Cheeger Inequalities for Vertex Expansion and Reweighted Eigenvalues

Tsz Chiu Kwok[*],   Lap Chi Lau[†],   Kam Chuen Tung[‡]

Abstract

The classical Cheeger's inequality relates the edge conductance ϕ of a graph and the second smallest eigenvalue λ_2 of the Laplacian matrix. Recently, Olesker-Taylor and Zanetti discovered a Cheeger-type inequality $\psi^2/\log|V| \lesssim \lambda_2^* \lesssim \psi$ connecting the vertex expansion ψ of a graph $G = (V, E)$ and the maximum reweighted second smallest eigenvalue λ_2^* of the Laplacian matrix.

In this work, we first improve their result to $\psi^2/\log d \lesssim \lambda_2^* \lesssim \psi$ where d is the maximum degree in G, which is optimal up to a constant factor. Also, the improved result holds for weighted vertex expansion, answering an open question by Olesker-Taylor and Zanetti.

Building on this connection, we then develop a new spectral theory for vertex expansion. We discover that several interesting generalizations of Cheeger inequalities relating edge conductances and eigenvalues have a close analog in relating vertex expansions and reweighted eigenvalues. These include:

- An analog of Trevisan's result that relates the bipartite vertex expansion ψ_B of a graph and the maximum reweighted lower spectral gap ζ^* of the adjacency matrix. This implies the first approximation algorithm for bipartite vertex expansion.

- An analog of higher-order Cheeger's inequalities that relates the k-way vertex expansion ψ_k of a graph and the maximum reweighted k-th smallest eigenvalue λ_k^* of the Laplacian matrix. This implies the first approximation algorithm for k-way vertex expansion.

- An analog of improved Cheeger's inequality that relates the vertex expansion ψ and the reweighted eigenvalues λ_2^* and λ_k^*. This provides an improved bound for ψ using λ_2^*, when the k-way vertex expansion ψ_k is large for a small k.

Finally, inspired by this connection, we present negative evidence to the 0/1-polytope edge expansion conjecture by Mihail and Vazirani. We construct 0/1-polytopes whose graphs have very poor vertex expansion. This implies that the fastest mixing time to the uniform distribution on the vertices of these 0/1-polytopes is almost linear in the graph size. This does not provide a counterexample to the conjecture, but this is in contrast with known positive results which proved poly-logarithmic mixing time to the uniform distribution on the vertices of subclasses of 0/1-polytopes.

1 Introduction

The connection between vertex expansion and reweighted eigenvalue is discovered through the study of the fastest mixing time problem introduced by Boyd, Diaconis and Xiao [BDX04]. In the fastest mixing time problem, we are given an undirected graph $G = (V, E)$ and a target probability distribution $\pi : V \to \mathbb{R}$. The task is to find a time-reversible transition matrix $P \in \mathbb{R}^{|V| \times |V|}$ supported on the edges of G, so that the stationary distribution of random walks with transition matrix P is π. The objective is to find such a transition matrix that minimizes the mixing time to the stationary distribution π. It is well-known that the mixing time to the stationary distribution is approximately inversely proportional to the spectral gap $1 - \alpha_2(P)$ of the time-reversible transition matrix P, where $1 = \alpha_1(P) \geq \alpha_2(P) \geq \cdots \geq \alpha_{|V|}(P) \geq -1$ are the eigenvalues of P. The fastest mixing time problem is thus formulated as follows in [BDX04] by the maximum spectral gap achievable through such a "reweighting" P of the input graph G.

Definition 1.1 (Maximum Reweighted Spectral Gap [BDX04]). *Given an undirected graph $G = (V, E)$ and a probability distribution π on V, the maximum reweighted spectral gap is defined as*

$$
\begin{aligned}
\lambda_2^*(G) \ := \ \max_{P \geq 0} \quad & 1 - \alpha_2(P) \\
\text{subject to} \quad & P(u, v) = P(v, u) = 0 && \forall uv \notin E \\
& \sum_{v \in V} P(u, v) = 1 && \forall u \in V \\
& \pi(u)P(u, v) = \pi(v)P(v, u) && \forall uv \in E.
\end{aligned}
$$

The graph is assumed to have a self-loop on each vertex, to ensure that the optimization problem for $\lambda_2^*(G)$ is always feasible. In the context of Markov chains, this corresponds to allowing a non-negative holding probability on each vertex.

The last constraint is the time reversible condition to ensure that the transition matrix P corresponds to random walks on an undirected graph (where the edge weight of uv is $\pi(u)P(u, v)$) and that the stationary distribution of P is π. Note that $\lambda_2^*(G) = \max_{P \geq 0}(1 - \alpha_2(P)) = \max_{P \geq 0} \lambda_2(I - P)$, which is the maximum reweighted second smallest eigenvalue of the normalized Laplacian matrix of G (where the edge weight of uv is $\pi(u)P(u, v)$) subject to the above constraints.

Boyd, Diaconis and Xiao showed that this optimization problem can be written as a semidefinite program and thus $\lambda_2^*(G)$ can be computed in polynomial time. Subsequently, the fastest mixing time problem has been studied in various work (see [Roc05, BDSX06, BDPX09, FK13, CA15] and more references in [OZ22]), but no general characterization was known. Roch [Roc05] showed that the vertex expansion $\psi(G)$ is an upper bound on the optimal spectral gap $\lambda_2^*(G)$.

Definition 1.2 (Weighted Vertex Expansion). *Let $G = (V, E)$ be an undirected graph and π be a probability distribution on V. For a subset $S \subseteq V$, let $\partial S := \{v \notin S \mid \exists u \in S \text{ with } uv \in E\}$ be the vertex boundary of S, and $\pi(S) := \sum_{v \in S} \pi(v)$ be the weight of S. The weighted vertex expansion*

of a set $S \subseteq V$ and of a graph G are defined as

$$\psi(S) := \frac{\pi(\partial S)}{\pi(S)} \quad and \quad \psi(G) := \min\left\{1, \min_{S \subseteq V: 0 < \pi(S) \le 1/2} \psi(S)\right\}.^1$$

When π is the uniform distribution, $\psi(S)$ is the usual vertex expansion $|\partial S|/|S|$.

Recently, Olesker-Taylor and Zanetti [OZ22] discovered an elegant Cheeger-type inequality for vertex expansion and the maximum reweighted spectral gap, showing that small vertex expansion is qualitatively the only obstruction for the fastest mixing time to be small. Note that their result only holds when π is the uniform distribution.

Theorem 1.3 (Cheeger Inequality for Vertex Expansion [OZ22])**.** *For any undirected graph $G = (V, E)$ and the uniform distribution $\pi = \vec{1}/|V|$,*

$$\frac{\psi(G)^2}{\log |V|} \lesssim \lambda_2^*(G) \lesssim \psi(G).$$

In terms of the fastest mixing time $\tau^*(G)$ to the uniform distribution, $\frac{1}{\psi(G)} \lesssim \tau^*(G) \lesssim \frac{\log^2 |V|}{\psi^2(G)}$. (See Section 2 for definitions for random walks and mixing time.)

Unlike Cheeger's inequality for edge conductance where $\phi(G)^2 \lesssim \lambda_2(G) \lesssim \phi(G)$, it is noted in [OZ22] that the $\log |V|$ term might not be completely removed: Louis, Raghavendra and Vempala [LRV13] proved that it is NP-hard to distinguish between $\psi(G) \le \epsilon$ and $\psi(G) \gtrsim \sqrt{\epsilon \log d}$ for every $\epsilon > 0$ where d is the maximum degree of the graph G, assuming the small-set expansion conjecture of Raghavendra and Steurer [RS10].

Besides the fastest mixing time problem, we note that these "reweighting problems" relating vertex expansion and reweighted eigenvalues are also well motivated in the study of approximation algorithms. One example is a conjecture of Arora and Ge [AG11, Conjecture 12], which roughly states that, if a graph G has almost perfect vertex expansion for every set, then there exists a reweighted doubly stochastic matrix P of the adjacency matrix of G so that P has few eigenvalues less than $-\frac{1}{16}$. They proved that if the conjecture was true, then there is an improved subexponential time algorithm for coloring 3-colorable graphs. Another example is a conjecture of Steurer [Ste10, Conjecture 9.2], which is also known to be related to a reweighting problem between vertex expansion and the graph spectrum, that if true would imply an improved subexponential time approximation algorithm for the sparsest cut problem.

1.1 Our Results

First we improve and generalize the result of Olesker-Taylor and Zanetti. Then we build on this new connection to develop a spectral theory for vertex expansion. Finally we present 0/1-polytopes with poor vertex expansion and discuss the implications to the 0/1-polytope expansion conjecture.

---

[1]When π is the uniform distribution, $\min_{S \subseteq V: 0 < \pi(S) \le 1/2} \psi(S)$ is always at most 1. For general π, however, this could be arbitrarily large. Consider for example a star graph where the center has most of the π-weight. Therefore, we need to put an upper bound of 1 on $\psi(G)$, as otherwise $\psi(G)$ cannot be bounded by eigenvalues of the normalized Laplacian matrix which are always upper bounded by 2.

1.1.1 Optimal Cheeger Inequality for Vertex Expansion

Olesker-Taylor and Zanetti [OZ22] posed the problem of reducing the $\log|V|$ factor in Theorem 1.3 to $\log d$, and also the problem of generalizing their result to weighted vertex expansion. Our first result provides a positive answer to these two questions.

Theorem 1.4 (Cheeger Inequality for Weighted Vertex Expansion). *For any undirected graph* $G = (V, E)$ *with maximum degree* d *and any probability distribution* π *on* V,

$$\frac{\psi(G)^2}{\log d} \lesssim \lambda_2^*(G) \lesssim \psi(G).$$

In terms of the fastest mixing time $\tau^*(G)$ *to the stationary distribution,* $\frac{1}{\psi(G)} \lesssim \tau^*(G) \lesssim \frac{\log d \cdot \log \pi_{\min}^{-1}}{\psi^2(G)}$.

In Section 8, we show that the $\log d$ factor in Theorem 1.4 is optimal, by exhibiting graphs G with $\lambda_2^*(G) \asymp \frac{\psi(G)^2}{\log d}$. Note that the tightness result does not rely on the small-set expansion hypothesis.

We note that Louis, Raghavendra and Vempala [LRV13] gave an SDP approximation algorithm for vertex expansion with the same approximation guarantee, but their SDP is different from and stronger than that in Definition 1.1 (see Lemma 3.10), and so it does not have the natural interpretation as the reweighted second eigenvalue and does not imply the result on fastest mixing time. The proof of Theorem 1.4 is based on the techniques in [LRV13, BHT00], which we will discuss in detail in Section 3.1.2.

1.1.2 Maximum Reweighted Lower Spectral Gap and Bipartite Vertex Expansion

Trevisan [Tre09] proved that the lower spectral gap $1 + \alpha_{\min}(G)$ of the normalized adjacency matrix of $G = (V, E)$ is small if and only if there is a subset $S \subseteq V$ which is an almost bipartite component in G with small edge conductance $\phi(S)$. We define the analogous notions for vertex expansion and for reweighted lower spectral gap.

Definition 1.5 (Bipartite Vertex Expansion). *Given an undirected graph* $G = (V, E)$, *the bipartite vertex expansion of* G *is defined as*

$$\psi_B(G) := \min\left\{1, \min_{\emptyset \neq S \subseteq V} \left\{\psi(S) \mid G[S] \text{ is an induced bipartite graph}\right\}\right\}.$$

Definition 1.6 (Maximum Reweighted Lower Spectral Gap). *Given an undirected graph* $G = (V, E)$ *and a probability distribution* π *on* V, *the maximum reweighted lower spectral gap is defined as*

$$
\begin{aligned}
\zeta^*(G) \ :=\ &\max_{P \geq 0} \quad \lambda_{\min}(D_P + P) \\
&\text{subject to} \quad P(u, v) = P(v, u) = 0 && \forall uv \notin E \\
&\qquad\qquad\quad \sum_{v \in V} P(u, v) \leq 1 && \forall u \in V \\
&\qquad\qquad\quad \pi(u)P(u, v) = \pi(v)P(v, u) && \forall uv \in E.
\end{aligned}
$$

3

where D_P is the diagonal matrix of row sums of P such that $D_P(u,u) = \sum_{v \in V} P(u,v)$ for $u \in V$. We note that this program is slightly different from that in *Definition 1.1*, and the main reason is that self-loops should not be allowed in this problem. We will explain more about this in *Section 4*.

We prove an analog of Trevisan's result that the maximum reweighted lower spectral gap is small if and only if there is an induced bipartite subgraph on S with small vertex expansion $\psi(S)$.

Theorem 1.7 (Cheeger Inequality for Bipartite Vertex Expansion). *For any undirected graph $G = (V, E)$ with maximum degree d and any probability distribution π on V,*

$$\frac{\psi_B(G)^2}{\log d} \lesssim \zeta^*(G) \lesssim \psi_B(G).$$

This is the first approximation algorithm for bipartite vertex expansion to our knowledge. Finding a two-colorable set with small vertex expansion is one of the three ways in Blum's coloring tools [Blu94] to make progress in designing approximation algorithms for coloring 3-colorable graphs. Indeed, it is in this context that Arora and Ge [AG11] made the reweighting conjecture mentioned in the introduction. Theorem 1.7 does not imply anything new about approximating graph coloring, but we hope that it is a step towards answering Arora and Ge's conjecture.

1.1.3 Higher-Order Cheeger Inequality for Vertex Expansion

Lee, Oveis Gharan and Trevisan [LOT12] and Louis, Raghavendra, Tetali and Vempala [LRTV12] proved the higher-order Cheeger inequalities, which state that the k-th smallest eigenvalue $\lambda_k(G)$ of the normalized Laplacian matrix of $G = (V, E)$ is small if and only if the k-way edge conductance $\phi_k(G)$ is small. More precisely, they proved that $\lambda_k(G) \lesssim \phi_k(G) \lesssim k^2 \sqrt{\lambda_k}$ and $\lambda_{\frac{k}{2}}(G) \lesssim \sqrt{\lambda_k \log k}$. We consider the analogous notion of k-way vertex expansion.

Definition 1.8 (k-Way Vertex Expansion). *Given an undirected graph $G = (V, E)$ and a probability distribution π on V, the k-way vertex expansion of G is defined as*

$$\psi_k(G) := \min \left\{ 1, \min_{S_1, \ldots, S_k \subseteq V} \max_{1 \leq i \leq k} \psi(S_i) \right\},$$

where the minimum is taken over pairwise disjoint subsets $S_1, \ldots, S_k$ of V.

Definition 1.9 (Maximum Reweighted k-th Smallest Eigenvalue). *Given an undirected graph $G = (V, E)$ and a probability distribution π on V, the maximum reweighted k-th smallest eigenvalue of the normalized Laplacian matrix of G is defined as $\lambda_k^*(G) := \max_{P \geq 0} \lambda_k(I - P)$, where P is subject to the same constraints stated in *Definition 1.1*.*

We prove an analog of higher-order Cheeger inequalities that the maximum reweighted k-th smallest eigenvalue is small if and only if the k-way vertex expansion is small. As in previous work [LOT12, LRTV12], there is a better approximation guarantee if we consider only $\frac{k}{2}$-way vertex expansion.

Theorem 1.10 (Higher-Order Cheeger Inequality for Vertex Expansion). *For any undirected graph* $G = (V, E)$ *with maximum degree* d *and any probability distribution* π *on* V,

$$\lambda_k^*(G) \lesssim \psi_k(G) \lesssim k^{\frac{9}{2}} \log k \sqrt{\log d \cdot \lambda_k^*(G)} \quad \text{and} \quad \psi_{\frac{k}{2}}(G) \lesssim \sqrt{k} \log k \sqrt{\log d \cdot \lambda_k^*(G)}.$$

Chan, Louis, Tang and Zhang [CLTZ18] developed a spectral theory for hypergraphs and proved a higher-order Cheeger inequality for hypergraph (edge) expansion. Through a reduction from vertex expansion to hypergraph expansion, they proved that $\psi_{\frac{k}{2}}(G) \lesssim k^{\frac{5}{2}} \log k \log \log k \cdot \log d \cdot \sqrt{\xi_k}$ for graphs with bounded ratio between the maximum degree and the minimum degree, where $\xi_k \lesssim \psi_k(G)$ is a relaxation for k-way vertex expansion. Compared to their result, Theorem 1.10 does not require the assumption about the maximum degree and the minimum degree of G, and has a better approximation ratio for $\frac{k}{2}$-way vertex expansion. Furthermore, Theorem 1.10 provides the first true approximation algorithm for k-way vertex expansion $\psi_k(G)$ to our knowledge.

