Theoretical Bounds on the Network Community Profile from Low-rank Semi-definite Programming

Yufan Huang [1] **C. Seshadhri** [2] **David Gleich** [1]

Abstract

We study a new connection between a technical measure called μ-conductance that arises in the study of Markov chains for sampling convex bodies and the network community profile that characterizes size-resolved properties of clusters and communities in social and information networks. The idea of μ-conductance is similar to the traditional graph conductance, but disregards sets with small volume. We derive a sequence of optimization problems including a low-rank semi-definite program from which we can derive a lower bound on the optimal μ-conductance value. These ideas give the first theoretically sound bound on the behavior of the network community profile for a wide range of cluster sizes. The algorithm scales up to graphs with hundreds of thousands of nodes and we demonstrate how our framework validates the predicted structures of real-world graphs.

1. Introduction

One of the central themes of network science is the discovery of peculiar properties that are not exhibited by random or geometric graphs. Over the past decade, network science has built a rich repository of data sets derived from social network, communication networks, biological data, internet trace data, and more. Early measurements on these networks demonstrated skewed degree distributions, high clustering coefficients, and community structure (Barabási & Albert, 1999; Watts & Strogatz, 1998; Newman, 2003; 2006). These measurements led to fundamentally new mechanisms that explain the networks (Leskovec et al., 2007; Seshadhri et al., 2012; Bonato et al., 2014). Accurately capturing and understanding these properties is critical to understanding the limits of what is possible with rich empirical data in graph-based learning (Seshadhri et al., 2020).

But many important network quantities are computationally intractable in the worst case and are only computed by heuristics. It is of critical importance to have rigorous theory that can guarantee the accuracy of these measurements.

We focus on one of the most significant network characteristics: the cluster structure (Flake et al., 2000; Newman, 2006; von Luxburg et al., 2012). Finding tightly connected sets of vertices with few connections outside is a central task in network analysis. This is often measured by the conductance. The conductance of a set S of vertices is the normalized fraction of edges that leave the set (the normalization is more involved; we give a formal definition later).

An important development in the cluster structure of real-world networks was the discovery of set size versus conductance relationships (Leskovec et al., 2008; 2009; 2010; Gleich & Seshadhri, 2012; Jeub et al., 2015). The key finding in these studies is counter-intuitive: in most real-world datasets, we cannot find *large sets* of *small conductance*. An example of this structure is shown in Figure 1. This finding directly contradicts the behavior of conductance in graphs that are derived from nearest neighbors in a geometry or graphs commonly used in partitioning computational domains, where the smallest conductance values occur in large sets. Moreover, the definition of minimum conductance is typically biased towards large sets (see equation (1)), but real-world networks exhibit the opposite behavior.

The key finding is the behavior of the *network community profile (NCP)*. The NCP plots, for each s, the minimum conductance among sets of size (technically volume) s. (Refer to Figure 1.) Observe how the plot (the blue line) slopes upward after an initial dip. This trend is consistent across many real-world networks. The NCP of a typical geometric graph slopes downwards. Currently, the NCPs are generated entirely through principled heuristic computations. Hence, it is difficult to guarantee the characteristic real-world behavior of the NCP curve without appropriate theoretical bounds. Our proposed algorithm is the first that can actually give a lower bound on the minimum conductance at a fixed size s.

[1]Computer Science, Purdue University, West Lafayette, IN, United States [2]Computer Science, University of California Santa Cruz, Santa Cruz, CA, United States. Correspondence to: Yufan Huang <huan1754@purdue.edu>, David Gleich <dgleich@purdue.edu>.

Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

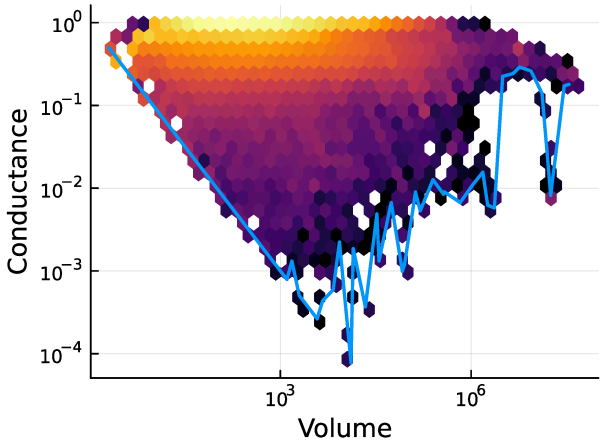


Figure 1. This is a heatmap over size (measured by set volume) and conductance values from around 630k sets identified by a seeded PageRank conductance minimizing procedure. This picture shows the expected and standard behavior of the size-resolved set conductance in a real-world graph (LIVEJOURNAL social network with around 5M vertices and 40M undirected edges). Namely, we see that the smallest conductance sets are those that are also small (volume less than 10^5 although the graph volume is 10^8). As sets get larger, their conductance grows. The lower envelope of these measurements is called the network community profile (NCP). The key insight is that for real-world networks, these network community profiles go up and to the right. This is a consistent trend for many real-world social and information networks. A weakness with this empirical finding is that there are no *lower bounds* for seeded PageRank that would certify the behavior of the overall profile. Our paper provides those bounds.

Main contributions. The primary motivating question for our paper is: *can we design theoretically rigorous algorithms that give practically viable bounds for NCP?*

1. Our main conceptual contribution is a connection between the technical notion of μ-conductance from Markov chain theory (Lovasz & Simonovits, 1990) and the empirical observations from social and information networks focused of the NCP. We discover that the interesting parts of the NCP basically correspond to a plot of μ-conductance. While this is easy to see (in hindsight), our insight provides us with an array of technical tools to address our main question.

2. We begin with a spectral relaxation to compute the μ-conductance. Unlike the standard relaxation for conductance that leads to the second eigenvalue, the μ-conductance program is non-convex. We give a further convex relaxation using semi-definite programs (SDPs). Unfortunately, this program would require super-quadratic time to solve and is practically infeasible. We give a computationally viable low rank formulation (but non-convex) of the SDP. We prove that locally optimal KKT points of this optimization problem yield rigorous lower bounds for the NCP points. (This is stated in Theorem 3.1, our main theoretical result.) We note that this is first theoretically sound and practically viable lower bound for the NCP.

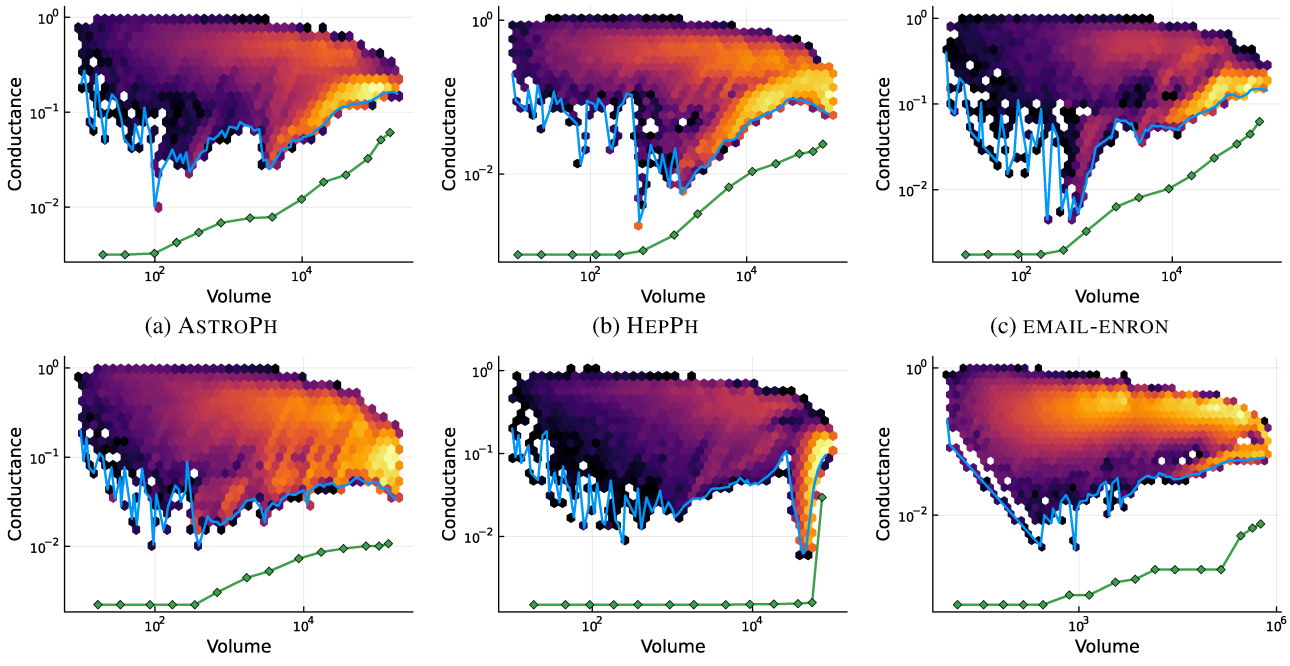
3. The low rank SDP can be solved with over 200k nodes. Using our algorithms (and Theorem 3.1), we provide the first validation on the shape of the NCP on real-world data, as first discovered in Leskovec et al. (2009). Even though our bound is loose, our lower bound validates the characteristic NCP plot and tracks the upward increase in conductance for larger set volume (Figure 2). We are able to distinguish somewhat anomalous graphs such as Deezer (Figure 2e) with a "flat" NCP, known to occur in particularly dense real-world networks (Jeub et al., 2015). Furthermore, with our new tool, we are able to study, with theoretical confidence, the NCP as the graph is "peeled" by k-core analysis.

High-level outline. We give a high-level outline of the main ideas in our paper, and explain the chain of theoretical insights that lead to the final practical lower bounds. Our starting point is the notion of μ-conductance, discovered in the context of mixing bound for random walks in high dimensional bodies (Lovasz & Simonovits, 1990). It is defined formally in equation (2). Simply put, the μ-conductance value is the minimal conductance restricted to sets with a μ-fraction of the total graph. It was originally proposed to improve the bounds on performance of volume sampling algorithms for convex bodies and study Markov chains where small sets need not have large conductance. *One can essentially generate the NCP by computing μ-conductance for varying values of μ where μ fixes the volume scale for the sets under consideration.*

But, for a given μ, how to study the optimal μ-conductance? A natural starting point is the classic spectral relaxation for the conductance of the graph. The conductance is co-$\mathbb{NP}$ hard to compute, but one can consider a continuous relaxation (Spectral Cut). This relaxation is convex and the optimal objective is the second eigenvalue, or spectral gap, of the Laplacian. Since the μ-conductance is a constrained version of conductance, we can adapt that constraint into the spectral relaxation. That yields the program (3). Note that these programs optimize over vectors, rather than sets. The spectral program for μ-conductance has extra constraints bounding each entry of the vector. These extra constraints lead to a non-convex program, showing how computing μ-conductance is significantly harder than conductance.

We now make a further relaxation, wherein we replace the vector by a positive semi-definite matrix. This relaxation leads to the semi-definite program (SDP) given in (4). SDPs are convex programs, but the number of variables is quadratic in the number of vertices. This program is infeasible for graphs with even tens of thousands of nodes.

To get a practically viable optimization problem, we formulate a low rank version of the SDP, stated in (5). But this problem is non-convex and cannot be solved globally.



