A Spectral Algorithm with Applications to Exploring Data Graphs Locally

Michael W. Mahoney * Lorenzo Orecchia † Nisheeth K. Vishnoi ‡

Abstract

We provide an optimization characterization of a *local* version of the traditional spectral optimization problem used for finding graph cuts. Rather than computing an approximation to the best partition in the entire input graph, we are motivated by the problem of computing an approximation to the best partition *near* an input seed set. Such a primitive seems quite useful in improving and refining clusters locally in many settings such as data graphs, machine learning and image segmentation. From a theoretical perspective, we show that the solution to our non-convex optimization relaxation for this problem, referred to as LocalSpectral, may be computed as the solution to a system of linear equations. Interestingly, this also provides an optimization characterization of a generalization of Personalized PageRank vectors, which could be of independent interest. We also show that how to obtain Cheeger-like quality-of-approximation guarantees when we round the solution to LocalSpectral; and that we can use these ideas to refine and improve given partitions in an input graph. We empirically illustrate the use of LocalSpectral in the analysis of data graphs.

1 Introduction

In this paper we consider the following problem: given a graph G, a cut $(T, \bar{T})$ in it, find a cut of minimum conductance[1] in G which is well-correlated with T or certify that there is none. The conductance of a set T is defined as $|E(T,\bar{T})|/\mathrm{vol}(T)\cdot\mathrm{vol}(\bar{T})$, where $\mathrm{vol}(T)$ is the sum of degrees of the nodes in T. Though one can imagine many applications of this primitive, the main application of this we consider is to explore clusters around a given *seed* set of nodes in data graphs. Before we make the notion of correlation precise we discuss the other choices we make. Graph conductance is a common measure of the quality of a cluster in several areas such as machine learning, data analysis, image segmentation and to infer the quality of a community in social graphs. This justifies our choice of the quantity we would like to minimize (conductance) among all cuts that are κ-correlated with the input cut [1, 2, 3].

Unfortunately, this makes our problem NP-hard as it generalizes the problem of computing the least conductance cut in a graph [4]. There are several ways to write efficient relaxations for graph conductance. Among these, the spectral method which relaxes the graph conductance problem to an eigenvalue problem enjoys popularity and success in the machine learning and data analysis community. Here, the algorithm is to compute the second eigenvector of the normalized laplacian of the input graph and output the best *sweep cut* of this eigenvector [5]. Given the pervasiveness of spectral methods, we choose to relax our problem in the spectral world by augmenting the standard spectral program with an extra *correlation constraint* around the input seed set. This allows us to build up on our understanding and use the tools available for spectral methods.

*Department of Mathematics, Stanford University, Stanford, CA 94305. mmahoney@cs.stanford.edu.

†Computer Science Division, UC Berkeley, Berkeley, CA, 94720. orecchia@eecs.berkeley.edu.

‡Microsoft Research, Bangalore, India. nisheeth.vishnoi@gmail.com.

[1] Conductance of a cut is the same as its normalized-cut value up to a constant.

The correlation. Given a cut $(T, \bar{T})$ in a graph G, a natural vector in $\mathbb{R}^V$ to associate with it is its characteristic vector; this takes value 1 for nodes in T and 0 otherwise. One way to capture its correlation with another cut $(U, \bar{U})$ is by the inner product of the characteristic vectors of the two cuts. It turns out that a more refined vector to associate with a cut is the vector obtained after removing from the characteristic vector its projection along the all 1 vector. Thus, the notion of correlation we consider is, roughly[2], the inner product of two such vectors for two cuts. To make it well defined we square the inner product and we normalize vectors so that correlation is in $[0, 1]$.

More generally, however, this seed vector need not have a simple interpretation as a subset of nodes. In this latter case, one might be interested in computing cuts that are correlated with certain structural parameters of the graph, *e.g.*, the degree-distribution, that are defined on the set of nodes.

Our relaxation. First observe that, given a graph $G = (V, E)$ where the degree of a vertex i is d_i, the standard spectral algorithm asks for a vector x which minimizes ${\sum_{ij \in E}(x_i - x_j)^2}/{\sum_{i \in V} d_i x_i^2}$ over vectors x such that $\sum_{i \in V} d_i x_i = 0$. Given the seed vector s and a correlation parameter κ, we add an extra constraint on x to model our optimization problem: we will be interested in optimizing the above when, in addition, $(\sum_{i \in V} d_i x_i s_i)^2 \geq \kappa$. Here, $(\sum_{i \in V} d_i x_i s_i)^2$ is precisely the correlation between the seed vector s and the variable vector x. We refer to this optimization problem as $\mathsf{LocalSpectral}(G, s, \kappa)$. Even though this formulation is simple and natural, to the best of our knowledge, it has not been considered in the literature before this work. In addition to providing this optimization formulation of a local version of the traditional spectral partitioning problem, in this paper we provide both a theoretical analysis of our formulation and an empirical evaluation of how the method can be applied in practice. We first outline our theoretical results.

Main theoretical contributions. First note that $\mathsf{LocalSpectral}$ is a non-convex optimization problem. This leaves us with the task of solving it which was just an eigenvalue computation in the case of standard spectral methods. Somewhat interestingly, our first result shows that the optimal of $\mathsf{LocalSpectral}$, under reasonable assumptions on s and κ, reduces to solving a system of linear equation. In fact, we are able to show that the optimal of $\mathsf{LocalSpectral}$ is a Generalized Personalized PageRank (GPPR) vector. Given a vector s and a *teleportation* constant α the Personalized PageRank vector of s with respect to the graph G is roughly defined as $(L_G - \alpha I)^{-1} s$. The optimal solution to $\mathsf{LocalSpectral}$ is proved to satisfy such an equation (see Theorem 3). Here L_G is the combinatorial laplacian of G. Interestingly, the proof of this theorem goes via relaxing the non-convex optimization problem as a natural semi-definite programming problem and invoking strong duality to establish that the relaxation is tight. Not only does this seem elegant from a theoretical perspective, but in addition it provides a systematic way to implement and interpret empirical evaluations.

Our next result then invokes general version of Mihail's proof of Cheeger Inequality [6] to prove that one can do a sweep cut rounding on the GPPR produced by $\mathsf{LocalSpectral}$ to obtain a cut which is *quadratically* within the optimal value. The ease with which this applies is another place where the power or spectral methods is to be felt (see Theorem 4.)

Finally we show, again using strong duality of the same SDP used in the proof of the PageRank characterization, that the optimal value of $\mathsf{LocalSpectral}$ lower bounds the conductance of cuts as a function of how well they are correlated with the seed vector (see Theorem 5). This primitive can be used to explore and possibly improve cuts around a given seed cut.

[2]The inner product we consider is with respect to degrees of the graph: for vector x, y, their inner product is $\sum_{i \in V} d_i x_i y_i$.

Main empirical contributions. The goal of our empirical analysis is to illustrate how our method can be applied to explore cuts in an input data graph. We consider a coauthorship network of scientists as compiled by Newman [2]. This graph consists of 379 nodes and 914 edges, with each node representing an author and each edge representing a coauthorship relation, and is a standard benchmark for the detection of communities because it combines certain structural heterogeneities that are typical of larger social graphs with a size that permits a more detailed and comprehensive analysis.

Our main observations are the following. First, we show how varying the teleportation parameter allows us to detect low-conductance cuts at different volumes around a seed vertex and how this information, aggregated over multiple choices of teleportation, can improve our understanding of the network structure in the neighborhood of the seed. Secondly, we bring evidence in support of our definition of GPPR by displaying specific instances in which GPPR is more effective than the Personalized PageRank at detecting low-conductance cuts at a given volume. Finally, we demonstrate how our method can find low-conductance cuts that are well-correlated to a input vector by giving an interesting application to the detection of sparse peripheral regions of the network. We believe our method may find applications in leveraging the additional feature data, which often comes with the vertices of a data network, to find interesting and meaningful cuts.

Related work. There are several lines of work that are related to ours. Here we only briefly discuss them and a more detailed comparison appears in Appendix 6.2. First, researchers have been interested in the cut improvement problem: given as input a graph and a cut, find a subset that is a better cut. This began with Gallo, Grigoriadis and Tarjan [7] and culminated with Andersen and Lang [8], who gave a general algorithm that uses a small number of single-commodity maximum-flows to find low-conductance cuts not only inside the input subset T, but among all cuts which are well-correlated with $(T, \bar{T})$. Our work may be viewed as a spectral analogue of this line of work. Second, local graph partitioning was introduced by Spielman and Teng [9], who were interested in finding a low-conductance cut in a graph in time nearly-linear in the volume of the output cut. They used random walk based methods to do this; and subsequently this result was improved by Andersen, Chung and Lang [10] by doing certain Personalized PageRank based random walks. Our approach is an axiomatic and optimization based, while this line of work was motivated by local graph partitioning and does not address the problem of discovering cuts near given seed vectors. Third, these flow-improvement and local partitioning methods have been used to characterize the clustering and community structure in large networks [11, 3, 12]. Our optimization program and empirical results suggest that this line of work can be extended to ask much more refined questions about graph structure near a seed vector.

Organization of the paper. In the next section, we will describe background and notation, and then, in Section 3, we will describe our formulation of local spectral graph partitioning as an optimization problem. Then, in Section 4, we will describe our main theoretical results, and in Section 5 we will provide an empirical evaluation of our main results.

2 Background and Notation

Graphs and Laplacians. Let $G = (V, E)$ be a simple, connected, and undirected graph with $V = [n]$ vertices and $m = |E|$ edges. Then, $A_G \in \mathbb{R}^{V \times V}$ will denote its adjacency matrix; D_G will denote the diagonal degree matrix of G, *i.e.*, $D_G(i, i) = d_i$, where d_i is the degree of vertex

i, and $D_G(i,j) = 0$, for all $i \neq j$; $L_G \stackrel{\text{def}}{=} D_G - A_G$ will denote the (combinatorial) laplacian of G; and $\mathcal{L}_G \stackrel{\text{def}}{=} D^{-1/2} L_G D^{-1/2}$ will denote the (normalized) laplacian of G. Throughout this paper, we will assume G is connected, in which case the eigenvalues of $\mathcal{L}_G$ are $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_n$. We will denote by $\lambda_2(G)$ as this second eigenvalue of the normalized laplacian of G. If $u_1, \ldots, u_n$ are the corresponding eigenvectors of $\mathcal{L}_G$, then we define $v_i \stackrel{\text{def}}{=} D^{-1/2} u_i$ and think of them as the associated eigenvectors of L_G.