1.1.4 Improved Cheeger Inequality for Vertex Expansion

Kwok, Lau, Lee, Oveis Gharan, and Trevisan [KLLOT13] proved an improved Cheeger inequality that $\phi(G) \lesssim k\lambda_2(G)/\sqrt{\lambda_k(G)}$ for any $k \geq 2$. This shows that $\lambda_2(G)$ is a tighter approximation to $\phi(G)$ when $\lambda_k(G)$ is large for a small k. The result provides an explanation for the good empirical performance of the spectral partitioning algorithm.

We prove an analogous result that if the $\lambda_k^*(G)$ is large for a small k, then $\lambda_2^*(G)$ is a tighter approximation to the vertex expansion $\psi(G)$. The following result is close to the tight result in [KLLOT13] for edge conductance as we will elaborate in Remark 6.5.

Theorem 1.11 (Improved Cheeger Inequality for Vertex Expansion). *For any undirected graph* $G = (V, E)$ *with maximum degree* d, *and for any probability distribution* π *on* V *and any* $k \geq 2$,

$$\lambda_2^*(G) \lesssim \psi(G) \lesssim \frac{k^{\frac{3}{2}} \cdot \lambda_2^*(G) \cdot \log d}{\sqrt{\lambda_k^*(G)}}.$$

We remark that the reweighting used in $\lambda_2^*(G)$ and $\lambda_k^*(G)$ may be different. Through Theorem 1.10, we obtain the following corollary that only depends on the graph structure: If the k-way vertex expansion $\psi_k(G)$ is large for a small k, then $\lambda_2^*(G)$ is a tighter approximation to $\psi(G)$.

1.1.5 Vertex Expansion of 0/1-Polytopes

Mihail and Vazirani (see [FM92]) conjectured that the graph $G = (V, E)$ (i.e. 1-skeleton) of any 0/1-polytope is an edge expander, such that $|\delta(S)|/|S| \geq 1$ for every subset $S \subseteq V$ with $|S| \leq |V|/2$, where $\delta(S)$ denotes the set of edges between S and $V \setminus S$. This conjecture would imply fast mixing time of random walks to the stationary distribution, with applications in designing fast sampling algorithms for many classes of combinatorial objects. The conjecture is proved to be correct in several cases [FM92, Kai04, ALOV19], most notably the recent resolution of the matroid expansion conjecture [ALOV19] by Anari, Liu, Oveis Gharan and Vinzant.

In all these positive results, the Markov chain can be set up so that the stationary distribution is the uniform distribution, with the mixing time to the stationary distribution poly-logarithmic in the graph size. Then the fast sampling algorithms can also be used to obtain an approximate counting algorithm on the number of vertices in the given 0/1-polytope, with poly-logarithmic runtime in the graph size. Therefore, sampling from the uniform distribution is usually the setting of interest.

Inspired by the connection between fastest mixing time and vertex expansion, we consider a variant of Mihail and Vazirani's conjecture: Is the graph of every 0/1-polytope a vertex expander? Perhaps surprisingly, we show that there are 0/1-polytopes whose graphs are very poor vertex expanders.

Theorem 1.12 (0/1-Polytopes with Poor Vertex Expansion). *Let π be the uniform distribution. For any $k > 2$ and any $n > 2k$ sufficiently large, there is a 0/1-polytope $Q = Q_{n,k} \subseteq \{0,1\}^n$ with $O(n^k)$ vertices and*

$$\psi(Q) \lesssim \frac{(4k)^k}{n^{k-2}}.$$

Theorem 1.12 and Theorem 1.3 together imply that even the fastest mixing time of the reversible random walks on some 0/1-polytopes is almost linear in the graph size.

Corollary 1.13 (Torpid Mixing to Uniform Distribution). *For any constant $k > 2$, there exists a 0/1-polytope Q such that any reversible Markov chain on its graph $G_Q = (V, E)$ with stationary distribution $\vec{1}/|V|$ has mixing time $\Omega(|V|^{1-\frac{2}{k}})$.*

While Theorem 1.12 does not provide a counterexample to the conjecture of Mihail and Vazirani, it shows that even if the conjecture is true, there are 0/1-polytopes for which random walks cannot be used for efficient uniform sampling and for efficient approximate counting.

Remark 1.14. *After posting the first version of this paper on arXiv, we recently found out that Gillmann [Gil07, Chapter 3.2] has already constructed examples of 0/1-polytopes whose graphs have poor vertex expansion. The polytopes $Q \subseteq \{0,1\}^n$ constructed have $2^{(h(c)+o(1))n}$ vertices and satisfy*

$$\psi(Q) \lesssim 2^{-(h(c)-2c)n},$$

where $h(x) := -x \log x - (1-x) \log(1-x)$ is the binary entropy function and $c := 1/5$ (correspondingly $h(c) = 0.7219...$). Applying Theorem 1.3, this would imply a fastest mixing time bound of $\Omega(|V|^{0.4459...})$. By choosing smaller values of c, an almost linear fastest mixing time bound can be obtained as in Corollary 1.13.

1.2 Related Work

In this subsection, we review previous spectral approaches for vertex expansion and compare them to the current approach using reweighted eigenvalues. For previous results about Cheeger's inequalities for edge conductances mentioned in the introduction, they will be discussed in the corresponding technical sections.

Second Eigenvalue and Vertex Expansion: There are classical results in spectral graph theory relating vertex expansions and (ordinary) eigenvalues. For any graph $G = (V, E)$ with maximum

degree d, let $\lambda_2'(G)$ be the second smallest eigenvalue of the (unnormalized) Laplacian matrix, it is known that

$$\psi(G) \geq \frac{2\lambda_2'(G)}{d + 2\lambda_2'(G)} \quad \text{and} \quad \lambda_2'(G) \geq \frac{\psi(G)^2}{4 + 2\psi(G)^2},$$

where the first inequality is the "easy" direction proved by Tanner [Tan84] and Alon and Milman [AM85], and the second inequality is the "hard" direction proved by Alon [Alo86]. These imply that $\lambda_2'(G)$ can be used to give an $O(\sqrt{d \cdot \psi(G)})$-approximation algorithm to $\psi(G)$. Compared to Cheeger's inequality for edge conductance that $\phi(G)^2/2 \leq \lambda_2(G) \leq 2\phi(G)$ where $\lambda_2(G)$ is the second smallest eigenvalue of the normalized Laplacian matrix, there is an extra factor d between the upper and lower bounds.

Spectral Formulation: Bobkov, Houdré and Tetali [BHT00] defined an interesting "spectral" quantity called λ_∞ (see Definition 3.5), which satisfies an exact analog of Cheeger's inequality for symmetric vertex expansion:

$$\frac{1}{2}\psi_{\text{sym}}(G)^2 \leq \lambda_\infty \leq 2\psi_{\text{sym}}(G),$$

where the symmetric vertex boundary of a set $S \subseteq V$ is defined as $\partial_{\text{sym}}(S) := \partial(S) \cup \partial(V - S)$ and the symmetric vertex expansion of S is defined as $\psi_{\text{sym}}(S) := |\partial_{\text{sym}}(S)|/|S|$, and the symmetric vertex expansion of a graph G is defined as $\psi_{\text{sym}}(G) := \min_{S:|S|\leq|V|/2} \psi_{\text{sym}}(S)$. However, it is not known how to compute λ_∞ efficiently, and it is recently shown to be NP-hard to compute λ_∞ by Farhadi, Louis and Tetali [FLT20].

Semidefinite Programming Relaxations: Louis, Raghavendra and Vempala [LRV13] gave a semidefinite programming relaxation sdp_∞ for λ_∞, and proved that for any graph $G = (V, E)$ with maximum degree d,

$$\frac{\psi_{\text{sym}}(G)^2}{\log d} \lesssim \mathsf{sdp}_\infty \lesssim \psi_{\text{sym}}(G).$$

Then, by constructing a graph H such that $\psi_{\text{sym}}(H) = \Theta(\psi(G))$, they reduce vertex expansion to symmetric vertex expansion and obtain a Cheeger's inequality for $\psi(G)$, one that is of the same form as in Theorem 1.4 for $\lambda_2^*(G)$. We will show in Lemma 3.10 that $\lambda_2^*(G)$ and sdp_∞ are different and sdp_∞ is a stronger relaxation such that $\lambda_2^*(G) \leq \mathsf{sdp}_\infty$.

The current best known approximation algorithm for vertex expansion $\psi(G)$ is an $O(\sqrt{\log |V|})$ SDP-based approximation algorithm by Feige, Hajiaghayi and Lee [FHL08]. This is an extension of the $O(\sqrt{\log |V|})$ SDP-based approximation algorithm for edge conductance $\phi(G)$ by Arora, Rao, and Vazirani [ARV09]. The SDP formulation of [ARV09] is known to be strictly more powerful than the spectral formulation by the second eigenvalue.

Even though $\lambda_2^*(G)$, sdp_∞ and the SDP in [FHL08] are all semidefinite programming relaxations for $\psi(G)$ and satisfy similar inequalities, we note that the approach of using reweighted eigenvalues has some additional features. One important feature is that $\lambda_2^*(G)$ is closely related to fastest mixing time. This allows one to develop a spectral theory for vertex expansion that relates (i) vertex expansion, (ii) reweighted eigenvalues and (iii) fastest mixing time, which parallels the classical spectral graph theory that relates (i) edge conductance, (ii) eigenvalues and (iii) mixing time. Another feature is that it allows one to extend known generalizations of Cheeger inequalities to the

vertex expansion setting, and as a consequence to obtain approximation algorithms for bipartite vertex expansion and k-way vertex expansion.

Spectral Hypergraph Theory: Louis [Lou15] and Chan, Louis, Tang, Zhang [CLTZ18] developed a spectral theory for hypergraphs. They defined a continuous time diffusion process on a hypergraph $H = (V, E)$ and used it to define the Laplacian operator and its eigenvalues $\gamma_1 \leq \gamma_2 \leq \ldots \leq \gamma_{|V|}$. The formulation is similar to the one in [BHT00] for vertex expansion, and they proved that there is an exact analog of Cheeger's inequality for hypergraphs:

$$\frac{1}{2}\phi(H) \leq \gamma_2 \leq \sqrt{2\phi(H)},$$

where $\phi(H)$ is the hypergraph edge conductance of H. As in [BHT00], the quantity γ_2 is not polynomial time computable, and a semidefinite programming relaxation similar to that in [LRV13] is used to design a $O(\sqrt{\phi(G) \log r})$-approximation algorithm for hypergraph edge conductance where r is the maximum size of a hyperedge. Using this spectral theory, they prove an analog of higher-order Cheeger inequality for hypergraph edge conductance, and also an approximation algorithm for small-set hypergraph edge conductance. Through a reduction from vertex expansion to hypergraph edge conductance, they obtain an analog of higher-order Cheeger inequality for vertex expansion as mentioned earlier after Theorem 1.10 and also an approximation algorithm for small-set vertex expansion. This theory also relates (i) expansion, (ii) eigenvalues and (iii) mixing time, and so the work in [Lou15, CLTZ18] is closest to the current work.

Compared to the theory in [Lou15, CLTZ18] for hypergraphs and for vertex expansion through reduction, we note that the current approach using reweighted eigenvalues is more direct and effective for vertex expansion. The reduction in [CLTZ18, Fact 3] from vertex expansion $\psi(G)$ of graph G with maximum degree $d_{\max}$ and minimum degree $d_{\min}$ to edge conductance $\phi(H)$ only satisfies

$$d_{\min} \cdot \phi(H) \leq \psi(G) \leq d_{\max} \cdot \phi(H),$$

and so the approximation ratio depends on the ratio between the maximum degree and the minimum degree. The current approach using reweighted eigenvalues does not have this dependency and also proves stronger bounds in k-way vertex expansion as discussed after Theorem 1.10. Also, the definitions of the hypergraph diffusion process and its eigenvalues are quite technically involved and require considerable effort to make rigorous [CTWZ17]. We believe that the definitions of reweighted eigenvalues are more intuitive and more closely related to ordinary eigenvalues. Also, reweighted eigenvalues have close connections to other important problems such as fastest mixing time and the reweighting conjectures in approximation algorithms.

1.3 Techniques

From a technical perspective, the advantage of relating reweighted eigenvalues to vertex expansions is that many ideas relating eigenvalues to edge conductances can be carried over to the new setting. So, many steps in our proofs are natural extensions of previous arguments, and we focus our discussion here on the new elements.

Vertex Expansion: The proof of Theorem 1.3 by Olesker-Taylor and Zanetti is based on the dual characterization of Definition 1.1 in Proposition 3.1, due to Roch [Roc05], and it has two main steps. In the first step, they used the Johnson-Lindenstrass lemma to project the SDP solution into a $O(\log |V|)$-dimension solution, and then further reduce it to a 1-dimensional "spectral" solution by taking the best coordinate. This is the step where the $\log |V|$ factor is lost. In the second step, they introduced an interesting new concept called the "matching conductance", and used some combinatorial arguments about greedy matchings for the analysis of Cheeger rounding on Roch's dual program.

In our proof of Theorem 1.4, we also use Roch's dual characterization and follow the same two steps. In the first step, we use the Gaussian projection method in [LRV13] to reduce the SDP solution to a 1-dimensional solution directly, and adapt their analysis to show that only a factor of $\log d$ is lost. In the second step, we bypass the concept of matching conductance and do a more traditional analysis of Cheeger rounding as in Bobkov, Houdré and Tetali [BHT00]. It turns out that this analysis works smoothly for weighted vertex conductance, while the approach using matching conductance faced some difficulty as described in [OZ22]. A new element in our proof is the introduction of an intermediate dual program using graph orientation, which is important in the analysis of both steps. In Section 3, we will review the background from [OZ22, Roc05, LRV13, BHT00] and give a more detailed comparison and overview.

Bipartite Vertex Expansion: The proof of Theorem 1.7 for bipartite vertex expansion follows closely the proof of Theorem 1.4 and Trevisan's result [Tre09], once the correct formulation in Definition 1.6 is found.

Multiway Vertex Expansion: For the proof of higher-order Cheeger inequality for vertex expansion in Theorem 1.10, one technical issue is that we do not know of a convex relaxation for the maximum reweighted k-th smallest eigenvalue in Definition 1.9. Instead, we define a related quantity $\sigma_k^*(G)$ called the maximum reweighted sum of the k smallest eigenvalues in Definition 5.1, which can be written as a semidefinite program. We show in Proposition 5.2 that this quantity has a nice dual characterization that satisfies the sub-isotropy condition. This allows us to adapt the techniques in [LOT12] to decompose the SDP solution into k disjointly supported SDP solutions with small objective values, so that we can apply Theorem 1.4 to find k disjoint sets with small vertex expansion. We will review the background in [LOT12] needed for the proof in Section 5.3.

Improved Cheeger Inequality: The proof of improved Cheeger inquality for vertex expansion is similar to that in [KLLOT13], which has two main steps. The first step is to prove that if the 1-dimensional solution to Roch's dual program is close to a k-step function, then Cheeger rounding performs well. The second step is to prove that if the 1-dimensional solution to Roch's dual program is far from a k-step function, then we can construct an SDP solution to σ_k^* with small objective value, which proves that λ_k^* is small. Therefore, if λ_k^* is large, then the 1-dimensional solution must be close to a k-step function, and hence Cheeger rounding performs well. One interesting aspect in this proof is to relate the performance of a rounding algorithm of one SDP (in this case $\lambda_2^*(G)$) to the objective value of another SDP (in this case $\sigma_k^*(G)$).

Vertex Expansion of $0/1$-Polytopes: The examples in Theorem 1.12 for $0/1$-polytope is obtained by a simple probabilistic construction. The graph of a $0/1$-polytope is defined by the set of points chosen in $\{0, 1\}^n$. Let L be the set of points with k ones, and let R be the set of points

9

with $(n-k)$ ones. We prove that if we choose a random set M of points with $n/2$ ones and set $|M| \asymp 4^k n^2$, then with high probability there are no edges between L and R in the resulting polytope, and so M is a small vertex separator of L and R where each has $\binom{n}{k}$ points. The proof is by elementary geometric arguments about the edges of a polytope, and a simple result bounding the number of linear threshold functions in the boolean hypercube $\{0, 1\}^n$.

