

New Stopping Criteria for Spectral Partitioning

James P Fairbanks, Anita Zakrzewska, and David A Bader

School of Computational Science and Engineering

Georgia Institute of Technology

Atlanta, Georgia

Email: {james.fairbanks,azakrzewska3,bader}@gatech.edu

Abstract—Spectral partitioning (clustering) algorithms use eigenvectors to solve network analysis problems. The relationship between numerical accuracy and network mining quality is insufficiently understood. We show that analyzing numerical accuracy and network mining quality together leads to an algorithmic improvement. Specifically, we study spectral partitioning using sweep cuts of approximate eigenvectors of the normalized graph Laplacian. We introduce a novel, theoretically sound, parameter free stopping criterion for iterative eigensolvers designed for graph partitioning. On a corpus of social networks, we validate this stopping criterion by showing the number of iterations is reduced by a factor of 4.15 on average, and the conductance is increased by only a factor of 1.24 on average. Regression analysis of these results shows that the decrease in the number of iterations needed is greater for problems with a small spectral gap, thus our stopping criterion helps more on harder problems. Experiments show that alternative stopping criteria are insufficient to ensure low conductance partitioning on real world networks. While our method guarantees partitions that satisfy the Cheeger Inequality, we find that it typically beats this guarantee on real world graphs.

I. INTRODUCTION

Large graphs are found in many domains including the analysis of social networks, document collections, and transportation networks. Finding good partitions of these graphs is a challenging network mining task. For a recent survey of the graph partitioning problem see [2]. A tutorial on spectral clustering for data analysis can be found in [23]. Iterative methods for network analysis need good stopping criteria to ensure that a high quality solution is found as quickly as possible.

In this work we introduce a new stopping criterion for computing eigenvectors which combines conductance and estimates of numerical error. Under the standard assumptions on eigensolvers, in Theorem 2 of Section III, we analyze eigenvalue accuracy in the context of spectral partitioning to derive a condition on approximate eigenvectors that provides the same theoretical guarantees as sweep cuts of the exact eigenvectors. Section IV shows this new stopping criterion reduces the number of iterations compared to traditional stopping criteria on real world networks from the Newman [16] and SNAP [14] collections. We provide a parameter free convergence criterion that is theoretically sound and empirically verified.

This work fits into a larger framework of studying how knowledge of the network mining task can shape our choice of numerical procedure. It is already common to see the computer architecture or networking capabilities of a distributed system shape the choice of numerical algorithm [8]. Some applications

use numerical solutions such that errors compound, but other applications use the numerical solution in a way robust to error. For instance in spectral partitioning, one spends many cycles computing high accuracy numerical solutions to an equation only to round the solution vector to binary values. When designing efficient algorithms, one must consider the accuracy requirements of the application. This paper shows that for spectral partitioning the numerical accuracy required is much lower than that which is typically assumed by users of these solvers. The default solver accuracy for Matlab's eigs is machine ϵ times $\|A\|$ which is approximately 10^{-15} . This accuracy is often necessary for many scientific applications, but is not necessary for this particular network analysis application.

The goal of this paper is to understand the effect of eigensolver accuracy on clustering quality. Stopping criterion for iterative methods are an important facet of this relationship. Without a good stopping criterion, an iterative method will either take too few iterations and fail to solve the problem, or take too many iterations and waste time. Theorem 2 presents a stopping criterion with quality guarantee, while Section IV validates its performance on real world datasets.

Graphs can have multiple partitions of similar quality and throughout this paper we assume that an application finds any of them to be sufficient. In this paper we show that, using new stopping criteria, we can compute approximate eigenvectors which induce nearly optimal graph partitions. These approximate eigenvectors are computed faster than approximation using classical stopping criteria based on residual tolerances. Therefore, we decrease running time while sacrificing little quality. Many of these insights can generalize to other applications where a numerical method solves a data mining problem, such as using personalized pagerank [5] to rank vertices in a graph, commute times [6] to compute a metric distance on the vertices, or the heat equation on a graph [3] to construct low conductance local cuts.

A. Contributions

This paper makes the following contributions:

- 1) A novel parameter free stopping criterion (Theorem 2) for spectral partitioning with both theoretical and experimental support.
- 2) Demonstrations that alternative stopping criteria are too weak to ensure high quality partitions.

- 3) Evidence that our method works when restricting to balanced sweep cuts.
- 4) Guidance on practical choices for the residual tolerance parameter of eigensolvers for graph partitioning.

B. Related Work

Many data analysis algorithms are phrased as optimization problems with numerical solutions [12]. The solution to the original data analysis problem depends on the accuracy of the solution to the induced numerical problem, and many theoretical results quantify the relationship between the exact solution to the numerical problem and the quality of the solution to the data analysis problem. However, there is little work evaluating the quality of a data analysis solution produced by an approximate solution to the numerical problem. In this paper, we address this topic for spectral partitioning. Spectral partitioning is performed in two steps. First, one or more vectors approximating some eigenvectors of a graph matrix are computed and then those vectors are used to partition the graph. The eigenvector computation step is often treated as a primitive operation without considering the trade-off between runtime and accuracy. This is the case in [9], which evaluates the running time and quality, in terms of conductance, of both spectral and other partitioning algorithms. Pothen et al. [18], when applying spectral partitioning to distributed memory sparse matrix computation, recognized the value of low accuracy solutions. Understanding the relationship between the error in eigenvalue approximation and the error in the original data mining problem is important. This work addresses the effect of the error in the computed eigenvalue on the quality of the sweep cuts of the computed eigenvector. Allowing more error in the eigenvector computation improves runtime performance possibly at the cost of partition quality. If the partition quality is not too large, trading quality for performance can be useful, especially for computationally expensive problems on large datasets.

Iterative methods can generate solutions to arbitrary approximation factors. Both runtime and solution accuracy increase with the number of iterations performed. Iterative methods for solving the eigenvector problem $A\mathbf{x} = \lambda\mathbf{x}$ have been shown to provide fast approximate solutions [13], [19]. For example, the implicitly restarted Arnoldi method (IRAM) allows one to solve for a small number of eigenvalues of a linear operator A [20]. A function that evaluates the action of A on arbitrary vectors along with $O(n(k+p) + (k+p)^2)$ space is sufficient to use the method. A practical implementation of the Arnoldi method, which is commonly used in practice, can be found in ARPACK [13].