(a) ASTROPH (b) HEPPH (c) EMAIL-ENRON

(d) FACEBOOK-PAGE (e) DEEZER (f) DBLP

Figure 2. Using the theory and algorithms proposed in this paper, we show the empirical network community profile along with our new lower bound for 6 real-world networks (the green line is our lower bound). This is the first guarantee on the behavior of these profiles that establishes a smooth transition from the sets of small conductance to sets of larger conductance. Note that this does not occur for all networks. For example, the Deezer network displays a flat profile until μ becomes really large, and our results confirm that we should not expect better small conductance sets. The gap between the measured conductances is expected because our analysis only gives a rough, yet informative, lower bound.

We now arrive at the deepest technical insight in our result. Consider locally optimal KKT points of the low rank (non-convex) SDP. We do a careful comparison of the KKT conditions of the convex program (4) and the low rank, non-convex (5). We discover that KKT points of (5) satisfy all KKT conditions of (4), barring one dual feasibility constraint. The violation of this constraint gives a bound on how far the low-rank KKT points are from the original SDP optimum. So, we can subtract out this violation from the objective of the low-rank KKT point, and get a provably correct lower bound on the SDP objective (which is a lower bound on the μ-conductance).

Our code to solve these problems is available from

`https://github.com/luotuoqingshan/mu-conductance-low-rank-sdp`.

Potential implications for random walks. Random walks are a central tool in modern network analysis. A common practice in graph-based learning and embedding is to use a random process to sample a region of the graph (Perozzi et al., 2014; Grover & Leskovec, 2016; Tang et al., 2015). Likewise, there are many results that attempt to estimate quantities based on a random sample of a graph (Leskovec & Faloutsos, 2006; Ahmed et al., 2010; Ribeiro & Towsley, 2010; Maiya & Berger-Wolf, 2011; Ribeiro & Towsley, 2012; Ahmed et al., 2014). Many of these results have a theoretical bound that depends on the mixing time of the random walk (Dasgupta et al., 2014; Chierichetti et al., 2016; Chierichetti & Haddadan, 2018), which is bounded by the conductance. As the NCPs show, the minimum conductance may be quite small, but only because of sets of small size. So global properties of the graph might not be affected by such small sets. Our new μ-conductance theory suggests that the standard mixing time bounds (based on conductance) may be quite pessimistic when the sampling involves a large set in the graph. It is likely that μ-conductance gives a better estimate of the mixing time for many applications and studying this is an exciting direction for future work.

2. Preliminaries and Technical Setting

Throughout the paper, we work with undirected graphs. Our methods and definitions are applicable to graphs with non-negative edge weights. Assume, without loss of generality, the vertices V are labeled from 1 to n. Let E be the set of edges (we assume both (i, j) and (j, i) are in E for an undirected graph). Let $\boldsymbol{A}$ be the symmetric adjacency matrix where $A_{i,j} = A_{j,i}$ is equal to the edge weight, or is just 1 for an unweighted graph, and $A_{i,j}$ is 0 for each non-edge. Let S be a set of vertices in the graph and let $\bar{S}$ be the complement set $\bar{S} = V \setminus S$. The notation ∂S indicates is the number of edges (or total edge weight) needed to separate the set

S from the rest of the graph: $\partial S = \sum_{(i,j)\in E, i\in S, j\in\bar{S}} A_{i,j}$. The notation $\mathrm{Vol}(S)$ is the sum of edges involving vertices in S: $\mathrm{Vol}(S) = \sum_{(i,j)\in E, i\in S} A_{i,j}$. By convention, we set $\mathrm{Vol}(G) = \mathrm{Vol}(V)$. We write $\mathbf{1}$ for the vector all ones, so $\mathrm{Vol}(G) = \mathbf{1}^\mathsf{T}\boldsymbol{A}\mathbf{1}$ and $\mathrm{Tr}(\cdot)$ denotes the trace.

The conductance of a set of vertices is

$$\phi(S) = \frac{\partial S}{\min\{\mathrm{Vol}(S), \mathrm{Vol}(\bar{S})\}}. \tag{1}$$

In principle, minimizing conductance finds sets where $\mathrm{Vol}(S)$ is large and ∂S is small. Thus, it is interesting that empirical NCPs suggest that the *best* conductance sets are not the largest. The μ-conductance of a graph is

$$\begin{aligned}\phi_\mu(G) = \ & \underset{S\subset V}{\text{minimize}} \ \ \phi(S) \\ & \text{subject to} \ \ \mu\mathrm{Vol}(G) \le \mathrm{Vol}(S) \le \mathrm{Vol}(G)/2.\end{aligned} \tag{2}$$

Here we adopt a slightly different definition from Lovasz and Simonovits's original paper. The definitions are similar in spirit as they both neglect sets with volume smaller than a specific volume but the original one involves a perturbed conductance. Note that if the set of smallest conductance in the graph G is *large* with $\mathrm{Vol}(S) \approx \mathrm{Vol}(G)/2$, then there is no difference between the μ-conductance and conductance values. It is only for graphs with hypothetical real-world NCP structure that we expect to see interesting behavior from μ-conductance.

2.1. Cheeger Inequalities and Spectral Cuts

The Cheeger inequality gives a two-sided bound to the set of best conductance in a graph via an eigenvector computation (Chung, 2007; Cheeger, 1969). Our manuscript focuses on lower bounding the conductance of sets, rather than upper-bounding them, so we are only concerned with one side of the Cheeger inequality. The eigenvector computation uses the Laplacian matrix $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$, where $\boldsymbol{D}$ is a diagonal matrix of row-sums of $\boldsymbol{A}$, that is, $\boldsymbol{D} = \mathrm{Diag}(\mathbf{d})$ where $\mathbf{d} = \boldsymbol{A}\mathbf{1}$. Formally, let

$$\begin{aligned}\lambda_2 = \ & \underset{\mathbf{x}\in\mathbb{R}^V}{\text{minimize}} \ \ \mathbf{x}^\mathsf{T}\boldsymbol{L}\mathbf{x} \\ & \text{subject to} \ \ \mathbf{x}^\mathsf{T}\boldsymbol{D}\mathbf{x} = 1, \mathbf{x}^\mathsf{T}\mathbf{d} = 0.\end{aligned} \qquad \text{(Spectral Cut)}$$

The value λ_2 is the second smallest generalized eigenvector of $\boldsymbol{L}\mathbf{x} = \lambda\boldsymbol{D}\mathbf{x}$. This eigenvector problem is called a spectral cut because $\mathbf{x}^T\boldsymbol{L}\mathbf{x}$ computes a cut in the graph that we have relaxed over the space of eigenvectors. It is well-known that

$$\lambda_2/2 \le \min_{S\subset V} \phi(S).$$

2.2. Network Community Profiles

Network community profiles are typically computed by running either seeded PageRank (Andersen et al., 2006), a flow improvement algorithm (Lang & Rao, 2004; Andersen & Lang, 2008), or a customized procedure (Gleich & Seshadhri, 2012) over a large number of random seeds with parameters designed to explore a variety of set sizes as in (Leskovec et al., 2009; Jeub et al., 2015). Formally, the NCP is the lower envelop of the size-vs-conductance over all sets in the graph (see the lower bound in Figure 1). We find it useful to display a heatmap over all sets sampled in addition to the lower envelop. Further related concepts are the spectral profile and balanced cuts, see Appendix B.1,B.2.

Algorithm 1 MuConductanceLowRankSDPLowerBound

Require: A graph G, a scalar μ, and rank parameter k
Ensure: A lower bound on $\phi_\mu(G)$
1: Compute a KKT point of (5) (e.g. using an Augmented Lagrangian and LBFGSB as in Section 4).
2: Let $\boldsymbol{Y}^*$ be the solution of (5) at KKT
3: Let θ be the value from Lemma 3.5, found via an eigenvalue computation.
4: **Return** $\frac{1}{2}(\mathrm{Tr}(\boldsymbol{Y}^*\boldsymbol{L}\boldsymbol{Y}^*) - \theta \cdot \min\{1, \frac{(1-\mu)n}{\mu\mathrm{Vol}(G)}\})$.

3. Main Theorem

The main theorem of our paper is a computable and informative lower bound on the μ-conductance of a graph.

Theorem 3.1. *Let G be a connected, undirected graph. Fix $0 \le \mu \le 1/2$. Let $\boldsymbol{Y}^*$ and θ be from Algorithm 1. Then*

$$\tfrac{1}{2}(\mathrm{Tr}(\boldsymbol{Y}^*\boldsymbol{L}\boldsymbol{Y}^*) - \theta \cdot \min\{1, \frac{(1-\mu)n}{\mu Vol(G)}\}) \le \phi_\mu(G).$$

This theorem yields an a posteriori bound as we have no a priori guarantee on the value of θ. In practice, θ is small, around 10^{-3} or 10^{-4} in most cases.

To prove the main theorem, we work through successive transformations of optimization problems that produce lower bounds on μ-conductance. The first is a spectral program akin to (Spectral Cut). This is relaxed into a computable SDP. That does not scale to larger problems, and so we translate it into a (non-convex) low-rank SDP. The low-rank SDP can only be locally optimized. Consequently, we derive an a posteriori bound by showing that any local minimizer of the low-rank SDP problem is related to a perturbed SDP.

3.1. A Spectral Program for μ-conductance

The problem (Spectral Cut) is equivalently stated $\min \frac{\mathbf{x}^\mathsf{T}\boldsymbol{L}\mathbf{x}}{\mathbf{x}^\mathsf{T}\boldsymbol{D}\mathbf{x}}$ s.t. $\mathbf{x}^T\mathbf{d} = 0$. This form makes a more direct relationship with conductance since if $\mathbf{x}_S$ is an indicator vector for a set S, $\mathbf{x}_S^\mathsf{T}\boldsymbol{L}\mathbf{x}_S = \partial S$, and $\mathbf{x}_S^\mathsf{T}\boldsymbol{D}\mathbf{x}_S$ is $\mathrm{Vol}(S)$. To satisfy $\mathbf{x}_S^\mathsf{T}\mathbf{d} = 0$ and $\mathbf{x}_S^\mathsf{T}\boldsymbol{D}\mathbf{x}_S = 1$, as in (Spectral Cut) we

(a) $\mu = 0.03$ (b) $\mu = 0.3$

Figure 3. We show a vector from the rank-1 approximation of optimal SDP solution $\boldsymbol{X}$ on a synthetic graph with 537 nodes and 1327 edges. The vector is shown as colored markers and as sorted values. The graph is constructed to have a small, good conductance set at the center. This shows that as μ increases, the solution vector delocalizes over the entire network to respond to other sets of reasonably small conductance.

shift and re-scale $\mathbf{x}_S$ to

$$\boldsymbol{\psi}_S = \sqrt{\frac{\mathrm{Vol}(G)}{\mathrm{Vol}(S)\mathrm{Vol}(\bar{S})}}\left(\mathbf{x}_S - \frac{\mathrm{Vol}(S)}{\mathrm{Vol}(G)}\mathbf{1}\right).$$

The problem with spectral cut is that *if* the set of minimal conductance is small, the solution $\mathbf{x}$ is often highly localized. In order to model μ-conductance, consider a set S with volume about $\mu\mathrm{Vol}(G)$ and further consider the scaled and shifted indicator vector $\boldsymbol{\psi}_S$ on this set. Then we find that $|x_i| \geq \sqrt{\frac{\mu}{(1-\mu)\mathrm{Vol}(G)}}$ and $|x_i| \leq \sqrt{\frac{1-\mu}{\mu\mathrm{Vol}(G)}}$. This suggests that if we expect $\mathbf{x}$ to indicate a large set, something where $\min(\mathrm{Vol}(S), \mathrm{Vol}(\bar{S}))$ large, then the elements of $\mathbf{x}$ should be small, but not too small, and delocalized. Thus we add constraints to spectral cut (Spectral Cut) to bound the entries, either the infinity norm or maximum of $\mathbf{x}$ and to separate small entries around zero. This should help spread the mass of $\mathbf{x}$ over the graph as in Figure 3 (where we look at the solution based on the forthcoming SDP). Consequently, we pose the following modified spectral cut

$$\begin{aligned} \lambda_\mu = \quad & \underset{\mathbf{x}\in\mathbb{R}^V}{\text{minimize}} && \mathbf{x}^\mathsf{T}\boldsymbol{L}\mathbf{x} && (3)\\ & \text{subject to} && \mathbf{x}^\mathsf{T}\boldsymbol{D}\mathbf{x} = 1, \mathbf{x}^\mathsf{T}\mathbf{d} = 0 && (a,\ b)\\ & && \|\mathbf{x}\|_\infty \leq \sqrt{\frac{1-\mu}{\mu\mathrm{Vol}(G)}} && (c)\\ & && |\mathbf{x}_i| \geq \sqrt{\frac{\mu}{(1-\mu)\mathrm{Vol}(G)}} && (d) \end{aligned}$$

In particular, the parameter μ in this program corresponds to the one in μ-conductance. Notice that for any set S with volume less than $\mu\mathrm{Vol}(G)$, $\boldsymbol{\psi}_S$ is not in the feasible region of (3). Although the optimal solution of (3) does not have to follow the form of $\boldsymbol{\psi}_S$, we believe constraints (c) and (d) will rule out small and localized sets.