For two vectors $x, y \in \mathbb{R}^n$, and the degree matrix D for a graph G, we define $\langle x, y\rangle_D \stackrel{\text{def}}{=} \sum_{i=1}^n x_i y_i d_i$. Given a subset of vertices $S \subseteq V$, we denote by 1_S the indicator vector of S in $\mathbb{R}^n$ and by 1 the vector in $\mathbb{R}^n$ having all entries set equal to 1. In addition, we denote by $\mathcal{S}_D$ the set of vectors $\{x \in \mathbb{R}^n : \langle x, 1\rangle_D = 0\}$ where D is the diagonal matrix of degrees of the vertices of the instance graph G. We consider the following definition of the complete graph K_n on the vertex set V $A_{K_n} \stackrel{\text{def}}{=} \frac{1}{\text{vol}(G)} DJD$. Notice that this is not the standard complete graph, but a weighted version of it, where the weights depend on D. With this scaling we have $D_{K_n} = D$. Hence, the laplacian of the complete graph defined in this manner becomes $L_{K_n} = D - \frac{1}{\text{vol}(G)} DJD$.

Graph Conductance. For a set of vertices $S \subseteq V$ in a graph, the *volume* of S is $\text{vol}(S) \stackrel{\text{def}}{=} \sum_{i \in S} d_i$, in which case $\text{vol}(G)$ is defined as $\text{vol}(V) = 2m$. In this paper, the *conductance* $\phi(S)$ *of a cut* $(S, \bar{S})$ [3] is $\phi(S) \stackrel{\text{def}}{=} \text{vol}(G) \cdot \frac{|E(S,\bar{S})|}{\text{vol}(S)\cdot\text{vol}(\bar{S})}$. The *conductance of the graph* G is $\phi(G) = \min_{S \subseteq V} \phi(S)$. Note that $\phi(G)$ is defined so that it is *dimension-less* quantity in the sense that it does not change if each edge in the graph is replaced by the same number of multiple edges.

Matrices. For a symmetric matrix A, we will use $A \succeq 0$ to denote that it is positive semi-definite. Moreover, given two symmetric matrices A and B, the expression $A \succeq B$ will mean $A - B \succeq 0$. Further, for two $n \times n$ matrices A and B, we let $A \circ B$ denote $\text{Tr}\,(A^T B)$. Finally, for a matrix A, let A^+ denote its (uniquely defined) Moore-Penrose pseudo-inverse.

Generalized PageRank Vectors. Given a graph $G = (V, E)$, a number $\alpha \in (-\infty, \lambda_2(G))$ and any vector $s \in \mathbb{R}^n$ such that $s \in \mathcal{S}_D$, a *Generalized Personalized PageRank (GPPR)* vector is any vector of the form $(L_G - \alpha L_{K_n})^+ Ds$.

3 Local Spectral Partitioning as an Optimization Problem

3.1 Background on Spectral Algorithm for Partitioning Graphs

Recall that the basic global graph partitioning problem is: given as input a graph $G = (V, E)$ find a set of nodes $S \subseteq V$ to solve $\phi(G) = \min_{S \subseteq V} \phi(S)$. Spectral methods approximate the solution to this problem by solving the relaxed problem Spectral presented in Figure 1.

Recall that $x^T L_G x$ counts the number of edges crossing the cut and $\langle x, x\rangle_D = 1$ encodes a variance constraint. It is easy to see that the optimal value of this program, if G is connected, is $\lambda_2(G)$ and the optimal vector to this program is $D^{-1/2} u_2$, where u_2 is the eigenvector corresponding to the smallest non-zero eigenvalue ($\lambda_2(G)$) of $\mathcal{L}_G$. Thus, the running time of Spectral is that of solving the eigenvector problem $\mathcal{L}_G y = \lambda_2(G) y$, More importantly, this program is a relaxation of the conductance problem and a good cut can be recovered from a simple *sweep* of

[3] The conductance of a set S (or cut $(S, \bar{S})$) is often defined as $\phi'(S) = |E(S,\bar{S})| / \min\{\text{vol}(S), \text{vol}(\bar{S})\}$. This notion is equivalent to that $\phi(G)$, in that value $\phi(G)$ thereby obtained for the conductance of the graph G differs by no more than a factor of 2, depending on which notion we use for the conductance of a set.

$$\begin{array}{ll} \text{minimize} & x^T L_G x \\ \text{s.t.} & \langle x, x\rangle_D = 1 \\ & \langle x, 1\rangle_D = 0 \end{array} \qquad \begin{array}{ll} \text{minimize} & x^T L_G x \\ \text{s.t.} & x^T L_{K_n} x = 1 \\ & \langle x, s\rangle_D^2 \geq \kappa \end{array}$$

Figure 1: Left: Spectral relaxation $\mathsf{Spectral}(G)$. Right: Our relaxation $\mathsf{LocalSpectral}(G, s, \kappa)$.

the optimal vector v_2. This is captured by the following celebrated result often referred to as Cheeger's Inequality.

Theorem 1 (Cheeger's Inequality). *For a connected graph G, $\lambda_2(G)/2 \leq \phi(G) \leq \sqrt{8\lambda_2(G)}$.*

The first part of the inequality is rather straightforward and establishes the fact that $\lambda_2(G)$ provides a lower bound to $\phi(G)$. The second part of the inequality above is the non-trivial part which shows that $\lambda_2(G)$ is a good relaxation to $\phi(G)$ and there are many proofs known for it (see e.g. [5]). A particularly interesting proof of this was found due to Mihail [6] which is able to round any *test vector* rather than v_2 and achieve the same guarantee as Cheeger's Inequality. We record this in the following theorem which we will use later.

Theorem 2 (Sweep Cut Rounding). *Let x be a vector such that $\langle x, 1\rangle_D = 0$. Then there is a t for which the set of vertices $S := \mathsf{SweepCut}_t(x) \stackrel{\text{def}}{=} \{i : x_i \geq t\}$ satisfies $\frac{x^T L_G x}{x^T D x} \geq \phi^2(S)/8$.*

3.2 Geometric Correlation

In order to state our proposed optimization problem to find cuts around a given cut, it will be useful to associate vectors with cuts so as to provide a geometric interpretation of when two cuts (or equivalently sets of nodes) are close. To do so, given a set $T \subseteq V$, we define the unit vector s_T as $s_T(i) = \sqrt{\mathrm{vol}(T)\mathrm{vol}(\bar{T})/2m} \cdot 1/\mathrm{vol}(T)$, if $i \in T$, and $s_T(i) = -\sqrt{\mathrm{vol}(T)\mathrm{vol}(\bar{T})/2m} \cdot 1/\mathrm{vol}(\bar{T})$ if $i \in \bar{T}$; or equivalently, $s_T \stackrel{\text{def}}{=} \sqrt{\frac{\mathrm{vol}(T)\mathrm{vol}(\bar{T})}{2m}} \left(\frac{1_T}{\mathrm{vol}(T)} - \frac{1_{\bar{T}}}{\mathrm{vol}(\bar{T})}\right)$. It is easy to check that this is well defined: one can replace s_T by $s_{\bar{T}}$ and the correlation remains the same with any other set. Defined this way, it immediately follows that $\langle s_T, 1\rangle_D = 0$ and that $\langle s_T, s_T\rangle_D = 1$ (see Appendix 6.3.1 for proofs). Thus, $s_T \in \mathcal{S}_D$ for $T \subseteq V$, and s_T can be seen as an appropriately normalized version of the vector consisting of the uniform distribution over T minus the uniform distribution over $\bar{T}$. [4] Relatedly, we will consider the following measure of correlation between two sets of nodes, or equivalently between two cuts, say a cut $(T, \bar{T})$ and a cut $(U, \bar{U})$: $K(T, U) \stackrel{\text{def}}{=} \langle s_T, s_U\rangle_D^2$. The proofs of the following simple facts regarding $K(T, U)$ are omitted: $K(T, U) \in [0, 1]$; $K(T, U) = 1$ if and only if $T = U$ or $\bar{T} = U$; $K(T, U) = K(\bar{T}, U)$; and $K(T, U) = K(T, \bar{U})$. Note that although we have described this notion of geometric correlation in terms of vectors of the form $s_T \in \mathcal{S}_D$ that represent partitions $(T, \bar{T})$, we will also be interested in applying similar ideas to other vectors $s \in \mathcal{S}_D$ for which there is not such a simple interpretation in terms of cuts.

3.3 Proposed Local Spectral Optimization Problem

Given as input a graph $G = (V, E)$, a seed set $T \subseteq V$, and a *correlation* parameter $\kappa \geq 0$, our relaxation is the program $\mathsf{LocalSpectral}$ in Figure 1. Let $\lambda(G, s, \kappa)$ denote the optimal value of $\mathsf{LocalSpectral}(G, s, \kappa)$. From the perspective of this as an optimization problem, $\mathsf{LocalSpectral}$ can

[4] Notice also that $s_T = -s_{\bar{T}}$; thus, since in the following we are only going to consider quadratic functions of s_T, we can consider both s_T and $s_{\bar{T}}$ to be representative vectors for the cut $(T, \bar{T})$.

be viewed as being generated from $\mathsf{Spectral}$ by augmenting the latter with a *locality* constraint of the form $\langle x, s\rangle_D^2 \geq \kappa$. Due to its form, this constraint may be viewed as a *correlation* condition, in which the solution is required to be well-connected or to lie near the *seed* vector s. We will be mostly interested in seed vectors s which correspond to s_T for some small set T. Thus, just as an optimal vector x for the standard spectral relaxation of conductance involves minimizing the *mixing* in G (*i.e.*, $x^T L_G x$) for a given amount of *variance* (*i.e.*, $\langle x, x\rangle_D$), with $\mathsf{LocalSpectral}$ we want to do the same but subject to the additional constraint that the solution vector x is well-correlated (*i.e.*, $\langle x, s\rangle_D^2 \geq \kappa$) with the input seed vector s. Alternatively, in the special case that the seed set S consists of a single node, $\mathsf{LocalSpectral}$ can be seen as a spectral relaxation of the problem: given as input a graph $G = (V, E)$, an input node v, and a positive integer k, find a set of nodes $T \subseteq V$ to solve $\min_{T\subseteq V: v\in S, \mathrm{vol}(T)\leq k} \phi(T)$, *i.e.*, find the best conductance set of nodes of volume no greater than k that contains the input node v. This is captured by the following (proof in Appendix 6.3.1).