Although eigenvectors produced by these methods are approximations, the impact of the error of these approximation techniques on the error of the original data mining solution has not been sufficiently studied. Eigenvectors of a kernel matrix can be approximated by the power method with k-means applied to these approximations to cluster the graph [1]. The k-means objective function is well approximated by such approximate eigenvectors. The bounds given in [1] depend on

using the k eigenvectors to partition into k parts and depend on the k th spectral gap. On the approximate eigenvectors, k-means is faster and sometimes more accurate in terms of normalized mutual information compared to using exact eigenvectors.

Instead of using k-way partitioning, our paper focuses on partitioning into two clusters based on sweep cuts of a single approximate eigenvector. Because two way partitioning can be used recursively to find small communities, this paper focuses on splitting a graph into two parts. The effects on multilevel partitioning [10], multiway partitioning, and local methods to improve the cut are beyond the scope of this work.

Other work focuses on the impact of probabilistic sampling error on data mining quality. In the context of Gram (kernel) matrices, Huang et al. [7] study the effect of perturbing the original data points on the spectral partitioning method. A similar topic is pursued in [24], where data points are quantized to reduce bandwidth in a distributed system. This paper differs by treating the data as correctly observed and evaluates error in the iterative solver.

In the problem of graph partitioning, multiple good partitions may exist. In a resource constrained environment, one would like to be able to recover one of these near optimal partitions while expending as few resources as possible. We will show that finding vectors which produce these near optimal partitions is much less expensive than highly accurate approximations to the eigenvectors.

II. DEFINITION AND NOTATION

Let A denote the adjacency matrix of an undirected graph with entries $a_{i,j}$ equal to 1 if vertex i is adjacent to vertex j . Use I to represent the identity matrix and $\mathbf{1}$ to represent the vector of ones. If D is the diagonal matrix whose entries are the degrees $d_i = \sum_j a_{i,j} = A\mathbf{1}$, then $L = D - A$ is the combinatorial Laplacian. This paper will focus on the *normalized adjacency matrix* $\hat{A} = D^{-1/2}AD^{-1/2}$ and the *normalized Laplacian* $\hat{L} = I - \hat{A}$. The equation $\hat{L}\mathbf{x} = \lambda\mathbf{x}$ is satisfied by pairs λ, \mathbf{x} which are called the eigenvalue $\lambda \in \mathbb{R}$ and eigenvector $\mathbf{x} \in \mathbb{R}^n$. When the matrix is a graph Laplacian \hat{L} , all eigenvalues are nonnegative. When the graph is connected, \hat{L} has a unique 0 eigenvalue and we sort the eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \dots \lambda_n$ and label the associated eigenvectors $\mathbf{q}_1, \mathbf{q}_2 \dots \mathbf{q}_n$. Since the graphs are undirected, there exists an orthonormal basis of eigenvectors. This can be expressed as the matrix decomposition $\hat{L} = Q\Lambda Q^T$, where $QQ^T = I$ and Λ is a diagonal matrix with nonnegative entries.

For a matrix A and a unit vector \mathbf{x} we define two numerical quantities of interest, the error and the residual. The error of \mathbf{x} with respect to \mathbf{q}_2 is measured as $e = \|\mathbf{x} - \mathbf{q}_2\mathbf{q}_2^T\mathbf{x}\|$, i.e. the norm of the projection of the vector onto the eigenspace. The residual r is defined as $\|\hat{L}\mathbf{x} - \mu\mathbf{x}\|$ where $\mu = \frac{\mathbf{x}^T\hat{L}\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$ is the Rayleigh Quotient of \mathbf{x} . The goal of an eigensolver is to produce an \mathbf{x} that satisfies $e = 0$. Since this cannot be done exactly, we use the error and residual to evaluate the quality of a solution. In practice, the error cannot be measured directly,

and we must rely on theorems relating the residual to the error. The standard approach, taken in ARPACK, is to iterate until r is less than a prescribed tolerance chosen by the user.

We measure the quality of a (vertex) cut of the graph using conductance, which is the surface area to volume ratio of a subset of vertices in the graph. Let S be a subset of the vertex set and use $\bar{S} = V \setminus S$ to denote set complement. These two sets are 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 . The conductance of a cut S is thus given by the formula [4]:

$$\phi(S) = \frac{\sum_{i \in S, j \notin S} a_{i,j}}{\min(\text{vol}(S), \text{vol}(\bar{S}))}$$

Sweep cuts of a vector \mathbf{x} are cuts of the form $S_x^t = \{i \mid x_i > x_t\}$. Any cut can be expressed as the sweep cut of some vector such as the vector with a 1 at index i if $i \in S$ otherwise -1 . We call the conductance of a vector the conductance of the minimal sweep cut of that vector: $\phi(\mathbf{x}) = \min_t \phi(S_x^t)$. Thus the graph partitioning problem can be represented as minimizing $\phi(\mathbf{x})$ over all \mathbf{x} . This work focuses on the spectral partitioning algorithms that use a sweep cut of the eigenvector \mathbf{q}_2 of \hat{L} to partition the graph.

III. EIGENVALUE ACCURACY AND CHEEGER'S INEQUALITY

Cheeger's inequality is a commonly used bound that allows one to guarantee that the small eigenvectors of the graph Laplacian will reveal low conductance cuts of the graph. In this section we examine the effects of approximation of the eigenvector on the guarantee provided by Cheeger's inequality. When finding low conductance cuts, one is solving the problem $\min_{\mathbf{x} \in \{-1,1\}^n} \frac{\mathbf{x}^T L \mathbf{x}}{\mathbf{x}^T D \mathbf{x}}$. This minimization is relaxed to $\min_{\mathbf{x} \perp \mathbf{1}} \frac{\mathbf{x}^T L \mathbf{x}}{\mathbf{x}^T D \mathbf{x}}$, which is solved by solutions to the generalized eigenvalue problem $L\mathbf{x} = \lambda D\mathbf{x}$. A change of variables relates solutions of this generalized eigenvalue problem to solutions of the eigenvalue problem $\hat{L}\mathbf{x} = \mu\mathbf{x}$. A pair λ, \mathbf{y} solves the generalized eigenequation $L\mathbf{y} = \lambda D\mathbf{y}$, if and only if the pair $\lambda, \mathbf{x} = D^{-\frac{1}{2}}\mathbf{y}$ solves the eigenequation $\hat{L}\mathbf{x} = \lambda\mathbf{x}$. Also, the vector $D^{\frac{1}{2}}\mathbf{1}$ lies in the kernel of \hat{L} as $\hat{L}D^{\frac{1}{2}}\mathbf{1} = L\mathbf{1} = 0$. With this change of variables, computational tools for symmetric eigenvalue problems solve the relaxed minimization problem. Cheeger's Inequality bounds the relaxation error in terms of the exact eigenvalue λ_2 .