In addition, we have $\lim_{\mu\to 0^+}\lambda_\mu = \lambda_2$. Even stronger, for all $\mu \leq$ some μ^*, we have $\lambda_\mu = \lambda_2$. Thus as μ gets close to 0, program (3) simplifies to program (Spectral Cut).

Lemma 3.2. *For* $0 \leq \mu \leq 1/2$, $\frac{\lambda_\mu}{2} \leq \phi_\mu(G)$.

This is analogous to "easy side" of the Cheeger inequality that creates a vector from the optimal set. The full proof of this is in Appendix A.1.

3.2. A Semi-definite Program for μ-conductance

We are not aware of any existing techniques to directly solve the problem in the form (3). However, it can be relaxed into a semi-definite program (SDP).

$$\begin{aligned} \lambda_\mu^{\mathrm{sdp}} = \quad & \underset{\boldsymbol{X}\succeq 0}{\text{minimize}} && \mathrm{Tr}(\boldsymbol{L}\boldsymbol{X}) && (4)\\ & \text{subject to} && \mathrm{Tr}(\boldsymbol{D}\boldsymbol{X}) = 1, \mathrm{Tr}(\mathbf{d}\mathbf{d}^\mathsf{T}\boldsymbol{X}) = 0\\ & && \mathrm{Diag}(\boldsymbol{X}) \leq \tfrac{1-\mu}{\mu}\tfrac{\mathbf{1}}{\mathrm{Vol}(G)}\\ & && \mathrm{Diag}(\boldsymbol{X}) \geq \tfrac{\mu}{1-\mu}\tfrac{\mathbf{1}}{\mathrm{Vol}(G)}. \end{aligned}$$

The derivation of this relaxation is standard (we walk through it in Appendix A.2 for completeness). It follows from replacing $\mathbf{x}$ with the symmetric positive semi-definite matrix $\boldsymbol{X} = \mathbf{x}\mathbf{x}^T$ and writing the constraints in an equivalent fashion, then relaxing over all symmetric positive semi-definite matrices. This gives the expected result

Lemma 3.3. *For* $0 \leq \mu \leq 1/2$, *we have* $\lambda_\mu^{\mathrm{sdp}} \leq \lambda_\mu$.

Interestingly, when $\mu = \frac{1}{2}$, our (4) is equivalent to the minimum bisection SDP lower bound (Burer & Monteiro, 2003) used in Leskovec et al. (2009, section 5.2) up to scale, which is a previously known lower bound for network community profiles at exactly half the volume. Formally, we have the following relationship.

Lemma 3.4.

$$\frac{1}{2}\lambda_{1/2}^{\mathrm{sdp}} = \frac{2}{\mathrm{Vol}(G)}C_G.$$

where C_G is the optimum of the minimum bisection SDP. The proof is included in Appendix A.3.

3.3. A Low-rank Program for μ-conductance

The problem with (4) is that it has n^2 variables, which means the running time will be worse than $O(n^3)$ in most cases, and may be $O(n^6)$ in the worst case. Thus it's difficult to get a high-precision solution on graphs with more than a few thousand nodes, which makes it impractical for graphs with tens or hundreds of thousand nodes. However, notice that this program only contains $O(n)$ constraints, thus this program admits an optimal solution with rank at most $O(\sqrt{n})$ (Lemon et al., 2016). This motivates us to change this program to a low-rank SDP formulation via Burer-Monterio method (Burer & Monteiro, 2003). Under some mild assumptions, the Burer-Monterio method has good optimality and convergence guarantee (Boumal et al., 2016; Cifuentes, 2021; Cifuentes & Moitra, 2022). We factorize the positive semi-definite matrix $\boldsymbol{X}$ into $\boldsymbol{Y}\boldsymbol{Y}^\mathsf{T}$ and introduce slack variables $\mathbf{s}$ to simplify the inequality constraints to simple

bounding box constraints. After these transformations, we arrive at the low-rank program

$$\begin{aligned}\lambda_\mu^{\text{lrsdp}} = \underset{\boldsymbol{Y}\in\mathbb{R}^{n\times k}}{\text{minimize}}\ & \operatorname{Tr}(\boldsymbol{Y}^\mathsf{T}\boldsymbol{L}\boldsymbol{Y}) && (5)\\ \text{subject to}\ & \operatorname{Tr}(\boldsymbol{Y}^\mathsf{T}\boldsymbol{D}\boldsymbol{Y}) = 1, \|\boldsymbol{Y}^\mathsf{T}\mathbf{d}\|_F^2 = 0 && (e,f)\\ & \operatorname{Diag}(\boldsymbol{Y}\boldsymbol{Y}^\mathsf{T}) + \mathbf{s} = \tfrac{1-\mu}{\mu\operatorname{Vol}(G)}\mathbf{1} && (g)\\ & \mathbf{s} \geq \mathbf{0} && (h)\\ & \mathbf{s} \leq \tfrac{1-2\mu}{\mu(1-\mu)}\tfrac{\mathbf{1}}{\operatorname{Vol}(G)}. && (i)\end{aligned}$$

Here k is the rank parameter we can tune and we know if $k = \Omega(\sqrt{n})$, then $\lambda_\mu^{\text{lrsdp}} = \lambda_\mu^{\text{sdp}}$.

3.4. Establishing an Overall Bound

However, the drawback of (5) is non-convexity, which makes it hard to be solved globally. Instead we consider the KKT points of (5). Since (5) is not convex, satisfying KKT conditions of it is no longer sufficient for global optimality. But if we compare the KKT conditions of (4) and (5) closely, we observe that the KKT points of (5) directly satisfy all KKT conditions of (4) except one dual feasibility condition. And the violation of this condition characterizes how far the KKT points of the low-rank program is away from the optimum of the SDP. Formally, let $\lambda \in \mathbb{R}, \beta \in \mathbb{R}, \boldsymbol{\gamma} \in \mathbb{R}^n, \mathbf{g} \in \mathbb{R}^n, \boldsymbol{\ell} \in \mathbb{R}^n$ be Lagrangian multipliers corresponding to constraints $(e), (f), (g), (h), (i)$, then we have the following important observation.

Lemma 3.5. *For a primal-dual pair $\boldsymbol{Y}^*, \mathbf{s}^*, \lambda^*, \beta^*, \boldsymbol{\gamma}^*, \mathbf{g}^*, \boldsymbol{\ell}^*$ satisfying the KKT conditions of (5), denote*

$$\theta = -\min\{0, \lambda_{min}(\boldsymbol{L} - \lambda^*\boldsymbol{D} - \beta^*\mathbf{d}\mathbf{d}^\mathsf{T} - \operatorname{Diag}(\boldsymbol{\gamma}^*))\},$$

then we have

$$\operatorname{Tr}(\boldsymbol{Y}^{*\mathsf{T}}\boldsymbol{L}\boldsymbol{Y}^*) - \theta\cdot\min\{1, \frac{(1-\mu)n}{\mu Vol(G)}\} \leq \lambda_\mu^{\text{sdp}}.$$

Basically this Lemma states that if the dual variable $\boldsymbol{Z} = \boldsymbol{L} - \lambda\boldsymbol{D} - \beta\mathbf{d}\mathbf{d}^\mathsf{T} - \operatorname{Diag}(\boldsymbol{\gamma})$ is not positive semi-definite, then we can still lower bound the optimum of the SDP (4) by subtracting a quantity related to this violation from the objective of (5). The full proof of this is in Appendix A.4.

Summing up all the Lemmas we get, we now have

$$\begin{aligned}&\frac{1}{2}(\operatorname{Tr}(\boldsymbol{Y}^{*\mathsf{T}}\boldsymbol{L}\boldsymbol{Y}^*) - \theta\cdot\min\{1, \frac{(1-\mu)n}{\mu\operatorname{Vol}(G)}\})\\ &\quad\leq \frac{1}{2}\lambda_\mu^{\text{sdp}} \leq \frac{1}{2}\lambda_\mu \leq \phi_\mu(G).\end{aligned}$$

This concludes the proof of Theorem 3.1.

4. Methods

In order to solve the non-convex low-rank SDP (5), we use an augmented Lagrangian approach. The augmented Lagrangian method is an iterative algorithm where in each iteration we minimize a function including the original objective, the estimated Lagrangian multipliers, and the penalty term which drives the solution into feasible region. It has been shown in practice that the augmented Lagrangian method achieves good performance in solving low-rank SDP problems (Burer & Monteiro, 2003).