1.4 Concurrent Work

Jain, Pham, and Vuong [JPV22] independently published a proof of Theorem 1.4 for the uniform distribution case. Their approach is based on a better analysis of dimension reduction for maximum matching, which is quite different from our approach as we bypassed the concept of matching conductance in [OZ22].

2 Preliminaries

Given two functions f, g, we use $f \lesssim g$ to denote the existence of a positive constant $c > 0$, such that $f \leq c \cdot g$ always holds. We use $f \asymp g$ to denote $f \lesssim g$ and $g \lesssim f$. We use $f \approx_k g$ to denote $f/g = 1 + o_k(1)$. For positive integers k, we use $[k]$ to denote the set $\{1, 2, \ldots, k\}$. For a function $f : X \to \mathbb{R}$, $\text{supp}(f)$ denotes the domain subset on which f is nonzero. For an event E, $\mathbb{1}[E]$ denotes the indicator function that is 1 when E is true and 0 otherwise.

Graphs

Let $G = (V, E)$ be an undirected graph. Throughout this paper, we use $n := |V|$ to denote the number of vertices and $m := |E|$ to denote the number of edges in the graph. If uv is an edge in G, we either write $uv \in E$ or use the notation $u \sim v$. The degree of a vertex v, denoted by $\deg(v)$, is the number of edges incident to v. The maximum degree of a graph is defined as $\max_{v \in V} \deg(v)$. We usually associate G with a probability distribution $\pi : V \to \mathbb{R}$ on the set of vertices, and we write $\pi(S) := \sum_{v \in S} \pi(v)$ for a subset $S \subseteq V$. We assume without loss that $\pi(u) > 0$ for all $u \in V$.

Let $S \subseteq V$ be a subset of vertices. The edge boundary of S is defined as $\delta(S) := \{uv \in E \mid u \in S, v \notin S\}$. The volume of S is defined as $\text{vol}(S) := \sum_{v \in S} \deg(v)$. The edge conductance of S is defined as $\phi(S) := |\delta(S)|/|S|$. The vertex boundary of S is defined as $\partial S := \{v \in V \setminus S \mid \exists u \in S \text{ with } uv \in E\}$. The π-weighted vertex expansion of S is defined as $\psi(S) := \pi(\partial S)/\pi(S)$, and when π is the uniform distribution $\psi(S) = |\partial S|/|S|$ is the usual vertex expansion. The induced edge set of S is defined as $E[S] := \{uv \in E \mid u \in S \text{ and } v \in S\}$.

Let $G = (V, \overrightarrow{E})$ be a directed graph. If uv is a directed edge in G, we either write $uv \in \overrightarrow{E}$ or use the notation $u \to v$. The indegree of a vertex v is defined as $\deg^{\text{in}}(v) := |\{u \in V \mid u \to v\}|$. We will define some directed analogs of vertex expansion in Section 3 and Section 4. They are not standard definitions and so we defer them to the respective sections.

10

Linear Algebra

Let $M \in \mathbb{R}^{n \times n}$ be a matrix. When M is symmetric, the spectral theorem states that M admits an orthonormal eigendecomposition $M = UDU^{-1}$, where D is a diagonal matrix and U is a unitary matrix such that $U^{-1}U = I_n$ where I_n is the $n \times n$ identity matrix.

Two matrices $M, N \in \mathbb{R}^{n \times n}$ are said to be cospectral if they are both diagonalizable, and their eigenvalues are the same. There are two cases of cospectral matrices that we will use.

Fact 2.1. *Let $M, N \in \mathbb{R}^{n \times n}$. Suppose that M is diagonalizable and that M and N are similar (i.e. $M = X^{-1}NX$ for some invertible matrix $X \in \mathbb{R}^{n \times n}$). Then, N is also diagonalizable, and M and N are cospectral.*

Fact 2.2. *Let $M, N \in \mathbb{R}^{n \times n}$. Suppose that there exist $A, B \in \mathbb{R}^{n \times n}$ such that $M = AB$ and $N = BA$. If M is diagonalizable, then N is also diagonalizable, and M and N are cospectral.*

Given that M is symmetric, we say that M is positive semidefinite (PSD) if $v^T M v \geq 0$ for all $v \in \mathbb{R}^n$, and we write $M \succcurlyeq 0$. Equivalently, M is PSD if all its eigenvalues are nonnegative. Also equivalently, M is PSD if there exists X such that $M = X^T X$. Let $x_i \in \mathbb{R}^n$ be the i-th column of X. Then M is called the Gram matrix of $x_1, \ldots, x_n \in \mathbb{R}^n$ as $M(i, j) = \langle x_i, x_j \rangle$ for all $i, j \in [n]$.

The trace of a matrix $M \in \mathbb{R}^{n \times n}$ is defined as $\operatorname{tr}(M) := \sum_{i=1}^n M(i, i)$. We will often use the fact that $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ for two matrices of compatible dimensions.

Random Walks

Given a finite state space X, a Markov chain on X is represented by a matrix $P \in \mathbb{R}^{X \times X}$, where $P(u, v)$ is the probability of traversing from state u to state v in one step. Thus, P has nonnegative entries and satisfies $\sum_{v \in X} P(u, v) = 1$ for all $u \in X$. A distribution $\pi : X \to \mathbb{R}$ is said to be a stationary distribution of P if $\pi^T P = \pi^T$.

A transition matrix P is said to be time-reversible with respect to π if $\pi(u)P(u, v) = \pi(v)P(v, u)$ for any $u, v \in X$. Note that this implies that π is a stationary distribution of P. The time reversibility condition can be written as $\Pi P = P^T \Pi$, where $\Pi := \operatorname{diag}(\pi)$. Thus, $\Pi^{1/2} P \Pi^{-1/2}$ is symmetric, hence diagonalizable with eigenvalues $1 = \alpha_1(P) \geq \alpha_2(P) \geq \cdots \geq \alpha_n(P) \geq -1$. As P is similar to $\Pi^{1/2} P \Pi^{-1/2}$, they have the same eigenvalues by Fact 2.1. The spectral gap of P is defined as $1 - \alpha_2(P)$.

For $\epsilon \in (0, 1)$, we define the ϵ-mixing time $\tau_{\mathrm{mix}}(P, \epsilon)$ of P to be the smallest $t \in \mathbb{N}$ such that $d_{TV}(\pi, \rho) \leq \epsilon$ for any initial distribution ρ. Here, $d_{TV}(\cdot, \cdot)$ is the total variation distance, defined as $d_{TV}(\rho_1, \rho_2) := \max_{S \subseteq V} |\rho_1(S) - \rho_2(S)|$ for any two distributions $\rho_1, \rho_2 : X \to \mathbb{R}_{\geq 0}$. The relaxation time $\tau_{\mathrm{rel}}(P)$ of P is defined as the reciprocal of the spectral gap, so $\tau_{\mathrm{rel}}(P) := \frac{1}{1 - \alpha_2(P)}$. Let $\pi_{\min} := \min_{u \in V} \pi(u)$. It is known that (see e.g. Chapter 12 of [LP17])

$$\left(\tau_{\mathrm{rel}}(P) - 1\right) \cdot \log \frac{1}{2\epsilon} \leq \tau_{\mathrm{mix}}(P, \epsilon) \leq \tau_{\mathrm{rel}}(P) \cdot \log \frac{1}{\epsilon \cdot \pi_{\min}}.$$

Because of this connection between the spectral gap and the mixing time of P, the optimization problem of maximizing the spectral gap of the random walk matrix is referred to as "fastest mixing time" in [BDX04].

Spectral Graph Theory

Given a graph $G = (V, E)$, its adjacency matrix $A = A(G)$ is a $n \times n$ matrix where the (u, v)-th entry is $\mathbb{1}_{uv \in E}$. The Laplacian matrix is defined as $L := D - A$, where $D := \text{diag}(\{\deg(v)\}_{v \in V})$ is the diagonal degree matrix. For a vector $x \in \mathbb{R}^n$, the Laplacian matrix has a useful quadratic form $x^T L x = \sum_{uv \in E} (x(u) - x(v))^2$.

The normalized adjacency matrix is defined as $\mathcal{A} = D^{-1/2} A D^{-1/2}$, and the normalized Laplacian matrix is defined as $\mathcal{L} := I - \mathcal{A}$. Observe that $\mathcal{A}$ is similar to the simple random walk matrix on G, so it is diagonalizable with eigenvalues $1 = \alpha_1(\mathcal{A}) \geq \alpha_2(\mathcal{A}) \geq \cdots \geq \alpha_n(\mathcal{A}) \geq -1$. Therefore, $\mathcal{L}$ is diagonalizable, and its eigenvalues are $0 = \lambda_1(\mathcal{L}) \leq \lambda_2(\mathcal{L}) \leq \cdots \leq \lambda_n(\mathcal{L}) \leq 2$. Note that we use α_i to denote the eigenvalues of the normalized adjacency matrix $\mathcal{A}$ and random walk matrix P, and we use λ_i to denote the eigenvalues of the normalized Laplacian matrix $\mathcal{L}$.

Let $\phi(G) := \min_{S \subseteq V : 0 < \pi(S) \leq 1/2} \phi(S)$ be the edge conductance of the graph G. Cheeger's inequality [Che70, AM85, Alo86] states that

$$\frac{\lambda_2}{2} \leq \phi(G) \leq \sqrt{2\lambda_2}.$$

This theorem is important because it connects (i) the spectral gap of the normalized Laplacian matrix, (ii) the edge conductance of the graph and (iii) the mixing time of random walks.

Convex Optimization

Consider optimization programs in the standard form

$$\mathcal{P} := \quad \min_{x \in \Omega} \quad f(x)$$
$$\text{subject to} \quad g_i(x) \leq 0 \quad \forall i \in [l]$$
$$h_j(x) = 0 \quad \forall i \in [p]$$

where $\Omega \subseteq \mathbb{R}^n$ and $f, g_i, h_j : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$. To define its Lagrangian dual, consider

$$\Lambda(x, \lambda, \mu) := f(x) + \sum_{i \in [l]} \lambda_i g_i(x) + \sum_{j \in [p]} \mu_j h_j(x)$$

defined on $\Omega \times \mathbb{R}^l \times \mathbb{R}^p$. The dual program is

$$\mathcal{D} := \max_{\lambda \geq 0, \mu} \inf_{x \in \Omega} \Lambda(x, \lambda, \mu).$$

Weak duality always holds, that is, $\mathcal{D} \leq \mathcal{P}$. We say that strong duality holds if $\mathcal{D} = \mathcal{P}$.

Linear programs are optimization programs where $\Omega = \mathbb{R}^n$ and f, g_i, h_j are all affine functions. It is well-known that strong duality always holds for linear programs.

Semidefinite programs (SDP) are optimization programs where the ambient space is $\mathbb{R}^{n \times n}$, $\Omega := \{X \in \mathbb{R}^{n \times n} : X \succcurlyeq 0\}$, and f, g_i, h_j are all affine functions. Unlike linear programs, there are SDP's where strong duality does not hold. We will use Von Neumann's minimax theorem in Section 5 to establish strong duality for the SDP for multiway vertex expansion.

Theorem 2.3 (Von Neumann's Minimax Theorem (see [Sim95])). *Let X, Y be compact convex sets. If f is a real-valued continuous function on $X \times Y$ with $f(x, \cdot)$ concave on Y for all $x \in X$ and $f(\cdot, y)$ convex on X for all $y \in Y$, then*

$$\min_{x \in X} \max_{y \in Y} f(x, y) = \max_{y \in Y} \min_{x \in X} f(x, y).$$

Several eigenvalue optimization problems can be formulated as semidefinite programs. Boyd, Diaconis and Xiao [BDX04] showed that the maximum reweighted spectral gap in Definition 1.1 can be written as a semidefinite program. Roch [Roc05] showed that the maximum reweighted lower spectral gap problem in Definition 1.6 can be written as a semidefinite program; see Proposition 4.1. For the higher-order Cheeger inequality for vertex expansion, we will use the following proposition in writing the maximum reweighted sum of k smallest eigenvalue problem in Definition 5.1 as a semidefinite program.

Proposition 2.4. *Let $X \in \mathbb{R}^{n \times n}$ be a symmetric matrix and let $1 \le k \le n$. Suppose the eigenvalues of X are $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$. Then, $\lambda_1 + \lambda_2 + \cdots + \lambda_k$ is the value of the following semidefinite program:*

$$
\begin{aligned}
\min_{Y \in \mathbb{R}^{n \times n}} \quad & \operatorname{tr}(XY) \\
\text{subject to} \quad & 0 \preccurlyeq Y \preccurlyeq I_n \\
& \operatorname{tr}(Y) = k.
\end{aligned}
$$

Polytopes

A polytope $\emptyset \ne Q \subseteq \mathbb{R}^n$ is a solution set to affine inequalities, i.e. $Q = \{x \in \mathbb{R}^n : Ax \le b\}$ for some $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Given a point set $X \subseteq \mathbb{R}^n$, its convex hull $\operatorname{conv}(X) \subset \mathbb{R}^n$ is the set of all convex combinations of points in X. Equivalently, $\operatorname{conv}(X)$ is the smallest convex set containing X. A basic result is that the convex hull of a finite point set in $\mathbb{R}^n$ is a polytope.

Given a polytope $Q \subseteq \mathbb{R}^n$. A subset $F \subseteq Q$ is a face of Q if, for any $x, y \in Q$, if $tx + (1-t)y \in F$ for some $t \in (0, 1)$ then $x, y \in F$. If F is a face of Q, then it is the intersection of Q and an affine subspace $Y \subseteq \mathbb{R}^n$. The dimension of a face F is the dimension of the smallest (by inclusion) affine subsapce $Y \subseteq \mathbb{R}^n$ such that $F = Q \cap Y$.

Dimension-0 faces are called extreme points or vertices, whereas dimension-1 faces are called edges. The graph $G_Q = (V, E)$ of Q has the dimension-0 faces as the vertices and the dimension-1 faces as the edges. This is also called the 1-skeleton of the polytope. The following fact about the non-existence of edges between two vertices of a polytope will be useful.

Proposition 2.5 ([KR03])**.** *Let $Q \subseteq \mathbb{R}^n$ be a bounded polytope with vertex set $F_0 = F_0(Q)$. Let $x \neq y \in F_0$, and let $L(x, y)$ be the line segment with endpoints x and y. Suppose that $\mathrm{conv}\big(F_0 \setminus \{x, y\}\big) \cap L(x, y)$ is nonempty. Then, $L(x, y)$ is not an edge of Q.*

We will also use the following version of hyperplane separation theorem.

Proposition 2.6. *Let $M \subseteq \mathbb{R}^n$ be convex and let $x \in \mathbb{R}^n$ be such that $x \notin M$. Then there exists an affine function $l : \mathbb{R}^n \to \mathbb{R}$ such that $l(x) = 0$ and $l(y) < 0$ for all $y \in M$.*

3 Optimal Cheeger Inequality for Vertex Expansion

The goal of this section is to prove Theorem 1.4. We will first review the proofs in [OZ22, LRV13] in Section 3.1, and then present how to combine their proofs with a graph orientation idea to prove Theorem 1.4 in Section 3.2.

3.1 Background

We will first review the proofs by Olesker-Taylor and Zanetti [OZ22] in Section 3.1.1, and then the proofs by Louis, Raghavendra and Vempala in [LRV13] in Section 3.1.2.

In this subsection, the stationary distribution π is assumed to be the uniform distribution. This will slightly simplify the presentation and was also the setting considered in previous works.

3.1.1 Review of [OZ22]

Recall the fastest mixing time problem formulated in Definition 1.1. When π is the uniform distribution, the problem is to find a doubly stochastic reweighted matrix P of G that minimizes the second largest eigenvalue of P.

The starting point is the following dual characterization of the primal program in Definition 1.1 obtained by Roch [Roc05], which is stated in the form for a general distribution π that we will use.