Lemma 1. *For $v \in V$, $\mathsf{LocalSpectral}(G, s_{\{v\}}, 1/k)$ is a relaxation of the problem of finding a minimum conductance cut T in G which contains the vertex v and is of volume at most k.*

Also note the following simple fact whose proof appears in Appendix 6.3.1.

Fact 1. *If $\langle s, 1\rangle_D = 0$, then any optimal solution $x^\star$ to $\mathsf{LocalSpectral}(G, s, \kappa)$ can be assumed to satisfy $\langle x^\star, 1\rangle_D = 0$.*

Since we will only plug in s in $\mathsf{LocalSpectral}$ such that $\langle s, 1\rangle_D = 0$, for the optimal solution $x^\star$, $x^{\star T} L_{K_n} x^\star = x^{\star T} D x^\star$. This is important, and explains our choice of picking s such that $\langle s, 1\rangle_D = 0$, as then we can straightforwardly apply Theorem 2.

4 Our Results

In this section, we show how $\mathsf{LocalSpectral}$ can be used to make statements about the quality of cuts *near* an input cut (in a manner analogous to how the standard spectral comes with provable guarantees on the quality of cuts it finds) in an efficient manner. Further discussion and interpretations of our techniques and results appear in Appendix 6.1. Formally, we prove the following theorems. First is a characterization of the optimal solution to $\mathsf{LocalSpectral}$ (proof in Section 4.1).

Theorem 3 (PageRank Characterization). *Let G be a connected graph and $s \in \mathbb{R}^n$ be such that $\langle s, 1\rangle_D = 0$, and $\langle s, D^{-1/2} u_2\rangle_D \neq 0$, where D is the degree matrix of G and u_2 is the second eigenvector of $\mathcal{L}_G$. Further, let $1 > \kappa \geq 0$ be a correlation parameter. Then, if $x^\star$ be an optimal solution to $\mathsf{LocalSpectral}(G, s, \kappa)$, there exists some $\gamma \in (-\infty, \lambda_2(G))$ and a constant c such that $x^\star = c(L_G - \gamma L_{K_n})^+ D s$. Further, for any $\varepsilon > 0$, an optimal solution to $\mathsf{LocalSpectral}(G, s, \kappa)$ of value at most $(1+\varepsilon)\cdot\lambda(G, s, \kappa)$ can be computed in time $\tilde{O}(m/\sqrt{\lambda_2(G)}\cdot\log(1/\varepsilon))$ using the Conjugate Gradient Method, or in time $\tilde{O}(m\log(1/\varepsilon))$ using the Spielman-Teng linear-equation solver [9].*

The next theorem follows from Theorem 2 and Theorem 3 and we omit the obvious proof.

Theorem 4 (Finding a Cut). *Let G be a graph and $s \in \mathbb{R}^n$ be such that $\langle s, 1\rangle_D = 0$ and $\langle s, D^{-1/2} u_2\rangle_D \neq 0$, where D is the degree matrix of G and u_2 is the second eigenvector of $\mathcal{L}_G$. Further, let $\kappa \geq 0$ be a correlation parameter. Then, if $x^\star$ be an optimal solution to $\mathsf{LocalSpectral}(G, s, \kappa)$, one can find a cut in G of conductance at most $\sqrt{8\cdot\lambda(G, s, \kappa)}$. Given the vector $x^\star$, this rounding takes time $O(n\log n)$ via computing a Sweep Cut of $x^\star$.*

minimize	$L_G \circ X$	maximize	$\alpha + \kappa\beta$
subject to	$L_{K_n} \circ X = 1$	subject to	$L_G \succeq \alpha L_{K_n} + \beta(Ds)(Ds)^T$
	$(Ds)(Ds)^T \circ X \geq \kappa$		$\beta \geq 0$

Figure 2: Left: Primal SDP relaxation of $\mathsf{LocalSpectral}(G, s, \kappa)$ $\mathsf{SDP}_p(G, s, \kappa)$. Right: Dual SDP relaxation of $\mathsf{LocalSpectral}(G, s, \kappa)$ $\mathsf{SDP}_d(G, s, \kappa)$

Our framework also easily allows us to also give a lower bound on conductance of cuts depending on their correlation with the *seed* vector. Roughly, it allows us to lower bound the conductance of an arbitrary cut T when the seed vector corresponds to a cut U in terms of the correlation between U and T. This is formalized in the next Theorem whose proof appears in Appendix 6.3.3.

Theorem 5 (Cut Improvement). *Let G be a graph and $s \in \mathbb{R}^n$ be such that $\langle s, 1\rangle_D = 0$, where D is the degree matrix of G. Further, let $\kappa \geq 0$ be a correlation parameter. Then for all sets $T \subseteq V$ such that $\kappa' \stackrel{\text{def}}{=} \langle s, s_T\rangle_D^2$, $\phi(T) \geq \lambda(G, s, \kappa)$ if $\kappa \leq \kappa'$ and $\phi(T) \geq \kappa'/\kappa \cdot \lambda(G, s, \kappa)$ if $\kappa' \leq \kappa$. In particular if $s = s_U$ for some $U \subseteq V$, then note that $\kappa' = K(U, T)$.*

4.1 Generalized Personalized PageRank as a Solution to LocalSpectral

In this section we prove Theorem 3 which shows that the Generalized Personalized PageRank vector of an undirected graph arises as an optimal solution to LocalSpectral modulo some claims whose proofs are moved to Appendix 6.3.2 and 6.3.3. The proof of Theorem 3 will involve a relaxation of the non-convex program LocalSpectral to a (convex) semi-definite program (SDP), *i.e.*, the variables will be distributions over vectors rather than the vectors themselves; and that the proof will imply that strong duality holds for this non-convex problem.

Proof. [of Theorem 3] Although the program $\mathsf{LocalSpectral}(G, s, \kappa)$ is *not* convex, one can relax it to $\mathsf{SDP}_p(G, s, \kappa)$ of Figure 2. Then, one can observe that strong duality holds for this SDP relaxation. Using strong duality and the complementary slackness conditions implied by it, one can argue that the $\mathsf{SDP}_p(G, s, \kappa)$ has a rank one unique optimal solution under the conditions of the theorem. This implies that the optimal solution of $\mathsf{SDP}_p(G, s, \kappa)$ is the same as the optimal solution of LocalSpectral. Combining this with the complementary slackness condition obtained from the dual $\mathsf{SDP}_d(G, s, \kappa)$ of Figure 2, one can derive that the optimal rank one solution is, up to a constant, GPPR vector with seed vector Ds and parameter γ, which must be set to ensure primal feasibility.
In more detail, the proof will proceed by establishing a sequence of claims. Consider $\mathsf{SDP}_p(G, s, \kappa)$ and its dual $\mathsf{SDP}_d(G, s, \kappa)$ (Figure 2).

Claim 1. *The primal $\mathsf{SDP}_p(G, s, \kappa)$ is a relaxation of the vector program $\mathsf{LocalSpectral}(G, s, \kappa)$.*

Claim 2. *Strong duality holds between $\mathsf{SDP}_p(G, s, \kappa)$ and $\mathsf{SDP}_d(G, s, \kappa)$.*

Claim 3. *The feasibility and complementary slackness conditions for a primal-dual pair $X^\star, \alpha^\star, \beta^\star$ listed in Figure 3 are sufficient for them to be an optimal solution.*

Claim 4. *These feasibility and complementary slackness conditions, coupled with the assumptions of the theorem, imply that $X^\star$ must be rank 1 and $\beta^\star > 0$.*

Now we complete the proof of the theorem. From the claim it follows that, $X^\star = x^\star x^{\star T}$ where $x^\star$ satisfies the equation $(L_G - \alpha^\star L_{K_n} - \beta^\star(Ds)(Ds)^T)x^\star = 0$. From the second complementary slackness condition in Figure 3, and the fact that $\beta^\star > 0$, we obtain that $\langle x^\star, s\rangle_D = \pm\sqrt{\kappa}$. Thus,

$$\begin{aligned} L_{K_n} \circ X^\star &= 1 \\ (Ds)(Ds)^T \circ X^\star &\geq \kappa \\ L_G - \alpha^\star L_{K_n} - \beta^\star (Ds)(Ds)^T &\succeq 0 \\ \beta^\star &\geq 0 \end{aligned} \qquad \begin{aligned} \alpha^\star(L_{K_n} \circ X^\star - 1) &= 0 \\ \beta^\star((Ds)(Ds)^T \circ X^\star - \kappa) &= 0 \\ X^\star \circ (L_G - \alpha^\star L_{K_n} - \beta^\star (Ds)(Ds)^T) &= 0 \end{aligned}$$

Figure 3: Left: Feasibility conditions. Right: Complementary slackness conditions.

$x^\star = \pm\beta^\star\sqrt{\kappa}(L_G - \alpha^\star L_{K_n})^+ Ds$. This proves that the optimal of our program is (up to a constant scaling) a GPPR. □

5 Empirical Evaluation

Algorithm description and implementation. In this section we show how to use the theoretical results, in particular, Theorem 3 and 4, to guide our empirical study. We refer to our cut-finding algorithm as LocalCut. The input parameters of LocalCut are a seed vector s, a teleportation parameter α and a size factor c. The execution of LocalCut starts by computing the GPPR of seed s at teleportation α. By Theorem 3, the resulting vector $x^\star$ is an optimal solution to LocalSpectral(G, s, κ_α) for some choice of κ_α. The resulting vector $x^\star$ is then rounded by outputting its sweep cut of least conductance to obtain good approximations as guaranteed by Theorem 4. In the case of a seed vector corresponding to a single vertex v, one can infer from Lemma 1 that $x^\star$ yields a lower bound to the conductance of cuts that contain v and have less than a certain volume k_α, while the sweep-cut rounding does not give a specific guarantee on the volume of the output cut. However, we found that it was often possible to find a low-conductance cut in the range dictated by k_α. To exploit this observation, we use a new input parameter, a *size factor* $c > 0$, that regulates the maximum volume of the sweep cuts considered when s represents a single vertex. In this case, LocalCut considers only sweep cuts of volume at most $c \cdot k_\alpha$ that contain the vertex v [5]. This new input parameter turns out to be extremely useful in exploring cuts at different sizes, as it neglects sweep cuts of low conductance at large volume and allows us to pick out more local cuts around the seed vertex.