Let σ be the coefficient for the penalty term and $\lambda, \beta, \boldsymbol{\gamma}$ be the Lagrangian multipliers defined in Section 3.4. The augmented Lagrangian for (5) without the bounding box constraint (h) and (i) is

$$\begin{aligned}&\mathcal{L}_A(\boldsymbol{Y}, \mathbf{s}; \lambda, \beta, \boldsymbol{\gamma}, \sigma)\\ &= \operatorname{Tr}(\boldsymbol{Y}^\mathsf{T}\boldsymbol{L}\boldsymbol{Y}) - \lambda(\operatorname{Tr}(\boldsymbol{Y}^\mathsf{T}\boldsymbol{D}\boldsymbol{Y}) - 1) - \beta(\mathbf{d}^\mathsf{T}\boldsymbol{Y}\boldsymbol{Y}^\mathsf{T}\mathbf{d})\\ &\quad - \boldsymbol{\gamma}^\mathsf{T}(\operatorname{Diag}(\boldsymbol{Y}\boldsymbol{Y}^\mathsf{T}) + \mathbf{s} - \tfrac{(1-\mu)}{\mu}\tfrac{\mathbf{1}}{\operatorname{Vol}(G)})\\ &\quad + \frac{\sigma}{2}\Big((\operatorname{Tr}(\boldsymbol{Y}^\mathsf{T}\boldsymbol{D}\boldsymbol{Y}) - 1)^2 + (\mathbf{d}^\mathsf{T}\boldsymbol{Y}\boldsymbol{Y}^\mathsf{T}\mathbf{d})^2\\ &\qquad + \|\operatorname{Diag}(\boldsymbol{Y}\boldsymbol{Y}^T) + \mathbf{s} - \tfrac{(1-\mu)}{\mu}\tfrac{\mathbf{1}}{\operatorname{Vol}(G)}\|_2^2\Big).\end{aligned}$$

In each iteration, we solve the following subproblem

$$\begin{aligned}\underset{\boldsymbol{Y},\mathbf{s}}{\text{minimize}}\ & \mathcal{L}_A(\boldsymbol{Y}, \mathbf{s}; \lambda, \beta, \boldsymbol{\gamma}, \sigma)\\ \text{subject to}\ & \mathbf{0} \leq \mathbf{s} \leq \tfrac{1-2\mu}{\mu(1-\mu)}\tfrac{\mathbf{1}}{\operatorname{Vol}(G)}\end{aligned} \quad (6)$$

using a Limited-Memory BFGS method with bound constraints on variables (Byrd et al., 1995). Since L-BFGS-B is a quasi-Newton Method, it requires the gradient of $\mathcal{L}_A$ with regard to variables $\boldsymbol{Y}$ and $\mathbf{s}$. Let

$$\mathbf{u} = \operatorname{Diag}(\boldsymbol{Y}\boldsymbol{Y}^\mathsf{T}) + \mathbf{s} - \frac{\mu}{(1-\mu)\operatorname{Vol}(G)}\mathbf{1},$$

we have

$$\begin{aligned}\nabla_{\boldsymbol{Y}}\mathcal{L}_A &= 2\boldsymbol{L}\boldsymbol{Y} - 2(\lambda - \sigma(\operatorname{Tr}(\boldsymbol{Y}^\mathsf{T}\boldsymbol{D}\boldsymbol{Y}) - 1))\boldsymbol{D}\boldsymbol{Y}\\ &\quad - 2(\beta - \sigma\mathbf{d}^\mathsf{T}\boldsymbol{Y}\boldsymbol{Y}^\mathsf{T}\mathbf{d})\mathbf{d}\mathbf{d}^\mathsf{T}\boldsymbol{Y}\\ &\quad - 2\left((\boldsymbol{\gamma} - \sigma\mathbf{u})\mathbf{1}^\mathsf{T}\right)\circ\boldsymbol{Y},\\ \nabla_{\mathbf{s}}\mathcal{L}_A &= -\boldsymbol{\gamma} + \sigma\mathbf{u}\end{aligned}$$

where $\circ$ is the element-wise or Hadamard product.

After each solve, we update the multipliers and penalty parameters following Alg 17.4 of Nocedal & Wright (1999).

Initialization and the rank parameter k. As L-BFGS-B is a quasi-Newton method, convergence is faster when the starting point is close to the optimal solution. We initialize $\boldsymbol{Y}$ by the k eigenvectors corresponding to the k smallest non-zero eigenvalues of normalized Laplacian $\boldsymbol{D}^{-1/2}\boldsymbol{L}\boldsymbol{D}^{-1/2}$. This is based on the observation that when $k = 1$, program (5) degenerates to program (3) and the Fiedler vector remains the optimal solution for small μ.

Comparison against SDP solvers. For small enough problems, we can solve both the SDP (4) as well as the low-rank

SDP (5). Therefore we compare our LRSDP with them on two small synthetic graphs with 85 and 537 vertices. We intentionally construct the two synthetic graphs with a dense core that has minimal conductance and localizes the Fiedler vector. (See Figure 3 and discussion of the construction in Appendix C.1.) We compare the solvers for different μs on each graph and the results are summarized in Table 1. These show that our LRSDP has objective values extremely close to solving the SDPs directly and is *much* faster.

Non-monotonic results. The results from μ-conductance must be monotonic. That is, for $\mu_1 \geq \mu_2$, we must have $\phi_{\mu_1} \geq \phi_{\mu_2}$ by set inclusion properties of the μ-conductance function. Because we have a lower bound, we found scenarios where the lower bounds were not monotonically increase in μ. Since our investigations typically involve multiple values of μ, we simply adjust the bounds to reflect the *tightest* lower bound from any value of μ that we computed. Practically, this corresponds to taking a stepwise maximum over the experimental results.

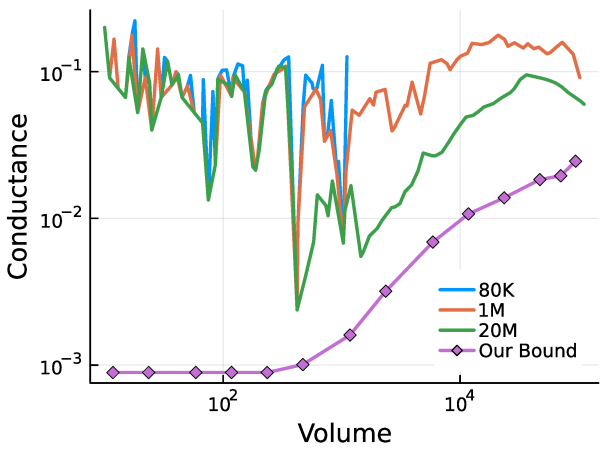


Figure 4. Gap shrinking effect illustrated on HepPh. The upper three line plots are the NCPs determined by different number of sets. This shows the more sets we search using seeded PageRank, the smaller the gap between the NCP and our lower bound.

5. Experiments

In this section, we revisit the lower bounds from the introduction (Figure 2). We then explore how the *running time* of our programs is affected by graph size, μ, and rank parameter k. Further, although directly tracking the true NCP is co-$\mathbb{NP}$ hard, we are still able to study the gap between the true NCP and our lower bound by a squeeze bound or *gap shrinking* analysis. In the end we do one interesting k-core analysis on one graph using our algorithm, which reveals the potential for use in other network analysis tasks.

5.1. Computational Details

When solving the subproblem (6) in our augmented Lagrangian procedure, we use L-BFGS-B with $m = 3$. We set the default tolerance of stationarity condition and primal feasibility condition of our augmented Lagrangian as 10^{-5}. For each dataset, we pick a set of μs varying from 10^{-6} to 0.4 which is dense enough to form an informative lower bound curve. We exhaustively try k from $\{1, 3, 5, 10\}$. To generate the NCP plots, we empirically sample a large number of sets from a seeded PageRank based method (Andersen et al., 2006). Specifically, we randomly sample a large collection of seeds and then try different ε ranging from 10^{-2} to 10^{-8}. For each seeded PageRank we get, we perform a sweepcut to get several sets with good μ-conductance.

5.2. Summary of Key Findings from Introduction

The main figure for our experiments is Figure 2. This shows the lower bounds on the NCPs produced by our procedure. We test our procedure on AstroPh , HepPh (Leskovec et al., 2007), Email-Enron (Leskovec et al., 2009), Facebook-Page (Rozemberczki et al., 2019), Deezer (Rozemberczki & Sarkar, 2020) and DBLP (Boldi & Vigna, 2004; Boldi et al., 2011). Their sizes are in Table 2. We can see although there is a gap between our lower bound and the NCP generated by seeded PageRank, our algorithm provides an informative lower bound which mirrors the trend of the NCP plot.

5.3. Running Time

We illustrate the effects of graph size, μ, and rank parameter k on the running time of our program. The results are summarized in Table 3. We can clearly observe that with graph size or rank parameter increasing, the running time increases but roughly linearly. This is expected because each iteration of L-BFGS-B takes linear time with regard to number of variables. Also, we can see with μ increasing, the running time tends to increase as well. The intuition is that with μ increasing, the feasible region of the low-rank program shrinks, which makes optimization harder. Also, our initialization favors smaller μ.

5.4. Gap Shrinking

Our theory gives a lower bound on the μ-conductance scores. To study how close our lower bound can be to the true μ-conductance score which is co-$\mathbb{NP}$ hard to compute, we study how close an upper bound of the μ-conductance score can be to our lower bound. This kind of squeeze bound gives an indirect way to estimate the real gap. In order to explore how tight our lower bound is, in other words how small the gap can be, for the HepPh graph, we dramatically increase the number of samples of sets from seeded PageRank. The results are summarized in Figure 4. We see that our lower bound is not that loose: about 1/3 off.

Table 1. To validate that the LRSDP (5) and SDP (4) are similar on problems where we can compute both, we examine their objective values on two small synthetic graphs (e.g. Figure 3). We choose two established SDP solvers, SCS (O'Donoghue et al., 2016) and Mosek (ApS, 2022). This shows that LRSDP gives nearly identical results and is much faster. Here LB stands for the lower bound provided by our low-rank program, which theoretically should be a lower bound for objective value of all SDP solutions. Empirically some objective values are lower than this bound because numerically they do not strictly satisfy all primal feasibility conditions.

NODES	EDGES	μ	OBJECTIVE VALUE			BOUND	TIME		
			LRSDP	SCS	MOSEK	LB	LRSDP	SCS	MOSEK
85	193	0.01	0.004407	0.004407	0.004406	0.004398	0.7 S	16.7 S	5.3 S
		0.05	0.004510	0.004511	0.004508	0.004499	2.1 S	18.4 S	4.9 S
		0.25	0.007318	0.007223	0.007314	0.007292	1.8 S	18.2 S	6.0 S
537	1327	0.01	0.001092	0.001089	0.001081	0.001083	17.8 S	1.6 HRS	16.9 HRS
		0.03	0.001115	0.001113	0.001092	0.001056	17.3 S	12.2 HRS	15.6 HRS
		0.1	0.001444	0.001440	0.001428	0.001390	21.0 S	56.7 MIN	13.4 HRS
		0.3	0.002733	0.002732	0.002731	0.002720	11.8 S	1.8 HRS	18.8 HRS

Table 2. Network Datasets. We report the number of vertices and edges of the largest connected component with self-loops removed.

DATASET	$\|V\|$	$\|E\|$
HEPPH	11,204	117,619
ASTROPH	17,903	196,972
FACEBOOK-PAGE	22,470	170,823
DEEZER-EUR	28,281	92,752
EMAIL-ENRON	33,696	180,811
DBLP	226,413	716,460

5.5. Investigation with k-cores

In order to show the potential of applying our method to broader network analysis tasks, we apply our low-rank program to analyze the NCP of k-cores of a graph (Seidman, 1983). The inspiration for this study is a discussion over whether the NCP represents a signal or noise mode of a graph (Zhang & Rohe, 2018). The core number of a vertex in a graph is the largest integer k such that the process of repeatedly removing vertices with degree less than k will not delete this vertex from the graph. So the 1-core is the entire graph. The 2-core is there result of sequentially deleting all degree 1 nodes. By analyzing the NCP of k-cores with various k, we can have a deeper understanding of the structure of a network. The results are summarized in Figure 5. These show that the NCP structure is preserved for Email-Enron up through the 5-core and is largely preserved at the 7-core. While this single experiment does not to resolve the question of signal vs. noise for the NCP, it does show how our tools could be used to study it.

5.6. Comparison with Other Lower Bounds

Besides our μ-conductance lower bound, there are two previously known lower bounds for network community profiles mentioned in (Leskovec et al., 2009), one spectral bound induced by Cheeger inequality and the Fiedler vector that is independent of volume and the other is given by the minimum bisection SDP. To get a comprehensive understanding of how our lower bound behaves compared with existing lower bounds we compare on two graphs. As is shown in Lemma 3.4, the minimum bisection SDP lower bound is actually equivalent to ours at $\mu = \frac{1}{2}$, here we directly solve our low-rank SDP at $\mu = \frac{1}{2}$ instead of solving the minimum bisection SDP. The results on AstroPh and HepPh graphs are shown in Figure 6. These show that we smoothly interpolate between the bounds as expected.