Proposition 3.1 (Dual Program for Fastest Mixing [Roc05, OZ22])**.** *Given an undirected graph $G = (V, E)$ and a probability distribution π on V, the following semidefinite program is dual to the primal program in Definition 1.1 with strong duality $\lambda_2^*(G) = \gamma(G)$ where*

$$
\gamma(G) := \min_{f:V \to \mathbb{R}^n, \ g:V \to \mathbb{R}_{\geq 0}} \quad \sum_{v \in V} \pi(v) g(v)
$$

$$
\text{subject to} \quad \sum_{v \in V} \pi(v) \|f(v)\|^2 = 1
$$

$$
\sum_{v \in V} \pi(v) f(v) = \vec{0}
$$

$$
g(u) + g(v) \geq \|f(u) - f(v)\|^2 \qquad \forall uv \in E.
$$

We note that this is equivalent to the dual program given in [BDX04], but Roch's program is written in a vector program form that will be more convenient for rounding. In Section 5, we will use von Neumann's minimax theorem to derive a generalization of Proposition 3.1 for proving the higher-order Cheeger inequality for vertex expansion.

The proof of Theorem 1.3 has two main steps. The first step is to project the above dual program to the following one-dimensional "spectral" program.

Definition 3.2 (One-Dimensional Dual Program for Fastest Mixing [OZ22]). *Given an undirected graph $G = (V, E)$ and a probability distribution π on V, $\gamma^{(1)}(G)$ is defined to be the program:*

$$\gamma^{(1)}(G) := \min_{f:V \to \mathbb{R}, \ g:V \to \mathbb{R}_{\geq 0}} \quad \sum_{v \in V} \pi(v) g(v)$$
$$\text{subject to} \quad \sum_{v \in V} \pi(v) f(v)^2 = 1$$
$$\sum_{v \in V} \pi(v) f(v) = 0$$
$$g(u) + g(v) \geq (f(u) - f(v))^2 \qquad \forall uv \in E.$$

Olesker-Taylor and Zanetti use the Johnson-Lindenstrauss lemma to first project the solution in Proposition 3.1 to $O(\log n)$ dimensions with constant distortion, and then take the best coordinate to obtain a 1-dimensional solution with the following guarantee. Note that this step works for any probability distribution π on V.

Proposition 3.3 ([OZ22], Proposition 2.9). *For any undirected graph $G = (V, E)$ and any probability distribution π on V,*
$$\gamma(G) \leq \gamma^{(1)}(G) \lesssim \log |V| \cdot \gamma(G).$$

In the second step, Olesker-Taylor and Zanetti observed that the dual program in Definition 3.2 is similar to the weighted vertex cover problem with edge weights $(f(u) - f(v))^2$ for each edge $uv \in E$, which is equivalent to the fractional matching problem by linear programming duality. To analyze Definition 3.2, they introduced an interesting new concept called "matching conductance", and used some combinatorial arguments about greedy matching as well as some spectral arguments to prove the following Cheeger-type inequality.

Theorem 3.4 ([OZ22], Theorem 2.10). *For any undirected graph $G = (V, E)$ and the uniform distribution $\pi = \vec{1}/|V|$,*
$$\psi(G)^2 \lesssim \gamma^{(1)}(G) \lesssim \psi(G).$$

Combining Proposition 3.1 and Proposition 3.3 and Theorem 3.4 gives
$$\psi(G)^2 \lesssim \gamma^{(1)}(G) \lesssim \log |V| \cdot \gamma(G) = \log |V| \cdot \lambda_2^*(G) \quad \text{and} \quad \lambda_2^*(G) = \gamma(G) \leq \gamma^{(1)}(G) \lesssim \psi(G),$$

proving Theorem 1.3.

Note that the proof of the second step only works when π is the uniform distribution. Olesker-Taylor and Zanetti discussed some difficulty in generalizing their combinatorial arguments to the weighted setting, and left it as an open question to prove Theorem 3.4 for any probability distribution π.

3.1.2 Review of [LRV13]

Our proof is based on the techniques in [LRV13] which we review here. Their algorithm is based on the following "spectral" formulation λ_∞ by Bobkov, Houdré and Tetali [BHT00], which is for the uniform distribution π.

Definition 3.5 (λ_∞ in [BHT00]). *Given an undirected graph $G = (V, E)$,*

$$\lambda_\infty(G) := \min_{x:V \to \mathbb{R}, \ x \perp \vec{1}} \frac{\sum_{u \in V} \max_{v:(v,u) \in E} \ (x(u) - x(v))^2}{\sum_{u \in V} x(u)^2}.$$

Bobkov, Houdré and Tetali [BHT00] proved an exact analog of Cheeger's inequality for symmetric vertex expansion that $\frac{1}{2}\psi_{\mathrm{sym}}(G)^2 \le \lambda\infty(G) \le 2\psi_{\mathrm{sym}}(G)$. We will use some of their arguments to prove a similar statement in Theorem 3.15 in Section 3.2.

The issue is that λ_∞ is not known to be efficiently computable, and indeed recently Farhadi, Louis and Tetali [FLT20] proved that it is NP-hard to compute $\lambda_\infty(G)$ exactly. To design an approximation algorithm for $\psi(G)$, Louis, Raghavendra and Vempala [LRV13] defined the following semidefinite programming relaxation for λ_∞, which we denote by sdp_∞.

Definition 3.6 (sdp_∞ in [LRV13]). *Given an undirected graph $G = (V, E)$,*

$$\mathsf{sdp}_\infty(G) := \min_{f:V \to \mathbb{R}^n, \ g:V \to \mathbb{R}} \quad \sum_{v \in V} g(v)$$

$$\text{subject to} \quad \sum_{v \in V} \|f(v)\|^2 = 1$$

$$\sum_{v \in V} f(v) = \vec{0}$$

$$g(v) \ge \|f(u) - f(v)\|^2 \qquad \forall u \in V \text{ with } uv \in E.$$

The rounding algorithm in [LRV13] is to project the solution to sdp_∞ into a 1-dimensional solution by setting $x(v) = \langle f(v), h \rangle$ where $h \sim N(0,1)^n$ is a random Gaussian vector. They proved that the 1-dimensional solution is a $O(\log d)$-approximation to sdp_∞ where d is the maximum degree of the graph.

Theorem 3.7 ([LRV13], Lemma 9.6). *For any undirected graph $G = (V, E)$ with maximum degree d,*

$$\mathsf{sdp}_\infty(G) \le \lambda_\infty \lesssim \log d \cdot \mathsf{sdp}_\infty(G).$$

For the analysis, they used the following properties of Gaussian random variables, for which we will also use in our proofs and so we state them here. The first fact is for the analysis of the numerator and the second fact is for the analysis of the denominator of λ_∞.

Fact 3.8 ([LRV13], Fact 9.7). *Let $Y_1, Y_2, \ldots, Y_d$ be d Gaussian random variables with mean 0 and variance at most σ^2. Let Y be the random variable defined as $Y := \max\{Y_i \mid i \in [d]\}$. Then*

$$\mathbb{E}[Y] \le 2\sigma\sqrt{\log d}.$$

Fact 3.9 ([LRV13], Lemma 9.8). *Suppose $z_1, \ldots, z_m$ are Gaussian random variables (not necessarily independent) such that $\mathbb{E}[\sum_{i=1}^m z_i^2] = 1$. Then*

$$\Pr\left[\sum_{i=1}^m z_i^2 \geq \frac{1}{2}\right] \geq \frac{1}{12}.$$

3.2 Proof of Theorem 1.4

We follow the same two-step plan as in [OZ22]. We will prove in Proposition 3.14 in Section 3.2.2 that $\gamma^{(1)}(G) \lesssim \gamma(G) \cdot \log d$ for any probability distribution π. Note that this already improves Theorem 1.3 to the optimal bound, when π is the uniform distribution. Then, we will prove in Theorem 3.15 in Section 3.2.3 that $\psi(G)^2 \lesssim \gamma^{(1)}(G) \lesssim \psi(G)$ for any probability distribution π on V. As in [OZ22], combining Proposition 3.1 and Proposition 3.14 and Theorem 3.15 gives Theorem 1.4.

3.2.1 Dual Program on Graph Orientation

To extend the techniques in [LRV13, BHT00] to prove the two steps, we will introduce a "directed" program $\overrightarrow{\gamma}(G)$ to bring $\gamma(G)$ in Proposition 3.1 closer to $\mathsf{sdp}_\infty(G)$ in Definition 3.6.

Observe that the two SDP programs $\gamma(G)$ and $\mathsf{sdp}_\infty(G)$ have very similar form. The only difference is that the last constraint in Proposition 3.1 only requires that $g(u) + g(v) \geq \|f(u) - f(v)\|^2$ for $uv \in E$, while the last constraint in Definition 3.6 has a stronger requirement that $\min\{g(u), g(v)\} \geq \|f(u) - f(v)\|^2$ for $uv \in E$. So sdp_∞ is a stronger relaxation than $\gamma(G) = \lambda_2^*(G)$.

Lemma 3.10. *For any undirected graph $G = (V, E)$ and any probability distribution π on V,*

$$\lambda_2^*(G) \leq \mathsf{sdp}_\infty(G).$$

For our analysis of $\lambda_2^*(G)$, we consider the following "directed" program $\overrightarrow{\gamma}(G)$ where the last constraint is $\max\{g(u), g(v)\} \geq \|f(u) - f(v)\|^2$ for $uv \in E$. We also state the corresponding 1-dimensional version as in Definition 3.2 in the following definition.

Definition 3.11 (Directed Dual Programs for $\gamma(G)$). *Given an undirected graph $G = (V, E)$ and a probability distribution π on V,*

$$\overrightarrow{\gamma}(G) := \min_{f:V \to \mathbb{R}^n, \ g:V \to \mathbb{R}_{\geq 0}} \quad \sum_{v \in V} \pi(v) g(v)$$

$$\text{subject to} \quad \sum_{v \in V} \pi(v) \|f(v)\|^2 = 1$$

$$\sum_{v \in V} \pi(v) f(v) = \vec{0}$$

$$\max\{g(u), g(v)\} \geq \|f(u) - f(v)\|^2 \qquad \forall uv \in E.$$

$\overrightarrow{\gamma}^{(1)}(G)$ is defined as the 1-dimensional program of $\overrightarrow{\gamma}(G)$ where $f : V \to \mathbb{R}$ instead of $f : V \to \mathbb{R}^n$.

Note that $\overrightarrow{\gamma}(G)$ is not a semidefinite program because of the max constraint, but $\gamma(G)$ and $\overrightarrow{\gamma}(G)$ are closely related and $\overrightarrow{\gamma}(G)$ is only used in the analysis as a proxy for $\gamma(G)$.

Lemma 3.12. *For any undirected graph $G = (V, E)$ and any probability distribution π on V,*

$$\gamma(G) \le \overrightarrow{\gamma}(G) \le 2\gamma(G) \quad \text{and} \quad \gamma^{(1)}(G) \le \overrightarrow{\gamma}^{(1)}(G) \le 2\gamma^{(1)}(G).$$

Proof. As $g \ge 0$, any feasible solution f, g to $\overrightarrow{\gamma}(G)$ is a feasible solution to $\gamma(G)$ and so the first inequalities follow. On the other hand, for any feasible solution f, g to $\gamma(G)$, note that $f, 2g$ is a feasible solution to $\overrightarrow{\gamma}(G)$ and so the second inequalities follow. $\qquad\square$

The reason that we call $\overrightarrow{\gamma}(G)$ the "directed" program is as follows. For each edge $uv \in E$, the constraint in $\mathsf{sdp}_\infty(G)$ requires both $g(u)$ and $g(v)$ to be at least $\|f(u) - f(v)\|^2$, while the constraint in $\overrightarrow{\gamma}(G)$ only requires at least one of $g(u)$ or $g(v)$ to be at least $\|f(u) - f(v)\|^2$. We think of $\overrightarrow{\gamma}(G)$ as assigning a direction to each edge and requiring that $g(v) \ge \|f(u) - f(v)\|^2$ for each directed edge $u \to v$. Then, we can rewrite the programs $\overrightarrow{\gamma}(G)$ and $\overrightarrow{\gamma}^{(1)}(G)$ by eliminating the variables $g(v)$ for $v \in V$, by minimizing over all possible orientations of the edge set E.

Lemma 3.13 (Directed Dual Programs Using Orientation for $\gamma(G)$). *Let $G = (V, E)$ be an undirected graph and π be a probability distribution on V. Let $\overrightarrow{E}$ be an orientation of the undirected edges in E. Then*

$$\overrightarrow{\gamma}(G) = \min_{f:V \to \mathbb{R}^n} \min_{\overrightarrow{E}} \quad \sum_{v \in V} \pi(v) \max_{u:uv \in \overrightarrow{E}} \|f(u) - f(v)\|^2$$

$$\text{subject to} \quad \sum_{v \in V} \pi(v) \|f(v)\|^2 = 1$$

$$\sum_{v \in V} \pi(v) f(v) = \vec{0}.$$

Similarly, $\overrightarrow{\gamma}^{(1)}(G)$ can be written in the same form with $f : V \to \mathbb{R}$ instead of $f : V \to \mathbb{R}^n$.

Proof. In one direction, given an orientation $\overrightarrow{E}$, we can define $g(v) := \max_{u:uv \in \overrightarrow{E}} \|f(u) - f(v)\|^2$, so that f, g is a feasible solution to $\overrightarrow{\gamma}(G)$ as stated in Definition 3.11 with the same objective value.

In the other direction, given a solution f, g in Definition 3.11, we can define an orientation $\overrightarrow{E}$ of E so that each directed edge uv satisfies $g(v) \ge \|f(u) - f(v)\|^2$. Note that $g(v) \ge \max_{u:uv \in \overrightarrow{E}} \|f(u) - f(v)\|^2$, and setting it to be an equality would satisfy all the constraints and not increase the objective value as $g \ge 0$. $\qquad\square$

This formulation will be useful in both the Gaussian projection step for Proposition 3.14 and the threshold rounding step for Theorem 3.15.

3.2.2 Gaussian Projection

The following proposition is an improvement of Proposition 3.3 in [OZ22]. The formulation in Lemma 3.13 allows us to use the expected maximum of Gaussian random variables in Fact 3.8 to analyze the projection as was done in [LRV13].

Proposition 3.14 (Gaussian Projection for $\gamma(G)$). *For any undirected graph $G = (V, E)$ with maximum degree d and any probability distribution π on V,*

$$\gamma(G) \leq \gamma^{(1)}(G) \lesssim \gamma(G) \cdot \log d.$$

Proof. We will prove that $\overrightarrow{\gamma}(G) \leq \overrightarrow{\gamma}^{(1)}(G) \lesssim \log d \cdot \overrightarrow{\gamma}(G)$, and the proposition will follow from Lemma 3.12. The first inequality is immediate as $\overrightarrow{\gamma}^{(1)}(G)$ is a restriction of $\overrightarrow{\gamma}(G)$, so we focus on proving the second inequality.

Let $f : V \to \mathbb{R}^n$ and $\overrightarrow{E}$ be a solution to $\overrightarrow{\gamma}(G)$ as stated in Lemma 3.13. As in [LRV13], we construct a 1-dimensional solution $y \in \mathbb{R}^n$ to $\overrightarrow{\gamma}^{(1)}(G)$ by setting $y(v) = \langle f(v), h \rangle$, where $h \sim N(0,1)^n$ is a Gaussian random vector with independent entries.

