonly used in practice.

Localized eigenvectors are essential to analysis of the ring of cliques. Cucuringu and Mahoney examine the network analysis implications of localized interior eigenvectors in the spectrum of the Co-voting network of US Senators [2]. The graph is defined with connections between members of the same session of Congress who vote together on the same bills and connections between individuals who are reelected to consecutive sessions. The first 41 eigenvectors are oscillatory across the congressional sessions with little variation between the vertices in the same session, but the next eigenvectors are small in magnitude on most sessions but take large positive values on members of one party and large negative values on members of the other party within a few sessions. Thus blends of the dominant eigenvectors indicate the sessions of congress. The ring of cliques also exhibits globally periodic extremal eigenvectors and localized interior eigenvectors due to its Kronecker product structure. We show that the ring of cliques, has a basis for an interior eigenspace with the nonzero entries of each vector completely restricted to an individual clique. This localization allows us to show that approximate eigenvectors recover the interesting combinatorial structure.

Other work focuses on the impact of errors in measurement on the behavior of data analysis algorithms. In the context of Gram (kernel) matrices, Huang et al.[8], studies the effect of perturbing the original data points on the spectral partitioning method. A similar line of investigation is pursued in [18] where data points are quantized to reduce bandwidth in a distributed system. This work connects approximation with performance. If one can demonstrate that data analysis accuracy is not affected too much, then one can use an algorithm which sacrifices accuracy to improve performance. Our paper treats the data as correctly observed and handles error in the iterative solver.

The impact of approximate numerical computing has been shown useful for several applications. In [1] eigenvectors of a kernel matrix are approximated with the power method and then k-means is applied to these approximations. The k-means objective function is well approximated when using approximate eigenvectors. The bounds given in [1] de-

pend on using the k eigenvectors to partition into k parts and depend on the kth spectral gap to control accuracy of approximation. Experiments also show that k-means on the approximate eigenvectors is faster and sometimes more accurate in terms of Normalized Mutual Information (NMI) compared to using exact eigenvectors. Our paper focuses on partitioning into two clusters based on sweep cuts of a single approximate eigenvector and makes a rigorous analysis of a model problem in order to understand how the numerical accuracy interacts with combinatorial structure of the data clusters. Pothen et al. [13], which used spectral partitioning for distributed memory sparse matrix computation, recognized the value of low-accuracy solutions. Approximate spectral coordinates are used to reorder matrices before conducting high accuracy linear solves. Our paper contributes to the understanding of how numerical approximation accuracy contributes to data analysis accuracy.

2 Blends of Eigenvectors

In order to understand the relationship between eigensolver error and graph partitioning, we study error bounds and the effect of pointwise error on the sweep cut procedure. Theorems 2 and 3 bound the error to a subspace in terms of the residual and quantities derived from the eigenvalues of the matrix. This control over the error is then used in 4 to relate eigenresidual to the conductance of a sweep cut of the graph. These results apply to general matrices. Although a small spectral gap implies poor control on the error to a single eigenvector, we derive a condition where low accuracy approximations effectively partition the graph. Section 3 applies these theorems to a special family of graphs to show that blends are faster to compute and provide nearly optimal partitions.

2.1 Converging to a Single Eigenspace

Let $M \in \mathbb{R}^{n \times n}$, $M = M^t$ be a general symmetric matrix. Consider the solutions to the equation $M\mathbf{v} = \lambda \mathbf{v}$. Because M is symmetric, there are n eigenvalues in $\mathbb{R}$ (counting multiplicities). The set of all eigenvalues is the spectrum $\lambda(M)$, which we order decreasingly as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. For $k = 1, \ldots, n$, let $\mathbf{v}_k$ be an eigenvector associated with λ_k, $M\mathbf{v}_k = \lambda_k \mathbf{v}_k$, such that $\mathbf{v}_k^t \mathbf{v}_l = 0$ whenever $l \neq k$. Define the eigenspace associated with λ_k as the invariant subspace associated with λ_k, that is

$$ \mathcal{X}_k := \left\{ \, \mathbf{x} \in \mathbb{R}^n \, : \, M\mathbf{x} = \lambda_k \mathbf{x} \, \right\}. $$

These definitions imply $\dim(\mathcal{X}_k) = \mathrm{mult}(\lambda_k)$ and $\mathcal{X}_k = \mathcal{X}_l$ when $\lambda_k = \lambda_l$.

Remark 1. *The results in this section are stated and proved in terms of generic symmetric matrix M because they apply beyond spectral graph theory. Spectral partitioning methods use the eigenvectors of $\hat{L} = I - \hat{A}$. Counting eigenvalues from largest to smallest starting with 1, we see $\lambda_2(\hat{A}) = \lambda_{n-1}(L)$ with the same eigenvectors. Letting $\mathbf{k} = D^{1/2} \mathbf{1} \left\| D^{1/2} \mathbf{1} \right\|^{-1}$,*

the normalized eigenvector of $\hat{A}$ associated with $\lambda_1(\hat{A}) = 1$, one can use $M = \hat{A} - \mathbf{k}\mathbf{k}^t$ in the results of this section. Subtracting $\mathbf{k}\mathbf{k}^t$ moves the eigenvalue 1 to 0, or $\mathbf{k}$ is an eigenvector associated with $0 \in \lambda(\hat{A} - \mathbf{k}\mathbf{k}^t)$. Thus, for this M and for k where $\lambda_k(M) > 0$, we have

$$\lambda_k(M) = \lambda_{k+1}(\hat{A}).$$

In particular, for the Fiedler eigenvalue, $\lambda_1(M) = \lambda_2(\hat{A}) = 1 - \lambda_{n-1}(\hat{L})$, and the Fiedler vectors in the associated eigenspace correspond to extremal eigenvalue $\lambda_1(M)$. Computationally, implementations of iterative methods approximating the eigenvectors of $\lambda_2(\hat{A})$ perform better with a routine applying the operator $\hat{A} - \mathbf{k}\mathbf{k}^t$.

Let $(\mathbf{x}, \mu)$, $\mathbf{x} \in \mathbb{R}^n$, $\mu \in \mathbb{R}$, be an approximate eigenpair of M with $\|\mathbf{x}\| = 1$ and $\mu = \mathbf{x}^T M \mathbf{x}$, the *Rayleigh quotient*, which minimizes the function $\|M\mathbf{x} - \theta\mathbf{x}\|$ over all real values θ. Define the two-norm of the eigenresidual as $\epsilon = \|M\mathbf{x} - \mu\mathbf{x}\|$. As in [12], we have a simple eigenvalue bound. By decomposing $\mathbf{x}$ in the eigenbasis $\mathbf{x} = \sum_{k=1}^{n} \alpha_k \mathbf{v}_k$, we see

$$\epsilon^2 = \|M\mathbf{x} - \mu\mathbf{x}\|^2 = \sum_{k=1}^{n} \alpha_k^2 (\lambda_k - \mu)^2 \geq \left(\sum_{k=1}^{n} \alpha_k^2\right) \left(\min_{1 \leq k \leq n} (\lambda - \mu)^2\right)$$

meaning there exists an eigenvalue λ_k within ϵ of μ,

$$\min_{1 \leq k \leq n} |\lambda_k - \mu| \leq \epsilon.$$

Also in [12], we have bounds estimating convergence to an eigenspace in angle. Define the eigengap for λ_k as $\delta_k = \min_{\lambda \in \lambda(M) \setminus \lambda_k} |\lambda_k - \lambda|$. Moreover, if ϵ is small compared to δ_k there exists a normalized eigenvector $\mathbf{v} \in \mathcal{X}_k$ with which $\mathbf{x}$ has a small angle,

$$\min_{\mathbf{v} \in \mathcal{X}_k} \sqrt{1 - \langle \mathbf{x}, \mathbf{v} \rangle^2} \leq \frac{\epsilon}{\delta_k}.$$

Instead of presenting a proof of this well-known result, we derive a similar bound for ℓ_2 and point-wise approximation to an eigenspace associated with an extremal eigenvector.

Theorem 2. *Consider approximate eigenpair $(\mathbf{x}, \mu)$ of symmetric $M \in \mathbb{R}^{n \times n}$ with $\|\mathbf{x}\| = 1$ and $\mu = \mathbf{x}^T M \mathbf{x}$. Assume*

$$|\mu - \lambda_1| < \min_{\lambda \in \lambda(M) \setminus \lambda_1} |\mu - \lambda|.$$

Given eigenresidual ϵ and eigengap δ_1, there exists an eigenvector $\mathbf{v} \in \mathcal{X}_1$, with $\|\mathbf{v}\| = 1$, and error bound

$$\|\mathbf{x} - \mathbf{v}\| \leq \frac{\sqrt{8}\epsilon}{\delta_1}.$$

Proof. Let $\alpha\mathbf{v}$ be the closest vector in $\mathcal{X}_1$ to $\mathbf{x}$. Decompose $\mathbf{x}$ into its eigenvector components within $\mathcal{X}_1$ and perpendicular to $\mathcal{X}_1$, $\mathbf{x} = \alpha\mathbf{v} + \sum_{\mathbf{v}_k \perp \mathcal{X}_1} \alpha_k \mathbf{v}_k$. Because μ is closer to λ_1 than any other eigenvalue, we have

$$\epsilon^2 \;=\; |\lambda_1 - \mu|^2 \alpha^2 + \sum_{\mathbf{v}_k \perp \mathcal{X}_k} |\lambda_k - \mu|^2 \alpha_k^2 \;\geq\; \sum_{\mathbf{v}_k \perp \mathcal{X}_k} |\lambda_k - \mu|^2 \alpha_k^2 \geq \frac{\delta_1^2}{4} \sum_{\mathbf{v}_k \perp \mathcal{X}_k} \alpha_k^2.$$

Rearranging gives

$$\frac{4\epsilon^2}{\delta_1^2} \;\geq\; \sum_{\mathbf{v}_k \perp \mathcal{X}_k} \alpha_k^2 = 1 - \alpha^2.$$

$$\|\mathbf{x} - \mathbf{v}\|^2 \;=\; \|\mathbf{x} - \alpha\mathbf{v}\|^2 + \|\mathbf{v} - \alpha\mathbf{v}\|^2 \leq \|\mathbf{x} - \alpha\mathbf{v}\|^2 + (1 - \alpha)^2 \leq 2(1 - \alpha^2) \leq \frac{8\epsilon^2}{\delta_1^2},$$

$\square$

This result implies,
$$\frac{1}{n} \sum_{i \in \mathcal{V}} |x_i - v_i|^2 \leq \|\mathbf{x} - \mathbf{v}\|^2 \leq \frac{8\epsilon^2}{n\delta_1^2}$$

so if ϵ^2/n is small compared to the δ_1^2, then the average average error squared is also small. Moreover we have a point-wise error bound,

$$\max_{i \in \mathcal{V}} |x_i - v_i| \leq \|\mathbf{x} - \mathbf{v}\| \leq \frac{\sqrt{8}\epsilon}{\delta_1}$$

For a large graph, it is typical that Fiedler eigenvalue is so close to the next-to-largest eigenvalue, that the error bounds demand an extremely small eigenresidual for convergence. Note that this error analysis is independent of the algorithm used to compute the solution. Choosing $\mathbf{x} = \mathbf{v}_3$ shows that the condition $|\mu - \lambda_1| < \min_{\lambda \in \lambda(M) \setminus \lambda_1} |\mu - \lambda|$ is necessary. Thus for matrices M with small δ_1, we need to remove this condition. When the δ_1 is small, a reasonable number of iterations of an eigensolver may produce a vector $\mathbf{x}$ with $\mathbf{x}^T M \mathbf{x}$ close to λ_1 which may not be very close to the true extremal eigenspace. In this case we examine convergence to a linear combination of eigenvectors associated with a range of eigenvalues.

2.2 Converging to a Subspace Spanned by Multiple Eigenspaces

This section generalizes the previous error bound to the distance between $\mathbf{x}$ and a subspace spanned by the eigenvectors associated with a range of eigenvalues. Assume that linear combinations of eigenvectors associated with a range of eigenvalues $[\lambda_q, \lambda_p]$ are satisfactory for subsequent data analysis algorithms. If the Rayleigh quotient is within $[\lambda_q, \lambda_p]$ and the eigenresidual is smaller than the distance between μ and any eigenvalue outside $[\lambda_q, \lambda_p]$, then the following theorem holds.