Theorem 1. *Cheeger Inequality [15] Let \hat{L} be the normalized Laplacian of a connected graph and D be the associated degree matrix. If \mathbf{x} is a vector orthogonal to $D^{\frac{1}{2}}\mathbf{1}$ such that $\hat{L}\mathbf{x} = \lambda_2\mathbf{x}$, then $\mathbf{y} = D^{-\frac{1}{2}}\mathbf{x}$ has a sweep cut S such that $\phi(S) = \phi(\mathbf{y}) \leq \sqrt{2\lambda_2}$.*

The Cheeger inequality guarantees that exact eigenvectors provide a sweep cut with conductance less than $\sqrt{2\lambda_2}$. The remainder of this section derives a stopping criterion providing the same guarantee for approximate eigenvectors. As one iterates a numerical solver, they can compute the conductance of each iterate. Once a partition with conductance less than $\sqrt{2\lambda_2}$ is found, the solver can stop while satisfying the same

guarantee provided by the exact solution. However, when running the solver, the true value of λ_2 is unknown. Thus one cannot use this as a stopping criterion directly. One way to create a stopping criterion is to use μ as an estimate of λ_2 then stop when $\phi(\mathbf{y}) \leq \sqrt{2\mu}$; however, this criteria is trivially satisfied [15], and does not ensure that $\phi(\mathbf{y}) \leq \sqrt{2\lambda_2}$ as $\phi(\mathbf{y})$ can fall between $\sqrt{2\mu}$ and $\sqrt{2\lambda_2}$. Section IV-D discusses further some alternative stopping criteria found to be unreliable for this problem.

In order to derive a stopping criterion that does guarantee $\phi(\mathbf{y}) \leq \sqrt{2\lambda_2}$, we find a lower bound on $\sqrt{2\lambda_2}$ computable from information available to the solver. Both the eigenresidual r and the Rayleigh Quotient μ are computable without knowledge of the true eigenvalue and thus form the inputs to the stopping criterion. Let $\psi(\mathbf{x}) = \sqrt{2(\mu - r)}$.

Theorem 2. *Let \hat{L} be the normalized graph Laplacian of a connected graph, and \mathbf{x} be a unit vector orthogonal to $D\mathbf{1}$ and $\mu = \mathbf{x}^T \hat{L} \mathbf{x}$ and $\mathbf{y} = D^{-\frac{1}{2}}\mathbf{x}$. If $\mu - \lambda_2 < |\mu - \lambda|$ for all other eigenvalues λ , then $\phi(\mathbf{y}) < \sqrt{2(\mu - r)} = \psi(\mathbf{x})$ is a stopping criterion that guarantees $\phi(\mathbf{y}) < \sqrt{2\lambda_2}$.*

Proof. First the fact that $\sqrt{2\alpha}$ is an increasing function gives, for any positive ϵ , $\mu - \epsilon < \lambda_2$ implies $\sqrt{2(\mu - \epsilon)} < \sqrt{2\lambda_2}$. We show that $r > |\mu - \lambda_2|$ directly. Using the eigendecomposition of $L = Q\Lambda Q^T$, let $\mathbf{z} = Q^T\mathbf{x}$. Since $\mathbf{x} \perp \mathbf{q}_1$, $z_1 = 0$. From the hypothesis that $|\mu - \lambda_2|$ is minimal, one sees

$$r^2 = \|(\Lambda - \mu I)\mathbf{z}\|^2 > (\lambda_2 - \mu)^2 \sum_i z_i^2 = (\lambda_2 - \mu)^2.$$

So $r > |\lambda_2 - \mu|$, and $\mu - r < \lambda_2$. Thus under these conditions we know that $\sqrt{2(\mu - r)} < \sqrt{2\lambda_2}$. Since, all terms on the left hand side are known to the solver at each iteration, this is a valid stopping criterion. \square

This stopping criterion is the first stopping criterion for spectral partitioning that does not require the implementation to specify a chosen parameter value. Unlike standard methods there is no choice of acceptable error that must be considered and no choice of tolerance exposed to the user. This simplifies practical application of this method for network analysis.

The assumption that μ is closer to λ_2 than to any other eigenvalues implies that r is less than the spectral gap, $\delta = \lambda_2 - \lambda_3$, of the matrix. Thus under the standard assumption that the Ritz value is close enough to the desired eigenvalue, we have a stopping criterion bounding $\phi(\mathbf{y})$ ¹.

When the eigenvalue is computed exactly, this bound coincides with the original Cheeger inequality. While this theorem does not imply that further iteration of the eigensolver will not reduce $\phi(\mathbf{y})$, it gives a condition under which the approximate eigenvector satisfies the same guarantee as the exact eigenvector. The experiments in Section IV-B show for many graphs, termination according to this new criterion leads to only a small increase in the conductance of a partition. For

¹This assumption is necessary as can be seen by examining the residual and Rayleigh quotient of \mathbf{q}_3

some graphs the partition produced by the lower accuracy approximate has lower conductance than the partition produced by the higher accuracy approximation.

A valid concern is that ensuring that one satisfies $\sqrt{2\lambda}$ is too weak of a guarantee for practitioners. We see that the final partitions are much better than the guarantee. The cycle and the hypercube show that for some graphs the eigenvectors achieve a partition with conductance $\Omega(\sqrt{2\lambda})$ [22]. This is the best guarantee possible for a method applicable to all graphs. If one knows the graph in question comes from a family where $\phi(G) \leq f(\lambda)$ where $f(x)$ is an increasing function, then one can apply the technique used in this paper by checking if the conductance of the current iterate is less than $f(\mu-r) < f(\lambda)$. Planar Graphs [21] are such a family of graphs. Experiments in Section IV show that our stopping criterion typically gives partitions outperforming the guarantee by at least a factor of $\lambda/5$.