First, consider the expected objective value of y to $\overrightarrow{\gamma}^{(1)}(G)$. For each max term in the summand,

$$\mathbb{E}\left[\max_{u:u\to v} \left(y(u) - y(v)\right)^2\right] = \mathbb{E}\left[\max_{u:u\to v} \left\langle f(u) - f(v), h \right\rangle^2\right] \leq 2 \max_{u:u\to v} \|f(u) - f(v)\|^2 \cdot \log d,$$

where the last inequality is by applying Fact 3.8 on normal random variable $\langle f(u) - f(v), h \rangle$ with variance $\|f(u) - f(v)\|^2$ for each of the at most d terms. By linearity of expectation, the expected objective value of $\overrightarrow{\gamma}^{(1)}(G)$ is

$$\mathbb{E}\left[\sum_{v \in V} \pi(v) \max_{u:u\to v} \left(y(u) - y(v)\right)^2\right] \leq 2 \log d \cdot \sum_{v \in V} \pi(v) \max_{u:u\to v} \|f(u) - f(v)\|^2 = 2 \log d \cdot \overrightarrow{\gamma}(G).$$

Therefore, by Markov's inequality,

$$\Pr\left[\sum_{v \in V} \pi(v) \max_{u:u\to v} \left(y(u) - y(v)\right)^2 \geq 48 \log d \cdot \overrightarrow{\gamma}(G)\right] \leq \frac{1}{24}.$$

Next, by applying Fact 3.9 with $z_v = \sqrt{\pi(v)} \cdot y(v)$, it follows that

$$\mathbb{E}\left[\sum_{v \in V} \pi(v) y(v)^2\right] = \sum_{v \in V} \pi(v) \|f(v)\|^2 = 1 \quad \implies \quad \Pr\left[\sum_{v \in V} \pi(v) y(v)^2 \geq \frac{1}{2}\right] \geq \frac{1}{12}.$$

Finally, since $\sum_{v \in V} \pi(v) f(v) = \vec{0}$, it holds that

$$\sum_{v \in V} \pi(v) y(v) = \sum_{v \in V} \pi(v) \langle f(v), h \rangle = \left\langle \sum_{v \in V} \pi(v) f(v), h \right\rangle = 0.$$

Therefore, with probability at least $\frac{1}{24}$, all of these events hold simultaneously. The second event

$$\sum_{v \in V} \pi(v) y(v)^2 \geq \frac{1}{2}$$

means that we can rescale y by a factor of at most $\sqrt{2}$, so that the constraint $\sum_{v \in V} \pi(v) y(v)^2 = 1$ is satisfied and the objective value is at most $96 \log d \cdot \overrightarrow{\gamma}(G)$. Hence we conclude that $\overrightarrow{\gamma}^{(1)}(G) \lesssim \overrightarrow{\gamma}(G) \cdot \log d$. $\qquad\square$

3.2.3 Cheeger Rounding for Vertex Expansion

We generalize Theorem 3.4 to weighted vertex expansion. Our proof does not use the concept of matching conductance in [OZ22], rather it is based on a more traditional analysis as in [BHT00] using the directed program $\overrightarrow{\gamma}^{(1)}(G)$ in Lemma 3.13.

Theorem 3.15 (Cheeger Inequality for Weighted Vertex Expansion). *For any undirected graph $G = (V, E)$ and any probability distribution π on V,*

$$\psi(G)^2 \lesssim \gamma^{(1)}(G) \lesssim \psi(G).$$

The organization is as follows. We will prove the easy direction in Lemma 3.16 in Appendix A. For the hard direction, we will work on $\overrightarrow{\gamma}^{(1)}(G)$ instead. First we do the standard preprocessing step to truncate the solution to have π-weight at most $1/2$. Then the main step is to define a modified vertex boundary condition for directed graphs and use it for the analysis of the standard threshold rounding. Finally we clean up the solution obtained from threshold rounding to find a set with small vertex expansion in the underlying undirected graph.

Lemma 3.16 (Easy Direction). *For any undirected graph $G = (V, E)$ and any probability distribution π on V,*

$$\gamma^{(1)}(G) \leq 2\psi(G).$$

We now turn to proving the hard direction. Given a solution $y : V \to \mathbb{R}$ to $\overrightarrow{\gamma}^{(1)}(G)$ in Lemma 3.13 satisfying $y \perp \pi$, we do the standard preprocessing step to truncate y to obtain a non-negative solution x with $\pi(\mathrm{supp}(x)) \leq 1/2$ and comparable objective value. Note that we no longer require that $x \perp \pi$. The proof of the following lemma is standard and is deferred to Appendix A.

Lemma 3.17 (Truncation). *Let $G = (V, E)$ be an undirected graph and π be a probability distribution on V. Given a solution y and $\overrightarrow{E}$ to $\overrightarrow{\gamma}^{(1)}(G)$ as stated in Lemma 3.13, there is a solution x and $\overrightarrow{E}$ with $x \geq 0$ and $\pi(\mathrm{supp}(x)) \leq 1/2$ and*

$$\frac{\sum_{v \in V} \pi(v) \max_{u : uv \in \overrightarrow{E}} (x(u) - x(v))^2}{\sum_{v \in V} \pi(v) x(v)^2} \leq 4 \overrightarrow{\gamma}^{(1)}(G).$$

For the standard threshold rounding, we define the appropriate vertex boundary $\overrightarrow{\partial} S$ for the analysis of the directed program $\overrightarrow{\gamma}^{(1)}(G)$. Note that, unlike ∂S, $\overrightarrow{\partial} S$ may contain vertices in S. A good interpretation is to think of $\overrightarrow{\partial} S$ as a vertex cover of the edge boundary $\delta(S)$ in the undirected sense.

Definition 3.18 (Directed Vertex Boundary and Expansion). *Let $G = (V, \overrightarrow{E})$ be a directed graph. For $S \subseteq V$, define the directed vertex boundary and the directed vertex expansion as*

$$\overrightarrow{\partial} S := \{v \in S \mid \exists u \notin S \text{ with } uv \in \overrightarrow{E}\} \cup \{v \notin S \mid \exists u \in S \text{ with } uv \in \overrightarrow{E}\} \quad \text{and} \quad \overrightarrow{\psi}(S) := \frac{\pi(\overrightarrow{\partial} S)}{\pi(S)}.$$

The main step is to prove that the standard threshold rounding will find a set S with small directed vertex expansion $\overrightarrow{\psi}(S)$.

Proposition 3.19 (Threshold Rounding for $\gamma(G)$). *Let $G = (V, E)$ be an undirected graph and π be a probability distribution on V. Given a solution x and $\overrightarrow{E}$ with $x \geq 0$ and*

$$\frac{\sum_{v \in V} \pi(v) \max_{u:u \to v} (x(u) - x(v))^2}{\sum_{v \in V} \pi(v) x(v)^2} \leq \gamma(G),$$

there is a set $S \subseteq \text{supp}(x)$ with $\overrightarrow{\psi}(S) \lesssim \sqrt{\gamma(G)}$.

Proof. For any $t \geq 0$, define $S_t := \{v \in V \mid x(v)^2 > t\}$. By a standard averaging argument,

$$\min_t \overrightarrow{\psi}(S_t) \leq \frac{\int_0^\infty \pi(\overrightarrow{\partial} S_t)\, dt}{\int_0^\infty \pi(S_t)\, dt}.$$

The denominator is

$$\int_0^\infty \pi(S_t)\, dt = \sum_{v \in V} \pi(v) \int_0^\infty \mathbb{1}[v \in S_t]\, dt = \sum_{v \in V} \pi(v) x(v)^2.$$

For the numerator, note that a vertex v is in $\overrightarrow{\partial} S_t$ if and only if $\min\{x(u)^2 \mid uv \in \overrightarrow{E}\} \leq t \leq \max\{x(u)^2 \mid uv \in \overrightarrow{E}\}$, where we recall the assumption that every vertex has a self loop, and so $vv \in \overrightarrow{E}$ and thus $\min\{x(u)^2 \mid uv \in \overrightarrow{E}\} \leq x(v)^2 \leq \max\{x(u)^2 \mid uv \in \overrightarrow{E}\}$. Hence the numerator is

$$\int_0^\infty \pi\left(\overrightarrow{\partial} S_t\right) dt$$

$$= \sum_{v \in V} \pi(v) \cdot \int_0^\infty \mathbb{1}\left[v \in \overrightarrow{\partial} S_t\right] dt$$

$$= \sum_{v \in V} \pi(v) \cdot \int_0^\infty \mathbb{1}\left[\min\left\{x(u)^2 \mid uv \in \overrightarrow{E}\right\} \leq t \leq \max\left\{x(u)^2 \mid uv \in \overrightarrow{E}\right\}\right] dt$$

$$= \sum_{v \in V} \pi(v) \left[\max_{\substack{u:u \to v \\ x(u) > x(v)}} \left\{x(u)^2 - x(v)^2\right\} + \max_{\substack{u:u \to v \\ x(u) < x(v)}} \left\{x(v)^2 - x(u)^2\right\}\right]$$

$$\leq 2 \sum_{v \in V} \pi(v) \left[\max_{u:u \to v} \left\{|x(u)^2 - x(v)^2|\right\}\right]$$

$$\leq 2 \sum_{v \in V} \pi(v) \left[\max_{u:u \to v} \left\{(x(u) - x(v))^2 + 2x(v) \cdot |x(u) - x(v)|\right\}\right]$$

$$\leq 2 \sum_{v \in V} \pi(v) \max_{u:u \to v} (x(u) - x(v))^2 + 4 \sqrt{\sum_{v \in V} \pi(v) x(v)^2 \cdot \sum_{v \in V} \pi(v) \max_{u:u \to v} (x(u) - x(v))^2},$$

21

where the second-last inequality is by $|x(u)^2 - x(v)^2| \leq |x(u) - x(v)| \cdot \left(|x(u) - x(v)| + 2|x(v)|\right)$, and the last inequality is by the Cauchy-Schwarz inequality.

Combining the numerator and the denominator bounds,

$$
\begin{aligned}
\frac{\int_0^\infty \pi\left(\overrightarrow{\partial} S_t\right) dt}{\int_0^\infty \pi(S_t)\, dt} &\leq \frac{\sum_{v \in V} \pi(v) \max_{u : u \to v} (x(u) - x(v))^2}{\sum_{v \in V} \pi(v) x(v)^2} + 2\sqrt{\frac{\sum_{v \in V} \pi(v) \max_{u : u \to v} (x(u) - x(v))^2}{\sum_{v \in V} \pi(v) x(v)^2}} \\
&= \gamma + 2\sqrt{\gamma} \lesssim \sqrt{\gamma},
\end{aligned}
$$

where the last inequality is by $\gamma \leq 2$ as was shown in the proof of the easy direction in Lemma 3.16. Therefore, $\min_t \overrightarrow{\psi}(S_t) \lesssim \sqrt{\gamma}$ and $S_t \subseteq \mathrm{supp}(x)$ by construction. $\qquad \square$

Finally, given a set S with small directed vertex expansion $\overrightarrow{\psi}(S)$, we show how to find a set $S' \subseteq S$ with small vertex expansion $\psi(S')$. This step is similar to the step in [OZ22, Proposition 2.2] from matching conductance to vertex expansion.

Lemma 3.20 (Postprocessing for Vertex Expansion). *Let $G = (V, \overrightarrow{E})$ be a directed graph. Given a set S with $\overrightarrow{\psi}(S) < 1/2$, there is a set $S' \subseteq S$ with $\psi(S') \leq 2\overrightarrow{\psi}(S)$ in the underlying undirected graph of G.*

Proof. From Definition 3.18, the observation is that all undirected edges in $\delta(S)$ are incident to at least one vertex in $\overrightarrow{\partial} S$. Define $S' := S - \overrightarrow{\partial} S$. Then observe that $\partial S' \subseteq \overrightarrow{\partial} S$, as there are no incoming edges to S' from $V - (S' \cup \overrightarrow{\partial} S)$ and all outgoing edges from S' go to $\overrightarrow{\partial}(S)$. This implies that

$$
\pi(\partial S') \leq \pi\left(\overrightarrow{\partial} S\right) = \overrightarrow{\psi}(S) \cdot \pi(S) \leq 2\overrightarrow{\psi}(S) \cdot \pi(S'),
$$

where the last inequality uses the assumption that $\overrightarrow{\psi}(S) = \pi(\overrightarrow{\partial} S)/\pi(S) < 1/2$ and so $\pi(S') \geq \pi(S) - \pi(\overrightarrow{\partial} S) \geq \pi(S)/2$. We conclude that $\psi(S') \leq 2\overrightarrow{\psi}(S)$. $\qquad \square$

We put together the steps and complete the proof of Theorem 3.15 in Appendix A.

4 Cheeger Inequality for Bipartite Vertex Expansion

The goal of this section is to prove Theorem 1.7, which relates the maximum reweighted lower spectral gap $\zeta^*(G)$ in Definition 1.6 and the bipartite vertex expansion $\psi_B(G)$ in Definition 1.5. The proof follows closely the proof of Theorem 1.4 in the previous section, so some steps will be stated without proofs, and the focus will be on the threshold rounding step.

4.1 Primal and Dual Programs

The primal program $\zeta^*(G)$ in Definition 1.6 has the following dual which is similar to $\gamma(G)$ in Proposition 3.1.

Proposition 4.1 (Dual Program for Lower Spectral Gap [Roc05]). *Given an undirected graph $G = (V, E)$ and a probability distribution π on V, the following semidefinite program is dual to the primal program in Definition 1.6 with strong duality $\zeta^*(G) = \nu(G)$ where*

$$\nu(G) := \min_{f:V \to \mathbb{R}^n,\ g:V \to \mathbb{R}_{\geq 0}} \quad \sum_{v \in V} \pi(v) g(v)$$

$$\text{subject to} \quad \sum_{v \in V} \pi(v) \|f(v)\|^2 = 1$$

$$g(u) + g(v) \geq \|f(u) + f(v)\|^2 \qquad \forall uv \in E.$$

There are two differences between $\gamma(G)$ in Proposition 3.1 and $\nu(G)$ in Proposition 4.1. One is that the constraint $g(u) + g(v) \geq \|f(u) - f(v)\|^2$ in $\gamma(G)$ is replaced by the constraint $g(u) + g(v) \geq \|f(u) + f(v)\|^2$ in $\nu(G)$, which are handled in a similar way. The other is that the constraint of $\sum_{v \in V} \pi(v) f(v) = \vec{0}$ in $\gamma(G)$ is not present in $\nu(G)$, and so it is slightly easier to work with $\nu(G)$, e.g. no truncation step needed.

The nice form of the dual program $\nu(G)$ is the main reason behind the definition of the primal program $\zeta^*(G)$. By the variational characterization of eigenvalues, the quadratic form of $D_P + P$, and the π-reversibility of P,

$$\lambda_{\min}(D_P + P) = \min_{x:V \to \mathbb{R}} \frac{x^T(D_P + P)x}{x^T x} = \min_{x:V \to \mathbb{R}} \frac{\sum_{uv \in E} \pi(u)P(u,v)(x(u) + x(v))^2}{\sum_{u \in V} \pi(u)x(u)^2},$$

and this is the intermediate form we need to derive the dual, see [Roc05].

Later, as in Section 3.2.1, we will define a directed dual program $\overrightarrow{\nu}(G)$, and the dual constraint $g \geq 0$ crucially enables us to relate the two programs. The dual constraint $g \geq 0$ comes from the primal constraint $\sum_{v \in V} P(u, v) \leq 1$, whereas if we use $\sum_{v \in V} P(u, v) = 1$ then g will be unconstrained.

For $\lambda_2^*(G)$, we sidestep the issue by adding self loops to each vertex of G. The non-negativity of g in the dual program $\gamma(G)$ follows indirectly from $g(u) + g(u) \geq \|f(u) - f(u)\|^2$, as now $(u, u) \in E$. Moreover, adding self loops does not change the vertex expansion of G. Therefore, $\lambda_2^*(G)$ can take the more natural form where P does correspond to a transition matrix. However, we cannot do the same for $\zeta^*(G)$, because the additional constraint on g becomes $g(u) + g(u) \geq \|f(u) + f(u)\|^2$ which changes the objective value, and also that adding self loops takes away the bipartiteness of subgraphs.

4.2 Proof of Theorem 1.7

We use the same two-step plan as in Section 3.2. In the first step, we project the solution to the dual program in Proposition 4.1 into a 1-dimensional solution to the following program.

Definition 4.2 (One-Dimensional Dual Program for Lower Spectral Gap). *Given an undirected*

23

graph $G = (V, E)$ and a probability distribution π on V, $\nu^{(1)}(G)$ is defined as the following program:

$$\nu^{(1)}(G) := \min_{f:V\to\mathbb{R},\ g:V\to\mathbb{R}_{\geq 0}} \quad \sum_{v\in V} \pi(v)g(v)$$

$$\text{subject to} \quad \sum_{v\in V} \pi(v)\|f(v)\|^2 = 1$$

$$g(u) + g(v) \geq \|f(u) + f(v)\|^2 \qquad \forall uv \in E.$$

As in Proposition 3.14, we use the Gaussian projection method in [LRV13] to prove the following guarantee.

Proposition 4.3 (Gaussian Projection for $\nu(G)$). *For any undirected graph $G = (V, E)$ with maximum degree d and any probability distribution π on V,*

$$\nu(G) \leq \nu^{(1)}(G) \lesssim \nu(G) \cdot \log d.$$

In the second step, we prove a Cheeger-type inequality relating $\psi_B(G)$ and $\nu(G)$.