Theorem 3. *Consider approximate eigenpair $(\mathbf{x}, \mu)$ of symmetric $M \in \mathbb{R}^{n \times n}$ with $\|\mathbf{x}\| = 1$ and $\mu = \mathbf{x}^t M \mathbf{x} \in [\lambda_q, \lambda_p]$. Define*

$$\delta_{p,q}(\mu) = \min_{\{k<p\} \cup \{k>q\}} |\lambda_k - \mu| \qquad and \qquad \mathcal{X}_p^q := \bigotimes_{k=p}^{q} \mathcal{X}_k.$$

Given eigenresidual $\epsilon = \|M\mathbf{x} - \mu\mathbf{x}\| \leq \min(\mu - \lambda_{q-1}, \lambda_{p+1} - \mu)$ there exists a vector $\mathbf{v} \in \mathcal{X}_p^q$, with $\|\mathbf{v}\| = 1$, ℓ_2 error bound,

$$\|\mathbf{x} - \mathbf{v}\| \leq \frac{\sqrt{2}\epsilon}{\delta_{p,q}(\mu)},$$

Proof. Let $\mathbf{x} = \sum_{k=1}^{n} \alpha_k \mathbf{v}_k$ and $\Pi\mathbf{x} = \sum_{k=p}^{q} \alpha_k \mathbf{v}_k$, the ℓ_2-orthogonal projection onto $\mathcal{X}_p^q$. Note $\sum_{k=1}^{n} \alpha_k^2 = 1$. In the case where $\|\Pi\mathbf{x}\| = 1$, we can let $\mathbf{v} = \mathbf{x}$ and see the bound is clearly satisfied. For $\|\Pi\mathbf{x}\| < 1$, we first demonstrate that $\|\mathbf{x} - \Pi\mathbf{x}\|$ is controlled by ϵ,

$$\epsilon^2 = \|M\mathbf{x} - \mu\mathbf{x}\|^2$$

$$\epsilon^2 = \sum_{k=1}^{n} \alpha_k^2 |\lambda_k - \mu|^2$$

$$\epsilon^2 = \sum_{\{k<p\} \cup \{k>q\}} \alpha_k^2 |\lambda_k - \mu|^2 + \sum_{k=p}^{q} \alpha_k^2 |\lambda_k - \mu|^2$$

$$\epsilon^2 \geq \delta_{p,q}(\mu)^2 \sum_{\{k<p\} \cup \{k>q\}} \alpha_k^2 + \sum_{k=p}^{q} \alpha_k^2 |\lambda_k - \mu|^2$$

$$\epsilon^2 \geq \delta_{p,q}(\mu)^2 \sum_{\{k<p\} \cup \{k>q\}} \alpha_k^2$$

$$\epsilon^2 \geq \delta_{p,q}(\mu)^2 \left(1 - \sum_{k=p}^{q} \alpha_k^2\right)$$

$$\frac{\epsilon}{\delta_{p,q}(\mu)} \geq \|\mathbf{x} - \Pi\mathbf{x}\|.$$

There is a unit vector in $\mathcal{X}_p^q$ that is also within some factor of ϵ to $\mathbf{x}$. Let $\mathbf{v} = \Pi\mathbf{x}/\|\Pi\mathbf{x}\|$, then $\|\mathbf{v} - \Pi\mathbf{x}\| = 1 - \|\Pi\mathbf{x}\|$. We have

$$
\begin{aligned}
\|\Pi\mathbf{x}\|^2 + \|(I - \Pi)\mathbf{x}\|^2 &= 1 \\
\|\Pi\mathbf{x}\|^2 &= 1 - \|(I - \Pi)\mathbf{x}\|^2 \\
\|\Pi\mathbf{x}\|^2 &\geq 1 - \frac{\epsilon^2}{\delta_{p,q}(\mu)^2} \\
\|\Pi\mathbf{x}\| &\geq \sqrt{1 - \frac{\epsilon^2}{\delta_{p,q}(\mu)^2}}.
\end{aligned}
$$

Using the inequality $a \leq \sqrt{a}$ for $a \in (0,1)$, we see

$$
\|\mathbf{v} - \Pi\mathbf{x}\| = 1 - \|\Pi\mathbf{x}\| \leq 1 - \sqrt{1 - \frac{\epsilon^2}{\delta_{p,q}(\mu)^2}} \leq \frac{\epsilon}{\delta_{p,q}(\mu)}.
$$

Then, because $(\mathbf{x} - \Pi\mathbf{x})^t(\Pi\mathbf{x} - \mathbf{v}) = 0$, we have

$$
\|\mathbf{x} - \mathbf{v}\|^2 = \|\mathbf{x} - \Pi\mathbf{x}\|^2 + \|\Pi\mathbf{x} - \mathbf{v}\|^2 \leq \frac{2\epsilon^2}{\delta_{p,q}(\mu)^2}.
$$

$\square$

Remark 2. *Note that the size of the* blend *gap, $\delta_{p,q}(\mu)$, is dependent on i) the size of $|\lambda_p - \lambda_q|$, ii) how internal Rayleigh quotient μ is within $[\lambda_q, \lambda_p]$, iii) and how far the exterior eigenvalues are, $|\lambda_q - \lambda_{q+1}|$ and $|\lambda_{p-1} - \lambda_p|$. For a problem where the spectrum is not known a priori, it is difficult to state an acceptable interval $[\lambda_q, \lambda_p]$ for accomplishing a given data mining task. Section 3.2 provides an example where one can choose $[\lambda_q, \lambda_p]$ a priori. The Congress graph has $\lambda_{41} - \lambda_{42} \geq 0.4$ and the first 41 eigenvectors indicate the natural clustering of the graph into sessions[2]. This analysis thus applies to this real world network.*

For our application, $p = 1$ and $\delta_{pq} = \mu - \lambda_{q+1}\left(\hat{A} - \mathbf{k}\mathbf{k}^t\right) \geq \lambda_q - \lambda_{q+1}$, which can be much larger than the spectral gap $\lambda_1(\hat{A} - \mathbf{k}\mathbf{k}^t) - \lambda_2(\hat{A} - \mathbf{k}\mathbf{k}^t)$. In Section 2.1 the goal of computation is a single eigenvector and the output of the approximation is a blend of eigenvectors, the coefficients of the output in the eigenbasis of the matrix describes the error introduced by approximation. In Section 2.2 the goal of computation is a blend of eigenvectors, and we improve the error bound when the spectral gap is small.

In order to relate the numerical accuracy to the conductance for general graphs we examine the impact of pointwise error on sweep cuts. For any prescribed conductance value ψ, we derive a condition on vectors $\mathbf{v}$ such that we can guarantee that small perturbations of $\mathbf{v}$ have conductance less than or equal to ψ. Let $S_{\mathbf{v}}(t)$ represent the sweep cut of $\mathbf{v}$ at t as in Equation (1).

Lemma 1. *For any graph G, vector $\mathbf{v} \in \mathbb{R}^n$ and scalar $\psi > 0$, define $T_\psi(\mathbf{v}) = \{t \mid \phi(S_\mathbf{v}(t)) \le \psi\}$. Let $g_\mathbf{v}(t) = \min_i |v_i - t|$ and $g_\mathbf{v} = \max_{t \in T_\psi(\mathbf{v})} g_\mathbf{v}(t)$. If $\|\mathbf{z}\|_\infty < g_\mathbf{v}$, then $\phi(\mathbf{v} + \mathbf{z}) \le \psi$.*

Proof. If $S_\mathbf{v}(t) = S_{\mathbf{v}+\mathbf{z}}(t)$ we can apply $\phi(\mathbf{v} + \mathbf{z}) < \phi(S_{\mathbf{v}+\mathbf{z}}(t)) = \phi(S_\mathbf{v}(t)) \le \psi$. $S_\mathbf{v}(t) = S_{\mathbf{v}+\mathbf{z}}(t)$ if and only if $sign(v_i + z_i - t) = sign(v_i - t)$ for all i. By checking cases, one can see that $\|\mathbf{z}\|_\infty < \min_i |v_i - t|$ is sufficient to guarantee that $v_i + z_i - t$ has the same sign as $v_i - t$. $\square$

Note that Lemma 1 is not a necessary condition as $\mathbf{z} = \frac{1}{2}(\mathbf{v} - t\mathbf{1})$ is a much larger perturbation of $\mathbf{v}$ such that $\phi(\mathbf{v} + \mathbf{z}) \le \psi$. Lemma 1 defines $g_\mathbf{v}$ as a measure of sensitivity of a single vector with respect to preserving sweep cuts of conductance less than or equal to ψ. For vectors $\mathbf{v}$ with small $g_\mathbf{v}$, a small perturbation can disrupt the sweep cuts which achieve conductance less than ψ. By defining the sensitivity of an invariant subspace appropriately, 4 provides a path to deriving a residual tolerance for arbitrary graphs. Denote by d_{min}, d_{max} the minimum and maximum degree of G.

Theorem 4. *Let G be a graph and $\psi > 0$. Define $V = \text{Span}\{D^{-\frac{1}{2}}\mathbf{v}_1 \ldots D^{-\frac{1}{2}}\mathbf{v}_q\}$, where for $j \in \{1 \ldots q\}$ $\hat{A}\mathbf{v}_j = \lambda_q \mathbf{v}_j$ and $\mathbf{v}_j$ are orthogonal. For any vector $\mathbf{x}$, let $\mu = \mathbf{x}^t \hat{A}\mathbf{x}$ and $\delta_{p,q}(\mu) = \min_{\{k<p\} \cup \{k>q\}} |\lambda_k - \mu|$. For any $\mathbf{q} \in V$, let $g_\mathbf{v}$ be defined as in Lemma 1. Define $g = \min_{\mathbf{v} \in V, \|\mathbf{v}\|_2 = 1} g_\mathbf{v}$. If $\left\|\hat{A}\mathbf{x} - \mu\mathbf{x}\right\| < \frac{1}{\sqrt{2}} \sqrt{\frac{d_{min}}{d_{max}}} \delta_{p,q}(\mu) g$, then $\phi\left(D^{-\frac{1}{2}}\mathbf{x}\right) \le \psi$.*

Proof. By 3 applied to $\mathbf{x}$, there is a unit vector $\mathbf{q} \in Span\{\mathbf{v}_1 \ldots \mathbf{v}_q\}$ such that $\|\mathbf{x} - \mathbf{q}\|_\infty \le \|\mathbf{x} - \mathbf{q}\|_2 < \sqrt{\frac{d_{min}}{d_{max}}} g$. Define $\mathbf{z} = (D^{-\frac{1}{2}}\mathbf{x} - \mathbf{v}) \|\mathbf{v}\|^{-1}$, where $\mathbf{v} = D^{-\frac{1}{2}}\mathbf{q} \in V$. By scaling and normalizing we see

$$\|\mathbf{z}\|_2 = \frac{\left\|D^{-\frac{1}{2}}\mathbf{x} - D^{-\frac{1}{2}}\mathbf{q}\right\|_2}{\left\|D^{-\frac{1}{2}}\mathbf{q}\right\|_2} < \sqrt{\frac{d_{max}}{d_{min}}} \|\mathbf{x} - \mathbf{q}\| < g$$

Since

$$\|\mathbf{z}\|_\infty < g = \min_{\mathbf{v} \in V, \|\mathbf{v}\|_2 = 1} \max_{t \in T_\mathbf{v}} \min_{i \in \{1 \ldots n\}} |v_i - t| < g_\mathbf{v},$$

Lemma 1 implies $\phi\left(D^{-\frac{1}{2}}\mathbf{x}\right) \le \psi$. $\square$

If one can bound the value of g from below, then this theorem gives a residual tolerance for the eigenvector approximation when using sweep cuts to partition the graph. Section 3.3 applies this theorem to the ring of cliques family of graphs.

This section connects the eigenresidual to the error when computing blends of eigenvectors, and quantifies the impact of error on the sweep cut procedure. If the eigengap is small enough, then one cannot guarantee that the Rayleigh quotient is closest to the Fielder value, thus one cannot guarantee that the computed eigenvector is close to the

desired eigenvector. In this small gap setting, a small eigenresidual indicates that the computed vector is close to the desired invariant subspace. 4 shows that vectors with small eigenresidual preserve low conductance sweep cuts for general graphs. 4 illustrates how the residual tolerance depends on both the blend gap $\delta_{p,q}(\mu)$ and the sensitivity g of the eigenvectors. The following section applies this theory to the ring of cliques in order to derive solver tolerances for graph partitioning.

3 The Ring of Cliques

To demonstrate the theory developed in Section 2 and to make our arguments as clear as possible, we employ a previously studied model problem, the ring of cliques [5]. Theorems 5 and 6 derive explicit formulas for all eigenvalues and eigenvectors. These formulas determine the relevant residual tolerance. Moreover, complete spectral knowledge gives a strong understanding the convergence properties of simple iterative methods.