Network. We tested LocalCut on a coauthorship network, which was previously used by Newman [13] to study community structure. This graph G consists of 379 nodes and 914 edges and its $\lambda_2(G)$ is 0.0029. A layout of this network is presented in Figure 7.2.

Among the vertices of this network, we selected three seed vertices, also displayed in Figure 7.2, to represent three different types of nodes: a *periphery node*, which belongs to a low-degree part of the graph and is surrounded by low-conductance of small volume; a *core node*, which belongs to a dense, high-conductance, part of the graph; and an *intermediate node*, which belongs to a regime between the core and the periphery. We note that this particular network displays a few very low-conductance cuts of large volume, partitioning the graph in a number of well-separated *global components*, each with its own core and periphery.

5.1 Empirical Results

In our empirical setup, we tested how varying the input parameters to LocalCut highlights different low-conductance cuts and how this information can be used to better understand some structural

[5] It is a simple consequence of our optimization characterization (Theorem 3), that the optimal vector has sweep cuts of volume at most k_α containing v.

properties of the network. Here, we only present our main findings regarding varying teleportation for a single-vertex seed vector and using a general vector as a seed. An expanded evaluation, including a study of size factor and other forms of seed vector, is presented in Appendix 7.

Teleportation. Figure 4 displays, for each of our three seed vertices, a plot of the volume and conductance of the cuts found by 1000 runs of LocalCut with $c = 2$ and α varying. We refer to this type of a plot as a *profile plot* and note that it is a specialization of the *network community profile* [3] to cuts including the seed vertex. The volume and conductance of the theoretical lower bound (from LocalSpectral) yielded by each run are also included. As a measure of comparison, we plot the volume and conductance of the cuts defined by the shortest-path balls around each seed. We immediately note that LocalCut found low-conductance cuts of different volumes around each seed vertex, outperforming the shortest-path algorithm by a factor of 4 in most cases.

For the core node, whose profile plot is shown in Figure 4(a), the volume of the output cuts grows smoothly as α is increased. For small α, the output cuts are forced to be small and, hence, display high conductance. Increasing the teleportation, the conductance progressively decreases, as the rounding starts to hit nodes in peripheral regions, whose inclusion only improves conductance as it increases the cut volume without adding many cut edges. This phenomena ends at $\alpha \approxeq 0.001$, when a cut of conductance close to the global optimum is found. This cut separates different global components of the network.

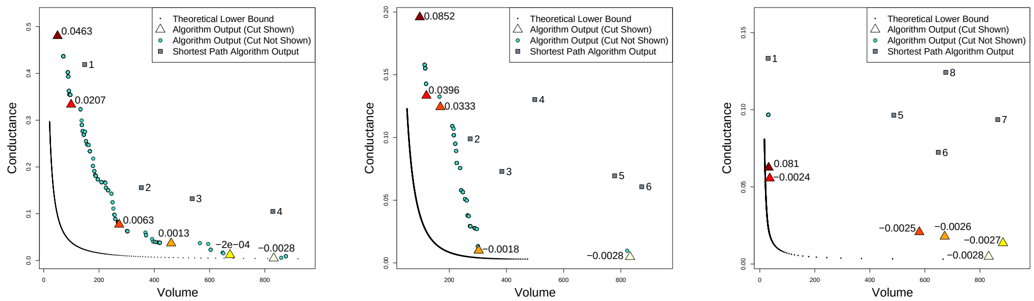
A similar interpretation applies to the profile plot of the intermediate node, shown in Figure 4(b). Here, however, the global component of the network containing the seed has smaller volume, around 300, and a very low conductance. As a result, further increasing the teleportation does not yield a larger cut, as the low-conductance cut at volume 300 still appears as a sweep cut and cannot be improved by cuts of only marginally larger volume. The overall effect is that the profile plot *jumps* from this cut to the much larger eigenvector sweep cut.

An extreme case in which an increase in teleportation does not yield larger cuts is that of the periphery node, whose profile plot is displayed in Figure 4(c). This vertex is contained in a small-volume cut of low conductance, and the only cuts of lower conductance in the network are those separating the global components. Hence, the teleportation must be greatly increased before the algorithm starts perceiving cuts at larger volumes.

Another striking difference between the results from the different seed nodes, is the range of the teleportation parameter necessary for the algorithm to access the cuts separating the global components of the network. For both the periphery node and the intermediate node, it was necessary to make α positive to detect such a cut. This highlights the usefulness of our definition of GPPR, as it allows us to capture the conductance of larger cuts than Personalized Pagerank.

In summary, we show various benefits of varying the teleportation parameter around a seed vertex. Besides representing low-conductance cuts at different volume scales, the profile plot, and in particular the presence of *jumps*, conveys useful information about the structure of the network in the neighborhood of the seed. In addition, each choice of teleportation provides a lower bound on conductance at a certain volume, which is of independent interest and can be also used as a benchmark for any other local methods.

General seed vector. Identifying peripheral areas that are well-separated from the rest of the graph is a useful primitive in studying the structure of social networks [3]. With this application in mind, we turn to an example of a general seed vector that represents a feature of the vertices. We exploit the fact that nodes in the periphery tend to have lower degree by using a seed vector that is a suitably normalized version of the degree vector. This choice biases LocalCut towards cuts that are well-correlated with low-degree vertices. A selection of the cuts found on this seed vector



(a) Core node (b) Intermediate node (c) Periphery node

Figure 4: Profile plots for varying α. Each profile plot shows results from 1000 runs of LocalCut, with $c = 2$ and α increasing in 0.01 increments starting at 0.0028. Cuts marked by a triangle are shown in the layouts of Figure 7.

when varying the teleportation with $c = 2$ is displayed in Figure 5 on a layout of the network. These cuts partition the network naturally into three well-separated regions: a sparse periphery in darker colors, a lighter-colored intermediate region, and a white dense core, where most high-degree vertices lie. We believe that this approach may be useful more generally in finding low-conductance cuts that are well-correlated with a known feature of the nodes. In particular, such a primitive should be helpful in community detection in networks where each vertex comes with extra meta-data or features which we believe to be correlated within communities of interest: this is the case for most social networks, where we may know affiliation or location information of every individual.

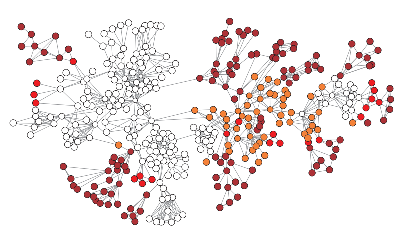


Figure 5: Selected cuts for varying α with the seed vertex equal to a normalized version of the degree vector. The cuts are displayed by assigning to each vertex a color corresponding to the smallest selected cut in which the vertex was included. Smaller cuts are darker, larger are lighter. The size of each vertex is an affine function of its degree, where smaller degree corresponds to larger size.

References

[1] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transcations of Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[2] M.E.J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, 74:036104, 2006.

[3] J. Leskovec, K.J. Lang, A. Dasgupta, and M.W. Mahoney. Statistical properties of community structure in large social and information networks. In *WWW '08: Proceedings of the 17th International Conference on World Wide Web*, pages 695–704, 2008.

[4] M. R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1979.

[5] F.R.K. Chung. *Spectral graph theory*, volume 92 of *CBMS Regional Conference Series in Mathematics*. American Mathematical Society, 1997.

[6] Milena Mihail. Conductance and convergence of markov chains-a combinatorial treatment of expanders. In *FOCS*, pages 526–531, 1989.

[7] G. Gallo, M. D. Grigoriadis, and R. E. Tarjan. A fast parametric maximum flow algorithm and applications. *SIAM J. Comput.*, 18(1):30–55, 1989.

[8] R. Andersen and K. Lang. An algorithm for improving graph partitions. In *SODA '08: Proceedings of the 19th ACM-SIAM Symposium on Discrete algorithms*, pages 651–660, 2008.

[9] D.A. Spielman and S.-H. Teng. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In *STOC '04: Proceedings of the 36th annual ACM Symposium on Theory of Computing*, pages 81–90, 2004.

[10] R. Andersen, F.R.K. Chung, and K. Lang. Local graph partitioning using PageRank vectors. In *FOCS '06: Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, pages 475–486, 2006.

[11] R. Andersen and K. Lang. Communities from seed sets. In *WWW '06: Proceedings of the 15th International Conference on World Wide Web*, pages 223–232, 2006.

[12] J. Leskovec, K.J. Lang, and M.W. Mahoney. Empirical comparison of algorithms for network community detection. In *WWW '10: Proceedings of the 19th International Conference on World Wide Web*, pages 631–640, 2010.

[13] M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. 74(3):036104–+, September 2006.

[14] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In *Proceedings of the 7th International Conference on World Wide Web*, pages 107–117, 1998.

[15] A. N. Langville and C. D. Meyer. Deeper inside PageRank. *Internet Mathematics*, 1(3):335–380, 2004.

[16] P. Berkhin. A survey on PageRank computing. *Internet Mathematics*, 2(1):73–120, 2005.

[17] T.H. Haveliwala. Topic-sensitive PageRank: A context-sensitive ranking algorithm for web search. *IEEE Transactions on Knowledge and Data Engineering*, 15(4):784–796, 2003.

[18] G. Jeh and J. Widom. Scaling personalized web search. In *WWW '03: Proceedings of the 12th International Conference on World Wide Web*, pages 271–279, 2003.

[19] Rohit Khandekar, Satish Rao, and Umesh Vazirani. Graph partitioning using single commodity flows. In *STOC '06: Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 385–390, New York, NY, USA, 2006. ACM.

[20] Lorenzo Orecchia, Leonard J. Schulman, Umesh V. Vazirani, and Nisheeth K. Vishnoi. On partitioning graphs via single commodity flows. In *STOC*, pages 461–470, 2008.

[21] Kevin Lang and Satish Rao. A flow-based method for improving the expansion or conductance of graph cuts. In *IPCO*, pages 325–337, 2004.

[22] Jure Leskovec, Kevin J. Lang, Anirban Dasgupta, and Michael W. Mahoney. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *CoRR*, abs/0810.1355, 2008.

[23] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, UK, 2004.

[24] H. A. van der Vorst. Bi-cgstab: a fast and smoothly converging variant of bi-cg for the solution of nonsymmetric linear systems. *SIAM J. Sci. Stat. Comput.*, 13(2):631–644, 1992.