Table 3. This table summarizes the running time on two graphs with a few different μ and k choices. We report the running time of the augmented Lagrangian method (ALM) for solving low-rank SDP and eigenvalue computation (EIGVAL) for calculating the dual feasibility violation separately.

GRAPH	μ	k	TIME	
			ALM	EIGVAL
HEPPH $\|V\| = 11204$ $\|E\| = 117619$	0.001	3	1.7 MIN	30.8 S
		5	3.4 MIN	48.1 S
		10	6.2 MIN	36.6 S
	0.1	3	1.8 HRS	21.4 MIN
		5	3.1 HRS	12.9 MIN
		10	6.9 HRS	23.0 MIN
DBLP $\|V\| = 226413$ $\|E\| = 716460$	0.001	3	21.8 HRS	3.1 HRS
		5	1.8 DAYS	3.5 HRS
		10	2.6 DAYS	8.7 HRS
	0.1	3	1.6 DAYS	1.9 HRS
		5	3.4 DAYS	33.6 MIN
		10	3.1 DAYS	5.6 HRS

5.7. Impact of Rank on the Lower Bound

The rank parameter k plays a key role in our solution. As is shown in Section 5.3, a higher k will slow down the computation. It also impacts the a posterori bound we achieve. We study this tradeoff here. The results on AstroPh and HepPh graphs are shown in Figure 7. We do not observe a strong pattern. Consequently, we recommend setting $k = 5$ as a pragmatic middle ground.

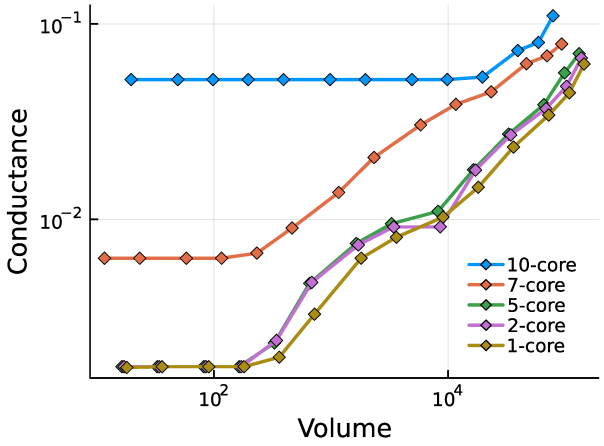


Figure 5. k-core analysis on Email-Enron.

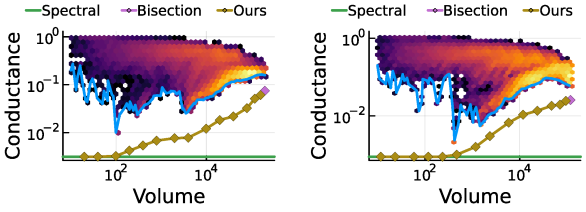


Figure 6. Comparison with spectral lower bounds (bottom green line) and minimum bisection SDP lower bounds (purple point at right) on two graphs Astro (Left) and HepPh (Right) graphs. We can observe that our μ-conductance is capable of offering a lower bound at more positions and provides the expected smooth interpolation between these bounds that had been missing from existing approaches.

6. Discussion and Future work

The theorem and algorithm here allows a complete characterization of the network community profile for graphs with over 200,000 vertices. This, in turn, has implications on random sampling on real-world graphs as discussed in the introduction. There were some theoretical datapoints known regarding bounds on the NCP. For instance, the position of the spectral partitioning set and the associated Cheeger inequality gives one point in the size-vs-conductance space. Another point arises from SDP-based methods (Leskovec et al., 2009) (Section 5.2) for bisection splits. Our tools are the first to interpolate between the two with robust bounds.

Our methods involve choosing a rank parameter, a value of μ, as well as tolerances associated with the L-BFGS-B based procedure. These can have non-trivial interactions. Typically, we find that the values of θ involved in the lower bound are small (think 10^{-4}). In a counter-intuitive observation, we found instances where using a weaker or higher tolerance values results in better or larger lower bounds on the μ-conductance value because the value of θ was changed.

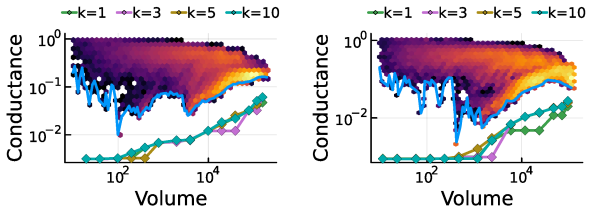


Figure 7. The effect of rank parameter k on the lower bound illustrated on Astro (Left) and HepPh (Right). We observe that different k exhibit comparable curves but extremely small k will get worse lower bounds.

In other scenarios, we found values of μ where we could not find a way to adjust rank and tolerance to make the value of θ small enough. This made the lower bound was extremely loose (or even negative in some scenarios). Our choice of overall parameters tends to minimize this.

Although we have focused on lower bounds in this manuscript (in the interest of space). In a related line of work line of work, we have developed a two-sided bound on (3), the μ-conductance spectral program (Huang & Gleich, 2023). This gives a full Cheeger-like characterization of this program. There are also numerous variants of Cheeger inequality (Louis et al., 2012; Kwok et al., 2013; Koutis et al., 2014; Zhu & Gleich, 2016), including those versions using multiple eigenvalues as well as more general weightings. Finding a multivector and multiset generalization of these results would be useful in a variety of scenarios.

At the moment, we are able to handle a variety of real-world graphs, but the runtime is still slow. Computations take days instead of hours. Scaling these algorithms up to the LiveJournal example from Figure 1 is another challenge, with many potential avenues including parallelization. Delocalized eigenvectors for spectral clustering also arise from the statistics perspective in terms of regularization (Amini et al., 2013; Zhang & Rohe, 2018). Many of these techniques involve directly regularizing the graph Laplacian by adding a small multiple of the all ones matrix, akin to the PageRank perturbation. We hope to study relationships between these regularization techniques and μ-conductance in the future, especially as they would help promise much faster runtimes via eigenvector techniques instead of low-rank SDPs.

Acknowledgments

We are grateful for the valuable suggestions from the reviewers. Huang and Gleich's work was funded in part by NSF CCF-1909528, IIS-2007481, DOE DE-SC0023162, and the IARPA Agile program. Seshadhri's was funded by NSF DMS-2023495, CCF-1839317, and CCF-1908384.

References

Ahmed, N., Neville, J., and Kompella, R. Reconsidering the foundations of network sampling. In *WIN 10*, 2010.

Ahmed, N. K., Duffield, N., Neville, J., and Kompella, R. Graph sample and hold: A framework for big-graph analytics. In *SIGKDD*, pp. 1446–1455. ACM, ACM, 2014.

Amini, A. A., Chen, A., Bickel, P. J., and Levina, E. Pseudo-likelihood methods for community detection in large sparse networks. *The Annals of Statistics*, 41 (4), August 2013. doi: 10.1214/13-aos1138. URL https://doi.org/10.1214/13-aos1138.

Andersen, R. and Lang, K. An algorithm for improving graph partitions. In *Proceedings of the 19th annual ACM-SIAM Symposium on Discrete Algorithms (SODA2008)*, pp. 651–660, January 2008.

Andersen, R., Chung, F., and Lang, K. Local graph partitioning using PageRank vectors. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, 2006. URL http://www.math.ucsd.edu/~fan/wp/localpartition.pdf.

ApS, M. *The MOSEK optimization toolbox for MATLAB manual. Version 10.0.*, 2022. URL http://docs.mosek.com/9.0/toolbox/index.html.

Barabási, A.-L. and Albert, R. Emergence of scaling in random networks. *Science*, 286:509–512, October 1999.

Boldi, P. and Vigna, S. The WebGraph framework I: Compression techniques. In *Proc. of the Thirteenth International World Wide Web Conference (WWW 2004)*, pp. 595–601, Manhattan, USA, 2004. ACM Press.

Boldi, P., Rosa, M., Santini, M., and Vigna, S. Layered label propagation: A multiresolution coordinate-free ordering for compressing social networks. In Srinivasan, S., Ramamritham, K., Kumar, A., Ravindra, M. P., Bertino, E., and Kumar, R. (eds.), *Proceedings of the 20th international conference on World Wide Web*, pp. 587–596. ACM Press, 2011.

Bonato, A., Gleich, D. F., Kim, M., Mitsche, D., Prałat, P., Tian, A., and Young, S. J. Dimensionality of social networks using motifs and eigenvalues. *PLoS ONE*, 9 (9):e106052, September 2014. doi: 10.1371/journal.pone.0106052.

Boumal, N., Voroninski, V., and Bandeira, A. S. The non-convex Burer–Monteiro approach works on smooth semidefinite programs. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, pp. 2765–2773, Red Hook, NY, USA, December 2016. Curran Associates Inc. ISBN 978-1-5108-3881-9.

Boyd, S., Boyd, S. P., and Vandenberghe, L. *Convex optimization*. Cambridge university press, 2004.

Burer, S. and Monteiro, R. D. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, 95(2):329–357, 2003.

Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. A limited memory algorithm for bound constrained optimization. *SIAM Journal on scientific computing*, 16(5):1190–1208, 1995.

Cheeger, J. A lower bound for the smallest eigenvalue of the laplacian. In *Proceedings of the Princeton conference in honor of Professor S. Bochner*, pp. 195–199, 1969.

Chierichetti, F. and Haddadan, S. On the complexity of sampling vertices uniformly from a graph. In *International Colloquium on Automata, Languages, and Programming (ICALP)*, pp. 149:1–149:13, 2018. doi: 10.4230/LIPIcs.ICALP.2018.149. URL https://doi.org/10.4230/LIPIcs.ICALP.2018.149.

Chierichetti, F., Dasgupta, A., Kumar, R., Lattanzi, S., and Sarlos, T. On sampling nodes in a network. In *Conference on the World Wide Web (WWW)*, 2016.

Chung, F. Four proofs of the cheeger inequality and graph partition algorithms. In *Proceedings of the ICCM*, pp. 751–772, 2007.

Cifuentes, D. On the Burer–Monteiro method for general semidefinite programs. *Optimization Letters*, 15(6):2299–2309, September 2021. ISSN 1862-4480. doi: 10.1007/s11590-021-01705-4.

Cifuentes, D. and Moitra, A. Polynomial time guarantees for the burer-monteiro method. In Oh, A. H., Agarwal, A., Belgrave, D., and Cho, K. (eds.), *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=lhLEGeBC-ru.

Dasgupta, A., Kumar, R., and Sarlos, T. On estimating the average degree. In *Conference on the World Wide Web (WWW)*, pp. 795–806, 2014.

Flake, G. W., Lawrence, S., and Giles, C. L. Efficient identification of web communities. In *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '00, pp. 150–160, New York, NY, USA, 2000. ACM. ISBN 1-58113-233-6. doi: 10.1145/347090.347121.

Gleich, D. F. and Seshadhri, C. Vertex neighborhoods, low conductance cuts, and good seeds for local community methods. In *KDD2012*, pp. 597–605, August 2012. doi: 10.1145/2339530.2339628.

Goel, S., Montenegro, R., Tetali, P., et al. Mixing time bounds via the spectral profile. *Electronic Journal of Probability*, 11:1–26, 2006.

Grover, A. and Leskovec, J. Node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pp. 855–864, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450342322. doi: 10.1145/2939672.2939754. URL https://doi.org/10.1145/2939672.2939754.