The ring of cliques has several attractive properties for analysis of spectral partitioning. The Community structure is as extreme as possible for a connected graph, so the solution is well-defined. Also, we can apply theorems about block circulant matrices [17] to produce closed form solutions to the eigenvector problem. This graph serves as a canonical example of when solving an eigenproblem accurately is unnecessarily expensive to achieve data analysis success. This example shows that it is possible for the combinatorial structure of the data to be revealed faster than the algebraic structure of the associated matrices. The graph is simple to partition accurately as there are many cuts relatively close to the minimum. Any robust partitioning algorithm will correctly recover the cliques in this graph. However, a Fiedler eigenvector is difficult to calculate with guarantees of point-wise accuracy when using non-preconditioned iterative methods. An algorithm that computes a highly accurate eigenpair will be inefficient on large problem instances. Sections 3.3 and 3.4 apply the tools from Section 2 in order to derive a residual tolerance sufficient for solving the ring of cliques. Section 3.5 bounds the number of power method iterations necessary to recover the ring of cliques, and Section 3.6 validates and illustrates these observations with an experiment.

3.1 Definition

A q-ring of b-cliques, $\mathcal{R}_{b,q}$, is parameterized by a block size b and a number of blocks q. Each block represents a clique of size b and all possible internal connections exist within each individual set of b vertices. For each block, there is a single vertex called the *corner* connected to the corners of the adjacent cliques. These q corners form a ring. Each block also has $(b-1)$ *internal* vertices that have no edges leaving the block. The adjacency matrix associated with $\mathcal{R}_{b,q}$ is a sum of tensor products of simple matrices (identity, cycle,

and rank-one matrices). We have

$$A = I_q \otimes (\mathbf{1}_b \mathbf{1}_b^t - I_b) + C_q \otimes (\mathbf{e}_1 \mathbf{e}_1^t),$$

where the I_k are identity matrices of dimension k, $\mathbf{1}_b$ is a constant vector with dimension b, $\mathbf{e}_1 \in \mathbb{R}^b$ is the cardinal vector with a one in its first entry and zero elsewhere, and C_q is the adjacency matrix of a cycle graph on q vertices. The matrix A and other matrices associated with this graph are *block-circulant*, which implies the eigenvectors are the Kronecker product of a periodic vector and the eigenvectors of a small eigenproblem defined by structure in the blocks. Figure 1 shows the structure of the graph, and Figure 2 shows the block structure of the adjacency matrix.
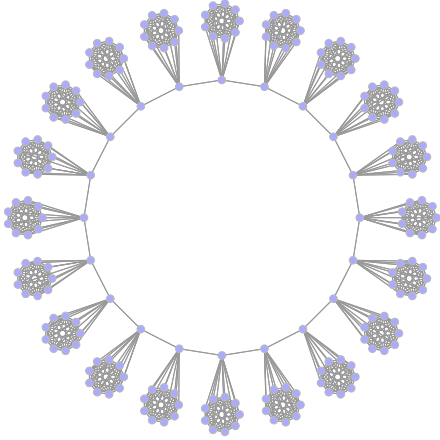
$$\begin{bmatrix} J_b & \mathbf{e}_1\mathbf{e}_1^t & 0 & \cdots & & \mathbf{e}_1\mathbf{e}_1^t \\ \mathbf{e}_1\mathbf{e}_1^t & J_b & \mathbf{e}_1\mathbf{e}_1^t & 0 & & \\ 0 & \mathbf{e}_1\mathbf{e}_1^t & J_b & \mathbf{e}_1\mathbf{e}_1^t & 0 & \\ \vdots & & \ddots & \ddots & \ddots & \\ & & 0 & \mathbf{e}_1\mathbf{e}_1^t & J_b & \mathbf{e}_1\mathbf{e}_1^t \\ \mathbf{e}_1\mathbf{e}_1^t & & & 0 & \mathbf{e}_1\mathbf{e}_1^t & J_b \end{bmatrix}$$

$$\text{where} \quad J_b = \mathbf{1}_b \mathbf{1}_b^t - I_b.$$

Figure 1: A drawing of $\mathcal{R}_{b,q}$ laid out to show strucure.

Figure 2: The adjacency matrix of $\mathcal{R}_{b,q}$ has block circulant structure.

Any partition that breaks a clique cuts at least $b - 2$ edges while any partition that does not break any cliques cuts at most q edges. The best partition is break the ring into two contiguous halves by cutting exactly two edges. There are $q/2$ partitions that achieve this minimal cut for even q. We will consider any of these equivalent. Any partition that breaks fewer than $b - 2$ edges will be regarded as a good, but not optimal cut. The fact that many partitions are close to optimal and then the vast majority of partitions are very far from optimal is a feature of this model problem.

3.2 Eigenpairs of ROC normalized Adjacency Matrix

Due to the block-circulant structure of the ring of cliques we are able to compute the eigenvalues and eigenvectors in a closed form. Let $\mathbf{c}^{k,q}, \mathbf{s}^{k,q}$ be the periodic vectors as
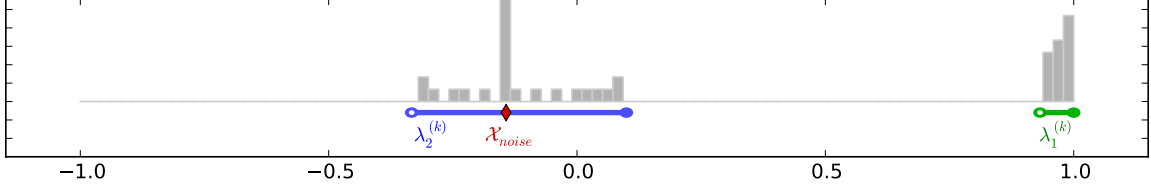
Figure 3: Distribution of eigenvalues of $\hat{A}$ for $\mathcal{R}_{b=8,q=32}$. The gray bars represent a histogram of the eigenvalues, including multiplcities. The red diamond represents a large multiplicity of $(n-2q)$ at $-(b-1)^{-1}$ corresponding to $\mathcal{X}_{noise}$ (see Theorem 5). There is a green interval near 1 containing the portion of the spectrum given by $\lambda_1^{(k)}$ and a blue interval near 0 containing $\lambda_1^{(k)}$ for $k = 0, 1, ..., \lfloor q/2 \rfloor$ (see Theorem 6). The open left endpoint of each interval signifies that the eigenvalue corresponding to $\lambda_{\lceil q/2 \rceil}^{(k)}$ is not present when q is odd.

defined in Equation (2).

$$\mathbf{c}_j^{k,q} = \cos\left(\frac{2\pi k}{q}j\right) \quad \mathbf{s}_j^{k,q} = \sin\left(\frac{2\pi k}{q}j\right) \quad \text{for } j = 1, \ldots, q \tag{2}$$

These vectors are periodic functions with k cycles through the interval $[-1, 1]$ sampled at the qth roots of unity. We employ results from [7] and [17] to derive the full spectrum of $\hat{A}$ and a full eigenbasis. In summary, there is an eigenvalue $-(b-1)^{-1}$ with a large multiplicity, $n - 2q = (b-2)q$. Furthermore, the eigenspace associated with $-(b-1)^{-1}$ can be represented in a basis that contains variation internal to each clique, that is with eigenvectors of the form $\mathbf{h} \otimes \mathbf{e}_i$ where $\mathbf{e}_i$ is the ith standard basis vector for each $i \in \{1 \ldots q\}$. For this reason, we call $\lambda_{noise} = -(b-1)^{-1}$ a noise eigenvalue. The positive eigenvalues are called signal eigenvalues. The signal eigenvectors have the form $\mathbf{p}^{k,q} \otimes (\xi \mathbf{e}_1 + \mathbf{g}_1)$, where $\mathbf{p}^{k,q}$ is either $\mathbf{s}^{k,q}$ or $\mathbf{c}^{k,q}$, $\mathbf{e}_1$ is one in its first entry, $\mathbf{g}_1$ is zero in its first entry and one elsewhere, and ξ is a scalar. All of the internal members of the cliques take the same value in any eigenvector associated with $\lambda_k(\hat{A}) \neq (b-1)^{-1}$. The slowly varying eigenvectors (associated with $\lambda_k(\hat{A}) \approx 1$) give nearly optimal partitions of the graph. Linear combinations of these slowly varying signal eigenvectors also give low conductance partitions. There are $q - 1$ non-localized eigenvectors with small positive or negative eigenvalues. These eigenvectors have the internal clique members and their corner with different sign which causes them to misclassify some of the corners. The distribution of the eigenvalues of $\mathcal{R}_{b,q}$ is illustrated in Figure 4. The rest of Section 3.2 contains formulas for the eigenpairs and the details of their derivations.

Theorem 5. ($\mathcal{R}_{b,q}$ **Noise Eigenpairs**) *There is an eigenspace $\mathcal{X}_{noise}$ of multiplicity $(n - 2q)$ associated with eigenvalue*

$$\lambda_{noise} = \frac{-1}{b-1}.$$

14

Any vector that is zero-valued on all corner vertices and sums to zero on each individual set of internal vertices is in $\mathcal{X}_{noise}$.

Proof. This is a specific case of locally supported eigenvectors [7] (LSEVs) brought on by a high level of local symmetry in $\mathcal{R}_{b,q}$. For each clique $\mathcal{K}$, let $\mathbf{x} = \mathbf{e}_i - \mathbf{e}_j$ for vertices i and j that are internal vertices of $\mathcal{K}$. Both $\mathbf{x}$ and $\hat{A}\mathbf{x}$ are zero valued outside of $\mathcal{K}$. Internally, due to $D^{-1/2}(\mathbf{e}_i - \mathbf{e}_j) = (b-1)^{-1/2}(\mathbf{e}_i - \mathbf{e}_j)$ and the orthogonality $\mathbf{1}_b^t(\mathbf{e}_i - \mathbf{e}_j) = 0$, we see

$$\hat{A}(\mathbf{e}_i - \mathbf{e}_j) = \frac{1}{b-1}(\mathbf{1}_b\mathbf{1}_b^t - I)(\mathbf{e}_i - \mathbf{e}_j) = \frac{-1}{b-1}(\mathbf{e}_i - \mathbf{e}_j).$$

Thus, $\mathbf{x}$ is an eigenvector of $\hat{A}$ associated with $-1/(b-1)$. There are $(b-2)$ such linearly independent eigenvectors for $\mathcal{K}$, and the same is true for all q cliques. Thus, we have a multiplicity of $q(b-2) = n - 2q$ for eigenvalue $\lambda_{noise} = (b-1)^{-1}$.

$\square$

These vectors are in the interior of the spectrum and thus are very well attenuated by the power-method[1]. The remaining eigenvectors must be constant on the internal nodes of the blocks because of orthogonality to the LSEVs which are spanned by $\mathbf{e}_i - \mathbf{e}_j$. In any vector $\mathbf{v}$ the projection of $\mathbf{v}$ onto the global eigenvectors defines a mean value for the elements of the blocks. Since all of the eigenvectors of interest are orthogonal to the constant vector, their entries must sum to zero. So the LSEVs cannot change the mean of a block. The remaining eigenvectors are given in 6.

---

[1]In a single iteration the shifted power method, $\mathbf{x}_{k+1} = (\hat{A} + (b-1)^{-1}I)\mathbf{x}_k$, perfectly eliminates all of the energy in the $(n-2q)$-dimensional eigenspace associated with $\lambda = -(b-1)^{-1}$. If the graph is perturbed with the addition and removal of a few edges, the eigenvectors become slightly less localized and the associated eigenvalues spread out to a short range of values and are not perfectly eliminated in a single iteration. However, the power method or a Krylov method will rapidly attenuate the energy in the associated eigenspaces.

Theorem 6. (ROC Signal Eigenpairs) *For $k = 0, ..., \lceil \frac{q}{2} \rceil - 1$, define*

$$\alpha_k = 2 \cos\left(\frac{2\pi k}{q}\right)$$

$$\beta_k = \frac{1}{2}\left(\alpha_k \sqrt{\frac{b-1}{b+1}} - \sqrt{b^2 - 1} + \sqrt{\frac{b+1}{b-1}}\right)$$

$$\xi_1^{(k)} = \beta_k + \sqrt{\beta_k^2 + (b-1)}$$

$$\xi_2^{(k)} = \beta_k - \sqrt{\beta_k^2 + (b-1)}$$

$$\lambda_1^{(k)} = \frac{\xi_1^{(k)}}{\sqrt{b^2 - 1}} + 1 - \frac{1}{b-1}$$

$$\lambda_2^{(k)} = \frac{\xi_2^{(k)}}{\sqrt{b^2 - 1}} + 1 - \frac{1}{b-1}$$

Let $\mathbf{1}_b$ and $\mathbf{1}_q$ be the vectors of all ones in $\mathbb{R}^b$ and $\mathbb{R}^q$, respectively. Also let $\mathbf{e}_1 \in \mathbb{R}^b$ have a one in its first entry, zero elsewhere and $\mathbf{g}_1 = \mathbf{1}_b - \mathbf{e}_1$. We have the following eigenpairs.