[25] Vladimir Batagelj, Vladimir Batagelj, Andrej Mrvar, and Andrej Mrvar. Pajek - analysis and visualization of large networks. In *Graph Drawing Software*, pages 77–103. Springer, 2003.

[26] T. Kamada and S. Kawai. An algorithm for drawing general undirected graphs. *Inf. Process. Lett.*, 31(1):7–15, 1989.

6 Appendix

6.1 Theoretical Discussions

Connection to Personalized PageRank. PageRank refers to a popular method to determine a global rank or global notion of importance for a node in a graph based on the link structure of the graph [14]; see also [15, 16] and references therein for details. There have been several extensions to the basic PageRank concept, including Topic-Sensitive PageRank [17] and Personalized PageRank [18]. In the same way that PageRank can be viewed as a way to express the quality of a web page over the entire web, Personalized PageRank expresses a link-based measure of page quality around user-selected pages. If we let $W = AD^{-1}$ be the transition matrix of the random walk over the graph, the Personalized PageRank $\rho_{\gamma,t}$ can be computed as the random walk

$$\rho_{\alpha,s} = \gamma \sum_{i=0}^{\infty} (1-\gamma)^i W^i s = \gamma \left(I - (1-\gamma)W\right)^{-1} t,$$

where the preference vector t [10] encodes the personalization and the parameter $\gamma \in (0,1)$ is known as teleportation. In applications to graph partitioning, for normalization reasons, the vector of interest is $D^{-1}\rho_{\gamma,t}$. This should be compared to our definition of Generalized Personalized PageRank (GPPR) vector, which, for $s \in \mathcal{S}_D$, can be written as

$$p_{\alpha,s} \stackrel{\text{def}}{=} (L_G - \alpha L_{K_n})^+ Ds = \frac{1}{1-\alpha} D^{-1} \left(I - \frac{1}{1-\alpha} W \right)^+ Ds \propto D^{-1} \rho_{(-\alpha/1-\alpha, Ds)}.$$

This expression converges for $\alpha \in (-\infty, \lambda_2(G))$ and generalizes the Personalized PageRank when $\alpha < 0$. We note that a GPPR vector arose out of our optimization formulation and was not something we designed into the optimization problem.

Interpretation of our main program. There are few interesting ways to view our local optimization problem of Figure 1 which would like to point out here. Recall that LocalSpectral may be interpreted as augmenting the standard spectral optimization program with a constraint that the output cut be well-correlated with the input seed set. To understand this program from the perspective of the dual, recall that the dual of LocalSpectral is given by the following.

$$\begin{aligned} \text{maximize} \quad & \alpha + \beta\kappa \\ \text{s.t.} \quad & L_G \succeq \alpha L_{K_n} + \beta \Omega_T \\ & \beta \geq 0 \end{aligned}$$

where $\Omega_T = D s_T s_T^T D$. Alternatively, by subtracting the second constraint of LocalSpectral from the first constraint, it follows that

$$x^T \left(L_{K_n} - L_{K_n} s_T s_T^T L_{K_n} \right) x \leq 1 - \kappa.$$

It can be shown that

$$L_{K_n} - L_{K_n} s_T s_T^T L_{K_n} = \frac{L_{K_T}}{\text{vol}(\bar{T})} + \frac{L_{K_{\bar{T}}}}{\text{vol}(T)},$$

where L_{K_T} is the D weighted complete graph on the vertex set T, and, thus, LocalSpectral is clearly equivalent to

$$\begin{aligned} \text{minimize} \quad & x^T L_G x \\ \text{s.t.} \quad & x^T L_{K_n} x = 1 \\ & x^T \left(\frac{L_{K_T}}{\text{vol}(\bar{T})} + \frac{L_{K_{\bar{T}}}}{\text{vol}(T)} \right) x \leq 1 - \kappa. \end{aligned}$$

The dual of this program is given by the following.

$$\begin{aligned} \text{maximize} \quad & \alpha - \beta(1-\kappa) \\ \text{s.t.} \quad & L_G \succeq \alpha L_{K_n} - \beta \left(\frac{L_{K_T}}{\text{vol}(\bar{T})} + \frac{L_{K_{\bar{T}}}}{\text{vol}(T)} \right) \\ & \beta \geq 0 \end{aligned}$$

From the perspective of this dual, this can be viewed as embedding a combination of a complete graph K_n and a weighted combination of complete graphs on the sets T and $\bar{T}$, *i.e.*, K_T and $K_{\bar{T}}$. Depending on the value of β, the latter terms clearly discourage cuts that substantially cut into T or $\bar{T}$, thus encouraging partitions that are near the input cut $(T, \bar{T})$.

Bounding the size of the output cut. Readers familiar with the spectral method may recall that given a graph with a small balanced cut, it is not possible, in general, to guarantee that the Sweep Cut procedure of Theorem 2 applied to the optimal of Spectral outputs a balanced cut. One may have to iterate several times before one gets a balanced cut. Our setting, building up on the spectral method, also suffers from this; we cannot hope, in general, to bound the size of the output cut (which is a Sweep Cut) in terms of the correlation parameter κ.

Interpretation as electrical flows. Here we briefly provide the connection of LocalSpectral to electrical flows. Let $s = s_T$. As shown, the optimal vector $x^\star$ of the program LocalSpectral satisfies $(L_G - \alpha L_{K_n})x^\star = cDs$ for some constant c. As a scaled version of $x^\star$ is equivalent for the purposes of our rounding we can consider the following system of linear equations instead

$$(L_G - \alpha L_{K_n})x = D\left(\frac{1_T}{\mathrm{vol}(T)} - \frac{1_{\bar{T}}}{\mathrm{vol}(\bar{T})}\right).$$

If $\alpha = 0$, then if we then consider an electrical network where a vertex $i \in T$ has $\frac{d_i}{\mathrm{vol}(T)}$ units of current fed into it and a vertex $i \in \bar{T}$ has $\frac{d_i}{\mathrm{vol}(\bar{T})}$ extracted from it, we see that the resulting voltages must exactly obey the system of linear equations above. Hence, the optimal vector $x^\star$ in the case $\alpha = 0$ can be interpreted as this setting of voltages scaled appropriately.

6.2 Relation to Previous Work

Relation to local graph partitioning. Local graph partitioning was introduced in the seminal paper of Spielman and Teng [9]. Roughly, the goal here is to find a low-conductance cut in a graph in time nearly-linear in the volume of the output cut. They introduced random walk based methods towards this. This subroutine was used by them to give a nearly linear time algorithm for outputting balanced cuts matching the Cheeger Inequality in graphs up-to pollywog factors. In our language, a local graph partitioning algorithm would start a random walk at a seed node and truncate the walk after a suitably chosen number of steps, and output the nodes visited by the walk. This result was improved by Andersen, Chung and Lang [10] by doing a certain Personalized PageRank based random walks. However, [10] and subsequent papers building on it are motivated by local graph partitioning and do not address the problem of discovering cuts near given seed vectors. Further, their approach is more along the lines of Spielman and Teng while ours is an axiomatic and optimization based.

Relation to cut-improvement algorithms. Many algorithms for finding minimum-conductance cuts, such as those in [19, 20], use as a crucial building block a primitive that takes as input a cut $(T, \bar{T})$ and attempts to find a lower-conductance cut that is *well correlated* with $(T, \bar{T})$. This primitive is referred to as a *cut-improvement algorithm* [21, 8] as its original purpose was limited to post-processing cuts output by other algorithms. Recently, cut-improvement algorithms have also been used to find low conductance in specific regions of large graphs [22]. Given a notion of correlation between cuts, cut-improvement algorithms typically produce approximation guarantees of the following form: for any cut $(C, \bar{C})$ that is ε-correlated with the input cut, the cut output by the algorithm has conductance upper-bounded by a function of the conductance of $(C, \bar{C})$ and ε.

Gallo, Grigoriadis and Tarjan [7] were the first to show that one can find a subset of an input set $T \subseteq V$ with minimum conductance in polynomial time. Lang and Rao [21] implement a closely related algorithm and demonstrate its effectiveness at refining cuts output by other methods. This line of research culminates in the work of Andersen and Lang [8], who give a more general algorithm that uses a small number of single-commodity maximum-flows to find low-conductance cuts not only inside the input subset T, but among all cuts which are well-correlated with $(T, \bar{T})$.

6.3 Proofs

In this section we present all the proofs which were moved out of the main paper due to space limitations. We present them based on the section they appeared in the main body.

6.3.1 Basic Facts from Section 3

Fact 2. $\langle s_T, 1\rangle_D = 0$.

Proof.

$$\begin{aligned}\langle s_T, 1\rangle_D &= \sqrt{\frac{\mathrm{vol}(T)\mathrm{vol}(\bar{T})}{2m}} \sum_{i\in V}\left(\frac{1_T(i)d_i}{\mathrm{vol}(T)} - \frac{1_{\bar{T}}(i)d_i}{\mathrm{vol}(\bar{T})}\right) \\ &= \sqrt{\frac{\mathrm{vol}(T)\mathrm{vol}(\bar{T})}{2m}} \left(\frac{\mathrm{vol}(T)}{\mathrm{vol}(T)} - \frac{\mathrm{vol}(\bar{T})}{\mathrm{vol}(\bar{T})}\right) = 0.\end{aligned}$$

□

Fact 3. $\langle s_T, s_T\rangle_D = 1$.

Proof.

$$\begin{aligned}\langle s_T, s_T\rangle_D &= \frac{\mathrm{vol}(T)\mathrm{vol}(\bar{T})}{2m} \sum_{i\in V}\left(\frac{1_T(i)}{\mathrm{vol}(T)} - \frac{1_{\bar{T}}(i)}{\mathrm{vol}(\bar{T})}\right)^2 d_i \\ &\overset{1_T(i)1_{\bar{T}}(i)=0}{=} \frac{\mathrm{vol}(T)\mathrm{vol}(\bar{T})}{2m} \sum_{i\in V}\left(\frac{1_T(i)d_i}{\mathrm{vol}^2(T)} + \frac{1_T(i)d_i}{\mathrm{vol}^2(\bar{T})}\right) \\ &= \frac{\mathrm{vol}(T)\mathrm{vol}(\bar{T})}{2m} \left(\frac{\mathrm{vol}(T)}{\mathrm{vol}^2(T)} + \frac{\mathrm{vol}(\bar{T})}{\mathrm{vol}^2(\bar{T})}\right) \\ &= \frac{\mathrm{vol}(T)\mathrm{vol}(\bar{T})}{2m} \left(\frac{\mathrm{vol}(T)+\mathrm{vol}(\bar{T})}{\mathrm{vol}(T)\mathrm{vol}(\bar{T})}\right) = \frac{\mathrm{vol}(G)}{2m} = 1.\end{aligned}$$

□

Fact 4. $s_T^T L_G s_T = \phi(T)$.