Huang, Y. and Gleich, D. F. A cheeger inequality for size-specific conductance. *arXiv*, cs.DM:2303.11452, 2023.

Jeub, L. G. S., Balachandran, P., Porter, M. A., Mucha, P. J., and Mahoney, M. W. Think locally, act locally: Detection of small, medium-sized, and large communities in large networks. *Phys. Rev. E*, 91:012821, January 2015. doi: 10.1103/PhysRevE.91.012821.

Koutis, I., Miller, G., and Peng, R. A generalized cheeger inequality. *arXiv preprint arXiv:1412.6075*, 2014.

Kwok, T. C., Lau, L. C., Lee, Y. T., Oveis Gharan, S., and Trevisan, L. Improved cheeger's inequality: Analysis of spectral partitioning algorithms through higher order spectral gap. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pp. 11–20, 2013.

Lang, K. and Rao, S. A flow-based method for improving the expansion or conductance of graph cuts. In Bienstock, D. and Nemhauser, G. (eds.), *Integer Programming and Combinatorial Optimization*, volume 3064 of *Lecture Notes in Computer Science*, pp. 325–337. Springer Berlin Heidelberg, 2004. ISBN 978-3-540-22113-5. doi: 10.1007/978-3-540-25960-2_25.

Lemon, A., So, A. M.-C., Ye, Y., et al. Low-rank semidefinite programming: Theory and applications. *Foundations and Trends® in Optimization*, 2(1-2):1–156, 2016.

Leskovec, J. and Faloutsos, C. Sampling from large graphs. In *Knowledge Data and Discovery (KDD)*, pp. 631–636. ACM, 2006.

Leskovec, J., Kleinberg, J., and Faloutsos, C. Graph evolution: Densification and shrinking diameters. *ACM transactions on Knowledge Discovery from Data (TKDD)*, 1 (1):2–es, 2007.

Leskovec, J., Lang, K. J., Dasgupta, A., and Mahoney, M. W. Statistical properties of community structure in large social and information networks. In *Proceedings of the 17th international conference on World Wide Web*, pp. 695–704, 2008.

Leskovec, J., Lang, K. J., Dasgupta, A., and Mahoney, M. W. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*, 6(1):29–123, September 2009. doi: 10.1080/15427951.2009.10129177.

Leskovec, J., Lang, K. J., and Mahoney, M. Empirical comparison of algorithms for network community detection. In *Proceedings of the 19th international conference on World wide web*, pp. 631–640, 2010.

Louis, A., Raghavendra, P., Tetali, P., and Vempala, S. Many sparse cuts via higher eigenvalues. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pp. 1131–1140, 2012.

Lovasz, L. and Simonovits, M. The mixing rate of markov chains, an isoperimetric inequality, and computing the volume. In *Proceedings of the 31st Annual Symposium on Foundations of Computer Science*, SFCS '90, pp. 346–354 vol. 1, Washington, DC, USA, 1990. IEEE Computer Society. ISBN 0-8186-2082-X. doi: 10.1109/FSCS.1990.89553.

Maiya, A. S. and Berger-Wolf, T. Y. Benefits of bias: Towards better characterization of network sampling. In *Knowledge Data and Discovery (KDD)*, pp. 105–113, 2011.

Newman, M. E. J. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003. doi: 10.1137/S003614450342480.

Newman, M. E. J. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, May 2006. doi: 10.1073/pnas.0601602103.

Nocedal, J. and Wright, S. J. *Numerical optimization*. Springer, 1999.

O'Donoghue, B., Chu, E., Parikh, N., and Boyd, S. Conic optimization via operator splitting and homogeneous self-dual embedding. *Journal of Optimization Theory and Applications*, 169(3):1042–1068, June 2016. URL http://stanford.edu/~boyd/papers/scs.html.

Perozzi, B., Al-Rfou, R., and Skiena, S. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD

'14, pp. 701–710, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450329569. doi: 10.1145/2623330.2623732. URL https://doi.org/10.1145/2623330.2623732.

Raghavendra, P., Steurer, D., and Tetali, P. Approximations for the isoperimetric and spectral profile of graphs and related parameters. In *Proceedings of the Forty-second ACM Symposium on Theory of Computing*, STOC '10, pp. 631–640, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0050-6. doi: 10.1145/1806689.1806776. URL http://doi.acm.org/10.1145/1806689.1806776.

Ribeiro, B. and Towsley, D. Estimating and sampling graphs with multidimensional random walks. In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*, pp. 390–403. ACM, 2010.

Ribeiro, B. and Towsley, D. On the estimation accuracy of degree distributions from graph sampling. In *Annual Conference on Decision and Control (CDC)*, pp. 5240–5247. IEEE, 2012.

Rozemberczki, B. and Sarkar, R. Characteristic Functions on Graphs: Birds of a Feather, from Statistical Descriptors to Parametric Models. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM '20)*, pp. 1325–1334. ACM, 2020.

Rozemberczki, B., Allen, C., and Sarkar, R. Multi-scale attributed node embedding, 2019.

Seidman, S. B. Network structure and minimum degree. *Social Networks*, 5(3):269–287, 1983. ISSN 0378-8733. doi: 10.1016/0378-8733(83)90028-X.

Seshadhri, C., Kolda, T. G., and Pinar, A. Community structure and scale-free collections of Erdös-Rényi graphs. *Physical Review E*, 85(5):056109, May 2012. doi: 10.1103/PhysRevE.85.056109.

Seshadhri, C., Sharma, A., Stolman, A., and Goel, A. The impossibility of low-rank representations for triangle-rich complex networks. *Proceedings of the National Academy of Sciences*, 117(11):5631–5637, 2020.

Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., and Mei, Q. Line: Large-scale information network embedding. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15, pp. 1067–1077, Republic and Canton of Geneva, CHE, 2015. International World Wide Web Conferences Steering Committee. ISBN 9781450334693. doi: 10.1145/2736277.2741093. URL https://doi.org/10.1145/2736277.2741093.

Vandenberghe, L. and Boyd, S. Semidefinite programming. *SIAM review*, 38(1):49–95, 1996.

von Luxburg, U., Williamson, R. C., and Guyon, I. Clustering: Science or art? In Guyon, I., Dror, G., Lemaire, V., Taylor, G., and Silver, D. (eds.), *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, volume 27 of *Proceedings of Machine Learning Research*, pp. 65–79, Bellevue, Washington, USA, 02 Jul 2012. PMLR. URL http://proceedings.mlr.press/v27/luxburg12a.html.

Watts, D. and Strogatz, S. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998. doi: 10.1038/30918.

Zhang, Y. and Rohe, K. Understanding regularized spectral clustering via graph conductance. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper/2018/file/2a845d4d23b883acb632fefd814e175f-Paper.pdf.

Zhu, Y. and Gleich, D. F. A parallel min-cut algorithm using iteratively reweighted least squares. *Parallel Computing*, 59:43–59, November 2016. doi: 10.1016/j.parco.2016.02.003.

A. Proofs

A.1. Proof of Lemma 3.2

The first lemma gives a lower bound of ϕ_μ with respect to λ_μ, the optimum of program (3).

Lemma A.1. *Given $G = (V, E)$ and $\mu \in [0, \frac{1}{2}]$, we have*

$$\frac{1}{2}\lambda_\mu \le \phi_\mu(G).$$

Proof. The basic idea is to find a test vector $\mathbf{y}$ in the feasible region of program (3) satisfying $\mathbf{y}^\mathsf{T}\boldsymbol{L}\mathbf{y} \le 2\phi_\mu(G)$. Notice that if ϕ_μ is achieved by the set T, then vector

$$\boldsymbol{\psi}_T = \sqrt{\tfrac{\mathrm{Vol}(G)}{\mathrm{Vol}(T)\mathrm{Vol}(\bar{T})}}\left(\mathbb{1}_T - \tfrac{\mathrm{Vol}(T)}{\mathrm{Vol}(G)}\mathbf{1}\right)$$

is naturally in the feasible region of (3), where $\mathbb{1}_T$ is the indicator vector for set T. As

$$\begin{aligned}\boldsymbol{\psi}_T^\mathsf{T}\boldsymbol{L}\boldsymbol{\psi}_T &= \frac{|\partial T|\mathrm{Vol}(G)}{\mathrm{Vol}(T)\mathrm{Vol}(\bar{T})}\\ &\le \frac{2|\partial T|}{\min(\mathrm{Vol}(T), \mathrm{Vol}(\bar{T}))}\\ &= 2\Phi^\mu,\end{aligned}$$

we have $\lambda^\mu \le \boldsymbol{\psi}_T^\mathsf{T}\boldsymbol{L}\boldsymbol{\psi}_T \le 2\Phi^\mu$. □

Lemma A.1 implies that the optimal value of program (3) can function as a lower bound for the μ-conductance. Furthermore, if we solve program (3) for different μs, then the curve of λ^μs with respect to corresponding μ can be a lower bound for the network community profile.

A.2. Proof of Lemma 3.3

We verify all steps of the relaxation from (3) to (4) as the proof of Lemma.

From (3), let $\mathbf{x}$ be the variable and let $\boldsymbol{X}$ be the rank-1 symmetric positive definite matrix $\boldsymbol{X} = \mathbf{x}\mathbf{x}^\mathsf{T}$. Then $\mathbf{x}^\mathsf{T}\boldsymbol{D}\mathbf{x} = 1$ is equivalent to $\mathrm{Tr}(\mathbf{x}^\mathsf{T}\boldsymbol{D}\mathbf{x}) = \mathrm{Tr}(\boldsymbol{D}\mathbf{x}\mathbf{x}^\mathsf{T}) = \mathrm{Tr}(\boldsymbol{D}\boldsymbol{X}) = 1$. Likewise, $\mathbf{x}^\mathsf{T}\mathbf{d} = 0$ is equivalent to $(\mathbf{x}^\mathsf{T}\mathbf{d})^2 = \mathrm{Tr}(\mathbf{x}\mathbf{x}^\mathsf{T}\mathbf{d}\mathbf{d}^\mathsf{T}) = 0$. Finally, for the inequality constraints, $\|\mathbf{x}\|_\infty \le \alpha$ is equivalent to $x_i^2 \le \alpha^2$ for all i, and a similar statement holds for the lower bound on $|x_i|$ too. Thus we arrive at

$$\begin{aligned}\lambda_\mu = \quad & \underset{\boldsymbol{X}=\mathbf{x}\mathbf{x}^\mathsf{T}}{\text{minimize}} && \mathrm{Tr}(\boldsymbol{L}\boldsymbol{X}) \\ & \text{subject to} && \mathrm{Tr}(\boldsymbol{D}\boldsymbol{X}) = 1 \\ & && \mathrm{Tr}(\mathbf{d}\mathbf{d}^\mathsf{T}\boldsymbol{X}) = 0 \\ & && \mathrm{Diag}(\boldsymbol{X}) \le \tfrac{1-\mu}{\mu}\tfrac{\mathbf{1}}{\mathrm{Vol}(G)} \\ & && \mathrm{Diag}(\boldsymbol{X}) \ge \tfrac{\mu}{1-\mu}\tfrac{\mathbf{1}}{\mathrm{Vol}(G)}.\end{aligned} \tag{7}$$

Note that this problem is directly equivalent to (3) because of the rank-1 condition $\boldsymbol{X} = \mathbf{x}\mathbf{x}^\mathsf{T}$. Thus we get (4) and Lemma 3.3 by relaxing the variable $\boldsymbol{X} = \mathbf{x}\mathbf{x}^\mathsf{T}$ to be a symmetric positive definite matrix.