IV. EXPERIMENTS

Iterative eigensolvers such as IRAM were developed to solve problems from physics and engineering. These methods are designed to quickly minimize error and residual. However when using eigenvectors for graph partitioning, minimizing the error to the true eigenvectors is less important than finding a vector \mathbf{x} minimizing $\phi(\mathbf{x})$. This leads to an experiment showing the conductance of the optimal sweep cut approaches the minimal value before the eigenresidual is small and that our approach returns such a vector with a low conductance.

A. Experimental Design

While spectral partitioning finds a sweep cut with low conductance, it does not guarantee the minimum possible conductance. Therefore, for our experiments, we compute a baseline eigenvector and partition using the standard approach of iterating until the residual is within an application determined tolerance. We use a tolerance of $\|Ax - \mu x\| < 10^{-6}$, but stop if the number of iterations reaches 800. Let I_F denote the corresponding number of eigensolver iterations. The conductance of the sweep cut of this approximate eigenvector is our baseline conductance ϕ_F . For our experiments, we compare this standard approach using residual tolerance to our stopping criterion from Theorem 2. The restart parameter and maximum number of iterations is chosen manually at 15 to balance time and memory constraints. For 34 of the graphs, the first time $\phi(\mathbf{x}) < \psi(\mathbf{x})$, the hypothesis of Theorem 2 is unsatisfied, but by taking one more step the number of such graphs drops to 17. Because stopping at the next iteration after $\phi(\mathbf{x}) < \psi(\mathbf{x})$ has a small impact on the average number of iterations needed and leads to a large decrease in the average conductance that we find, we use this iteration, represented by I_C , in our experiments. ϕ_C represents the conductance of the sweep cut after I_C iterations.

We conduct experiments on matrices from the Newman [16] and the SNAP [14] collections. These include graphs from co-purchasing, citation, co-authorship, road, autonomous systems,

and online networks². Table I includes the size of each graph, the first 2 eigenvalues, and the eigengap $|\lambda_2 - \lambda_3|$ to show the difficulty of the problem instances. From error analysis in [11], [17] we know that when the small eigenvalues of \hat{L} are close together, the eigenvectors are difficult to compute. The graphs range from the small and well conditioned N/lesmis to the large and ill conditioned S/web-Google.

For the purpose of experimentation, we compute functions of each iterate³. These include the conductance ϕ , the Rayleigh quotient μ , and the residual $r = \|Ax - \mu x\|$. Table II shows these measurements in detail for a single graph. We also show the lower bound on the Cheeger bound $\psi(x) = \sqrt{2(\mu-r)}$ from Theorem 2. From Theorem 2 we know that if $\phi < \psi$ then $\phi < \sqrt{2\lambda}$, which means that our partition is within the guarantee provided by the Cheeger inequality. A dash in Table II indicates $\mu - r < 0$ and the stopping criteria cannot be satisfied.

Because $\hat{L} = I - \hat{A}$, the eigenvalues of \hat{L} are the same as one minus the eigenvalues of \hat{A} with the same eigenvectors. We solve for the largest eigenvalues of \hat{A} and then compute the corresponding smallest eigenvalues of \hat{L} . This technique avoids the need for a linear solver for $\hat{L}x = b$. We also work with vectors orthogonal to $\mathbf{q}_1 = D^{-1/2}\mathbf{1} \|D^{-1/2}\mathbf{1}\|^{-1}$ by iterating with the linear operator $M = \hat{A} - \mathbf{q}_1\mathbf{q}_1^T$. The desired eigenvalue and eigenvector of the Laplacian correspond to the largest eigenvalue of M . This improves the performance of the solver while computing the appropriate Laplacian eigenvalues and eigenvectors.

B. Summary of Results

Table III compares, for all graphs, our stopping criterion to the standard criterion of iterating until $\|r\| < tol$. Values for each graph are averaged across 10 runs with the last row showing the average across all graphs. The values in Table III differ from those seen in Figure 1 because the latter shows results from individual runs. We can see that stopping at I_C results in a conductance less than five times the final conductance for every graph. On average across all graphs the conductance resulting from our approach is only 1.24 times greater and only 0.24 times as many iterations are needed or a reduction by a factor of 4.15. This is the primary result of the experiments. Our stopping criteria reduces the number of iterations significantly without creating a large loss in the achieved conductance.

Table II shows a sample of iterations in detail for the S/web-Google graph. Although we show such detailed results for only one graph, these observations generalize to many of the large graphs seen in the experiment. Table II also shows that the approximate eigenvalue μ is monotonically decreasing, but the minimal conductance of a sweep cut is not. We can see many iterations with $\mu - r < 0$, and few iterations with $\mu - r > 0$ and

²We make adjacency matrices symmetric by taking $A + A^T$. Because we can find connected components faster than solving the eigenequation, we restrict to the largest connected component of each graph

³Requesting the approximate eigenvector at each step also prompts using an alternative implementation of IRAM over the standard ARPACK.

TABLE I

SHOWS THE SIZE OF EACH GRAPH ALONG WITH SOME EIGENVALUES. SMALL VALUES OF $\lambda_3 - \lambda_2$ INDICATE DIFFICULT PROBLEMS. GRAPHS FROM THE NEWMAN AND SNAP COLLECTIONS ARE ABBREVIATED WITH N AND S, RESPECTIVELY.