Proof.

$$s_T^T L_G s_T = \frac{\mathrm{vol}(T)\mathrm{vol}(\bar{T})}{2m} \sum_{ij\in E}\left(\frac{1_T(i)}{\mathrm{vol}(T)} - \frac{1_{\bar{T}}(i)}{\mathrm{vol}(\bar{T})} - \frac{1_T(j)}{\mathrm{vol}(T)} + \frac{1_{\bar{T}}(j)}{\mathrm{vol}(\bar{T})}\right)^2$$

If i, j are both in T or both in $\bar{T}$, then $\left(\frac{1_T(i)}{\mathrm{vol}(T)} - \frac{1_{\bar{T}}(i)}{\mathrm{vol}(\bar{T})} - \frac{1_T(j)}{\mathrm{vol}(T)} + \frac{1_{\bar{T}}(j)}{\mathrm{vol}(\bar{T})}\right)^2 = 0$. Else, if the edge i, j crosses the cut $(T, \bar{T})$, then

$$\left(\frac{1_T(i)}{\mathrm{vol}(T)} - \frac{1_{\bar{T}}(i)}{\mathrm{vol}(\bar{T})} - \frac{1_T(j)}{\mathrm{vol}(T)} + \frac{1_{\bar{T}}(j)}{\mathrm{vol}(\bar{T})}\right)^2 = \left(\frac{1}{\mathrm{vol}(T)} + \frac{1}{\mathrm{vol}(\bar{T})}\right)^2 = \frac{\mathrm{vol}^2(G)}{\mathrm{vol}^2(T)\mathrm{vol}^2(\bar{T})}.$$

Hence, to complete the proof notice that

$$s_T^T L_G s_T = \frac{\mathrm{vol}(T)\mathrm{vol}(\bar{T})}{2m} \cdot \frac{|E(T,\bar{T})|\mathrm{vol}^2(G)}{\mathrm{vol}^2(T)\mathrm{vol}^2(\bar{T})} = \mathrm{vol}(G) \cdot \frac{|E(T,\bar{T})|}{\mathrm{vol}(T)\mathrm{vol}(\bar{T})} = \phi(T).$$

□

The following lemma describes the relationship between $K(T, U)$ and the dot product between the vectors s_T and s_U.

Lemma 2. *For any sets* $T, U \subseteq V$, $K(U,T) = \frac{\mathrm{vol}(U)\mathrm{vol}(\bar{U})}{\mathrm{vol}(T)\mathrm{vol}(\bar{T})} \cdot \left(\frac{\mathrm{vol}(T\cap U)}{\mathrm{vol}(U)} - \frac{\mathrm{vol}(T)\cap \bar{U}}{\mathrm{vol}(\bar{U})}\right)^2$.

Proof. First, notice that s_T can be written as

$$s_T = \sqrt{\frac{\mathrm{vol}(T)\mathrm{vol}(\bar{T})}{2m}} \left(\frac{2m \cdot 1_T}{\mathrm{vol}(T)\mathrm{vol}(\bar{T})} - \frac{1_V}{\mathrm{vol}(\bar{T})}\right).$$

Then, since $s_U^T D1 = 0$, it follows that

$$\begin{aligned} s_T^T D s_U &= \sqrt{\frac{\mathrm{vol}(U)\mathrm{vol}(\bar{U})}{\mathrm{vol}(T)\mathrm{vol}(\bar{T})}} \cdot \left(\frac{1_T^T D 1_U}{\mathrm{vol}(U)} - \frac{1_T^T D 1_{\bar{U}}}{\mathrm{vol}(\bar{U})}\right) \\ &= \sqrt{\frac{\mathrm{vol}(U)\mathrm{vol}(\bar{U})}{\mathrm{vol}(T)\mathrm{vol}(\bar{T})}} \cdot \left(\frac{\mathrm{vol}(T\cap U)}{\mathrm{vol}(U)} - \frac{\mathrm{vol}(T)\cap \bar{U}}{\mathrm{vol}(\bar{U})}\right). \end{aligned}$$

Finally, squaring both sides of this expression completes the proof. □

Proof. [of Lemma 1] Note that if we let $x = s_T$ in $\mathsf{LocalSpectral}(G, s_{\{v\}}, 1/k)$, then $s_T^T L_G s_T = \phi(S)$ and $\langle s_T, s_{\{v\}}\rangle_D^2 = \frac{d_v(2m-\mathrm{vol}(T))}{\mathrm{vol}(T)(2m-d_v)} \geq 1/k$. Here we have used the easy observation that for any sets T, U $\langle s_T, s_U\rangle_D^2 = K(T,U) = \frac{\mathrm{vol}(U)\mathrm{vol}(\bar{U})}{\mathrm{vol}(T)\mathrm{vol}(\bar{T})}\left(\frac{\mathrm{vol}(T\cap U)}{\mathrm{vol}(U)} - \frac{\mathrm{vol}(T\cap\bar{U})}{\mathrm{vol}(\bar{U})}\right)^2$. □

Proof. [of Fact 1] Let $x^\star$ be any optimal solution to $\mathsf{Localspectral}(G, s, \kappa)$. Consider $\tilde{x} \stackrel{\mathrm{def}}{=} x^\star - \frac{\langle x^\star, 1\rangle_D}{2m}1$. It follows from the assumption that $\langle s, 1\rangle_D = 0$, that $\tilde{x}$ satisfies all the constraints of $\mathsf{Localspectral}(G, s, \kappa)$ and has objective value same as that of $x^\star$. □

6.3.2 Proofs of Claims in Theorem 3

Claim 5. *[same as Claim 3] The following feasibility and complementary slackness conditions are sufficient for a primal-dual pair* $X^\star, \alpha^\star, \beta^\star$ *to be an optimal solution. The feasibility conditions are:*

$$\begin{aligned} L_{K_n} \circ X^\star &= 1 & (1)\\ (Ds)(Ds)^T \circ X^\star &\geq \kappa & (2)\\ L_G - \alpha^\star L_{K_n} - \beta^\star (Ds)(Ds)^T &\succeq 0 & (3)\\ \beta^\star &\geq 0, & (4) \end{aligned}$$

and the complementary slackness conditions are:

$$\begin{aligned} \alpha^\star(L_{K_n} \circ X^\star - 1) &= 0 & (5)\\ \beta^\star((Ds)(Ds)^T \circ X^\star - \kappa) &= 0 & (6)\\ X^\star \circ (L_G - \alpha^\star L_{K_n} - \beta^\star (Ds)(Ds)^T) &= 0. & (7) \end{aligned}$$

Proof. This follows from the convexity of $\mathsf{SDP}_p(G, s, \kappa)$ and Slater's condition [23]. □

Proof. [of Claim 1] Consider a vector x that is a feasible solution to $\mathsf{LocalSpectral}(G, s, \kappa)$, and note that $X = xx^T$ is a feasible solution to $\mathsf{SDP}_p(G, s, \kappa)$. □

Proof. [of Claim 2] Since $\mathsf{SDP}_p(G, s, \kappa)$ is convex, if suffices ro verify that Slater's constraint qualification condition is true for this primal SDP. Consider $X = ss^T$. Then, $(Ds)(Ds)^T \circ ss^T = (s^T Ds)^2 = 1 > \kappa$. □

Proof. [of Claim 4] We start by establishing simple facts.

Fact 5. $\alpha^\star \le \lambda_2(G)$. *Moreover if* $\lambda_2 = \alpha^\star$ *then* $\langle v_2, s\rangle_D = 0$.

Proof. Let $v_2 = D^{-1/2}u_2$ where u_2 is the unit length eigenvector corresponding to $\lambda_2(G)$ of $\mathcal{L}_G$. Plugging in v_2 in Equation (4), we obtain that

$$v_2^T L_G v_2 - \alpha^\star - \beta^\star \langle v_2, s\rangle_D^2 \ge 0.$$

But $v_2^T L_G v_2 = \lambda_2(G)$ and $\beta^\star \ge 0$. Hence, $\lambda_2(G) \ge \alpha^\star$. It follows that if $\lambda_2 = \alpha^\star$ then $\langle v_2, s\rangle_D = 0$. □

Fact 6. *We may assume that the optimal* $X^\star$ *satisfies* $1^T D^{1/2} X^\star D^{1/2} 1 = 0$.

Proof. Identical to Fact 1 by removing from $X^\star$. □

Now we return to the proof of the claim. If we assume $\langle v_2, s\rangle_D \neq 0$, then we know that $\alpha^\star < \lambda_2(G)$ from Fact 5. Note that since G is connected and $\alpha^\star < \lambda_2(G)$, $L_G - \alpha^\star L_{K_n}$ has rank exactly $n-1$. From the complementary slackness condition (7) we can deduce that the image of $X^\star$ is in the kernel of $L_G - \alpha^\star L_{K_n} - \beta^\star (Ds)(Ds)^T$. But $\beta^\star (Ds)(Ds)^T$ is a rank one matrix and since $\langle s, 1\rangle_D = 0$, it reduces the rank of $L_G - \alpha^\star L_{K_n}$ by one precisely when $\beta^\star > 0$. If $\beta^\star = 0$ then $X^\star$ must be 0 which is not possible if $\mathsf{SDP}_p(G, s, \kappa)$ is feasible. Hence, the rank of $L_G - \alpha^\star L_{K_n} - \beta^\star (Ds)(Ds)^T$ must be exactly $n-2$ and since $X^\star$ cannot have any component along the all ones vector, $X^\star$ must be rank one. This proves the claim. □

6.3.3 Proof of Theorem 5

From Theorem 3 it follows that $\lambda(G, s, \kappa)$ is same as the optimal value of $\mathsf{SDP}_p(G, s, \kappa)$ which, by strong duality, is the same as the optimal value of $\mathsf{SDP}_d(G, s, \kappa)$. Let $\alpha^\star, \beta^\star$ be the optimal dual values to $\mathsf{SDP}_d(G, s, \kappa)$. Then from the dual feasibility constraint $L_G - \alpha^\star L_{K_n} - \beta^\star (Ds)(Ds)^T \succeq 0$, it follows that

$$s_T^T L_G s_T - \alpha^\star s_T^T L_{K_n} s_T - \beta^\star \langle s, s_T\rangle_D^2 \ge 0.$$

Notice that since $\langle s_T, 1\rangle_D = 0$, $s_T^T L_{K_n} s_T = s_T^T D s_T = \langle s_T, s_T\rangle_D = 1$. Further, since $s_T^T L_G s_T = \phi(T)$, we obtain, if $\kappa \le \kappa'$, that

$$\phi(T) \ge \alpha^\star + \beta^\star \langle s, s_T\rangle_D^2 \ge \alpha^\star + \beta^\star \kappa = \lambda(G, s, \kappa).$$

If on the other hand, $\kappa' \le \kappa$,

$$\phi(T) \ge \alpha^\star + \beta^\star \langle s, s_T\rangle_D^2 \ge \alpha^\star + \beta^\star \kappa \ge {}^{\kappa'}\!/_{\kappa} \cdot (\alpha^\star + \beta^\star \kappa) = {}^{\kappa'}\!/_{\kappa} \cdot \lambda(G, s, \kappa).$$

Finally observe that if $s = s_U$ for some $U \subseteq V$, then $\langle s_U, s_T\rangle_D^2 = K(U, T)$. Note that strong duality was used here.