(i) *For $k = 0$, we have 2 eigenvalues of $\hat{A}$, $\lambda_1^{(0)}$ and $\lambda_2^{(0)}$, each with multiplicity 1. The associated (unnormalized) eigenvectors are*

$$\mathbf{v}_1^{(0)} = \sqrt{b+1}(\mathbf{1}_q \otimes \mathbf{e}_1) + \sqrt{b-1}(\mathbf{1}_q \otimes \mathbf{g}_1)$$

and

$$\mathbf{v}_2^{(0)} = (b-1)^{3/2}(\mathbf{1}_q \otimes \mathbf{e}_1) - \sqrt{b+1}(\mathbf{1}_q \otimes \mathbf{g}_1)$$

respectively.

(ii) *For each $k = 1, ..., \lceil \frac{q}{2} \rceil - 1$, we have 2 eigenvalues of $\hat{A}$, $\lambda_1^{(k)}$ and $\lambda_2^{(k)}$, each with multiplicity 2. Two independent (unnormalized) eigenvectors associated with $\lambda_1^{(k)}$ are*

$$\mathbf{v}_{1,1}^{(k)} = \mathbf{c}^{k,q} \otimes \left(\xi_1^{(k)} \mathbf{e}_1 + \mathbf{g}_1\right) \quad and \quad \mathbf{v}_{1,2}^{(k)} = \mathbf{s}^{k,q} \otimes \left(\xi_1^{(k)} \mathbf{e}_1 + \mathbf{g}_1\right).$$

Two independent (unnormalized) eigenvectors associated with $\lambda_2^{(k)}$ are

$$\mathbf{v}_{2,1}^{(k)} = \mathbf{c}^{k,q} \otimes \left(\xi_2^{(k)} \mathbf{e}_1 + \mathbf{g}_1\right) \quad and \quad \mathbf{v}_{2,2}^{(k)} = \mathbf{s}^{k,q} \otimes \left(\xi_2^{(k)} \mathbf{e}_1 + \mathbf{g}_1\right).$$

(iii) If q is even, then for $k = \frac{q}{2}$, we have 2 eigenvalues of $\hat{A}$, $\lambda_1^{(q/2)}$ and $\lambda_2^{(q/2)}$, each with multiplicity 1. The associated (unnormalized) eigenvectors are

$$\mathbf{v}_1^{(q/2)} = \mathbf{c}^{q/2,q} \otimes \left(\xi_1^{(q/2)} \mathbf{e}_1 + \mathbf{g}_1 \right)$$

and

$$\mathbf{v}_2^{(q/2)} = \mathbf{s}^{q/2,q} \otimes \left(\xi_2^{(q/2)} \mathbf{e}_1 + \mathbf{g}_1 \right)$$

respectively.

Note if values of $\lambda_p^{(k)}$ and $\lambda_q^{(l)}$ coincide for $(p,k) \neq (q,l)$ the eigenvalue multiplicities add up.

Proof. Let $D_b = \text{diag}((b+1)\mathbf{1} - 2\mathbf{g}_1)$. Matrix $\hat{A}$ can be organized into blocks that all co-commute with each other implying the eigenvectors are tensor products involving the eigenvectors of the blocks. We decompose $\hat{A} = (I_q \otimes B_1) + (C_q \otimes B_2)$, where I_q is the identity in $\mathbb{R}^q$, C_q is the adjacency matrix of a q- cycle, $(C_q)_{ij} = 1 \iff |i - j| = 1 \mod q$, $B_1 = D_b^{-1/2}(\mathbf{1}_b\mathbf{1}_b^t - I)D_b^{-1/2}$, and $B_2 = \frac{1}{b+1}\mathbf{e}_1\mathbf{e}_1^t$. Because any eigenvector $\mathbf{y}$ of C_q is also an eigenvector of I_q, eigenvectors of $\hat{A}$ have the form $\mathbf{y} \otimes \mathbf{z}$. Vectors $\mathbf{z}$ are derived by plugging various eigenvectors of C_q into $\mathbf{y}$ and solving for a set of constraints that $\mathbf{z}$ must satisfy for $(\mathbf{y} \otimes \mathbf{z})$ to be an eigenvector associated with $\hat{A}$.

We describe the eigendecomposition of C_q. For $k = 0, \dots, \lceil \frac{q}{2} \rceil - 1$, $\alpha_k = 2\cos(2\pi/k)$ is an eigenvalue of C_q. For $k = 0$, α_0 is simple, and $\mathbf{1}_b$ is the associated eigenvector. For $k = 1, ..., \lceil \frac{q}{2} \rceil$-1, α_k has multiplicity 2 and the associated 2-dimensional eigenspace is $\text{span}(\{\mathbf{c}^{k,q}, \mathbf{s}^{k,q}\})$, as defined in (2). If q is even, then $\alpha_{q/2}$ is a simple eigenvalue as well and the associated eigenvector is $\mathbf{c}^{q/2,q}$. Letting $\mathbf{y}$ be an eigenvector associated with α_k and using the properties of Kronecker products, we see

$$
\begin{aligned}
\hat{A}(\mathbf{y} \otimes \mathbf{z}) &= [(I_q \otimes B_1) + (C_q \otimes B_2)](\mathbf{y} \otimes \mathbf{z}) = [(I_q\mathbf{y} \otimes B_1\mathbf{z}) + (C_q\mathbf{y} \otimes B_2\mathbf{z})] \\
&= [(\mathbf{y} \otimes B_1\mathbf{z}) + (\alpha_k\mathbf{y} \otimes B_2\mathbf{z})] = [(\mathbf{y} \otimes B_1\mathbf{z}) + (\mathbf{y} \otimes \alpha_k B_2\mathbf{z})] \\
&= \mathbf{y} \otimes (B_1\mathbf{z} + \alpha_k B_2\mathbf{z}) = \mathbf{y} \otimes [(B_1 + \alpha_k B_2)\mathbf{z}]
\end{aligned}
$$

Here we see that if $\mathbf{z}$ is an eigenvector of $H_k := B_1 + \alpha_k B_2$, then $(\mathbf{y} \otimes \mathbf{z})$ is an eigenvector associated with $\hat{A}$. Observe that $H_k = D_b^{-1/2}(\mathbf{1}_b\mathbf{1}_b^t + \alpha_k\mathbf{e}_1\mathbf{e}_1^t - I)D_b^{-1/2}$ is a scaling of a rank-2 shift from the identity matrix, where we would expect 3 eigenvalues: 2 simple and one of multiplicity $(b-2)$.

We can easily verify that there is a $(b-2)$-dimensional eigenspace of H_k associated with $-1/(b-1)$. The tensor products of these vectors are an alternative basis associated with the locally supported eigenvectors from Theorem 5. The associated eigenspace of H_k is orthogonal to $\text{span}(\{\mathbf{e}_1, \mathbf{g}_1\})$. Due to eigenvector orthogonality, the last two eigenvectors

17

must be in the range of span($\{\mathbf{e}_1, \mathbf{g}_1\}$). Note that $D_b^p \mathbf{e}_1 = (b+1)^p \mathbf{e}_1$ and $D_b^p \mathbf{g}_1 = (b-1)^p \mathbf{g}_1$. We use this to solve for these eigenvectors and their associated eigenvalues in terms of α_k and b,

$$H_k(\xi \mathbf{e}_1 + \mathbf{g}_1) = \lambda(\xi \mathbf{e}_1 + \mathbf{g}_1)$$
$$(\mathbf{1}_b \mathbf{1}_b^t + \alpha_k \mathbf{e}_1 \mathbf{e}_1^t - I_b)D_b^{-1/2}(\xi \mathbf{e}_1 + \mathbf{g}_1) = \lambda D_b^{1/2}(\xi \mathbf{e}_1 + \mathbf{g}_1)$$

The right-hand side expands to

$$\left(\lambda \xi \sqrt{b+1}\right) \mathbf{e}_1 + \left(\lambda \sqrt{b-1}\right) \mathbf{g}_1,$$

The left-hand side expands and simplifies to

$$\left(\frac{\xi}{\sqrt{b+1}} + \sqrt{b-1}\right)(\mathbf{e}_1 + \mathbf{g}_1) + \left(\frac{\xi \alpha_k}{\sqrt{b+1}} - \frac{\xi}{\sqrt{b+1}}\right)\mathbf{e}_1 + \left(\frac{-1}{\sqrt{b-1}}\right)\mathbf{g}_1 =$$
$$\left(\frac{\xi \alpha_k}{\sqrt{b+1}} + \sqrt{b-1}\right)\mathbf{e}_1 + \left(\frac{\xi}{\sqrt{b+1}} + \sqrt{b-1} - \frac{1}{\sqrt{b-1}}\right)\mathbf{g}_1 =$$

The coefficients of $\mathbf{e}_1$ and $\mathbf{g}_1$ must be equal, individually, because they are not linearly dependent. Equating the left-hand and right-hand sides and simplifying gives two nonlinear equations in ξ and λ,

$$\frac{\xi \alpha_k}{b+1} + \sqrt{\frac{b-1}{b+1}} = \xi \lambda \tag{3}$$

$$\frac{\xi}{\sqrt{b^2-1}} + 1 - \frac{1}{b-1} = \lambda. \tag{4}$$

Multiplying the second equation by ξ, setting the left-hand sides of both equations equal to eliminate λ, then simplifying, yields the following quadratic equation in ξ,

$$\xi^2 - \left(\alpha_k \sqrt{\frac{b-1}{b+1}} - \sqrt{b^2-1} + \sqrt{\frac{b+1}{b-1}}\right)\xi - (b-1) = 0,$$

which is easily solved. Define

$$\beta_k = \frac{1}{2}\left(\alpha_k \sqrt{\frac{b-1}{b+1}} - \sqrt{b^2-1} + \sqrt{\frac{b+1}{b-1}}\right) \qquad \text{and} \qquad \gamma_k = b-1.$$

Given ξ, λ is determined by the second equation in (4). The solution set to nonlinear equations (3)-(4) is then

$$\xi_1^{(k)} = \beta_k + \sqrt{\beta_k^2 + \gamma_k}, \qquad \lambda_1^{(k)} = \frac{\xi_1^{(k)}}{\sqrt{b^2-1}} + 1 - \frac{1}{b-1}, \qquad \text{and}$$

$$\xi_2^{(k)} = \beta_k - \sqrt{\beta_k^2 + \gamma_k}, \qquad \lambda_2^{(k)} = \frac{\xi_2^{(k)}}{\sqrt{b^2-1}} + 1 - \frac{1}{b-1}.$$

Thus we have local eigenpairs of H_k, $((\xi_1^{(k)}\mathbf{e}_1 + \mathbf{g}_1), \lambda_1^{(k)})$ and $((\xi_2^{(k)}\mathbf{e}_1 + \mathbf{g}_1), \lambda_2^{(k)})$. The local eigenpairs $((\xi_1^{(k)}\mathbf{e}_1 + \mathbf{g}_1), \lambda_1^{(k)})$ yield global eigenpairs of $\hat{A}$ of the form $((\mathbf{c}_k \otimes (\xi_1^{(k)}\mathbf{e}_1 + \mathbf{g}_1)), \lambda_1^{(k)})$ and $((\mathbf{s}_k \otimes (\xi_1^{(k)}\mathbf{e}_1 + \mathbf{g}_1)), \lambda_1^{(k)})$. Similarly, $((\xi_1^{(k)}\mathbf{e}_1 + \mathbf{g}_1), \lambda_2^{(k)})$ yield global eigenvectors of $\hat{A}$ associated with $\lambda_2^{(k)}$. This accounts for the last $2q$ eigenpairs of $\hat{A}$. $\qquad\square$

In order to illustrate these formulas, Figure 4 shows the computed eigenvectors for the graph $\mathcal{R}_{16,10}$ along with the eigenvalues. Eigenvectors associated with eigenvalues close to 1 have low conductance sweep cuts.
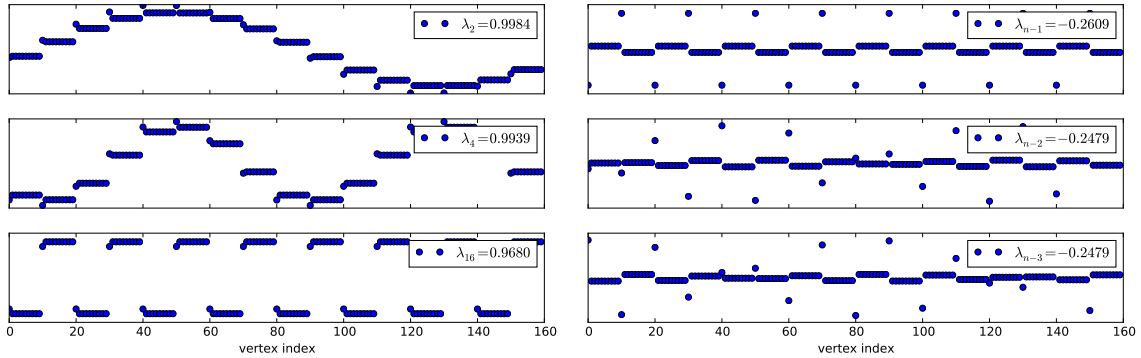

Figure 4: The eigenvectors of $\hat{A}$ are shown for $\mathcal{R}_{16,10}$. The eigenvectors with eigenvalues close to 1 (left) indicate the block structure with differing frequencies. The eigenvectors close to -1 (right) assign opposite signs to the internal vertices and corner vertices of each block.