name	$ V $	$ E $	λ_1	λ_2	δ
N/dolphins	62	318	0.03952	0.23435	0.19483
N/karate	34	156	0.13227	0.28705	0.15478
N/polbooks	105	882	0.03780	0.17589	0.13809
S/wiki-Vote	8297	103689	0.10055	0.16859	0.06804
N/celegansneural	297	2345	0.19524	0.25629	0.06105
S/p2p-Gnutella04	10879	39994	0.02189	0.08147	0.05958
S/p2p-Gnutella09	8114	26013	0.03515	0.08965	0.05450
N/football	115	1226	0.13680	0.18292	0.04612
S/p2p-Gnutella08	6301	20777	0.04037	0.08103	0.04066
S/p2p-Gnutella05	8846	31839	0.10959	0.14601	0.03642
N/polblogs	1490	19025	0.08144	0.10904	0.02760
S/p2p-Gnutella24	26518	65369	0.06600	0.09155	0.02555
S/p2p-Gnutella30	36682	88328	0.06124	0.08551	0.02427
N/adjnoun	112	850	0.35604	0.37559	0.01955
S/p2p-Gnutella06	8717	31525	0.12074	0.13797	0.01723
S/Oregon-1	11492	46818	0.03290	0.04820	0.01530
S/Oregon-2	11806	65460	0.02919	0.04191	0.01272
S/p2p-Gnutella25	22687	54705	0.02837	0.04100	0.01263
S/as-735	7716	26467	0.03349	0.04254	0.00905
S/soc-Epinions1	75888	508837	0.00479	0.01323	0.00844
S/cit-HepPh	34546	421578	0.01572	0.02300	0.00728
S/as-caida	31379	106762	0.01120	0.01826	0.00706
S/cit-HepTh	27770	352807	0.01839	0.02391	0.00552
N/netscience	1589	5484	0.00303	0.00850	0.00547
N/as-22july06	22963	96872	0.01936	0.02418	0.00482
S/soc-sign-epinions	131828	841372	0.01123	0.01575	0.00452
N/lesmis	77	508	0.08813	0.09222	0.00409
S/ca-AstroPh	18772	396160	0.00629	0.01038	0.00409
S/ca-HepTh	9877	51971	0.00312	0.00718	0.00406
N/hep-th	8361	31502	0.00558	0.00888	0.00330
S/soc-Slashdot0811	77360	905468	0.01247	0.01529	0.00282
S/wiki-Talk	2394385	5021410	0.01680	0.01941	0.00261
S/ca-HepPh	12008	237010	0.00178	0.00400	0.00222
S/email-EuAll	265214	420045	0.00010	0.00213	0.00203
N/cond-mat-2003	31163	240058	0.00427	0.00606	0.00179
S/p2p-Gnutella31	62586	147892	0.05989	0.06141	0.00152
S/soc-Slashdot0902	82168	948464	0.01174	0.01321	0.00147
S/email-Enron	36692	367662	0.00353	0.00454	0.00101
S/ca-CondMat	23133	186936	0.00719	0.00803	0.00084
S/soc-sign-Slashdot081106	77357	516575	0.02105	0.02175	0.00070
S/web-NotreDame	325729	1497134	0.00134	0.00182	0.00048
S/amazon0312	400727	3200440	0.00425	0.00361	0.00064
N/cond-mat-2005	40421	351382	0.00428	0.00484	0.00056
N/astro-ph	16706	242502	0.00328	0.00383	0.00055
S/amazon0302	262111	1234877	0.00029	0.00083	0.00054
S/amazon0601	403394	3387388	0.00036	0.00085	0.00049
S/soc-sign-Slashdot090216	81871	545671	0.01723	0.01769	0.00046
S/soc-sign-Slashdot090221	82144	549202	0.01723	0.01769	0.00046
N/power	4941	13188	0.00027	0.00055	0.00028
S/ca-GrQc	5242	28980	0.00187	0.00206	0.00019
N/cond-mat	16726	95188	0.00718	0.00729	0.00011
S/amazon0505	410236	3356824	0.00070	0.00078	0.00008
S/web-Google	916428	5105039	0.00043	0.00044	0.00001

$\phi < \psi$ so one can check $\mu - r > 0$ as a preliminary stopping criterion that is faster to evaluate than ϕ . This reduces the number of times conductance must be evaluated.

Table IV shows the convergence history for the graph Newman/polbooks, which is co-purchasing Network among political books. We see that 5 iterations of the ARNOLDI process gives a good conductance cut $\phi = 0.5250$. However, we see that the numerical accuracy is poor $r \approx 0.483$. At 10

TABLE II

SOME ITERATES OF THE SOLVER ARE SHOWN IN DETAIL FOR THE GRAPH S/WEB-GOOGLE. BECAUSE THEOREM 2 REQUIRES THAT $\mu - r \geq 0$, ψ IS NOT SUPPLIED WHEN THIS CONDITION IS NOT MET. WE FIND THAT $\mu - r > 0$ IS A GOOD PROXY FOR THE STOPPING CRITERION $\phi \leq \psi$, WHICH IMPLIES THAT ONE CAN USE THIS TEST IN ORDER TO REDUCE THE NUMBER OF TIMES THAT ϕ MUST BE COMPUTED.

iter	ϕ	μ	$\sqrt{2\mu}$	ψ	r
20	0.02283	0.00881	0.13276	–	0.01900
30	0.00571	0.00445	0.09435	–	0.01110
40	0.00399	0.00277	0.07438	–	0.00755
50	0.00297	0.00193	0.06209	–	0.00584
60	0.00203	0.00144	0.05372	–	0.00423
70	0.00153	0.00117	0.04837	–	0.00347
80	0.00143	0.00101	0.04493	–	0.00267
90	0.00100	0.00090	0.04248	–	0.00234
110	0.00100	0.00074	0.03859	–	0.00194
120	0.00062	0.00069	0.03703	–	0.00169
130	0.00062	0.00064	0.03579	–	0.00158
140	0.00062	0.00061	0.03483	–	0.00142
160	0.00062	0.00056	0.03339	–	0.00120
170	0.00071	0.00054	0.03288	–	0.00111
250	0.00071	0.00046	0.03030	–	0.00085
260	0.00083	0.00045	0.03006	–	0.00079
620	0.00083	0.00041	0.02878	0.02554	0.00009
630	0.00083	0.00041	0.02877	0.02552	0.00009

iterations the conductance ($\phi = 0.4762$) matches the optimal value and the numerical accuracy is still poor $r \approx 0.00161$. At 20 iterations the numerical accuracy is high $r \approx 10^{-7}$. We can see that between the 5th and 10th iteration, several vertices move across the sweep cut boundary (set Δ column), and that the ordering among the vertices changes (ρ column which is the Pearson correlation between the orderings). However, no vertices cross the sweep cut boundary between the 10th and 20th iteration despite a change in the ordering ($\rho \neq 1$). This indicates that cluster structure can be detected before numerically accurate eigenvectors are computed.

The convergence behavior in terms of conductance is not monotonic, which indicates that the conductance should be evaluated throughout the iterations. Table V shows the iteration details for the Newman/polblogs graph. We see that the optimal cut is found after 10 minor iterations of the Arnoldi process. After applying an additional 10 minor iterations, we see an increase in conductance⁴. One can improve the performance of a solver by checking the conductance at each step and taking the minimum.