Theorem 4.4. *For any undirected graph $G = (V, E)$ and any probability distribution π on V,*

$$\psi_B(G)^2 \lesssim \nu^{(1)}(G) \lesssim \psi_B(G).$$

Combining Proposition 4.1 and Proposition 4.3 and Theorem 4.4 gives

$$\psi_B(G)^2 \lesssim \nu^{(1)}(G) \lesssim \nu(G) \cdot \log d = \zeta^*(G) \log d \quad \text{and} \quad \zeta^*(G) = \nu(G) \leq \nu^{(1)}(G) \lesssim \psi_B(G),$$

proving Theorem 1.7. We will prove Proposition 4.3 and Theorem 4.4 in the following subsections.

4.3   Dual Program on Graph Orientation

As in Section 3.2.1, we introduce a directed program for the analysis of both steps.

Definition 4.5 (Directed Dual Programs for $\nu(G)$). *Given an undirected graph $G = (V, E)$ and a probability distribution π on V,*

$$\overrightarrow{\nu}(G) := \min_{f:V\to\mathbb{R}^n,\ g:V\to\mathbb{R}_{\geq 0}} \quad \sum_{v\in V} \pi(v)g(v)$$

$$\text{subject to} \quad \sum_{v\in V} \pi(v)\|f(v)\|^2 = 1$$

$$\max\{g(u), g(v)\} \geq \|f(u) + f(v)\|^2 \qquad \forall uv \in E.$$

$\overrightarrow{\nu}^{(1)}(G)$ is defined as the 1-dimensional program of $\overrightarrow{\nu}(G)$ where $f : V \to \mathbb{R}$ instead of $f : V \to \mathbb{R}^n$.

As in Lemma 3.12, we show that $\nu(G)$ and $\overrightarrow{\nu}(G)$ are closely related. The proof is the same as in Lemma 3.12 and is omitted, but note that $g \geq 0$ is needed.

Lemma 4.6. *For any undirected graph $G = (V, E)$ and any probability distribution π on V,*

$$\nu(G) \leq \overrightarrow{\nu}(G) \leq 2\nu(G) \quad \text{and} \quad \nu^{(1)}(G) \leq \overrightarrow{\nu}^{(1)}(G) \leq 2\nu^{(1)}(G).$$

As in Lemma 3.13, we use an orientation of the edges to eliminate the variables $g(v)$ for $v \in V$ in $\overrightarrow{\nu}(G)$. The proof is the same as in Lemma 3.13 and is omitted, but note that $g \geq 0$ is needed.

Lemma 4.7 (Directed Dual Programs Using Orientation for $\nu(G)$). *Let $G = (V, E)$ be an undirected graph and π be a probability distribution on V. Let $\overrightarrow{E}$ be an orientation of the undirected edges in E. Then*

$$\overrightarrow{\nu}(G) = \min_{f:V \to \mathbb{R}^n} \min_{\overrightarrow{E}} \quad \sum_{v \in V} \pi(v) \max_{u:uv \in \overrightarrow{E}} \|f(u) + f(v)\|^2$$
$$\text{subject to} \quad \sum_{v \in V} \pi(v) \|f(v)\|^2 = 1.$$

Similarly, $\overrightarrow{\nu}^{(1)}(G)$ can be written in the same form with $f : V \to \mathbb{R}$ instead of $f : V \to \mathbb{R}^n$.

Once we have this formulation using orientation, we can use the same proof as in Proposition 3.14 to show that $\overrightarrow{\nu}(G) \leq \overrightarrow{\nu}^{(1)}(G) \lesssim \log d \cdot \overrightarrow{\nu}(G)$, and thus Proposition 4.3 follows from Lemma 4.6 and we omit the proof. It remains to prove Theorem 4.4, which will be done in the next subsection.

4.4 Cheeger Rounding for Bipartite Vertex Expansion

The goal of this subsection is to prove Theorem 4.4. We will prove the following easy direction in Appendix B.

Lemma 4.8 (Easy Direction). *For any undirected graph $G = (V, E)$ and any probability distribution π on V,*

$$\nu^{(1)}(G) \leq 2\psi_B(G).$$

For the hard direction, we will work with $\overrightarrow{\nu}^{(1)}(G)$ instead. There is no need to do the truncation step as in Lemma 3.17, as there is no constraint about $\pi(S)$ of the output set S. The main step is to define a modified bipartite vertex expansion condition for directed graphs and use it for the analysis of the threshold rounding.

Let S_1, S_2 be two disjoint subsets of V. In the edge conductance setting, rephrasing using our terminology, Trevisan [Tre09] defined the "bipartite edge boundary" $\delta(S_1, S_2)$ as $E(S_1) \cup E(S_2) \cup \delta(S_1 \cup S_2)$ where $E(S_i)$ is the set of induced edges in S_i for $i \in \{1, 2\}$, and the "bipartite edge conductance" $\phi(S_1, S_2)$ as $|\delta(S_1, S_2)|/\operatorname{vol}(S_1 \cup S_2)$. We define the appropriate bipartite vertex boundary $\overrightarrow{\partial}(S_1, S_2)$ for vertex expansion and for directed graphs in the following definition. As in Definition 3.18, note that $\overrightarrow{\partial}(S_1, S_2)$ could contain vertices in $V - (S_1 \cup S_2)$. Again, a good intuition is to think of $\overrightarrow{\partial}(S_1, S_2)$ as a vertex cover of the edges in the bipartite edge boundary $\delta(S_1, S_2)$ in the undirected sense.

Definition 4.9 (Directed Bipartite Vertex Boundary and Expansion)**.** *Let $G = (V, \overrightarrow{E})$ be a directed graph. Let S_1, S_2 be two disjoint subsets of V. The directed bipartite vertex boundary of (S_1, S_2) is defined as*

$$
\begin{aligned}
\overrightarrow{\partial}(S_1, S_2) \quad &:= \quad \left\{ v \in S_1 \mid \exists u \in S_1 \text{ with } uv \in \overrightarrow{E}, \text{ or } \exists u \notin S_1 \cup S_2 \text{ with } uv \in \overrightarrow{E} \right\} \cup \\
&\qquad \left\{ v \in S_2 \mid \exists u \in S_2 \text{ with } uv \in \overrightarrow{E}, \text{ or } \exists u \notin S_1 \cup S_2 \text{ with } uv \in \overrightarrow{E} \right\} \cup \\
&\qquad \left\{ v \notin S_1 \cup S_2 \mid \exists u \in S_1 \cup S_2 \text{ with } uv \in \overrightarrow{E} \right\},
\end{aligned}
$$

and the directed bipartite vertex expansion as

$$
\overrightarrow{\psi}(S_1, S_2) := \frac{\pi\left(\overrightarrow{\partial}(S_1, S_2)\right)}{\pi(S_1 \cup S_2)}.
$$

An example of directed bipartite vertex expansion is provided in Figure 4.1.
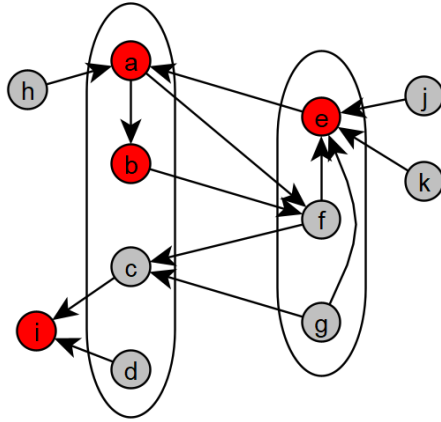

Figure 4.1: In the graph shown, the bipartition is $S_1 = \{a, b, c, d\}$ and $S_2 = \{e, f, g\}$. The vertices a, b, e, i, colored in red, are the vertices in $\overrightarrow{\partial}(S_1, S_2)$.

We prove that the threshold rounding defined in [Tre09], when applied on $\overrightarrow{\nu}^{(1)}(G)$, will give a set with small directed bipartite vertex expansion.

Proposition 4.10 (Threshold Rounding for $\nu(G)$)**.** *Let $G = (V, E)$ be an undirected graph and π be a probability distribution on V. Given a solution x and $\overrightarrow{E}$ to $\overrightarrow{\nu}^{(1)}(G)$, there is a polynomial time algorithm to find two disjoint subsets $S_1, S_2 \subseteq V$ with $\overrightarrow{\psi}(S_1, S_2) \lesssim \sqrt{\overrightarrow{\nu}^{(1)}(G)}$.*

Proof. For any $t > 0$, define $S_t := \{v \in V \mid x(v) > \sqrt{t}\}$ and $S_{-t} := \{v \in V \mid x(v) < -\sqrt{t}\}$ as in [Tre09]. By a standard averaging argument,

$$
\min_t \overrightarrow{\psi}(S_t, S_{-t}) \leq \frac{\int_0^\infty \pi\left(\overrightarrow{\partial}(S_t, S_{-t})\right) dt}{\int_0^\infty \pi(S_t \cup S_{-t}) dt}.
$$

26

The denominator is

$$\int_0^\infty \pi(S_t \cup S_{-t})\,dt = \sum_{v \in V} \pi(v) \int_0^\infty \mathbb{1}[v \in S_t \cup S_{-t}]\,dt = \sum_{v \in V} \pi(v)x(v)^2 = 1.$$

For the numerator, we consider when a vertex v is in $\overrightarrow{\partial}(S_t, S_{-t})$. Assume without loss that $x(v) \geq 0$; the other case is symmetric. There are two scenarios where $v \in \overrightarrow{\partial}(S_t, S_{-t})$:

1. The first scenario is when $uv \in E[S_t]$ in the undirected sense for some directed edge $u \to v$. For a fixed edge $u \to v$, this happens when $t < \max\{0, \min\{x(u), x(v)\}\}^2 \leq ((x(u) + x(v))/2)^2$, where the last inequality can be verified by considering the cases $x(v) \geq 0$ and $x(v) \leq 0$ separately. Therefore, the first scenario happens when

$$t < \max_{u:u \to v} \max\{0, \min\{x(u), x(v)\}\}^2 \leq \max_{u:u \to v} \left(\frac{x(u) + x(v)}{2}\right)^2.$$

2. The second scenario is when $uv \in \delta(S_t \cup S_{-t})$ in the undirected sense for some directed edge $u \to v$. For a fixed edge $u \to v$, this happens when $x(u)^2 \leq t < x(v)^2$ (so $u \notin S_t \cup S_{-t}$ and $v \in S_t$), or when $x(v)^2 \leq t < x(u)^2$ (so $u \in S_t$ and $v \notin S_t \cup S_{-t}$). Therefore, the second scenario happens when

$$\min_{u:u \to v} x(u)^2 \leq t < x(v)^2 \qquad \text{or} \qquad x(v)^2 \leq t < \max_{u:u \to v} x(u)^2.$$

Hence the numerator is

$$\int_0^\infty \pi\big(\overrightarrow{\partial}(S_t, S_{-t})\big)\,dt$$

$$= \sum_{v \in V} \pi(v) \cdot \int_0^\infty \mathbb{1}\Big[v \in \overrightarrow{\partial}(S_t, S_{-t})\Big]\,dt$$

$$\leq \sum_{v \in V} \pi(v) \left[\max_{u:u \to v}\left(\frac{x(u)+x(v)}{2}\right)^2 + \left(x(v)^2 - \min_{u:u \to v} x(u)^2\right) + \left(\max_{u:u \to v} x(u)^2 - x(v)^2\right)\right]$$

$$= \frac{1}{4}\overrightarrow{\nu}^{(1)}(G) + \sum_{v \in V} \pi(v)\left[\max_{\substack{u:u \to v \\ |x(u)|<|x(v)|}}\big\{x(v)^2 - x(u)^2\big\} + \max_{\substack{u:u \to v \\ |x(u)|>|x(v)|}}\big\{x(u)^2 - x(v)^2\big\}\right]$$

$$\leq \frac{9}{4}\overrightarrow{\nu}^{(1)}(G) + 2\sum_{v \in V} \pi(v)\left[\max_{\substack{u:u \to v \\ |x(u)|<|x(v)|}}|x(u)|\cdot|x(u)+x(v)| + \max_{\substack{u:u \to v \\ |x(u)|>|x(v)|}}|x(v)|\cdot|x(u)+x(v)|\right]$$

$$\leq \frac{9}{4}\overrightarrow{\nu}^{(1)}(G) + 2\sum_{v \in V} \pi(v)|x(v)|\max_{u:u \to v}|x(u)+x(v)|$$

$$\leq \frac{9}{4}\overrightarrow{\nu}^{(1)}(G) + 2\sqrt{\sum_{v \in V} \pi(v)x(v)^2 \cdot \sum_{v \in V} \pi(v)\max_{u:u \to v}\big(x(u)+x(v)\big)^2}$$

$$\lesssim \overrightarrow{\nu}^{(1)}(G) + \sqrt{\overrightarrow{\nu}^{(1)}(G)}.$$

We explain these steps one by one. The first inequality is by the two scenarios explained in detail above. The second equality uses the definition that $\overrightarrow{\nu}^{(1)}(G) = \sum_{v \in V} \pi(v) \max_{u:u \to v}(x(u) + x(v))^2$. For the second inequality, in the first max we write $x(v)^2 - x(u)^2 = (x(u) + x(v))^2 - 2x(u)(x(v) + x(u)) \leq (x(u) + x(v))^2 + 2|x(u)||x(v) + x(u)|$, and then take out $(x(u) + x(v))^2$ from the summation by using again the definition that $\overrightarrow{\nu}^{(1)}(G) = \sum_{v \in V} \pi(v) \max_{u:u \to v}(x(u) + x(v))^2$, while the second max is handled similarly. In the third inequality we replace $|x(u)|$ in the first max term by $|x(v)|$. The fourth inequality is by an application of the Cauchy-Schwarz inequality. The final inequality is by $\sum_{v \in V} \pi(v)x(v)^2 = 1$ in the constraint of $\overrightarrow{\nu}^{(1)}(G)$.

We showed in the easy direction in Lemma 4.8 that $\overrightarrow{\nu}^{(1)}(G) \leq 2$ and thus $\overrightarrow{\nu}^{(1)}(G) \lesssim \sqrt{\overrightarrow{\nu}^{(1)}(G)}$. We conclude that there exists (S_t, S_{-t}) with $\overrightarrow{\psi}(S_t, S_{-t}) \lesssim \sqrt{\overrightarrow{\nu}^{(1)}(G)}$. □

Finally, as in Lemma 3.20, given (S_1, S_2) with small directed bipartite vertex expansion, we show how to extract an induced bipartite graph with small vertex expansion.

Lemma 4.11 (Postprocessing for Bipartite Vertex Expansion). *Let $G = (V, \overrightarrow{E})$ be a directed graph. Given two disjoint subsets $S_1, S_2 \subseteq V$ with $\overrightarrow{\psi}(S_1, S_2) < 1/2$, there are $S_1' \subseteq S_1$ and $S_2' \subseteq S_2$ with $\psi(S_1', S_2') \leq 2\overrightarrow{\psi}(S_1, S_2)$ and $S_1' \cup S_2'$ is an induced bipartite graph in the underlying undirected graph of G.*

Proof. From Definition 4.9, the observation is that $\overrightarrow{\partial}(S_1, S_2)$ is a vertex cover of $E(S_1) \cup E(S_2) \cup \delta(S_1 \cup S_2)$. So, by setting $S_1' := S_1 - \overrightarrow{\partial}(S_1, S_2)$ and $S_2' := S_2 - \overrightarrow{\partial}(S_1, S_2)$, then (S_1', S_2') is an induced bipartite graph as there could be no edges induced in S_1' and no edges induced in S_2'. Also, $\partial(S_1' \cup S_2') \subseteq \overrightarrow{\partial}(S_1, S_2)$, as there could be no edges between $S_1' \cup S_2'$ and $V - (S_1 \cup S_2 \cup \overrightarrow{\partial}(S_1, S_2))$. Therefore,

$$\pi(\partial(S_1' \cup S_2')) \leq \pi\big(\overrightarrow{\partial}(S_1, S_2)\big) = \overrightarrow{\psi}(S_1, S_2) \cdot \pi(S_1 \cup S_2) \leq 2\overrightarrow{\psi}(S_1, S_2) \cdot \pi(S_1' \cup S_2'),$$

where the last inequality uses the assumption that $\overrightarrow{\psi}(S_1, S_2) = \pi\big(\overrightarrow{\partial}(S_1, S_2)\big)/\pi(S_1 \cup S_2) < 1/2$ and so $\pi(S_1' \cup S_2') \geq \pi(S_1 \cup S_2) - \pi\big(\overrightarrow{\partial}(S_1, S_2)\big) \geq \pi(S_1 \cup S_2)/2$. We thus conclude that $\psi(S_1', S_2') \leq 2\overrightarrow{\psi}(S_1, S_2)$. □

We complete the proof of Theorem 4.4 in Appendix B.