7 Empirical Evaluation

In this section we expand on Section 5. We choose to repeat a large part of that section here for the sake of readability of this section.

7.1 Algorithm Description and Implementation

We refer to our cut-finding algorithm as LocalCut. In this section we show how to use the theoretical results, in particular, Theorem 3 and 4, to guide our empirical study. We refer to our cut-finding algorithm as LocalCut. The input parameters of LocalCut are a seed vector s, a teleportation parameter α and a size factor c. The execution of LocalCut starts by computing the GPPR of seed s at teleportation α. By Theorem 3, the resulting vector $x^\star$ is an optimal solution to $\mathsf{LocalSpectral}(G, s, \kappa_\alpha)$ for some choice of κ_α. The resulting vector $x^\star$ is then rounded by outputting its sweep cut of least conductance to obtain good approximations as guaranteed by Theorem 4. In the case of a seed vector corresponding to a single vertex v, one can infer from Lemma 1 that $x^\star$ yields a lower bound to the conductance of cuts that contain v and have less than a certain volume k_α, while the sweep-cut rounding does not give a specific guarantee on the volume of the output cut. However, we found that it was often possible to find a low-conductance cut in the range dictated by k_α. To exploit this observation, we use a new input parameter, a *size factor* $c > 0$, that regulates the maximum volume of the sweep cuts considered when s represents a single vertex. In this case, LocalCut considers only sweep cuts of volume at most $c \cdot k_\alpha$ that contain the vertex v [6]. This new input parameter turns out to be extremely useful in exploring cuts at different sizes, as it neglects sweep cuts of low conductance at large volume and allows us to pick out more local cuts around the seed vertex.

We implemented LocalCut in a combination of MATLAB and C++, solving linear systems using the Stabilized Biconjugate Gradient Method [24] provided in MATLAB 2006b. On the graph of interest, the coauthorship network, and on a Dell PowerEdge 1950 machine with 2.33 GHz and 16GB of RAM, 1000 executions of the algorithm ran in less than 5 seconds.

7.2 Network

We tested LocalCut on a coauthorship network, which was previously used by Newman [13] to study community structure. This graph G consists of 379 nodes and 914 edges, where each node represents an author and each edge a coauthorship relation. The edges are unweighted. A layout of this network is presented in Figure 7.2. The spectral gap $\lambda_2(G)$ is 0.0029 and the best sweep cut of the corresponding eigenvector yields the cut separating the graph into the left and right parts shown in Figure 7.2.

The three highlighted vertices in Figure 7.2 were used as seed vertices for LocalCut. They were chosen to represent three different types of nodes: a *periphery node*, which belongs to a low-degree part of the graph and is surrounded by low-conductance cuts of small volume; a *core node*, which belongs to a dense, high-conductance, part of the graph; and an *intermediate node*, which belongs to a regime between the core and the periphery. We note that this particular network displays a few cuts of very low conductance and large volume, which partition the graph in a number of well-separated *global components*, each with its own core and periphery.

7.3 Empirical Results

In this section we expand on Section 5. We repeat a large part of that section here for sake of continuity and readability of this section. We studied how varying these parameters highlights different low-conductance cuts and how LocalCut can be applied to understand the community structure of the network. In our first two experiments, discussed in Sections 7.3.1 and 7.3.2, we used single-vertex seed vectors and analyzed the effects of independently varying α, c, and the

[6] It is a simple consequence of our optimization characterization, that the optimal vector has sweep cuts of volume at most k_α containing v.

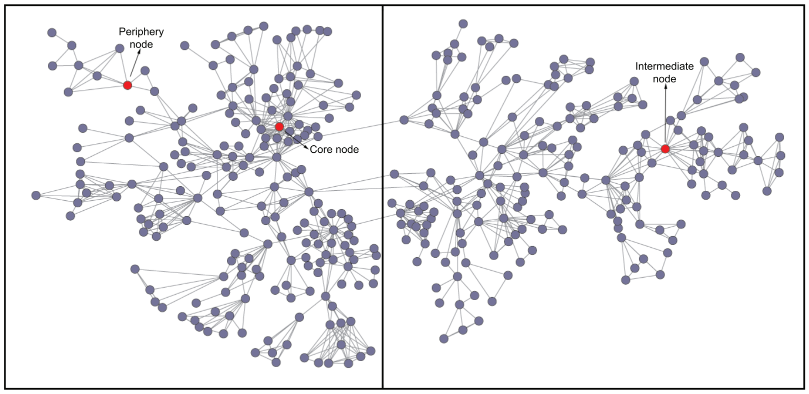


Figure 6: The coauthorship network. This layout was obtained in the Pajek [25] visualization software, using the Kamada-Kawai method [26] on each component of a partition provided by LocalCut and tiling the layouts at the end. The boxes show the two main global components of the network, which are displayed separately in Figures 7, 8 and 9.

location of the seed vertex. In the last set of experiments, presented in Section 7.4, we considered general seed vectors and found an elegant application to the detection of peripheral regions of the graph.

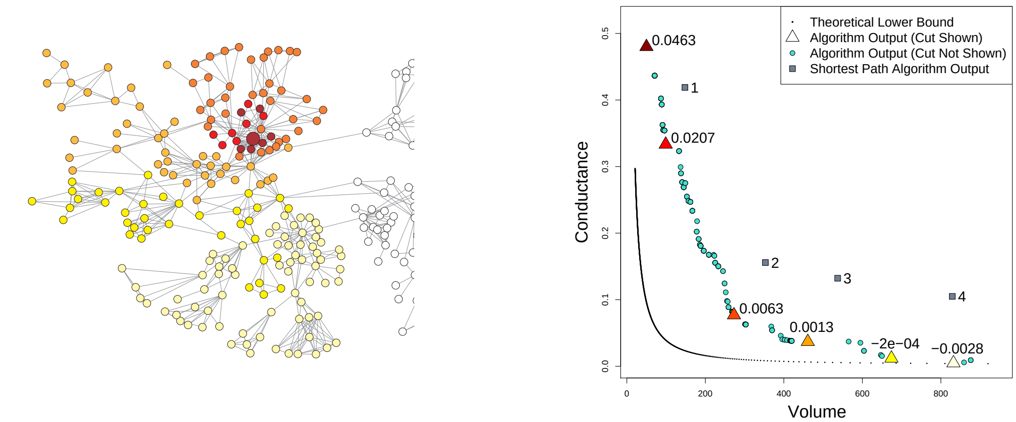
7.3.1 Teleportation

To study the dependance of LocalCut on the teleportation parameter α, for each of our three representative seed nodes we executed 1000 runs of LocalCut with $c = 2$ and α varying by .001 increments. Figure 7 displays, for each seed, a plot of the volume and conductance of the cuts found by each run of LocalCut. We refer to this type of plot as a *profile plot* and note that it is a specialization of the *network community profile* [3] to cuts including the seed vertex. The volume and conductance of the theoretical lower bound yielded by each run are also included. As a measure of comparison, we plot the volume and conductance of the cuts defined by the shortest-path balls around each seed. Next to the plot, we present an image of representative cuts detected by LocalCut.

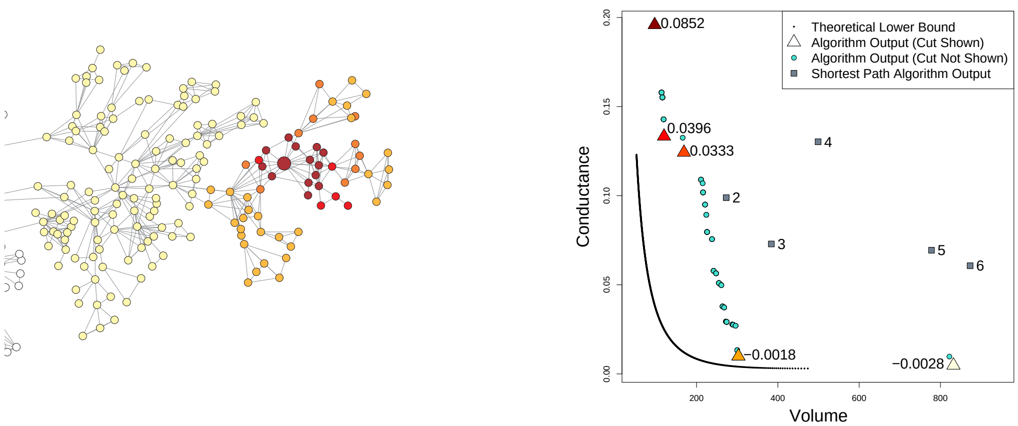
LocalCut found low-conductance cuts of different volumes around each seed vertex, outperforming the shortest-path algorithm by a factor of 4 in most cases. However, the results of LocalCut still lie away from the lower bound, which is also approximately a factor of 4 smaller at most volumes.

Recall that, given a teleportation parameter α, the rounding selects the cut of smallest conductance among sweep cuts of volume at most $c \cdot k_\alpha$, where k_α is the lower bound obtained from the optimization program. Increasing α, and consequently k_α, can lead LocalCut to pick out larger cuts, but does not guarantee this will happen. In particular, in many instances there may not be a way of marginally increasing the volume of a cut while lowering its conductance. In those cases, LocalCut may output the same small sweep cut for a range of teleportation parameters until $c \cdot k_\alpha$ becomes large enough that a larger, lower-conductance cut can be found.