C. Effect of Eigengap

We perform regression analysis on the experimental results. We know from standard error analysis that $e < \sqrt{2\frac{r}{\delta}}$ [11], [17]. This error bound tells us that the difficulty of an eigenproblem is controlled by the eigengap δ . Thus we focus on the dependence of our method on the eigengap δ . Regression analysis of iteratio, $\frac{I_F}{I_C}$, and phiratio, $\frac{\phi_C}{\phi_F}$, as a function of gap, δ is shown in Figure 1. One can see that there is a linear relationship between $\log \delta$ and $\frac{I_F}{I_C}$ showing that as the problem gets harder our stopping criterion saves more iterations. Based

⁴This matches the final conductance of ARPACK with a tolerance of 10^{-15} .

TABLE IV
THE COMBINATORIAL STRUCTURE IS REVEALED FASTER THAN
NUMERICAL STRUCTURE.

iter	stop	resid	μ	ϕ	$ S $	ρ	set Δ
5	8	0.048826	0.04177	0.05250	55	-	-
10	0	0.0016126	0.03781	0.04762	52	0.80038	3
20	0	7.8948e-07	0.03780	0.04762	52	0.90443	0

TABLE III

IMPROVEMENT SHOWN FOR EACH GRAPH AVERAGED OVER 10 RUNS. THE COLUMNS ARE THE GRAPH NAME, ITERATIONS UNTIL $\phi < \psi$, ITERATIONS UNTIL $r < tol$, RATIO OF $\phi(\mathbf{x}^{I_C})$ TO $\phi(\mathbf{x}^{I_F})$, THE CONDUCTANCE FOUND WHEN $r < tol$, AND THE RESIDUAL UPON EARLY TERMINATION. THIS TABLE SHOWS THE GRAPHS WHERE OUR CONVERGENCE CRITERION WAS REACHED WITHIN THE MAXIMUM NUMBER OF ITERATIONS.

Graph	I_C	I_F	$\frac{\Phi_C}{\Phi_F}$	Φ_F	Residual
N/adjnoun	10.0	42.0	1.0515	0.4615	0.0407
N/as-22july06	17.0	132.0	1.7490	0.0298	0.0363
N/astro-ph	54.5	315.5	1.2273	0.0046	0.0031
N/celegansneural	10.0	24.5	1.1272	0.2258	0.0345
N/cond-mat	43.0	487.5	0.9184	0.0152	0.0072
N/cond-mat-2003	67.0	225.5	1.0948	0.0064	0.0041
N/cond-mat-2005	59.0	329.5	1.1051	0.0064	0.0041
N/dolphins	14.5	20.0	1.0000	0.0682	0.0011
N/football	10.0	29.5	1.1107	0.1207	0.0254
N/hep-th	46.0	187.0	1.0331	0.0250	0.0067
N/karate	10.0	15.0	1.0000	0.1515	0.0049
N/lesmis	11.5	31.0	0.9916	0.1526	0.0187
N/netscience	46.5	108.5	1.1571	0.0048	0.0022
N/polblogs	12.0	30.0	0.7371	0.1250	0.0220
N/polbooks	14.5	20.0	0.9973	0.0476	0.0027
N/power	236.5	785.5	0.9240	0.0025	0.0003
S/Oregon-1	14.5	83.5	1.3502	0.0685	0.0372
S/Oregon-2	14.5	87.5	1.3659	0.0489	0.0379
S/amazon0302	233.5	493.0	1.0324	0.0008	0.0003
S/amazon0312	103.0	800.0	1.2125	0.0018	0.0017
S/amazon0505	147.0	800.0	1.0702	0.0011	0.0007
S/amazon0601	229.5	641.5	1.0724	0.0006	0.0004
S/as-735	11.5	95.0	1.3674	0.0651	0.0531
S/as-caida	30.5	115.5	1.0527	0.0302	0.0121
S/ca-AstroPh	37.5	124.5	1.1896	0.0102	0.0069
S/ca-CondMat	41.0	155.0	1.1845	0.0109	0.0063
S/ca-GrQc	60.0	199.0	1.4333	0.0025	0.0028
S/ca-HepPh	62.0	138.5	1.2651	0.0024	0.0019
S/ca-HepTh	58.5	168.5	1.7388	0.0036	0.0029
S/cit-HepPh	21.5	89.0	1.3945	0.0357	0.0250
S/cit-HepTh	18.5	143.0	1.6112	0.0312	0.0267
S/email-Enron	58.0	188.0	1.0037	0.0045	0.0029
S/email-EuAll	158.0	259.0	1.0000	0.0001	0.0001
S/p2p-Gnutella04	17.5	33.0	1.0000	0.0455	0.0124
S/p2p-Gnutella05	10.0	50.5	1.2796	0.1500	0.0857
S/p2p-Gnutella06	10.0	49.0	1.0991	0.2222	0.0947
S/p2p-Gnutella08	12.5	46.5	1.2493	0.0500	0.0486
S/p2p-Gnutella09	15.0	39.0	1.0000	0.0429	0.0205
S/p2p-Gnutella24	10.0	55.5	1.7662	0.1000	0.0875
S/p2p-Gnutella25	13.5	61.0	1.5241	0.0435	0.0511
S/p2p-Gnutella30	10.0	70.0	1.3673	0.1000	0.0762
S/p2p-Gnutella31	10.0	142.0	1.0657	0.1111	0.0878
S/soc-Epinions1	49.5	122.0	1.2661	0.0061	0.0050
S/soc-Slashdot0811	18.5	143.5	2.2980	0.0135	0.0305
S/soc-Slashdot0902	12.5	155.5	4.0594	0.0119	0.0477
S/soc-sign-Slashdot081106	21.5	131.0	1.8164	0.0238	0.0295
S/soc-sign-Slashdot090216	25.0	124.5	1.5721	0.0192	0.0178
S/soc-sign-Slashdot090221	25.5	121.0	1.2030	0.0192	0.0157
S/soc-sign-epinions	32.0	155.0	1.5445	0.0138	0.0132
S/web-Google	278.0	800.0	0.9728	0.0007	0.0004
S/wiki-Talk	13.5	156.5	2.0337	0.0321	0.0650
S/wiki-Vote	10.0	30.5	1.0000	0.1250	0.0580
Average	49.1	189.4	1.2432	0.0557	0.0246

TABLE V

CONDUCTANCE CAN INCREASE WITH ADDITIONAL ITERATIONS

iter	stop	resid	μ	ϕ	$ S $	ρ	set Δ
5	0	0.14468	0.15164	0.09436	687	-	-
10	0	0.03402	0.09225	0.08490	664	0.57271	73
20	0	0.00173	0.08144	0.12500	4	0.57252	561
30	0	8.4846e-07	0.08144	0.12500	4	0.57286	0

on the lack of a linear relationship between $\log \delta$ and $\frac{\phi_C}{\phi_F}$, there is no evidence to suggest that our method loses approximation quality on harder problems.