5 Higher-Order Cheeger Inequality for Vertex Expansion

The goal of this section is to prove Theorem 1.10. There are four main steps in the proof.

The first step is to reformulate the problem as a semidefinite program using the maximum reweighted sum of the k smallest eigenvalues $\sigma_k^*(G)$. Using von Neumann's minimax theorem, we construct the dual program of $\sigma_k^*(G)$ and see that it satisfies the so-called sub-isotropy condition. The main focus in this section will then be to relate $\sigma_k^*(G)$ and $\psi_k(G)$, rather than to relate $\lambda_k^*(G)$ and $\psi_k(G)$ directly.

The second step is to project the dual solution to $\sigma_k^*(G)$ into a low-dimensional solution. In this step, we use a similar apporach as in Section 3.2, by introducing an intermediate directed dual program and then using the Gaussian projection method. Also, we use a theorem in [LOT12] that proves that Gaussian projection approximately preserves the sub-isotropy condition.

The third step is to partition the low-dimensional solution into k disjointly supported functions each with small objective value. In this step, we closely follow the techniques in [LOT12], such as radial projection distance, smooth localization and random parititoning. We will review the techniques in [LOT12] in Section 5.3 before presenting our proofs.

The last step is to apply the Cheeger inequality for vertex expansion in Section 3 on these k functions to find disjoint sets with small vertex expansion. We will prove the easy direction and put together the steps to prove Theorem 1.10 in Section 5.4.

5.1   Primal and Dual Programs

As mentioned in Section 1.3, the maximum reweighted k-th smallest eigenvalue $\lambda_k^*(G)$ as formulated in Definition 1.9 is not a convex program. Instead, we will study the following related quantity.

Definition 5.1 (Maximum Reweighted Sum of k Smallest Eigenvalues)**.** *Given an undirected graph $G = (V, E)$ and a probability distribution π on V, the maximum reweighted sum of k smallest eigenvalues of the normalized Laplacian matrix of G is defined as $\sigma_k^*(G) := \max_{P \geq 0} \sum_{i=1}^{k} \lambda_k(I - P)$, where P is subject to the same constraints stated in Definition 1.1. Note that*

$$\lambda_k^*(G) \leq \sigma_k^*(G) \leq k \cdot \lambda_k^*(G).$$

We reformulate the primal program in Definition 5.1 as the semidefinite program in Proposition 5.2. The proof of Proposition 5.2 has a few small steps. First we rewrite the sum $\sum_{i=1}^{k} \lambda_k(I - P)$ as $\sum_{i=1}^{k} \lambda_k(I - \mathcal{Q})$ for a symmetric matrix $\mathcal{Q}$. Then we use Proposition 2.4 to write $\sum_{i=1}^{k} \lambda_k(I - \mathcal{Q})$ as a minimization problem using semidefinite programming. Next we apply von Neumann's minimax theorem to change the order of max-min to min-max. Then we do a change of variable and rewrite the program into a vector program form. Finally, we use linear programming duality to rewrite the inner maximization problem as a minimization problem as was done in [Roc05].

Proposition 5.2 (Dual Program for $\sigma_k^*(G)$)**.** *For any undirected graph $G = (V, E)$ with a self loop at each vertex and any probability distribution π on V, the following semidefinite program is dual to the primal program in Definition 5.1 with strong duality $\sigma_k^*(G) = \kappa(G)$ where*

$$\kappa(G) := \min_{f:V \to \mathbb{R}^n, \ g:V \to \mathbb{R}_{\geq 0}} \quad \sum_{v \in V} \pi(v) g(v)$$
$$\text{subject to} \quad g(u) + g(v) \geq \|f(u) - f(v)\|^2 \quad \forall uv \in E$$
$$\sum_{v \in V} \pi(v) f(v) f(v)^T \preccurlyeq I_n$$
$$\sum_{v \in V} \pi(v) \|f(v)\|^2 = k.$$

Proof. By Definition 5.1, $\sigma_k^*(G) = \max_{P \geq 0} \sum_{i=1}^k \lambda_k(I - P)$, where the maximum is over all P satisfying the constraints in Definition 1.1. Consider the sum of eigenvalues for a fixed P that satisfies the constraints. The time reversible constraint $\pi(u)P(u,v) = \pi(v)P(v,u)$ for all $uv \in E$ is equivalent to the matrix $Q := \Pi P$ being symmetric where $\Pi := \text{diag}(\pi)$. Let $\mathcal{Q} := \Pi^{-1/2}Q\Pi^{-1/2}$ be the normalized adjacency matrix of Q. Note that P and $\mathcal{Q}$ have the same spectrum, as $\mathcal{Q} = \Pi^{-1/2}Q\Pi^{-1/2} = \Pi^{1/2}P\Pi^{-1/2}$. Therefore $\sum_{i=1}^k \lambda_i(I - P) = \sum_{i=1}^k \lambda_i(I - \mathcal{Q})$ where $I - \mathcal{Q}$ is a symmetric matrix.

By Proposition 2.4, the sum of the k smallest eigenvalue of the symmetric matrix $I - \mathcal{Q}$ can be written as the following semidefinite program:

$$\sum_{i=1}^k \lambda_i(I - \mathcal{Q}) = \min_{Y \in \mathbb{R}^{n \times n}} \quad \text{tr}\left(Y \cdot (I - \mathcal{Q})\right)$$
$$\text{subject to} \quad 0 \preccurlyeq Y \preccurlyeq I$$
$$\text{tr}(Y) = k.$$

Note that $I - \mathcal{Q}$ is the normalized Laplacian matrix of Q. We consider the change of variable $Y = \Pi^{1/2}Z\Pi^{1/2}$, so as to rewrite the objective function in terms of $\Pi - Q$ which is the Laplacian matirx of Q:

$$\sum_{i=1}^k \lambda_i(I - \mathcal{Q}) = \min_{Z \in \mathbb{R}^{n \times n}} \quad \text{tr}\left(Z \cdot (\Pi - Q)\right)$$
$$\text{subject to} \quad 0 \preccurlyeq \Pi^{\frac{1}{2}}Z\Pi^{\frac{1}{2}} \preccurlyeq I$$
$$\text{tr}(\Pi^{\frac{1}{2}}Z\Pi^{\frac{1}{2}}) = k.$$

Therefore, the primal program for $\sigma_k^*(G)$ can be rewritten in terms of Q as follows:

$$\sigma_k^*(G) = \max_{Q \geq 0} \min_{Z \in \mathbb{R}^{n \times n}} \quad \text{tr}\left(Z \cdot (\Pi - Q)\right)$$
$$\text{subject to} \quad Q(u,v) = 0 \qquad\qquad \forall uv \notin E$$
$$\sum_{v \in V} Q(u,v) = \pi(u) \qquad\qquad \forall u \in V$$
$$Q(u,v) = Q(v,u) \qquad\qquad \forall uv \in E$$
$$0 \preccurlyeq \Pi^{\frac{1}{2}}Z\Pi^{\frac{1}{2}} \preccurlyeq I$$
$$\text{tr}\left(\Pi^{\frac{1}{2}}Z\Pi^{\frac{1}{2}}\right) = k.$$

Now we write the dual program by using von Neumann's minimax theorem in Theorem 2.3 to switch the max-min to min-max in the objective function. Note that von Neumann's theorem can be applied because the objective function is multiliner in Z and Q (hence concave in Q and convex in Z), the feasible region of Q is compact and convex as it is bounded and defined by linear constraints, and the feasible region of Z is compact and convex as it is bounded and defined by PSD

and trace constraints. Hence, we can switch the order of $\max_Q \min_Z$ to obtain the dual program by rewriting the objective function as

$$\min_{Z \in \mathbb{R}^{n \times n}} \max_{Q \geq 0} \ \operatorname{tr} \left(Z \cdot (\Pi - Q) \right).$$

Next we rewrite this dual program into a vector program form. As $Z \succcurlyeq 0$, we can write $Z = FF^T$ where F is an $n \times n$ matrix. We denote the i-th column of F by $f_i \in \mathbb{R}^n$ for $1 \leq i \leq n$ and think of it as an eigenvector, denote the v-th row of F by $f(v) \in \mathbb{R}^n$ and think of it as the spectral embedding of a vertex v, and denote the (v, i)-th entry of F by $f_i(v)$ for $1 \leq i \leq n$ and $v \in V$. As $\Pi - Q$ is the Laplacian matrix of Q, the quadratic form for a vector $x \in \mathbb{R}^n$ is $x^T(\Pi - Q)x = \sum_{uv \in E}(x(u) - x(v))^2 \cdot Q(u, v)$, and thus the objective function can be rewritten as

$$\operatorname{tr}(F^T(\Pi - Q)F) = \sum_{i=1}^n f_i^T(\Pi - Q)f_i = \sum_{i=1}^n \sum_{uv \in E}(f_i(u) - f_i(v))^2 \cdot Q(u, v) = \sum_{uv \in E} \|f(u) - f(v)\|^2 \cdot Q(u, v).$$

Note that $\Pi^{1/2} Z \Pi^{1/2} = (\Pi^{1/2} F)(F^T \Pi^{1/2})$ and $F^T \Pi F = (F^T \Pi^{1/2})(\Pi^{1/2} F)$ have the same spectrum by Fact 2.2. So the first constraint can be rewritten as

$$0 \preccurlyeq F^T \Pi F = \sum_{v \in V} \pi(v) f(v) f(v)^T \preccurlyeq I,$$

and the second constraint can be rewritten as

$$\operatorname{tr}(F^T \Pi F) = \operatorname{tr} \left(\sum_{v \in V} \pi(v) f(v) f(v)^T \right) = \sum_{v \in V} \pi(v) \|f(v)\|^2 = k.$$

Therefore, the dual program for $\sigma_k^*(G)$ can be rewritten as follows:

$$
\begin{aligned}
\kappa(G) \ := \ \min_{f: V \to \mathbb{R}^n} \max_{Q \geq 0} \quad & \sum_{uv \in E} \|f(u) - f(v)\|^2 \cdot Q(u, v) \\
\text{subject to} \quad & Q(u, v) = 0 && \forall uv \notin E \\
& \sum_{v \in V} Q(u, v) = \pi(u) && \forall u \in V \\
& Q(u, v) = Q(v, u) && \forall uv \in E \\
& \sum_{v \in V} \pi(v) f(v) f(v)^T \preccurlyeq I \\
& \sum_{v \in V} \pi(v) \|f(v)\|^2 = k.
\end{aligned}
$$

Finally, as in [Roc05], note that the inner maximization problem is just a linear program in Q. For a fixed embedding $f: V \to \mathbb{R}^n$, we can use the linear programming duality theorem to rewrite the inner maximization problem into the following minimization problem:

$$
\begin{aligned}
\min_{g: V \to \mathbb{R}_{\geq 0}} \quad & \sum_{v \in V} \pi(v) g(v) \\
\text{subject to} \quad & g(u) + g(v) \geq \|f(u) - f(v)\|^2 && \forall uv \in E,
\end{aligned}
$$

where $g(u)$ is a dual variable for the constraint $\sum_{v \in V} Q(u,v) = \pi(u)$. Recall that we assumed the graph has a self-loop $Q(v,v)$ at each vertex v so that the primal program is always feasible, and the primal variable $Q(v,v)$ gives the dual constraint $g(v) \geq 0$.

To summarize, we rewrite the max-min optimization problem in the primal program into a min-min optimization problem using von Neumann's minimax theorem and linear programming duality. The resulting program in the statement is a semidefinite program in the vector program form. $\square$

5.2   Gaussian Projection

The second step is to project a solution to $\kappa(G)$ in Proposition 5.2 into a low-dimensional solution and prove that several properties are approximately preserved. The projection algorithm is a high dimensional version of the simple Guassian projection algorithm in Section 3.2.2.

Definition 5.3 (Gaussian Projection). *Let $f : V \to \mathbb{R}^n$ be an embedding where each vertex v is mapped to a vector $f(v) \in \mathbb{R}^n$. Given an integer $h \leq n$, let Γ be an $h \times n$ matrix where each entry $\Gamma_{i,j}$ for $1 \leq i \leq h$ and $1 \leq j \leq n$ is an independent standard Gaussian random variable $N(0,1)$. The Gaussian projection $\bar{f} : V \to \mathbb{R}^h$ of f is an embedding of each vertex $v \in V$ to an h-dimensional vector defined as*

$$\bar{f}(v) = \frac{1}{\sqrt{h}} \cdot \Gamma\big(f(v)\big).$$

As in Section 3.2.2, we consider a related directed program $\overrightarrow{\kappa}(G)$ for the analysis of the Gaussian projection algorithm. The proof of the following lemma is the same as in Lemma 3.12 and Lemma 3.13 and is omitted.

Lemma 5.4 (Directed Dual Program Using Orientation for $\kappa(G)$). *Let $G = (V, E)$ be an undirected graph and π be a probability distribution on V. Let $\overrightarrow{E}$ be an orientation of the undirected edges in E. Define*

$$\overrightarrow{\kappa}(G) := \min_{f:V \to \mathbb{R}^n} \min_{\overrightarrow{E}} \quad \sum_{v \in V} \pi(v) \max_{u:uv \in \overrightarrow{E}} \|f(u) - f(v)\|^2$$

$$\text{subject to} \quad \sum_{v \in V} \pi(v) f(v) f(v)^T \preceq I_n$$

$$\sum_{v \in V} \pi(v) \|f(v)\|^2 = k.$$

Then $\kappa(G) \leq \overrightarrow{\kappa}(G) \leq 2\kappa(G)$.

To prove that several properties of f to $\overrightarrow{\kappa}(G)$ are preserved in $\bar{f}$, we define the following quantities. The first quantity is the objective value of $\overrightarrow{\kappa}(G)$, which is called the energy of the function f.

Definition 5.5 (Energy). *Given a directed graph $G = (V, \overrightarrow{E})$ and a probability distribution π on V, the energy of an embedding $f : V \to \mathbb{R}^h$ is defined as*

$$\mathcal{E}(f) = \sum_{v \in V} \pi(v) \max_{u:u \to v} \|f(u) - f(v)\|^2.$$

32

The second quantity is the LHS of the last constraint, which is called the mass of the function f.

Definition 5.6 (Mass). *Given an embedding $f : V \to \mathbb{R}^h$, the mass of f is defined as*

$$\mu(f) = \sum_{v \in V} \pi(v) \|f(v)\|^2 .$$

The final quantity is related to the constraint $\sum_{v \in V} \pi(v) f(v) f(v)^T \preccurlyeq I_n$, which is called the sub-isotropy condition for the vectors $\left\{ \sqrt{\pi(v)} f(v) \right\}_{v \in V}$. In [LOT12], the sub-isotropy condition is used to establish the following spreading property, which is used crucially in the spectral partitioning algorithm for k-way edge conductance.

Definition 5.7 (Spreading Property [LOT12]). *Let π be a probability distribution on V. For two parameters $\Delta \in [0,1]$ and $\eta \in [0,1]$, an embedding $f : V \to \mathbb{R}^h$ is called (Δ, η)-spreading if for every subset $S \subseteq V$,*

$$\mathrm{diam}_{d_f}(S) \leq \Delta \quad \implies \quad \sum_{v \in S} \pi(v) \|f(v)\|^2 \leq \eta \cdot \sum_{v \in V} \pi(v) \|f(v)\|^2 ,$$

where $\mathrm{diam}_{d_f}(S) := \max_{u,v \in S} d_f(u,v)$ is the diameter of the set S under the radial projection distance function d_f to be defined in Definition 5.12.

As we will state formally in Lemma 5.13 in the next subsection, any feasible solution f to $\vec{\kappa}(G)$ is $(\Delta, \frac{1}{k(1-\Delta^2)})$-spreading, so we can regard $\eta = \frac{1}{k(1-\Delta^2)}$ in the following. The precise parameters and also the definition of the radial projection distance are not important in this subsection.