Corollary 1. *The asymptotic expansions of the eigenvalues are as follows.*

(i) For the signal eigenpairs, we see

$$\lambda_1^{(k)} = 1 - \frac{4 - \alpha_k}{2(b^2 - 1)} + \frac{1}{4(b-1)^2} + \frac{\alpha_k^2}{4(b+1)^2} + \mathcal{O}(b^{-3})$$

(ii) For the non-signal (and non-noise) eigenpairs, we see

$$\lambda_2^{(k)} = \frac{\alpha_k - 1}{b+1} - \frac{\alpha_k^2}{4(b+1)^2} - \frac{\alpha_k}{2(b^2 - 1)} - \frac{1}{4(b-1)^2} + \mathcal{O}(b^{-3})$$

19

(iii) In particular, if q is even we have

$$\lambda_2 = \lambda_1^{(1)} \approx 1 - \left[2 - \cos\left(\frac{2\pi}{q}\right)\right]\frac{1}{b^2-1} + \frac{1}{4(b-1)^2} + \cos^2\left(\frac{2\pi}{q}\right)\frac{1}{(b+1)^2},$$

$$\lambda_q = \lambda_1^{(q/2)} \approx 1 - \frac{3}{b^2-1} + \frac{1}{4(b-1)^2} + \frac{1}{(b+1)^2},$$

$$\lambda_{q+1} = \lambda_2^{(0)} \approx \frac{1}{b+1} - \frac{1}{(b+1)^2} - \frac{1}{b^2-1} - \frac{1}{(b-1)^2},$$

$$\lambda_n = \lambda_2^{(q/2)} \approx \frac{-2}{b+1} - \frac{1}{(b+1)^2} + \frac{1}{b^2-1} - \frac{1}{(b-1)^2}.$$

In other words,

$$1 - \frac{C_2}{b^2} < \lambda_2 < 1, \qquad 1 - \frac{C_q}{b^2} < \lambda_q < 1,$$

$$0 < \lambda_{q+1} < \frac{C_{q+1}}{b}, \qquad -\frac{C_n}{b} < \lambda_n < 0,$$

where C_2, C_q, C_{q+1}, and C_n are positive quantities of order 1.

Proof. These are seen through the formulas in Theorem 6 and first-order Taylor expansion of $\sqrt{\beta_k^2 + (b-1)}$ about leading asymptotic term $\frac{1}{4}((b^2-1)^2)$ and simplification. Let $\theta = \sqrt{b-1}$ and $\eta = \sqrt{b+1}$. Then

$$\beta_k^2 = \frac{1}{4}\left(\frac{\alpha_k\theta}{\eta} - \theta\eta + \frac{\eta}{\theta}\right)^2 = \frac{\alpha_k^2\theta^2}{4\eta^2} + \frac{\theta^2\eta^2}{4} + \frac{\eta^2}{4\theta^2} - \frac{\alpha_k\theta^2}{2} + \frac{\alpha_k}{2} - \frac{\eta^2}{2}.$$

The first-order Taylor expansion we employ is $\sqrt{a^2+x} = a + \frac{1}{2}a^{-1}x + \mathcal{O}(a^{-3}x^2)$, which yields

$$
\begin{aligned}
\sqrt{\beta_k^2 + \theta^2} &= \sqrt{\left[\frac{\theta\eta}{2}\right]^2 + \left[\left(1-\frac{\alpha_k}{2}\right)\theta^2 - \frac{\eta^2}{2} + \frac{\alpha_k^2\theta^2}{4\eta^2} + \frac{\alpha_k}{2} + \frac{\eta^2}{4\theta^2}\right]} \\
&= \frac{\theta\eta}{2} + \frac{1}{\theta\eta}\left[\left(1-\frac{\alpha_k}{2}\right)\theta^2 - \frac{\eta^2}{2} + \frac{\alpha_k^2\theta^2}{4\eta^2} + \frac{\alpha_k}{2} + \frac{\eta^2}{4\theta^2}\right] + \mathcal{O}(b^{-2}) \\
&= \frac{\theta\eta}{2} + \left(1-\frac{\alpha_k}{2}\right)\frac{\theta}{\eta} - \frac{\eta}{2\theta} + \frac{\alpha_k^2\theta}{4\eta^3} + \frac{\alpha_k}{2\theta\eta} + \frac{\eta}{4\theta^3} + \mathcal{O}(b^{-2}).
\end{aligned}
$$

Now, we see

$$
\begin{aligned}
\xi_1^{(k)} &= \beta_k + \sqrt{\beta_k^2 + \theta^2} = \frac{1}{2}\left(\frac{\alpha_k\theta}{\eta} - \theta\eta + \frac{\eta}{\theta}\right) + \sqrt{\beta_k + \theta^2} \\
&= \frac{\theta}{\eta} + \frac{\alpha_k^2\theta}{4\eta^3} + \frac{\alpha_k}{2\theta\eta} + \frac{\eta}{4\theta^3} + \mathcal{O}(b^{-2}), \qquad \text{and} \\
\xi_2^{(k)} &= \beta_k - \sqrt{\beta_k^2 + \theta^2} = \frac{1}{2}\left(\frac{\alpha_k\theta}{\eta} - \theta\eta + \frac{\eta}{\theta}\right) - \sqrt{\beta_k + \theta^2} \\
&= -\theta\eta + (\alpha_k - 1)\frac{\theta}{\eta} + \frac{\eta}{\theta} - \frac{\alpha_k^2\theta}{4\eta^3} - \frac{\alpha_k}{2\theta\eta} - \frac{\eta}{4\theta^3} + \mathcal{O}(b^{-2}).
\end{aligned}
$$

Then, noting $\eta^{-1} - \theta^{-1} = -2\eta^{-1}\theta^{-1}$, we see

$$
\begin{aligned}
\lambda_1^{(k)} &= 1 - \frac{1}{\theta^2} + \frac{\xi_1^{(k)}}{\theta\eta} \\
&= 1 - \frac{1}{\theta^2} + \frac{1}{\eta^2} + \frac{\alpha_k}{2\theta^2\eta^2} + \frac{1}{4\theta^4} + \frac{\alpha_k^2}{4\eta^4} \\
&= 1 - \frac{4 - \alpha_k}{2\theta^2\eta^2} + \frac{1}{4\theta^4} + \frac{\alpha_k^2}{4\eta^4} + \mathcal{O}(b^{-3}).
\end{aligned}
$$

Substituting in for θ and η gives the result presented in (i). Similarly, we see (ii) via

$$
\lambda_2^{(k)} = 1 - \frac{1}{\theta^2} + \frac{\xi_2^{(k)}}{\theta\eta} = \frac{\alpha_k - 1}{\eta^2} - \frac{\alpha_k^2}{4\eta^4} - \frac{\alpha_k}{2\theta^2\eta^2} - \frac{1}{4\theta^4} + \mathcal{O}(b^{-3}).
$$

Lastly, (iii) is seen by plugging in for specific values of k and $\alpha_k = 2\cos\left(\frac{2\pi k}{q}\right)$.

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

Remark 3. *We observe several facts:*

- *The vector $D^{-1/2}\mathbf{v}_1^{(0)}$ is the constant vector. It causes no errors, but does not help to separate any of the cliques.*

- *The vectors $D^{-1/2}\mathbf{v}_{1,1}^{(k)}$ and $D^{-1/2}\mathbf{v}_{1,2}^{(k)}$ for $k = 1, \ldots, \lceil q/2 \rceil - 1$ assign the same sign to the corners as the internal members of each block and are associated with positive eigenvalues. Note that we can consider all these eigenvectors as signal eigenvectors,*

$$
\mathcal{X}_{signal} = span\left\{\mathbf{v}_1^{(1)}, \mathbf{v}_1^{(2)}, ..., \mathbf{v}_1^{(\lceil q \rceil - 1)}\right\}.
$$

Because $\mathcal{X}_{signal} \perp \mathcal{X}_{noise}$, all sweep cuts of vectors in $\mathcal{X}_{signal}$ keep internal vertices of each clique together.
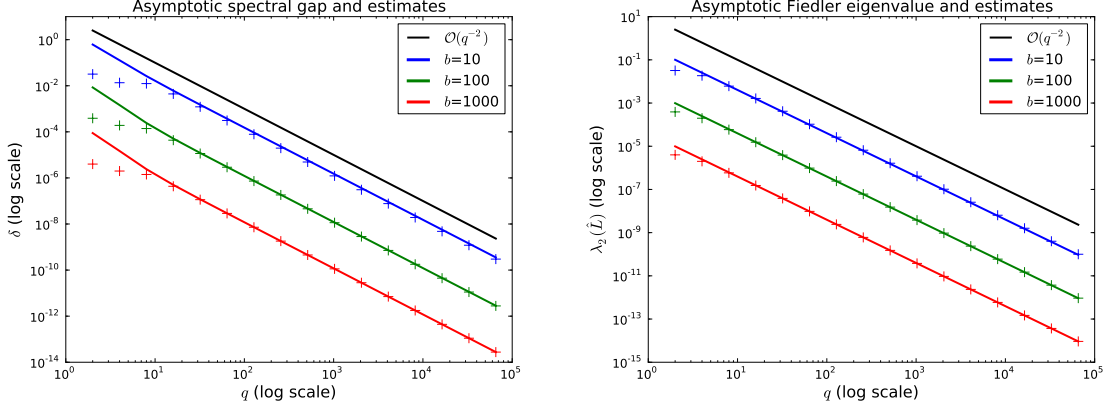
21

Figure 5: Asymptotic estimates of spectral gaps (left) and Fiedler eigenvalues (right) for rings of cliques with parameters $b = 10, 100, 1000$ and $q = 2, 4, ..., 8096$. Lines represent leading-order terms derived in Theorems 7 and 8 and '+' represent actual eigenvalues as given by the formulas in Theorem 6.

- If q is even, the vector $D^{-1/2}\mathbf{v}_2^{(q/2)}$ has values at the corners of opposite sign to the values of the internal vertices and the sign of each corner oscillates around the ring. This is the most oscillatory global eigenvector.

- The vectors $D^{-1/2}\mathbf{v}_{2,1}^{(k)}$ and $D^{-1/2}\mathbf{v}_{2,2}^{(k)}$ for $k = 1, \ldots, \lceil q/2 \rceil - 1$ assign opposite signs to the corners and the internal members of each block. If vectors make large contributions to the blend we compute, then there is potential to misclassify several of the corner vertices.

We use the previous result to derive asymptotic estimates of the eigengaps associated with eigenvalues near 1 and the size of the Fiedler eigenvalue. These asymptotic estimates are compact formulas in terms of b and q. Figure 5 verifies these estimates empirically.