A.3. Proof of Lemma 3.4

The minimum bisection SDP is

$$\begin{aligned}C_G = \quad & \underset{\boldsymbol{Y}\succeq 0}{\text{minimize}} && \tfrac{1}{4}\mathrm{Tr}(\boldsymbol{L}\boldsymbol{Y}) \\ & \text{subject to} && \mathrm{Tr}(\mathbf{d}\mathbf{d}^\mathsf{T}\boldsymbol{Y}) = 0 \\ & && \mathrm{Diag}(\boldsymbol{Y}) = \mathbf{1}.\end{aligned}$$

Proof of Lemma 3.4. When $\mu = \frac{1}{2}$, we notice that the two inequality constraints

$$\frac{\mu}{1-\mu}\frac{\mathbf{1}}{\mathrm{Vol}(G)} \le \mathrm{Diag}(\boldsymbol{X}) \le \frac{1-\mu}{\mu}\frac{\mathbf{1}}{\mathrm{Vol}(G)}$$

become the equality constraint $\mathrm{Diag}(\boldsymbol{X}) = \frac{\mathbf{1}}{\mathrm{Vol}(G)}$. Further, we can verify that $\mathrm{Tr}(\boldsymbol{D}\boldsymbol{X}) = 1$ is naturally satisfied when $\mathrm{Diag}(\boldsymbol{X}) = \frac{\mathbf{1}}{\mathrm{Vol}(G)}$. Therefore the only difference is scaling. If we let $\boldsymbol{X} = \mathrm{Vol}(G)\boldsymbol{Y}$ and scale C_G by $\frac{4}{\mathrm{Vol}(G)}$, we can see the two programs are exactly the same. Then we get $\lambda_{1/2}^{\mathrm{sdp}} = \frac{4}{\mathrm{Vol}(G)} C_G$. □

A.4. Proof of Lemma 3.5

To prove Lemma 3.5, we need to first make a few important observations. Here, for convenience, we relabel some programs with informative tags.

Remember we have the following SDP relaxation for program (3)

$$\begin{array}{rlll} \lambda_\mu^{\mathrm{sdp}} = & \text{minimize} & \mathrm{Tr}(\boldsymbol{L}\boldsymbol{X}) & (\mu\text{-conductance SDP}) \\ & \text{subject to} & \mathrm{Tr}(\boldsymbol{D}\boldsymbol{X}) = 1 & \\ & & \mathrm{Tr}(\mathbf{d}\mathbf{d}^\mathsf{T}\boldsymbol{X}) = 0 & \\ & & \mathrm{Diag}(\boldsymbol{X}) + \mathbf{s} = \frac{1-\mu}{\mu}\frac{\mathbf{1}}{\mathrm{Vol}(G)} & \\ & & \mathbf{0} \le \mathbf{s} \le \frac{1-2\mu}{\mu(1-\mu)}\frac{\mathbf{1}}{\mathrm{Vol}(G)} & \\ & & \boldsymbol{X} \succeq 0. & \end{array}$$

The Lagrangian dual is

$$\begin{array}{rlll} \lambda_\mu^{\mathrm{sdd}} = & \underset{\lambda,\beta,\boldsymbol{\gamma},\mathbf{g},\boldsymbol{\ell},\boldsymbol{Z}}{\text{maximize}} & \lambda + \frac{1-\mu}{\mu\mathrm{Vol}(G)}\boldsymbol{\gamma}^\mathsf{T}\mathbf{1} - \frac{1-2\mu}{\mu(1-\mu)\mathrm{Vol}(G)}\boldsymbol{\ell}^\mathsf{T}\mathbf{1} & (\mu\text{-conductance SDD}) \\ & \text{subject to} & \boldsymbol{L} - \lambda\boldsymbol{D} - \beta\mathbf{d}\mathbf{d}^\mathsf{T} - \mathrm{Diag}(\boldsymbol{\gamma}) - \boldsymbol{Z} = \mathbf{0} & \\ & & \boldsymbol{\ell} - \mathbf{g} - \boldsymbol{\gamma} = \mathbf{0} & \\ & & \boldsymbol{\ell} \ge \mathbf{0} & \\ & & \mathbf{g} \ge \mathbf{0} & \\ & & \boldsymbol{Z} \succeq 0. & \end{array}$$

They have the following relation.

Lemma A.2. *Strong duality holds between* (μ-conductance SDP) *and* (μ-conductance SDD)*, in other words,* $\lambda_\mu^{sdp} = \lambda_\mu^{sdd}$*, and the optimum of* ($\mu$-conductance SDP) *is achieved.*

Proof. This is a standard SDP duality claim (for example see Vandenberghe & Boyd (1996)) implied by the fact that (μ-conductance SDD) has a strictly feasible solution $\lambda = -1, \beta = -1, \boldsymbol{\gamma} = \mathbf{1}, \boldsymbol{\ell} = 2\mathbf{1}, \mathbf{g} = \mathbf{1}$. □

Observe that the objective and all constraints of (μ-conductance SDP) are affine with regard to variables $\boldsymbol{X}$ and $\mathbf{s}$, so the KKT conditions are *sufficient* for optimality (see Section 5.5.3 of Boyd et al. (2004) for example).

Lemma A.3. *The following KKT conditions are sufficient for a primal-dual pair* $\boldsymbol{X}^*, \mathbf{s}^*$ *and* λ^*, β^*, $\boldsymbol{\gamma}^*$, $\mathbf{g}^*$, $\boldsymbol{\ell}^*$, $\boldsymbol{Z}^*$ *to be an optimal solution. The primal feasibility conditions are*

$$\begin{aligned} &\mathrm{Tr}(\boldsymbol{D}\boldsymbol{X}^*) = 1 \\ &\mathrm{Tr}(\mathbf{d}\mathbf{d}^\mathsf{T}\boldsymbol{X}^*) = 0 \\ &\mathrm{Diag}(\boldsymbol{X}^*) + \mathbf{s}^* = \tfrac{1-\mu}{\mu Vol(G)}\mathbf{1} \\ &\mathbf{0} \le \mathbf{s}^* \le \tfrac{1-2\mu}{\mu(1-\mu)Vol(G)}\mathbf{1} \\ &\boldsymbol{X}^* \succeq 0, \end{aligned} \tag{8}$$

and dual feasibility conditions are

$$\begin{aligned} &\boldsymbol{L} - \lambda^*\boldsymbol{D} - \beta^*\mathbf{d}\mathbf{d}^\mathsf{T} - \mathrm{Diag}(\boldsymbol{\gamma}^*) - \boldsymbol{Z}^* = \mathbf{0} \\ &\boldsymbol{\ell}^* - \mathbf{g}^* - \boldsymbol{\gamma}^* = \mathbf{0} \\ &\boldsymbol{\ell}^* \ge \mathbf{0} \\ &\mathbf{g}^* \ge \mathbf{0} \\ &\boldsymbol{Z}^* \succeq 0, \end{aligned} \tag{9}$$

and the complementary slackness conditions are

$$\begin{aligned} &\mathbf{g}^{*\mathsf{T}}\mathbf{s}^* = 0 \\ &\boldsymbol{\ell}^{*\mathsf{T}}(\tfrac{1-2\mu}{\mu(1-\mu)Vol(G)}\mathbf{1} - \mathbf{s}^*) = 0 \\ &\mathrm{Tr}(\boldsymbol{X}^*\boldsymbol{Z}^*) = 0. \end{aligned} \tag{10}$$

Note that the stationarity conditions of program (μ-conductance SDP) is a subset of the dual feasibility conditions, so we do not list them out.

Recall the low-rank SDP we propose is as follows

$$\begin{aligned} \lambda_\mu^{\mathrm{lrsdp}} = \quad \underset{\boldsymbol{Y}\in\mathbb{R}^{n\times k},\mathbf{s}}{\text{minimize}} \quad & \mathrm{Tr}(\boldsymbol{Y}^\mathsf{T}\boldsymbol{L}\boldsymbol{Y}) \\ \text{subject to} \quad & \mathrm{Tr}(\boldsymbol{Y}^\mathsf{T}\boldsymbol{D}\boldsymbol{Y}) = 1 \\ & \mathrm{Tr}(\mathbf{d}\mathbf{d}^\mathsf{T}\boldsymbol{Y}\boldsymbol{Y}^\mathsf{T}) = 0 \\ & \mathrm{Diag}(\boldsymbol{Y}\boldsymbol{Y}^\mathsf{T}) + \mathbf{s} = \tfrac{1-\mu}{\mu\mathrm{Vol}(G)}\mathbf{1} \\ & \mathbf{0} \le \mathbf{s} \le \tfrac{1-2\mu}{\mu(1-\mu)\mathrm{Vol}(G)}\mathbf{1}. \end{aligned} \qquad (\mu\text{-conductance LRSDP})$$

Basically we just factorize $\boldsymbol{X}$ into $\boldsymbol{Y}\boldsymbol{Y}^\mathsf{T}$. So it is intuitive it has a strong connection with (μ-conductance SDP). In fact, it turns out, for a primal-dual pair $\boldsymbol{Y}^*, \mathbf{s}^*$ and $\lambda^*, \beta^*, \boldsymbol{\gamma}^*, \mathbf{g}^*, \boldsymbol{\ell}^*$ satisfying the KKT conditions of (μ-conductance LRSDP), let

$$\begin{aligned} \boldsymbol{X}^* &= \boldsymbol{Y}^*\boldsymbol{Y}^{*\mathsf{T}} \\ \boldsymbol{Z}^* &= \boldsymbol{L} - \lambda^*\boldsymbol{D} - \beta^*\mathbf{d}\mathbf{d}^\mathsf{T} - \mathrm{Diag}(\boldsymbol{\gamma}^*) \end{aligned}$$

then $\boldsymbol{X}^*, \mathbf{s}^*$ and $\lambda^*, \beta^*, \boldsymbol{\gamma}^*, \mathbf{g}^*, \boldsymbol{\ell}^*, \boldsymbol{Z}^*$ are a primal-dual pair which almost satisfies all KKT conditions of (μ-conductance SDD), except

$$\boldsymbol{Z}^* \succeq 0.$$

It's easy to verify the claim above because we have the following fact.

Lemma A.4. *For a primal-dual pair $\boldsymbol{Y}^*, \mathbf{s}^*$ and λ^*, β^*, $\boldsymbol{\gamma}^*$, $\mathbf{g}^*$, $\boldsymbol{\ell}^*$ to satisfy all KKT conditions of (μ-conductance LRSDP), they need to satisfy the stationarity conditions*

$$\begin{aligned} &(\boldsymbol{L} - \lambda^*\boldsymbol{D} - \beta^*\mathbf{d}\mathbf{d}^\mathsf{T} - \mathrm{Diag}(\boldsymbol{\gamma}^*))\boldsymbol{Y}^* = \mathbf{0} \\ &\boldsymbol{\ell}^* - \boldsymbol{\gamma}^* - \mathbf{g}^* = \mathbf{0}, \end{aligned} \tag{11}$$

and primal feasibility conditions

$$\begin{aligned} &\mathrm{Tr}(\boldsymbol{Y}^{*\mathsf{T}}\boldsymbol{D}\boldsymbol{Y}^*) = 1 \\ &\mathrm{Tr}(\mathbf{d}\mathbf{d}^\mathsf{T}\boldsymbol{Y}^*\boldsymbol{Y}^{*\mathsf{T}}) = 0 \\ &\mathrm{Diag}(\boldsymbol{Y}^*\boldsymbol{Y}^{*\mathsf{T}}) + \mathbf{s}^* = \tfrac{1-\mu}{\mu Vol(G)}\mathbf{1} \\ &\mathbf{0} \le \mathbf{s}^* \le \tfrac{1-2\mu}{\mu(1-\mu)Vol(G)}\mathbf{1}, \end{aligned} \tag{12}$$

and dual feasibility conditions

$$\begin{aligned} &\boldsymbol{\ell}^* \ge \mathbf{0} \\ &\mathbf{g}^* \ge \mathbf{0}, \end{aligned} \tag{13}$$

and the complementary slackness conditions

$$\begin{aligned} &\mathbf{g}^{*\mathsf{T}}\mathbf{s}^* = 0 \\ &\boldsymbol{\ell}^{*\mathsf{T}}(\tfrac{1-2\mu}{\mu(1-\mu)Vol(G)}\mathbf{1} - \mathbf{s}^*) = 0. \end{aligned} \tag{14}$$

Therefore if $\boldsymbol{Z}^* \succeq 0$ is violated, the objective of (μ-conductance LRSDP) at KKT points may deviate from $\lambda_\mu^{\mathrm{sdp}}$.