D. Alternative Stopping Criteria

An alternative approach to stopping criteria comes from examining changes in the objective function. A common approach in data mining is to stop when the objective function stops decreasing between iterations. This comes from an idea that changes in the objective function are monotonically decreasing in magnitude. This strategy works for some machine learning problems and appears to be a good convergence criteria, however in this case $\phi(\mathbf{x}^i)$ can stall out. More precisely, for some iteration i , $\phi(\mathbf{x}^i) - \phi(\mathbf{x}^{i+1}) = 0$ while there exists a j much larger than i such that $\phi(\mathbf{x}^j) < \phi(\mathbf{x}^i)$. Therefore,

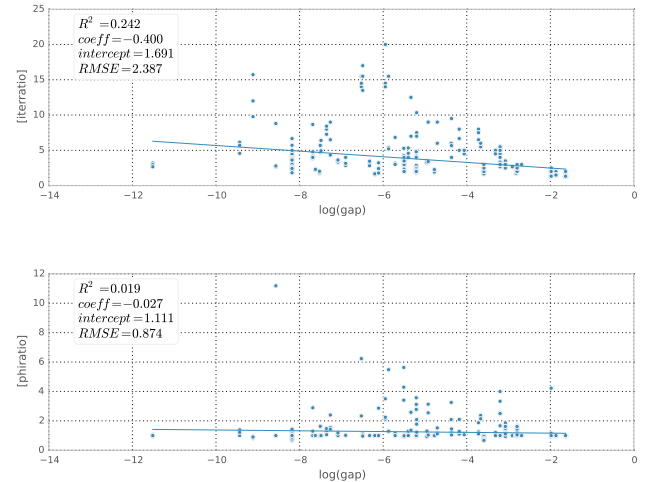


Fig. 1. Results for individual runs on each graph, not averaged, are plotted. The iteration savings $\frac{I_F}{I_C}$, shown on top, increase as the gap decreases and the loss in conductance $\frac{\phi_C}{\phi_F}$, shown on bottom, does not increase as the gap decreases.

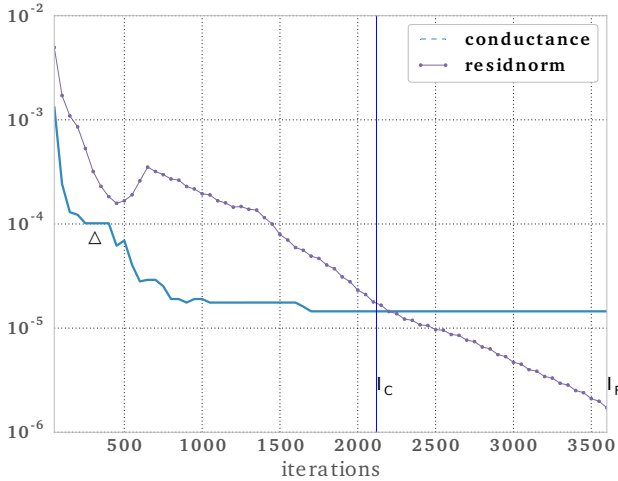


Fig. 2. The decrease in conductance and residual for SNAP/web-Notre-Dame is shown. The first flat region of the conductance curve (marked Δ) indicate where one would stop based on $\phi(\mathbf{x}^i) - \phi(\mathbf{x}^{i+1}) = 0$. This shows that stopping when the objective function stops decreasing does not achieve low conductance partitions for this problem. Using our criterion and stopping at I_C iterations does achieve a low conductance cut.

stopping when the conductance stops improving yields large loss in conductance. This occurs for ill conditioned problems such as the SNAP/Web-Notre-Dame graph as illustrated in Figure 2.

E. Balance

The theorems used in this paper do not guarantee the existence of balanced cuts, which are favored in applications. One can form a multi-way partition by recursively constructing 2-way partitions until reaching sufficiently small parts. If the smaller part of the 2-way partition contains more than pn vertices for large enough $0 < p < 1/2$, this creates a balanced, low depth recursion tree. In order to understand the effects of our stopping criterion on choosing a balanced cut we make a similar comparison as Section IV-B while restricting to balanced cuts. We introduce the notation ϕ^p as the best sweep cut which contains at least p percent of the vertices on each side. Table VI shows the minimum, maximum, and mean taken over all graphs, of the best sweep cut of the final iterate ϕ_F^0 , the best balanced sweep cut of the final iterate ϕ_F^{10} , and the ratio between the conductance of the best balanced sweep cut of the I_C -th iterate ϕ_C^{10} . The value of ϕ_C^0 and ϕ_F^0 for each graph is denoted ϕ_C and ϕ_F in Table III. For balanced cuts, the increase in conductance due to our stopping criterion is on average less than a factor of 1.3 and for over half the graphs our method finds the same conductance, which suggests that our methods works well in practice even with a balance condition. In contrast, for half the graphs, the ratio of conductance between the best cut and the best balanced cut for the final iteration is at least 4.5. In other words, the cost of restricting to a balanced cut exceeds the cost due to the novel stopping criterion.

TABLE VI

THE EFFECT OF RESTRICTING TO BALANCED CUTS IS SHOWN. THE MAX, MEAN, AND MIN ARE TAKEN OVER ALL GRAPHS. ϕ_F^{10} REPRESENTS THE CONDUCTANCE OF THE BEST CUT OF THE FINAL COMPUTED EIGENVECTOR WITH AT LEAST $\frac{p}{10}$ VERTICES ON EACH SIDE. ϕ_C^{10} REPRESENTS THE CONDUCTANCE OF THE BEST BALANCED CUT ACHIEVED WITH OUR STOPPING CRITERION.