Theorem 7. (ROC Asymptotic Eigengap) *For the graph $\mathcal{R}_{b,q}$ the spectral gap relevant for computing an eigenvector associated with $\lambda_2(\hat{L})$ is asymptotically $\mathcal{O}\left(b^{-2}q^{-2}\right)$ for large b and q.*

Proof. Because $\lambda_2(\hat{A}) = \lambda_3(\hat{A})$, the eigenvalues of interest are $\lambda_2(\hat{A}) = \lambda_1^{(1)}$ and $\lambda_4(\hat{A}) = \lambda_1^{(2)}$. We will take Taylor expansions and collect the leading-order terms to understand the asymptotic behavior. Define scalar function $f(x) = \sqrt{x + a}$, for constant a that we define below. Using the formulas in the previous result and a bit of algebra, we see $\lambda_1^{(1)} - \lambda_1^{(2)} =$

$$\frac{\xi_1^{(1)} - \xi_1^{(2)}}{\sqrt{b^2 - 1}} = \frac{\beta_1 - \beta_2 + \sqrt{\beta_1^2 + (b-1)} - \sqrt{\beta_2^2 + (b-1)}}{\sqrt{b^2 - 1}} = \frac{\beta_1 - \beta_2 + f(x_1) - f(x_2)}{\sqrt{b^2 - 1}} \quad (5)$$

22

with

$$x_k = \left(\frac{1}{4} - \frac{1}{2(b+1)}\right)\alpha_k^2 + \left(1 - \frac{b}{2}\right)\alpha_k \qquad \text{and} \qquad a = \frac{b^2}{4} + \frac{b}{2} - \frac{3}{2} + \frac{1}{2(b-1)}.$$

We expand the two differences in the numerator of (5) separately, concentrating on the $f(x_1) - f(x_2)$ first. First-order Taylor expansion of $f(x)$ at x_1 yields

$$f(x_2) = f(x_1) + f'(x_1)(x_2 - x_1) + \frac{f''(y)}{2}(x_2 - x_1)^2.$$

where $y \in (\min(x_1, x_2), \max(x_1, x_2))$. Rearranging and plugging in, we see

$$f(x_1) - f(x_2) = f'(x_1)(x_1 - x_2) - \frac{f''(y)}{2}(x_1 - x_2)^2 = \frac{x_1 - x_2}{2\sqrt{x_1 + a}} - \frac{(x_1 - x_2)^2}{4(y+a)^{3/2}}$$

Further, assume $q >> 4\pi k$ and use a third-order Taylor expansion at 0 to see $\alpha_k = 2 - 4(\pi k/q)^2 + \mathcal{O}(q^{-4})$. Similarly, $\alpha_k^2 = 4\cos^2(2\pi k/q) = 2(1+\cos(4\pi k/q)) = 4 - 16(\pi k/q)^2 + \mathcal{O}(q^{-4})$. Thus,

$$\begin{aligned}
(x_1 - x_2) &= \left(\frac{1}{4} - \frac{1}{2(b+1)}\right)\left(\frac{48\pi^2}{q^2}\right) + \left(1 - \frac{b}{2}\right)\left(\frac{12\pi^2}{q^2}\right) + \mathcal{O}\left(bq^{-4}\right) \\
&= \frac{-6\pi^2 b}{q^2} + \frac{24\pi^2}{q^2} + \mathcal{O}\left(bq^{-4}\right).
\end{aligned}$$

We return to $\beta_1 - \beta_2$ and expand using the cosine Taylor expansions.

$$\beta_1 - \beta_2 = \frac{\alpha_1 - \alpha_2}{2}\sqrt{\frac{b-1}{b+1}} = \frac{6\pi^2}{q^2}\sqrt{\frac{b-1}{b+1}} + \mathcal{O}(q^{-4})$$

Lastly, $(y+a)^{3/2}$ is $\mathcal{O}(b^3)$ and $(x_1 + a)^{1/2}$ is $\mathcal{O}(b)$, so

$$\lambda_1^{(1)} - \lambda_1^{(2)} = \frac{6\pi^2}{q^2(b+1)} + \frac{-3\pi^2 b}{q^2\sqrt{x_1 + a}\sqrt{b^2 - 1}} + \frac{12\pi^2}{q^2\sqrt{x_1 + a}\sqrt{b^2 - 1}} + \mathcal{O}(b^{-1}q^{-4}).$$

As $b \to \infty$ we see $2b^{-1}\sqrt{x_1 + a} \to 1$, so the first two terms cancel asymptotically and the third term is $\mathcal{O}\left(q^{-2}b^{-2}\right)$. $\qquad\square$

Theorem 8. (ROC Asymptotic Fiedler Eigenvalue) *Let b and q be large and the same order. For graph $\mathcal{R}_{b,q}$ the smallest nonzero eigenvalue of $\hat{L}$ is $\mathcal{O}(b^{-2}q^{-2})$.*

Proof. The eigenvalue of interest is $\lambda_2(\hat{L}) = 1 - \lambda_2(\hat{A}) = 1 - \lambda_1^{(1)}$. Define scalar function $g(z) = \sqrt{z + 1/4}$. Using the formulas in Theorem 6 and a bit of algebra, we see

$$1 - \lambda_1^{(1)} = \frac{1}{b-1} - \frac{\xi_1^{(1)}}{\sqrt{b^2 - 1}} = \frac{1}{b-1} - \frac{\beta_1 + \sqrt{\beta_1^2 + (b-1)}}{\sqrt{b^2 - 1}} = \frac{1}{b-1} - \frac{\beta_1}{\sqrt{b^2 - 1}} - g(z) \quad (6)$$

with
$$z = \frac{(1 - \alpha_1)b}{2(b^2 - 1)} + \frac{2\alpha_1 - 3}{2(b^2 - 1)} + \frac{1}{4(b-1)^2} + \frac{\alpha_1^2}{4(b+1)^2}.$$

We derive the result by demonstrating that the larger terms in (6) cancel. Expanding the second term in in the right-hand-side yields

$$\frac{\beta_1}{\sqrt{b^2 - 1}} = \frac{\alpha_1}{2(b+1)} - \frac{1}{2} + \frac{1}{2(b-1)} = -\frac{1}{2} + \frac{(1 + \alpha_1)b}{2(b^2 - 1)} + \frac{1 - \alpha_1}{2(b^2 - 1)}.$$

A second-order Taylor expansion of $g(z)$ at zero shows for each z, we see $g(z) = g(0) + g'(0)z + \frac{1}{2}g''(0)z^2 + \frac{1}{6}g'''(0)z^3 + \frac{1}{24}g^{(iv)}(y)z^4 = \frac{1}{2} + z - z^2 + 2z^3 + \mathcal{O}(z^4)$, where

$$g'(z) = \frac{1}{2\sqrt{z + 1/4}}, \qquad g''(z) = -\frac{1}{4(z + 1/4)^{3/2}}, \qquad \text{and} \qquad g'''(z) = \frac{3}{8(z + 1/4)^{5/2}}.$$

Plugging in, we see the terms of $g(z)$ up to order b^{-2} are

$$\frac{1}{2} + \frac{(1 - \alpha_1)b}{2(b^2 - 1)} + \frac{2\alpha_1 - 3}{2(b^2 - 1)} + \frac{1}{4(b-1)^2} + \frac{\alpha_1^2}{4(b+1)^2} - \frac{(1 - \alpha_1)^2 b^2}{4(b^2 - 1)^2}.$$

The constant terms in $\beta_1/\sqrt{b^2 - 1}$ and $g(z)$ cancel. The order b^{-1} terms in $1 - \lambda_1^{(1)}$ cancel to an order b^{-2} term,

$$\frac{1}{b - 1} - \frac{(1 + \alpha_1)b}{2(b^2 - 1)} - \frac{(1 - \alpha_1)b}{2(b^2 - 1)} = \frac{1}{b - 1} - \frac{b}{b^2 - 1} = \frac{1}{b^2 - 1}.$$

Combining fractions the order b^{-2} terms in $1 - \lambda_1^{(1)}$, are reduced to

$$\frac{1}{b^2 - 1} - \left[\frac{1 - \alpha_1}{2(b^2 - 1)} + \frac{2\alpha_1 - 3}{2(b^2 - 1)} + \frac{1}{4(b-1)^2} + \frac{\alpha_1^2}{4(b+1)^2} - \frac{(1 - \alpha_1)^2 b^2}{4(b^2 - 1)^2} \right]$$
$$= \frac{(2 - \alpha_1)b^2}{(b^2 - 1)^2} + \frac{(2\alpha_1^2 - 3)b}{2(b^2 - 1)^2} + \frac{-\alpha_1^2 + 2\alpha_1 - 9}{4(b^2 - 1)^2}$$

Factoring in the other b^{-3} terms and the cosine expansion $\alpha_1 = 2 - 4(\pi/q)^2 + \mathcal{O}(q^{-4})$, we see

$$\frac{4\pi^2 b^2}{q^2 (b^2 - 1)^2} + \mathcal{O}\left(b^{-4} + b^{-2} q^{-4} \right).$$

$\square$

Theorem 9. *For any vector $\mathbf{x} \in \mathcal{X}_{signal}$ of $\mathcal{R}_{b,q}$, $\phi\left(D^{-\frac{1}{2}} \mathbf{x} \right) \leq q \, \phi\left(\mathbf{v}_2 \right)$.*

24

Proof. For any $\mathbf{x} \in \mathcal{X}_2$, let $S, \bar{S}$ be the partition given by the optimal sweep cut of $D^{-\frac{1}{2}}\mathbf{x}$. Fiedler's nodal domain theorem implies at least one of $S, \bar{S}$ is a connected component. Because the eigenvectors are block constant, all vertices of each clique are assigned to the same side of the partition. These imply that $E(S, \bar{S}) = 2$. The optimal conductance partition is found when $\mathrm{Vol}\,(S) = \mathrm{Vol}\,(\bar{S}) = q\binom{b}{2}$. Thus $\phi_G = \phi(\mathbf{v}) = 2(q\binom{b}{2})^{-1}$.

For any $\mathbf{x} \in \mathcal{X}_{signal}$, the optimal sweep cut of $D^{-\frac{1}{2}}\mathbf{x}$ will partition the graph into two pieces one containing k blocks and the other containing $n-k$ blocks for some $k \leq \frac{n}{2}$. That is $\min(\mathrm{Vol}\,(S), \mathrm{Vol}\,(\bar{S})) = k\binom{b}{2}$. Since only edges between cliques are cut, $E(S, \bar{S}) \leq 2k$. Thus $\phi\left(D^{-\frac{1}{2}}\mathbf{x}\right) \leq 2\binom{b}{2}^{-1}$. By assigning adjacent blocks to alternating sides of the partition, we see the bound is tight.

$\square$

Notice that the smallest eigenvalue of $\hat{L}$ scales as $\mathcal{O}\left(b^{-2}q^{-2}\right)$, but the optimal conductance scales as $\mathcal{O}\left(b^{-2}q^{-1}\right)$, and that the worst case sweep cut partition of a blend has conductance $2\binom{b}{2}^{-1}$ independent of q. The remainder of this section shows that by accepting this factor of q in conductance, one gains tremendously in computational efficiency.

3.3   A residual tolerance for ring of cliques

In order to derive a residual tolerance for the ring of cliques, we show that for any vector in $Span\{\mathbf{v}_1 \ldots \mathbf{v}_q\}$ at least one block is sufficiently far from the other blocks in spectral coordinates.

Lemma 2. *Define* $\mathcal{B}_i = \{ib+2 \ldots ib+q\}$ *as the vertex blocks of* $\mathcal{R}_{b,q}$. *For any vector* $\mathbf{x}$, *let* $\mu_j = |\mathcal{B}_i|^{-1} \sum_{i \in \mathcal{B}_j} x_i$. *Let* $\mathcal{W}$ *be the span of* $\{\mathbf{e}_{ib+1} \mid i \in 0 \ldots q-1\} \cup \{\mathbf{e}_{\mathcal{B}_i} \mid i \in 0 \ldots q-1\}$. *For any vector* $\mathbf{x} \in \mathcal{W}$, $\|\boldsymbol{\mu}\|_\infty > n^{-1/2} \|\mathbf{x}\|_2$.

Proof. By construction of $\mathbf{x}$, $\mu_i = x_j$ for all $j \in \mathcal{B}_i$. Thus $\|\mathbf{x}\|_\infty = \|\boldsymbol{\mu}\|_\infty$. Equivalence of norms implies $\|\mathbf{x}\|_2 < \sqrt{n}\mu_q$. $\square$

We are able to apply 4 and derive a residual tolerance for recovering the ring of cliques.

Corollary 2. *If* $\mathbf{x}$ *is an approximate eigenvector of* $\mathcal{R}_{b,q}$ *with eigenresidual less than* $\frac{C}{q\sqrt{n}}$ *for some constant* C *and* $\mathbf{x}^t\mathbf{1} = 0$, *then* $\phi\left(D^{-\frac{1}{2}}\mathbf{x}\right) \leq 2\binom{b}{2}^{-1}$

Proof. In the setting of 4, choose $G = \mathcal{R}_{b,q}$, $\psi = 2\binom{b}{2}^{-1}$. In the notation of Lemma 2 applied to sorted $\mathbf{v}$, for all $\mathbf{v} \in V$, $g_v = \max_i(\mu_{i+1} - \mu_i) \geq q^{-1}\|\boldsymbol{\mu}\|_\infty \geq (q\sqrt{n})^{-1}$. For some $C \in \mathcal{O}(1)$, $\delta_{p,q}(\mu)g > \frac{C}{q\sqrt{n}}$. So 4 implies computing $\mathbf{x}$ to a residual tolerance of $\frac{C}{q\sqrt{n}}$ is sufficient to guarantee $\phi(\mathbf{x}) \leq 2\binom{b}{2}^{-1}$. $\square$

2 gives a sufficient condition on approximate eigenvectors of $\mathcal{R}_{b,q}$ such that $\mathbf{x}$ partitions the graph at least as well as any partition that recovers the cliques. 10 and 11 using analysis specialized for $\mathcal{R}_{b,q}$ in Section 3.4 to construct the minimal perturbation that causes the sweep cut proceedure to fail.