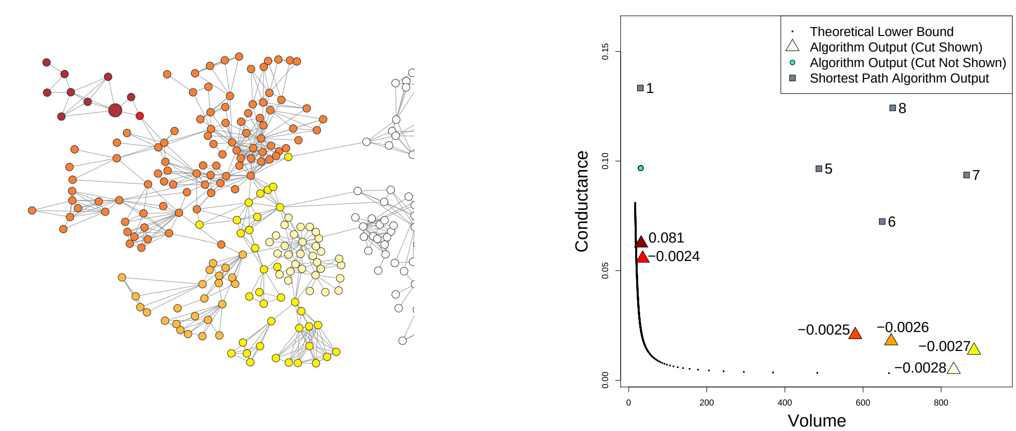
For the core node, whose profile plot is shown in Figure 7(a), the volume of the output cuts grows smoothly as α is increased. For small α, the output cuts are forced to be small and



(a) Selected cuts and profile plot for the core node.



(b) Selected cuts and profiles plot for the intermediate node.



(c) Selected cuts and profile plot for the periphery node.

Figure 7: Selected cuts and profile plots for varying α. The cuts are displayed by assigning to

hence display high conductance, as the region around the node is expander-like. Increasing the teleportation, the conductance progressively decreases, as the rounding starts to hit nodes in peripheral regions, whose inclusion only improves conductance as it increases the cut volume without adding many cut edges. This phenomena ends at $\alpha \approxeq 0.001$, when a cut of conductance close to the global optimum is found. This cut separates different global components of the network.

A similar interpretation applies to the profile plot of the intermediate node, shown in Figure 7(b). Here, however, the global component of the network containing the seed has smaller volume, around 300, and a very low conductance, so that the profile plot jumps from this cut to the much larger eigenvector sweep cut.

An extreme case in which an increase in teleportation does not yield larger cuts is that of the periphery node, whose profile plot is displayed in Figure 7(c). This vertex is contained in a small-volume cut of low conductance, and the only cuts of lower conductance in the network are those separating the global components. Hence, the teleportation must be greatly increased before the algorithm starts outputting cuts at larger volumes.

Another striking difference between the results from the different seed nodes, is the range of the teleportation parameter necessary for the algorithm to access the cuts separating the global components of the network. For both the periphery node and the intermediate node, it was necessary to make α positive to detect such a cut. This highlights the usefulness of our definition of generalized PageRank, as it allows us to capture the conductance of larger cuts than Personalized Pagerank and provides a smooth interpolation between Personalized Pagerank and the second eigenvector of the graph. Indeed, for $\alpha = 0.0028 \approxeq \lambda_2(G)$, LocalCut outputs the same cut as the eigenvector sweep cut for all three seeds.

In summary, we show various benefits of varying the teleportation parameter around a seed vertex. Besides representing low-conductance cuts at different volume scales, the profile plot, and in particular the presence of *jumps*, conveys useful information about the structure of the network in the neighborhood of the seed. In addition, each choice of teleportation provides a lower bound on conductance at a certain volume, which is of independent interest and can be also used as a benchmark for other local methods.

7.3.2 Output Size

We turn to analyzing the effect of varying the size factor c, for a fixed choice of teleportation parameter α. Varying c, like varying α, has the effect of producing low-conductance cuts at different volumes around the seed vertex. Moreover, it is possible by sufficiently increasing c to obtain some low-conductance large-volume cuts, even at lower teleportations. This is exemplified in Figure 8, which shows the result of varying c with the core node as the seed and $\alpha = .02$. When $c = 2$, this setting only yielded a cut of volume close to 100, as seen in Figure 7(a). By allowing larger c, cuts of larger volume were output.

While many of these cuts tend to have conductance slightly worse than the best found by varying the teleportation, this observation leaves open the possibility that a single choice of teleportation parameter α is sufficient to produce good low-conductance cuts at all volumes by varying c. This would allow us to only solve a single optimization problem and still find cuts of different volumes.

To answer this question, for each of the three seed nodes, we selected three choices of teleportation parameter and let c vary. The resulting output cuts for the core node as the seed are plotted in Figure 8. Plots for the other seeds are similar and not displayed. No single teleportation setting dominates the others: specifically, at volume 200 the lowest-conductance cut was produced with $\alpha = .02$, at volume 400 with $\alpha = .01$ and at volume 600 with $\alpha = 0$. However,

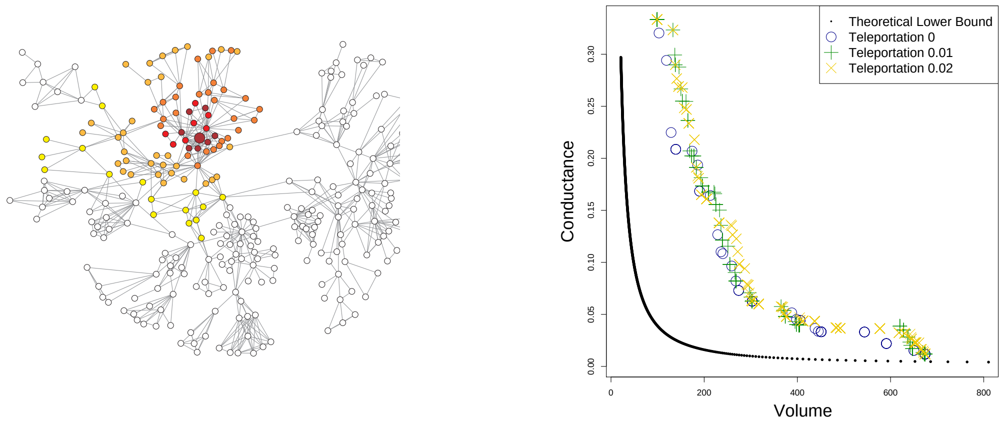


Figure 8: Selected cuts and profile plots for varying c with the core node as the seed and $c = 2$. The cuts are displayed by assigning to each vertex a color corresponding to the smallest selected cut in which the vertex was included. Smaller cuts are darker, larger are lighter. The seed vertex is shown larger. The profile plot shows results from 1000 runs of LocalCut, with varying c and $\alpha \in \{0, 0.01, 0.02\}$.

the results of all three settings roughly track each other. Overall, the highest choice of $\alpha = 0$ performed marginally better overall, recording lowest conductance cuts at both small and large volumes.

We conclude that the best results are achieved when we vary both the teleportation parameter and the size factor. Moreover, multiple teleportation choices have the advantage of yielding multiple lower bounds at different volumes.

7.4 Multiple Seeds and Correlation

Here we demonstrate the working of LocalCut on general seed vectors. In our first example, we consider a seed vector representing a subset of four nodes, located in different peripheral branches of the network. In Figure 9, we display a selection of the cuts found by varying the teleportation, with $c = 2$. The smaller cuts tend to contain the branches in which each seed node is found, while larger cuts incorporate nearby branches. Identifying peripheral areas that are well-separated from the rest of the graph is a useful primitive in studying the structure of social networks [3], and this experiment shows how LocalCut may be used in the context when some peripheral nodes are known.

With the same application in mind, we turn to an example of a general seed vector that represents a feature of the vertices. We exploit the fact that nodes in the periphery tend to have lower degree by using a seed vector that is a normalized version of the degree vector. This choice biases LocalCut towards cuts that are well-correlated with low-degree vertices. A selection of the cuts found on this seed vector when varying the teleportation with $c = 2$ is displayed in Figure 10. These cuts partition the network naturally into three well-separated regions: a sparse periphery in darker colors, a lighter-colored intermediate region, and a white dense core, where most high-degree vertices lie. This specific partition was used to produce the layout of Figure 7.2. We believe that this approach may be useful more generally in finding low-conductance cuts that are well-correlated with a known feature of the nodes.

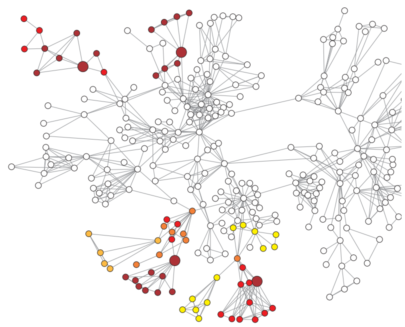


Figure 9: Selected cuts for varying α with the seed vector corresponding to a subset of 4 vertices lying in the periphery of the network. The cuts are displayed by assigning to each vertex a color corresponding to the smallest selected cut in which the vertex was included. Smaller cuts are darker, larger are lighter. The seed vertices are shown larger.

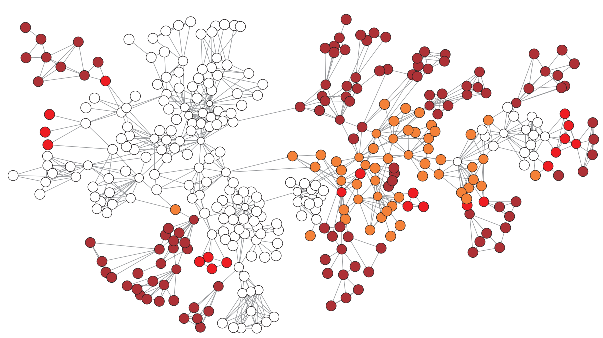


Figure 10: Selected cuts for varying α with the seed vertex equal to a normalized version of the degree vector. The cuts are displayed by assigning to each vertex a color corresponding to the smallest selected cut in which the vertex was included. Smaller cuts are darker, larger are lighter. The size of each vertex is an affine function of its degree, where smaller degree corresponds to larger size.