However, we observe that we can bound this deviation by the violation extent of $\boldsymbol{Z}^* \succeq 0$.

Denote

$$\theta = -\min\{0, \lambda_{\min}(\boldsymbol{L} - \lambda^*\boldsymbol{D} - \beta^*\mathbf{d}\mathbf{d}^\mathsf{T} - \mathrm{Diag}(\boldsymbol{\gamma}^*))\}.$$

If $\theta = 0$, then all KKT conditions of (μ-conductance SDP) are satisfied, which means $\boldsymbol{Y}^*, \mathbf{s}^*$ achieves global optimality.

If $\theta > 0$, we consider the following perturbed variant for (μ-conductance SDP).

$$\begin{array}{rll} \widehat{\lambda}_\mu^{\text{sdp}} = & \text{minimize} & \operatorname{Tr}((\boldsymbol{L}+\theta\boldsymbol{I})\boldsymbol{X}) \\ & \text{subject to} & \operatorname{Tr}(\boldsymbol{D}\boldsymbol{X}) = 1 \\ & & \operatorname{Tr}(\mathbf{d}\mathbf{d}^\mathsf{T}\boldsymbol{X}) = 0 \\ & & \operatorname{Diag}(\boldsymbol{X}) + \mathbf{s} = \frac{1-\mu}{\mu}\frac{\mathbf{1}}{\operatorname{Vol}(G)} \\ & & \mathbf{0} \le \mathbf{s} \le \frac{1-2\mu}{\mu(1-\mu)}\frac{\mathbf{1}}{\operatorname{Vol}(G)} \\ & & \boldsymbol{X} \succeq 0. \end{array} \qquad \text{(Perturbed } \mu\text{-conductance SDP)}$$

Basically we add θ into objective and keep feasible region unchanged.

Denote its dual optimum by $\widehat{\lambda}_\mu^{\text{sdd}}$, we can similarly show that strong duality holds, in other words $\widehat{\lambda}_\mu^{\text{sdp}} = \widehat{\lambda}_\mu^{\text{sdd}}$ and $\widehat{\lambda}_\mu^{\text{sdp}}$ is achieved.

Now for a primal-dual pair $\boldsymbol{Y}^*, \mathbf{s}^*$ and $\lambda^*, \beta^*, \boldsymbol{\gamma}^*, \mathbf{g}^*, \boldsymbol{\ell}^*$ satisfying all KKT conditions of (μ-conductance LRSDP), let

$$\begin{aligned} \boldsymbol{X}^* &= \boldsymbol{Y}^*\boldsymbol{Y}^{*\mathsf{T}} \\ \boldsymbol{Z}^* &= \boldsymbol{L} + \theta\boldsymbol{I} - \lambda^*\boldsymbol{D} - \beta^*\mathbf{d}\mathbf{d}^\mathsf{T} - \operatorname{Diag}(\boldsymbol{\gamma}^*), \end{aligned}$$

then the variables $\boldsymbol{X}^*, \mathbf{s}^*$ and multipliers $\lambda^*, \beta^*, \boldsymbol{\gamma}^*, \mathbf{g}^*, \boldsymbol{\ell}^*, \boldsymbol{Z}^*$ satisfy all the KKT conditions of (Perturbed μ-conductance SDP) but the following complementary slackness condition is violated

$$\operatorname{Tr}(\boldsymbol{Z}^*\boldsymbol{X}^*) = 0,$$

instead we have

$$\operatorname{Tr}(\boldsymbol{Z}^*\boldsymbol{X}^*) = \theta\operatorname{Tr}(\boldsymbol{X}^*).$$

Since all other conditions are satisfied, we know the dual value at this point is

$$\operatorname{Tr}\big((\boldsymbol{L}+\theta\boldsymbol{I})\boldsymbol{X}^*\big) - \operatorname{Tr}(\boldsymbol{Z}^*\boldsymbol{X}^*) = \operatorname{Tr}(\boldsymbol{L}^*\boldsymbol{X}^*).$$

Thus we know

$$\operatorname{Tr}(\boldsymbol{L}^*\boldsymbol{X}^*) \le \widehat{\lambda}_\mu^{\text{sdd}} = \widehat{\lambda}_\mu^{\text{sdp}},$$

which means the objective value at a KKT point of (μ-conductance LRSDP) is actually upper bounded by the optimum of the perturbed SDP (Perturbed μ-conductance SDP).

Indeed, we are able to bound the gap between $\widehat{\lambda}_\mu^{\text{sdp}}$ and λ_μ^{sdp}. Assume $\boldsymbol{X}_{\text{opt}}, \mathbf{s}_{\text{opt}}$ achieves the optimum of (μ-conductance SDP). Because the feasible region of (Perturbed μ-conductance SDP) is same with that of (μ-conductance SDP), we know that

$$\widehat{\lambda}_\mu^{\text{sdp}} \le \operatorname{Tr}((\boldsymbol{L}+\theta\boldsymbol{I})\boldsymbol{X}_{\text{opt}}) = \lambda_\mu^{\text{sdp}} + \theta\cdot\operatorname{Tr}(\boldsymbol{X}_{\text{opt}}) \le \lambda_\mu^{\text{sdp}} + \theta\cdot\min\{1, \frac{(1-\mu)n}{\mu\operatorname{Vol}(G)}\},$$

where the last inequality is due to the fact that $\operatorname{Tr}(\boldsymbol{D}\boldsymbol{X}_{\text{opt}}) = 1$ and $\operatorname{Diag}(\boldsymbol{X}_{\text{opt}}) \le \frac{1-\mu}{\mu\operatorname{Vol}(G)}\mathbf{1}$.

Therefore piecing all things together, we get

$$\operatorname{Tr}(\boldsymbol{L}^*\boldsymbol{X}^*) \le \widehat{\lambda}_\mu^{\text{sdp}} \le \lambda_\mu^{\text{sdp}} + \theta\cdot\min\{1, \frac{(1-\mu)n}{\mu\operatorname{Vol}(G)}\}.$$

We remark that the gap $\theta\cdot\min\{1, \frac{(1-\mu)n}{\mu\operatorname{Vol}(G)}\}$ has the potential to be further tightened, which brings us a better posterior bound. The intuition is that assuming $\boldsymbol{X}_{\text{opt}} = \boldsymbol{X}^*$, then we can turn it into $\theta\cdot\operatorname{Tr}(\boldsymbol{X}^*)$ where $\boldsymbol{X}^*$ is what we know because it is $\boldsymbol{Y}^*\boldsymbol{Y}^{*\mathsf{T}}$ and $\boldsymbol{Y}^*$ is the solution returned by our augmented Lagrangian method. In general, whenever there is some non-trivial relation between trace of $\boldsymbol{X}^*$ and $\boldsymbol{X}_{\text{opt}}$, we can get a non-trivial tighter bound. We also note that in a further literature review, we found that Lemma 3.5 can be derived from Boumal et al. (2016, Theorem 4).

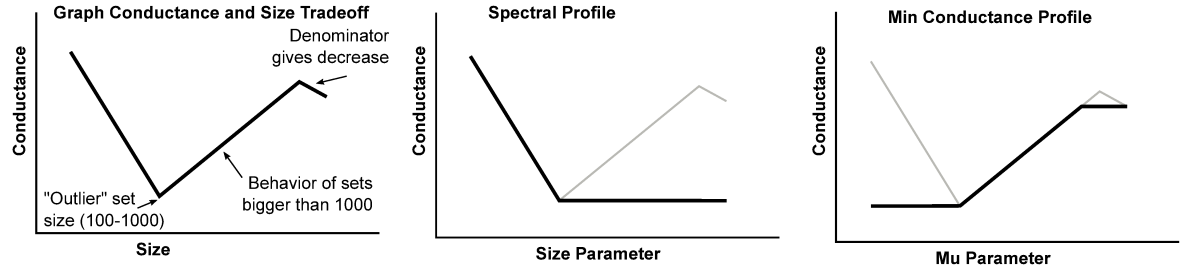


Figure 8. Three different notions of a network profile. The left figure shows a sketch of the Leskovec et al. (2009) network community profile, which measures the smallest conductance of sets with a given volume. The spectral profile (Goel et al., 2006; Raghavendra et al., 2010) is the dark line in the middle figure that measures relaxations of sets with volume up to r fraction of the maximum volume. The μ-conductance profile (Lovasz & Simonovits, 1990), the dark line in the right figure, measures the conductance of sets with at least a μ fraction of the total volume.

B. Other Related Work

B.1. Spectral Profile

The spectral profile or conductance profile (Goel et al., 2006; Raghavendra et al., 2010) is another related graph profile, with a key difference and distinction. These conductance profiles study the behavior of sets with size *up* to a given fraction of the volume. In our notation,

$$\phi^r_{\max}(G) = \begin{array}{ll} \underset{S \subset V}{\text{minimize}} & \phi(S) \\ \text{subject to} & \text{Vol}(S) \leq r\text{Vol}(G). \end{array} \tag{15}$$

Formally, they study spectral profiles, which are related to eigenvalues, but these are within a factor of 2 of the conductance values. We illustrate the differences between the profile we are interested in, these spectral profiles, and the network community profile (Leskovec et al., 2009; Jeub et al., 2015) in Figure 8. A network community profile just measures the minimum conductance of sets with a given size (technically we use the volume measure throughout this manuscript), which is *swept* over all possible sizes. (Typically, the given size is taken to be an approximation to make the curve look more smooth.) Here, we have shown a characteristic network community profile as described in Leskovec et al. (2009) (and illustrated in Figure 1) and annotated it with the features that give rise to the characteristic shape.

B.2. Balanced Cut

Balanced cut is another common problem that seeks to find a set, or group of sets, that are balanced with respect to the size of the graph. It has traditionally been important in parallel computing where balance implies equally distributed workloads. This is similar to μ-conductance with the size of the vertex set instead of volume. Although this is related to the NCP, the techniques for balanced cut tend to focus on good approximation algorithms. Since many of these techniques give approximation algorithms with unknown or hidden constants, they cannot directly translate into lower bounds. It is likely that a suitable adaptation of our techniques might also give lower bounds for balanced cuts as well.

C. Additional Information

C.1. Synthetic graph construction

The synthetic graphs we use are designed to have a dense core with a geometric like periphery. To do this, we first randomly picking coordinates of n points according to normal distribution in each dimension. Then we scale the coordinates of 90% of them by 1.5 and scale the coordinates of the left 10% by 0.1. This forms a small dense piece at the center. In the end we link each point to its five geometrically closest neighbors.