3.4 Minimal Perturbation

We want to find the minimal error at which a vector can make a mistake. The effects of the corner vertices only enter into the constants of the following results, and for clarity of exposition we omit handling of the corner vertices. 10 shows that no perturbation with norm less than $(1 + 2qn)^{-\frac{1}{2}}$ can induce a mistake in the sweep cut. 11 constructs a perturbation that induces a mistake in the sweep cut and has norm less than $b^{-\frac{1}{2}}$. For the parameter regime $q \in \mathcal{O}(1)$, the bounds in 2, 10, and 11 are all equivalent up to constant factors.

Using the same notation as Lemma 2, we say that a vector $\mathbf{y}$ recovers all the cliques of $\mathcal{R}_{b,q}$ if there is a threshold $t \in (\min_i y_i, \max_i y_i)$ such that for all $\mathcal{B}_i$ with $j, k \in \mathcal{B}_i$, $y_j < t$ if and only if $y_k < t$.

Theorem 10. *Let $\mathcal{W}$ be defined as in Lemma 2, and let $\mathbf{P}_{\mathcal{W}}$ be the orthogonal projector onto $\mathcal{W}$. For any vector $\mathbf{y}$ orthogonal to $\mathbf{1}$, define $\mathbf{z} = (\mathbf{I} - \mathbf{P}_{\mathcal{W}})\mathbf{y}$. If $\|\mathbf{z}\|_2 \leq (1 + 2qn)^{-\frac{1}{2}} \|\mathbf{y}\|_2$, then $\mathbf{y}$ recovers all the cliques of $\mathcal{R}_{b,q}$.*

Proof. Define $\mathbf{x} = \mathbf{P}_{\mathcal{W}}\mathbf{y}$. Without loss of generality, relabel the vertices and blocks such that $\mu_i \leq \mu_{i+1}$. Let $\alpha_i = \max_i z_i$ and $\beta_i = \min_i z_i$ for each $\mathcal{B}_i$. Note that $\alpha_i > 0$ and $\beta_i < 0$ since $\mathbf{z} \perp \mathbf{e}_{\mathcal{B}_i}$. The vector $\mathbf{y}$ recovers all the cliques if and only if there is a $\mathcal{B}_i$ where $\alpha_i - \beta_{i+1} \leq \mu_{i+1} - \mu_i$. In this case, a threshold can be chosen in $(\mu_i + \alpha_j, \mu_{i+1} + \beta_k)$. Suppose that $\mathbf{y}$ does not recover all the cliques, then for all $\mathcal{B}_i$ $\alpha_i - \beta_{i+1} > \mu_{i+1} - \mu_i$. This implies $\sum_{i=0}^{q-1} \alpha_i - \beta_{i+1} > \sum_{i=0}^{q-1} \mu_{i+1} - \mu_i$. Thus we can bound the 1-norm error as follows:

$$\|\mathbf{z}\|_1 \geq \sum_i (\alpha_i - \beta_{i+1}) \geq \mu_q - \mu_1 \geq n^{-\frac{1}{2}} \|\mathbf{x}\|_2 .$$

Since $\mathbf{z}$ must have at least q nonzero entries $\|\mathbf{z}\|_2 > (2qn)^{-\frac{1}{2}} \|\mathbf{x}\|_2$. Applying $\|\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{z}\|^2$, we see that $(1 + 2qn)^{-\frac{1}{2}} \|\mathbf{y}\| < \|\mathbf{z}\|_2$. $\square$

The proof of Theorem 10 yields a construction for the minimal perturbation of $\mathbf{x}$ that does not recover all the cliques.

Theorem 11. *For any unit vector $\mathbf{x} \in \mathcal{W}$ orthogonal to $\mathbf{1}$, there exists a perturbation $\mathbf{z}$ where $\|\mathbf{z}\| < b^{-\frac{1}{2}}, \mathbf{P}_{\mathcal{W}}\mathbf{z} = 0$ such that $\mathbf{y} = \mathbf{x} + \mathbf{z}$ does not recover all the cliques.*

Proof. For any $\mathbf{x} \in \mathcal{W}$, set $\alpha_0 = 0, \beta_{q-1} = 0, \alpha_i = -\beta_{i+1} = \frac{\mu_{i+1} - \mu_i}{2}$.

$$\|\mathbf{z}\|_2^2 = \sum_{i=0}^{q-1} \alpha_i^2 + \beta_{i+1}^2 = \frac{1}{2} \sum_{i=0}^{q-1} (\mu_{i+1} - \mu_i)^2 = b^{-1} \|\mathbf{x}\|_2^2 - 2 \sum_{i=0}^{q-1} \mu_{i+1}\mu_i < b^{-1} \|\mathbf{x}\|_2^2$$

$\square$

Theorem 10 implies that $(1 + 2qn)^{-\frac{1}{2}}$ is a sufficient accuracy to ensure recovery of all the cliques, and Theorem 11 implies that for some elements of the top invariant subspace accuracy less than $b^{-\frac{1}{2}}$ is necessary to ensure recovery of all the cliques from that vector.

Figure 6 lends validation to the formulas in 10. The experiment shown is to take a random (gaussian unit norm) linear combination of $\mathcal{X}_{signal}$, and then construct the minimal perturbation that makes a mistake. Figure 6 shows the minimum over all samples as a function of n. This experiment is conducted for three different parameter regimes, $q = 25$, $b = 25$, and $b = q = \sqrt{n}$. One can see that the lower bound from 10 is below the empirical observation, and that this lower bound is within a constant factor of the observed size of the minimal perturbation.

We can now apply Theorem 2 and Theorem 3 to determine the residual tolerance for an eigensolver for graph paritioning. The residual tolerance can be no larger than that for the the ring of cliques, but some graphs may require smaller tolerances.

3.5 The Power Method

From the eigenvalues and error tolerances above, one can determine an upper bound on the number of iterations required by the power method to recover all the cliques in $\mathcal{R}_{b,q}$.

Theorem 12. *Let $\mathbf{x}_0$ be sampled from $\mathcal{N}_n(0,1)$. Let $\mathbf{x}_k$ be the k-th iteration of the power method, $\mathbf{x}_{k+1} \longleftarrow \hat{A}\mathbf{x}_k/\|\mathbf{x}_k\|$ for $\mathcal{R}_{b,q}$. Let $\zeta = \left(\frac{e^{7/8}}{8}\right)^{q/2} + \left(\frac{2}{e}\right)^{(n-q)/2}$. There is a k^* of $\mathcal{O}\left(\log_b q\right)$ such that for $k \geq k^*$ a sweep cut based on $\mathbf{x}_k$ makes no errors with probability at least $1 - \zeta$.*

Proof. First, we bound $\|(\mathbf{I} - \mathbf{P}_{\mathcal{W}})\mathbf{x}_0\|^2$ and $\|\mathbf{P}_{\mathcal{W}}\mathbf{x}_0\|^2$ probabilistically. Each entry in $\mathbf{x}_0$ is independently sampled from $\mathcal{N}(0,1)$. For any orthonormal basis of $\mathbb{R}^n$, $\{\mathbf{v}_k\}_{k=1}^n$, the distribution of each $\mathbf{v}_k^t \mathbf{x}_0$ is also $\mathcal{N}(0,1)$. Therefore, the distribution of $\|\mathbf{P}_{\mathcal{W}}\mathbf{x}_0\|^2$ is a χ^2-distribution of order q, which has expected value q and cumulative distribution function $\gamma(q/2, z/2)/\Gamma(q/2)$, where $\Gamma(\cdot)$ is the gamma function and $\gamma(\cdot, \cdot)$ is the lower incomplete gamma function. Let $c_0 \in (0,1)$, using Chernoff bounds we have

$$p_0 := \text{Prob}\left[\|\mathbf{P}_{\mathcal{W}}\mathbf{x}_0\|^2 > c_0 q\right] = 1 - \frac{\gamma\left(\frac{q}{2}, \frac{c_0 q}{2}\right)}{\Gamma\left(\frac{q}{2}\right)} \geq 1 - \left(c_0 e^{1-c_0}\right)^{q/2}$$
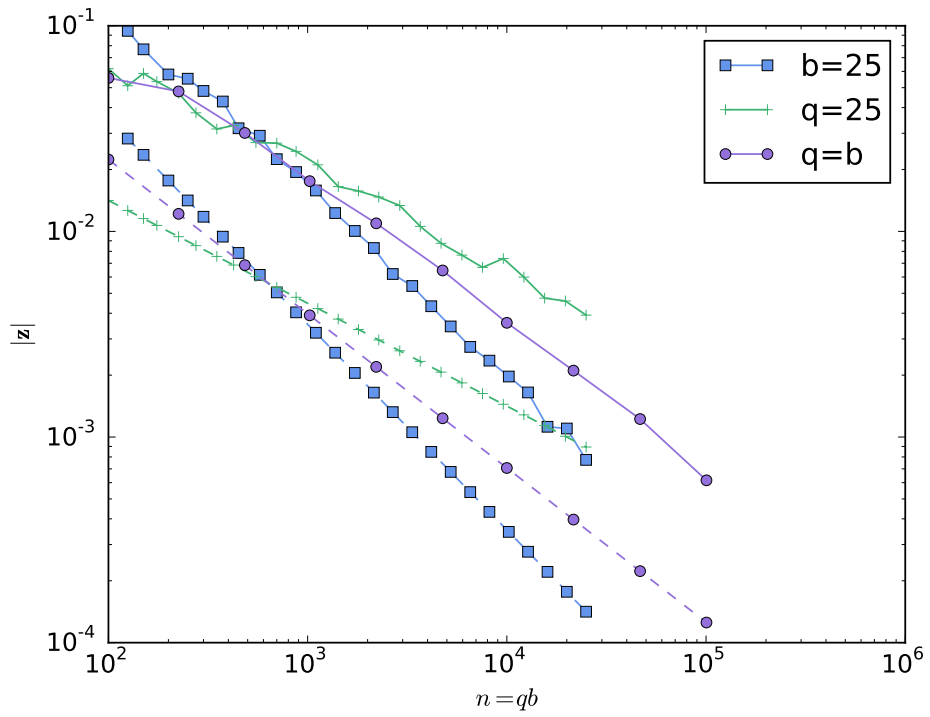
27

Figure 6: Empirical measurements of minimal error perturbations on a log-log scale. Lower bounds are shown in the same color with dashed lines.