	ϕ_F^0	ϕ_F^{10}	$\frac{\phi_C^{10}}{\phi_F^{10}}$	$\frac{\phi_F^{10}}{\phi_F^0}$
min	0.0001	0.0017	0.4111	1.0000
median	0.0455	0.2080	1.0000	4.5926
mean	0.0736	0.2732	1.2557	36.829
max	0.6133	0.8245	5.0514	1001

V. CONCLUSIONS

This work fits into a larger context of understanding the connection between numerical accuracy of solvers and network analysis quality. We measure the solver accuracy using the norm of the residual vector and the network analysis quality as the conductance of the resulting partition. We show that by analyzing of both measures, one can rigorously derive improved parameter free stopping criterion. This stopping criterion is empirically validated on real world networks. The result of our analysis is a large reduction in the number of iterations used to solve this network analysis problem with iterative methods. This leads to faster methods for large problems. Understanding the relationship between the numerical method and the network analysis method leads to an algorithmic improvement. We demonstrate this improvement on real world networks and using an advanced eigensolver. Empirically, simpler convergence criteria based on intuition do not achieve factor of two approximations to the prior work.

This paper draws the following quantitative conclusions.

- 1) Our stopping criterion for spectral partitioning leads to a 4.15 fold decrease in iteration with only a 1.24 fold increase in resulting conductance.
- 2) A practical choice for the residual tolerance parameter of eigensolvers for low conductance partitioning is 10^{-4} .
- 3) Alternative stopping criteria fail to ensure high quality solutions, as shown on the S/web-Google graph.
- 4) When imposing a balance condition of 10% on the cuts, stopping using our criterion increases the conductance by a factor of 1.4 on average compared to using the high fidelity eigenvectors.
- 5) Analyzing the performance of our method as a function of spectral gap δ indicates that our method reduces cost more on harder problems.

By analyzing the numerical accuracy of iterative methods along with the network analysis objective function, we are able to gain new insights. This opens new questions about the relationship to multiway cuts and higher dimensional techniques.

When applying the proof of $\psi \leq \sqrt{2\lambda_2}$, one must ensure $|\mu - \lambda_2| = \min_i |\mu - \lambda_i|$. While we show empirically that stopping when $\phi < \psi$ provides low conductance cuts, the proof used in the guarantee does not apply without that

hypothesis. For some difficult problems this condition is not ensured by the IRAM solver. Further study of approximating arbitrary linear combinations of the low energy eigenvectors will deepen our understanding of these techniques.

ACKNOWLEDGMENTS

The authors thank Geoffrey D. Sanders for helpful comments. This work is supported in part by the National Defense Science and Engineering Graduate Fellowship and by Defense Advanced Research Projects Agency (DARPA) under agreement #HR0011-13-2-0001 (DARPA PERFECT). The content, views and conclusions presented in this document do not necessarily reflect the position or the policy of DARPA or the U.S. Government, no official endorsement should be inferred.

REFERENCES

- [1] Christos Boutsidis, Alex Gittens, and Anju Kambandur. Spectral clustering via the power method - provably. In *Proceedings of The 32nd International Conference on Machine Learning*, pages 40–48, 2015.
- [2] Aydin Buluç, Henning Meyerhenke, Ilya Safro, Peter Sanders, and Christian Schulz. Recent advances in graph partitioning. *CoRR*, abs/1311.3144, 2013.
- [3] Fan Chung and Olivia Simpson. Computing heat kernel pagerank and a local clustering algorithm. *arXiv preprint arXiv:1503.03155*, 2015.
- [4] Fan RK Chung. *Spectral graph theory*, volume 92. American Mathematical Soc., 1997.
- [5] Gianna M. Del Corso, Antonio Gullí, and Francesco Romani. Fast pagerank computation via a sparse linear system. *Internet Math.*, 2(3):251–273, 2005.
- [6] Peter G. Doyle and J. Laurie Snell. *Random walks and electric networks*, volume 22. Carus Mathematical Monographs, Mathematical Association of America, Washington, DC, 1984.
- [7] Ling Huang, Donghui Yan, Nina Taft, and Michael I. Jordan. Spectral clustering with perturbed data. In *Advances in Neural Information Processing Systems 21*, pages 705–712. Curran Associates, Inc., 2008.
- [8] U Kang, Brendan Meeder, Evangelos E. Papalexakis, and Christos Faloutsos. Heigen: Spectral analysis for billion-scale graphs. *IEEE Transactions on Knowledge and Data Engineering*, 26(2):350–362, 2014.
- [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] George Karypis and Vipin Kumar. A fast and high quality multi-level scheme for partitioning irregular graphs. *SIAM J. Sci. Comput.*, 20(1):359–392, 1998.
- [11] Tosio Kato. On the upper and lower bounds of eigenvalues. *Journal of the Physical Society of Japan*, 4(4-6):334–339, 1949.
- [12] Efi Kokiofoulou, Jie Chen, and Yousef Saad. Trace optimization and eigenproblems in dimension reduction methods. *Numerical Linear Algebra with Applications*, 18(3):565–602, 2011.
- [13] Richard Lehoucq and Daniel Sorensen. Deflation techniques for an implicitly restarted Arnoldi iteration. *SIAM Journal on Matrix Analysis and Applications*, 17(4):789–821, 1996.
- [14] Jure Leskovec and Andrej Krevl. SNAP Datasets: Stanford large network dataset collection. <http://snap.stanford.edu/data>, June 2014.
- [15] Milena Mihail. Conductance and convergence of Markov chains—a combinatorial treatment of expanders. In *30th Annual Symposium on Foundations of Computer Science*, volume 89, pages 526–531, 1989.
- [16] Mark Newman. Newman Datasets. <http://www-personal.umich.edu/~mejn/netdata/>, April 2013.
- [17] Beresford N. Parlett. *The Symmetric Eigenvalue Problem*. Society for Industrial and Applied Mathematics, 1998.
- [18] 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.
- [19] Yousef Saad. *Numerical Methods for Large Eigenvalue Problems: Revised Edition*, volume 66. Siam, 2011.
- [20] Daniel. C. Sorensen. *Implicitly restarted Arnoldi/Lanczos methods for large scale eigenvalue calculations*. Hampton, VA, 1995.
- [21] Daniel A. Spielman and Shang-Hua Teng. Spectral partitioning works: Planar graphs and finite element meshes. In *Proceedings., 37th Annual Symposium on Foundations of Computer Science.*, pages 96–105. IEEE, 1996.
- [22] Luca Trevisan. Lecture notes on expansion, sparsest cut, and spectral graph theory. University of California, Berkeley, 2014.
- [23] Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.
- [24] 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*, pages 907–916, 2009.