Similarly, $\|(\mathbf{I} - \mathbf{P}_\mathcal{W})\mathbf{x}_0\|^2$ is from a χ^2-distribution of order order $n - q$, with expected value $n - q$ and known cdf. Let $c_1 \in (1, \infty)$, we have

$$p_1 := \text{Prob}\left[\|\mathbf{P}_\mathcal{W}\mathbf{x}_0\|^2 < c_1(n-q)\right] = \frac{\gamma\left(\frac{n-q}{2}, \frac{c_1(n-q)}{2}\right)}{\Gamma\left(\frac{n-q}{2}\right)} \geq 1 - \left(c_1 e^{1-c_1}\right)^{(n-q)/2}$$

The union of events $\left[\|\mathbf{P}_\mathcal{W}\mathbf{x}_0\|^2 > c_0 q\right]$ and $\left[\|\mathbf{P}_\mathcal{W}\mathbf{x}_0\|^2 < c_1(n-q)\right]$ is a subset of all possibilities for which $\left[\frac{\|(\mathbf{I}-\mathbf{P}_\mathcal{W})\mathbf{x}_0\|^2}{\|\mathbf{P}_\mathcal{W}\mathbf{x}_0\|^2} < \frac{c_1(n-q)}{c_0 q}\right]$ holds. Therefore, setting $c_0 = 1/8$ and $c_1 = 2$, we see

$$\text{Prob}\left[\frac{\|(\mathbf{I} - \mathbf{P}_\mathcal{W})\mathbf{x}_0\|^2}{\|\mathbf{P}_\mathcal{W}\mathbf{x}_0\|^2} < \frac{c_1(n-q)}{c_0 q}\right] > p_0 p_1 > 1 - \zeta$$

where $\zeta := \left(\frac{e^{7/8}}{8}\right)^{q/2} + \left(\frac{2}{e}\right)^{(n-q)/2}$ is a small positive constant when $q, b > 4$. Because a sweep cut does not depend on the norm of a vector, we consider the iteration, $\mathbf{x}_k \longleftarrow \frac{1}{\lambda_q}\hat{A}\mathbf{x}_{k-1}$ which is equivalent to the power method. Letting $\lambda^* = \max(|\lambda_{q+1}|, |\lambda_n|)$, this iteration accentuates vector components in the range of $\mathbf{P}_\mathcal{W}$ by a factor greater than 1 and attenuates those orthogonal to this space by factors less than λ^*/λ_q. If $\|\mathbf{P}_\mathcal{W}\mathbf{x}_0\|^2 > c_0 q$, then

$$\|\mathbf{x}_k\|^2 \geq \|\mathbf{P}_\mathcal{W}\mathbf{x}_k\|^2 \geq \|\mathbf{P}_\mathcal{W}\mathbf{x}_0\|^2 \geq c_0 q.$$

Also, if $\|(\mathbf{I} - \mathbf{P}_\mathcal{W})\mathbf{x}_0\|^2 \leq c_1(n-q)$, then

$$\|(\mathbf{I} - \mathbf{P}_\mathcal{W})\mathbf{x}_k\|^2 \leq \left(\frac{\lambda^*}{\lambda_q}\right)^{2k}\|(\mathbf{I} - \mathbf{P}_\mathcal{W})\mathbf{x}_0\|^2 \leq \left(\frac{\lambda^*}{\lambda_q}\right)^{2k}c_1(n-q).$$

Therefore, under the assumptions on $\mathbf{x}_0$, the k-th iteration satisfies

$$\frac{\|(\mathbf{I} - \mathbf{P}_\mathcal{W})\mathbf{x}_k\|}{\|\mathbf{x}_k\|} \leq \left(\frac{\lambda^*}{\lambda_q}\right)^k \sqrt{\frac{c_1}{c_0}(b-1)} = 4\left(\frac{\lambda^*}{\lambda_q}\right)^k \sqrt{b-1}.$$

By 10, if this ratio is less than $(1 + 2qn)^{-1/2}$, then $\mathbf{x}_k$ makes no errors. We see this is ensured by

$$k \geq k^* := \left\lceil \frac{\log 4 + \log(b-1) + \log(1 + 2qn)}{2\log(\lambda_q/\lambda^*)}\right\rceil.$$

Revisiting Corollary 1 we that $\lambda_q > 1 - C_q/b^2$ and $\lambda^* < \max(C_{q+1}, C_n)/b$ so $\lambda_q/\lambda^* = C^*b$, where C^* is an order 1 constant. Plugging this in we see that k^* is in $\mathcal{O}(\log_b q)$.

$\square$

3.6   Experiment

Here we show the results of a numerical experiment in order to lend intuition and validation to the theorems. Take $\mathcal{R}_{b,q}$ and a random seed vector $\mathbf{x}^{(0)}$. Then apply power iteration $\mathbf{x}^{(i)} = (\hat{A} - \mathbf{k}\mathbf{k}^t)\mathbf{x}^{(i-1)}$. Far $b = 20$ and $q = 30$ the relevant measures of convergence are shown in Table 1. Figure 7 illustrates the convergence behavior in terms of the conductance of all sweep cuts, and the reordered adjacency matrix represented in sparsity plots. Table 1 shows that the convergence to the Fiedler vector stalls after iteration 3, but convergence to the space orthogonal to $\mathcal{X}_{noise}$ continues unabated. Letting Π be the projection onto $\mathcal{X}_{noise}^{\perp}$, we measure $\|\Pi\mathbf{x}^{(i)}\|$ for each iteration. Applying 12, we calculate that $k^* = 5$ iterations will perfectly resolve the clique structure with probability at least $1 - \zeta = 0.99999998575$. After one iteration the sweep cut did not split any cliques, but only a single clique is shaved off. After 3 iterations a nearly optimal partition is found.
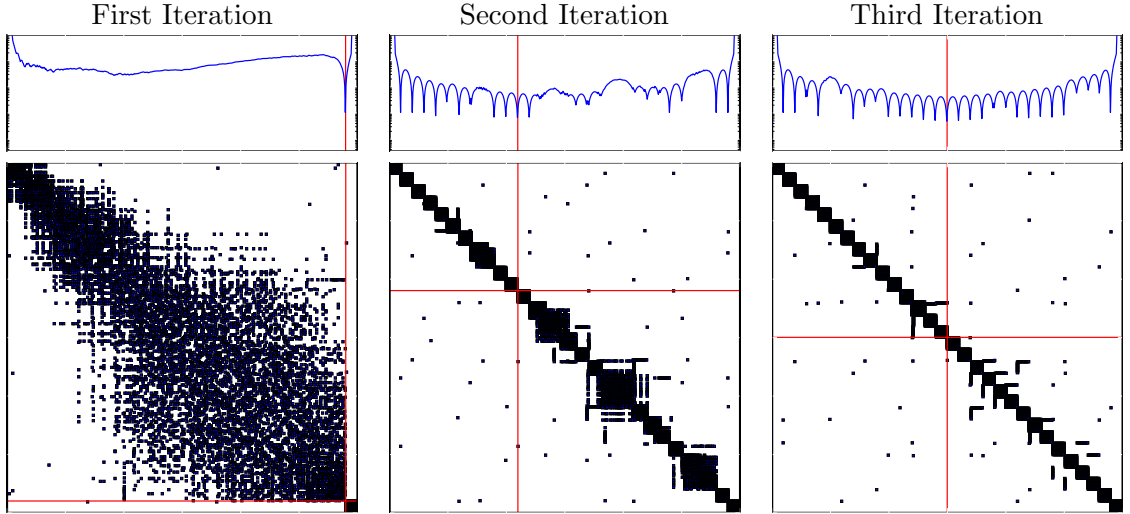
First Iteration	Second Iteration	Third Iteration

Figure 7: First several iterations of the power method applied to $\mathcal{R}_{b=20,q=30}$. Above: sweep conductance of A reordered by sorting the 1st (top-left), 2nd (top middle), and 3rd iterations (top-right). Horizontal axis represents which vertex to split at under the induced ordering; vertical axis is the conductance for each split on a log scale. Below: matrix sparsity plots of A reordered by sorting the 1st (bottom-left), 2nd (bottom-middle), and 3rd iterations (bottom-right). Red lines demonstrate which edges are cut for the optimal cut in each ordering.

i	ϵ	$\|\Pi\mathbf{x}^{(i)}\|$	μ	$\phi(\mathbf{x})$	$\sqrt{2\mu}$
0	1.24806e+01	2.25433e+01	1.52966e+00	2.03094e+00	1.74909e+00
1	9.20534e-01	5.02679e-02	8.37276e-01	1.05263e-02	1.29404e+00
2	1.02629e-01	1.45977e-02	5.67751e-02	6.68577e-03	3.36972e-01
3	1.08242e-02	8.10095e-04	1.01284e-02	4.89853e-03	1.42326e-01
4	1.01568e-02	4.28591e-05	9.87761e-03	4.89853e-03	1.40553e-01
5	1.01262e-02	2.26698e-06	9.85062e-03	4.89853e-03	1.40361e-01
6	1.01022e-02	1.19907e-07	9.82549e-03	4.89853e-03	1.40182e-01
7	1.00782e-02	6.34217e-09	9.80038e-03	4.89853e-03	1.40003e-01
8	1.00543e-02	3.35448e-10	9.77526e-03	4.89853e-03	1.39823e-01
9	1.00303e-02	1.77422e-11	9.75013e-03	4.89853e-03	1.39643e-01
10	1.00063e-02	9.38388e-13	9.72498e-03	4.89853e-03	1.39463e-01
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
∞	0.00000e+00	0.00000e+00	1.14176e-04	3.50018e-04	1.51113e-02
			$\lambda_2(\hat{L})$	$\phi(\mathbf{v}_2) = \phi_{\mathcal{G}}$	$\sqrt{2\lambda_2(\hat{L})}$

Table 1: Table corresponding to Figure 7 with 10 iterations. Convergence to the Fiedler eigenpair is slow, yet convergence to the orthogonal complement of $\mathcal{X}_{noise}$ is rapid (column 2).

31

4 Conclusions

When partitioning graphs where the spectral gap is small, computation of an accurate approximation to a Fiedler vector is difficult. In order to satisfy the needs of spectral partitioning without computing eigenvectors to high accuracy, we introduce spectral blends and in particular the *blend gap*. Section 2.2 controls the distance between an approximate eigenvector and an invariant subspace in terms of the eigenresidual and blend gap thereby showing that accurate approximation to a spectral blend is easier to compute than an accurate approximation of a single eigenvector. We provide a general tool for deriving residual tolerances based on the structure of the graph spectrum. In order to illustrate the utility of spectral blends, Section 3 studies a model problem and uses the theory of block cyclic matrices and locally supported eigenvectors to present a closed form for the eigenvalues and vectors. We show that any blend of large eigenvalue eigenvectors for the ring of cliques recovers a correct clustering. This indicates that for problems where there are multiple good partitions of the graph, spectral blends can be used to partition accurately. The eigendecomposition of the model problem provides error and residual tolerances for solving this problem with sweep cuts. 10 allows us to give guidance for error tolerances for spectral partitioning. One should solve the eigenproblem to a tolerance no greater than $\mathcal{O}\left(n^{-1}\right)$ for graphs of size n. 12 shows that for the ring of cliques where the number of clusters is polynomial in the sizes of the clusters, the number of power method steps taken to recover the clusters is $\mathcal{O}\left(1\right)$. Further research will be able to expand these results to more general graphs which have multiple good partitions.

References

[1] Christos Boutsidis, Alex Gittens, and Anju Kambandur. Spectral clustering via the power method provably. 2015.

[2] Mihai Cucuringu and Michael W. Mahoney. Localization on low-order eigenvectors of data matrices. *CoRR*, abs/1109.1355, 2011.

[3] Gianna M. Del Corso, Antonio Gullí, and Francesco Romani. Fast pagerank computation via a sparse linear system. *Internet Math.*, 2(3):251–273, 2005.

[4] Peter G. Doyle and J. Laurie Snell. *Random walks and electric networks*, volume 22. Carus Mathematical Monographs, Mathematical Association of America, Washington, DC, 1984.

[5] S. Fortunato and M. Barthelemy. Resolution limit in community detection. *Proceedings of the National Academy of Science*, 104:36–41, January 2007.

[6] Stephen Guattery and Gary L Miller. On the quality of spectral separators. *SIAM Journal on Matrix Analysis and Applications*, 19(3):701–719, 1998.

[7] Van Emden Henson and Geoffrey Sanders. Locally supported eigenvectors of matrices associated with connected and unweighted power-law graphs. 39:353–378, 2012.

[8] Ling Huang, Donghui Yan, Nina Taft, and Michael I. Jordan. Spectral clustering with perturbed data. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems 21*, pages 705–712. Curran Associates, Inc., 2008.

[9] Ravi Kannan, Santosh Vempala, and Adrian Vetta. On clusterings: Good, bad and spectral. *Journal of the ACM (JACM)*, 51(3):497–515, 2004.

[10] R. Lehoucq and D. Sorensen. Deflation techniques for an implicitly restarted arnoldi iteration. *SIAM Journal on Matrix Analysis and Applications*, 17(4):789–821, 1996.

[11] Milena Mihail. Conductance and convergence of markov chains-a combinatorial treatment of expanders. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 526–531. IEEE Comput. Soc. Press, 1989.

[12] B. Parlett. *The Symmetric Eigenvalue Problem*. Society for Industrial and Applied Mathematics, 1998.

[13] Alex Pothen, Horst D Simon, and Kang-Pu Liou. Partitioning sparse matrices with eigenvectors of graphs. *SIAM Journal on Matrix Analysis and Applications*, 11(3):430–452, 1990.

[14] Yousef Saad. *Numerical Methods for Large Eigenvalue Problems: Revised Edition*, volume 66. Siam, 2011.

[15] D. A. Spielman. Spectral graph theory: Conductance, the normalized laplacian, and cheeger's inequality, 2012.

[16] Daniel Spielman. Spectral graph theory. *Lecture Notes, Yale University*, pages 740–0776, 2009.

[17] Garry J Tee. Eigenvectors of block circulant and alternating circulant matrices. 2005.

[18] Donghui Yan, Ling Huang, and Michael I. Jordan. Fast approximate spectral clustering. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, pages 907–916, New York, NY, USA, 2009. ACM.