The goal of this subsection is to prove that the energy, the mass, and the spreading property of $f : V \to \mathbb{R}^n$ are approximately preserved in the projection $\bar{f} : V \to \mathbb{R}^h$ for a small h. In [LOT12], it was already proved that the mass and the spreading property of f are approximately preserved in $\bar{f}$.

Lemma 5.8 ([LOT12], Lemma 4.3). *Let π be a probability distribution on V. Let $f : V \to \mathbb{R}^n$ be an embedding that is (Δ, η)-spreading. Let $\bar{f} : V \to \mathbb{R}^h$ be a Gaussian projection of f as defined in Definition 5.3. For some value [2]*

$$h \lesssim \frac{1}{\Delta^2} \left(\log \left(\frac{1}{\eta \Delta} \right) \right),$$

with probability at least $1/2$, the following two properties hold simultaneously:

$$\mu(\bar{f}) \geq \frac{1}{2} \mu(f) \quad \text{and} \quad \bar{f} \text{ is } \left(\frac{\Delta}{4}, \left(1 + \Delta \right) \eta \right) \text{-spreading}.$$

We prove that the energy is also approximately preserved. We use the following proposition whose proof is deferred to Appendix C.

---

[2] Lemma 4.3 in [LOT12] was stated slightly differently. Their assumptions are that $f : V \to \mathbb{R}^k$ and $\eta \geq 1/k$, and their conclusion is that $h \lesssim \frac{1}{\Delta^2} \log(\frac{k}{\Delta})$. We note that the dependency on k in their conclusion is based on the substitution $\eta = 1/k$ in the bound on h we stated, which has no dependency on the ambient dimension n. Their proof, without the substitution $\eta = 1/k$, gives the bound we stated.

Proposition 5.9 (Expected Maximum of χ-Squared Distribution). *Let (Γ_{ij}) for $1 \leq i \leq d$ and $1 \leq j \leq m$ be Gaussian random variables with mean 0 and variance at most 1, and such that $\Gamma_{i1}, \Gamma_{i2}, \ldots, \Gamma_{im}$ are mutually independent for each $i \in [d]$. Let $Y_i := \frac{1}{m} \sum_{1 \leq j \leq m} \Gamma_{ij}^2$ and let $Y := \max_{1 \leq i \leq d} Y_i$. Then,*

$$\mathbb{E}[Y] \leq 4 \left(1 + \frac{1 + \log d}{m} \right).$$

The main result in this subsection is the following lemma which compares the energy, the mass, and the spreading property of f and of its Gaussian projection $\bar{f}$. Note that the second and the third items of the following lemma are directly from Lemma 5.8.

Lemma 5.10 (Dimension Reduction). *Let $G = (V, \overrightarrow{E})$ be a directed graph with maximum indegree d and π be a probability distribution on V. Let $f : V \rightarrow \mathbb{R}^n$ be an embedding that is (Δ, η)-spreading. Let $\bar{f} : V \rightarrow \mathbb{R}^h$ be a Gaussian projection of f as defined in Definition 5.3. By setting $h \lesssim \frac{1}{\Delta^2} \log(\frac{1}{\eta\Delta})$, with probability at least $1/4$, the following three properties hold simultaneously:*

$$\mathcal{E}(\bar{f}) \lesssim \left(1 + \frac{\log d}{h} \right) \cdot \mathcal{E}(f) \quad \text{and} \quad \mu(\bar{f}) \geq \frac{1}{2}\mu(f) \quad \text{and} \quad \bar{f} \text{ is } \left(\frac{\Delta}{4}, (1 + \Delta)\eta \right)-\text{spreading}.$$

Proof. The second and the third item are from Lemma 5.8. We will prove the first item holds with probability at least $3/4$ by using Proposition 5.9, and this would imply the lemma by union bound. Let Γ_i be the i-th row of Γ in Definition 5.3. For $uv \in \overrightarrow{E}$,

$$\left\| \bar{f}(u) - \bar{f}(v) \right\|^2 = \sum_{i=1}^{h} \left(\frac{1}{\sqrt{h}} \left\langle \Gamma_i, f(u) - f(v) \right\rangle \right)^2 = \frac{1}{h} \sum_{i=1}^{h} g_i^2,$$

where $g_i = \langle \Gamma_i, f(u) - f(v) \rangle$ is an independent Gaussian random variable with mean zero and variance $\| f(u) - f(v) \|^2$. Applying Proposition 5.9 on each $v \in V$ with indegree at most d,

$$\mathbb{E}\left[\max_{u:u \rightarrow v} \left\| \bar{f}(u) - \bar{f}(v) \right\|^2 \right] \leq 4 \left(1 + \frac{1 + \log d}{h} \right) \cdot \max_{u:u \rightarrow v} \| f(u) - f(v) \|^2.$$

By linearity of expectation and Markov's inequality,

$$\Pr\left[\sum_{u \in V} \pi(u) \max_{u:u \rightarrow v} \left\| \bar{f}(u) - \bar{f}(v) \right\|^2 \leq 16 \left(1 + \frac{1 + \log d}{h} \right) \cdot \sum_{u \in V} \pi(u) \max_{u:u \rightarrow v} \| f(u) - f(v) \|^2 \right] \geq \frac{3}{4},$$

implying that $\mathcal{E}(\bar{f}) \lesssim (1 + \frac{\log d}{h}) \cdot \mathcal{E}(f)$ with probability at least $3/4$. $\quad\square$

5.3 Spectral Partitioning

The third step is to show that given $\bar{f} : V \rightarrow \mathbb{R}^h$ in Lemma 5.10, we can construct l disjointly supported functions $\bar{f}_1, \ldots, \bar{f}_l : V \rightarrow \mathbb{R}^h$ with comparable energy and mass to that of $\bar{f}$.

Lemma 5.11 (Spectral Partitioning). *Let $G = (V, E)$ be an undirected graph and π be a probability distribution on V. Let $\overrightarrow{E}$ be an orientation of E and $\bar{f} : V \to \mathbb{R}^h$ be an embedding. Let l be the targeted number of disjointly supported functions where $1 \le l \le k$. Suppose that there exist $\Delta \in (0, 1)$ and $\delta \in (0, 1)$ such that*

$$\bar{f} \text{ is } \left(\frac{\Delta}{4}, \frac{1}{k(1 - \Delta)} \right) - \text{spreading} \qquad \text{and} \qquad \Delta \le 1 - \frac{2(l - 1)}{2(1 - \delta)k - 1}.$$

Then there exist embeddings $\bar{f}_1, \ldots, \bar{f}_l : V \to \mathbb{R}^h$ such that the supports of $\{\bar{f}_i\}_{i=1}^l$ are pairwise disjoint and

$$\mathcal{E}(\bar{f}_i) \lesssim \left(1 + \frac{h}{\Delta\delta} \right)^2 \cdot \mathcal{E}(\bar{f}) \qquad \text{and} \qquad \mu(\bar{f}_i) \gtrsim \frac{1}{k} \mu(\bar{f}).$$

The proof of this step follows closely the proof in [LOT12], so we will first review the ideas and results in [LOT12] before presenting the proof of Lemma 5.11.

5.3.1   Review of [LOT12]

In [LOT12], given the first k eigenvectors $f_1, \ldots, f_k : V \to \mathbb{R}$ of the normalized Laplacian matrix, the spectral embedding $f : V \to \mathbb{R}^k$ is defined for each vertex $v \in V$ as $f(v) = \big(f_1(v), \ldots, f_k(v) \big)$. Since the eigenvectors are orthonormal, the spectral embedding satisfies the isotropy condition $\sum_{v \in V} f(v)f(v)^T = I_n$. Lee, Oveis Gharan and Trevisan observed that the isotropy condition implies that not many points can be close in the radial projection distance defined below.

Definition 5.12 (Radial Projection Distance [LOT12]). *Let $G = (V, E)$ be a graph and $f : V \to \mathbb{R}^h$ be an embedding of the vertices. For each pair of vertices $u, v \in V$, the radial projection distance between u and v is defined as*

$$d_f(u, v) = \left\| \frac{f(u)}{\|f(u)\|} - \frac{f(v)}{\|f(v)\|} \right\|$$

if $\|f(u)\| > 0$ and $\|f(v)\| > 0$. Otherwise, if $f(u) = f(v) = 0$ then $d_f(u, v) := 0$, else $d_f(u, v) = \infty$.

More precisely, they proved the following bound on the spreading property in Definition 5.7 of the embedding f. Their result is stated for an embedding f satisfying the isotropy condition, but the same proof works for an embedding f satisfying the sub-isotropy condition in Proposition 5.2.

Lemma 5.13 (Sub-Isotropy Implies Spreading [LOT12](Lemma 3.2)). *Let $G = (V, E)$ be an undirected graph and π be a probability distribution on V. Suppose $f : V \to \mathbb{R}^h$ is an embedding with mass $\mu(f)$ in Definition 5.6. Then, for any $\Delta \in [0, 1)$,*

$$\sum_{u \in V} \pi(u)f(u)f(u)^T \preccurlyeq I_h \qquad \Longrightarrow \qquad f \text{ is } \left(\Delta, \frac{1}{\mu(f) \cdot (1 - \Delta^2)} \right) - \text{spreading}.$$

As the embedding is spreading, any subset of points with small diameter in radial projection distance cannot have too much mass. In order to construct many disjointly supported embeddings $f_1, \ldots, f_l : V \to \mathbb{R}^h$, the points in $\mathbb{R}^h$ are partitioned into many groups of small diameter using the following definition and theorem from metric geometry.

35

Definition 5.14 (Padded Decomposition [LOT12]). *Let (X, d_X) be a finite metric space. For $\Delta, \alpha, \delta > 0$, a random partition $\mathcal{P}$ of X is called (Δ, α, δ)-padded if*

- *each part in $\mathcal{P}$ has diameter at most Δ with respect to the distance function d_X;*

- $\Pr\left[B\left(x, \frac{\Delta}{\alpha}\right) \subseteq \mathcal{P}(x)\right] \geq \delta$ *for every $x \in X$, where $B\left(x, \frac{\Delta}{\alpha}\right)$ is the open ball of radius $\frac{\Delta}{\alpha}$ centered at x and $\mathcal{P}(x)$ is the part in $\mathcal{P}$ that contains x.*

Theorem 5.15 ([LOT12] Theorem 2.3, [GKL03]). *Let (X, d) be a finite metric space. If $X \subseteq \mathbb{R}^h$, then for every $\Delta > 0$ and $\delta \in (0, 1)$, X admits a $\left(\Delta, O(\frac{h}{\delta}), 1 - \delta\right)$-padded random partition.*

Let $\mathcal{P} = P_1 \sqcup P_2 \sqcup \ldots \sqcup P_m$ be a random partition sampled from Theorem 5.15. By the second property in Definition 5.14, there is only a small fraction of points close to the boundary of the partition. The points close to the boundary are removed to form $P_1' \sqcup P_2' \sqcup \ldots \sqcup P_m'$, so that the distance between each pair P_i' and P_j' is lower bounded by say 2ϵ. As each P_i' does not have too much mass, they can be grouped into disjoint sets $S_1, \ldots, S_k$ where each has mass at least $\frac{1}{2}$. Then the disjoint supported functions $f_1, \ldots, f_k$ are constructed on $S_1, \ldots, S_k$ by the following smooth localization procedure.

Lemma 5.16 (Smooth Localization [LOT12](Lemma 3.3)). *Let $G = (V, E)$ be an undirected graph and $f : V \to \mathbb{R}^h$ be an embedding. For any $S \subseteq V$ and any $\epsilon > 0$, there is a mapping $f' : V \to \mathbb{R}^h$ which satsifies the following three properties:*

1. $f'(v) = f(v)$ *for all $v \in S$,*

2. $\text{supp}(f') \subseteq N_\epsilon(S)$ *where $N_\epsilon(S) := \{v \in V \mid \exists u \in S \text{ with } d_f(u, v) \leq \epsilon\}$ denotes the set of vertices with radial projection distance at most ϵ from S,*

3. *for $uv \in E$, $\|f'(u) - f'(v)\| \leq \left(1 + \frac{2}{\epsilon}\right) \|f(u) - f(v)\|$.*

To summarize, the energy of each f_i is upper bounded by the third item in Lemma 5.16, and the mass of each f_i is lower bounded by $\frac{1}{2k}$ fraction of the total mass by the spreading property in Lemma 5.13.

5.3.2 Proof of Lemma 5.11

As our proof follows closely the steps in [LOT12], the review in the previous subsubsection also serves well as an overview of our proof.

Given an embedding $\bar{f} : V \to \mathbb{R}^h$ and a target $l \leq k$, we would like to find l disjoint subsets $S_1, \ldots, S_l$ of V such that

1. for $1 \leq i \leq l$, the mass $\mu(S_i) := \sum_{u \in S_i} \pi(u) \left\|\bar{f}(u)\right\|^2$ of each S_i is at least $\frac{1}{2k} \cdot \mu(V)$, where $\mu(V) = \mu(\bar{f})$ in Definition 5.6, and

2. for $1 \leq i \neq j \leq l$, the distance $d_{\bar{f}}(S_i, S_j) := \min_{u \in S_i, v \in S_j} d_{\bar{f}}(u, v)$ between S_i and S_j is at least 2ϵ for some $\epsilon > 0$ to be determined later, where $d_{\bar{f}}$ is the radial projection distance in Definition 5.12.

To this end, equip V with the pseudo-metric $d_{\bar{f}}$ and consider the metric space $(V, d_{\bar{f}})$. Let $\mathcal{P} = P_1 \sqcup P_2 \sqcup \cdots \sqcup P_m$ be a $\left(\frac{\Delta}{4}, \frac{ch}{\delta}, 1-\delta\right)$-padded random partition sampled from Theorem 5.15, where c is a universal constant and $\Delta \in (0,1)$ and $\delta \in (0,1)$ are to be determined. By the assumption that $\bar{f}$ is $\left(\frac{\Delta}{4}, \frac{1}{k(1-\Delta)}\right)$-spreading, the first property in Definition 5.14 implies that $\mu(P_i) \le \frac{1}{k(1-\Delta)} \cdot \mu(\bar{f})$ for $1 \le i \le m$. Let $U := \{x \in V \mid B(x, \frac{\Delta\delta}{4ch}) \not\subseteq P(x)\}$ be the set of points that are close to the boundaries of $\mathcal{P}$. The second property in Definition 5.14 implies that there exists a realization of $\mathcal{P}$ such that $\mu(V - U) \ge (1-\delta) \cdot \mu(\bar{f})$. We take such a realization $\mathcal{P} = P_1 \sqcup P_2 \sqcup \cdots \sqcup P_m$ and remove all points in U to obtain $P'_i := P_i - U$ for $1 \le i \le m$. By doing so, we end up with disjoint sets $P'_1, P'_2, \ldots, P'_m$ with the following properties:

1. $\mu(P'_i) \le \frac{1}{k(1-\Delta)} \cdot \mu(\bar{f})$ for $1 \le i \le m$,

2. $\sum_{i=1}^m \mu(P'_i) \ge (1-\delta) \cdot \mu(\bar{f})$,

3. $d_{\bar{f}}(P'_i, P'_j) \ge \frac{2\Delta\delta}{ch}$ for $i \ne j \in [m]$.

Next, we will merge some of the sets $P'_1, \ldots, P'_m$ to form disjoint sets $S_1, \ldots, S_l$ so that $\mu(S_i) \ge \frac{1}{2k} \cdot \mu(\bar{f})$ for $1 \le i \le l$. This can be done by a simple greedy process, where we sort the P'_i by nonincreasing mass, and put consecutive sets into a group S_j until $\mu(S_j) \ge \frac{1}{2k} \cdot \mu(\bar{f})$. By the first property and the greedy process, each group S_j produced has $\mu(S_j) \le \frac{1}{k(1-\Delta)} \cdot \mu(f)$. Hence, by the second property, the greedy process will succeed to produce at least l groups with mass at least $\frac{1}{2k} \cdot \mu(\bar{f})$ as long as

$$(1-\delta) \cdot \mu(\bar{f}) - (l-1) \cdot \frac{\mu(\bar{f})}{k(1-\Delta)} \ge \frac{\mu(\bar{f})}{2k} \qquad \Longleftrightarrow \qquad \Delta \le 1 - \frac{2(l-1)}{2(1-\delta)k - 1},$$

which is exactly the assumption we made in the statement about Δ. Therefore, we can produce $S_1, \ldots, S_l$ satisfying the two requirements $\mu(S_i) \ge \frac{1}{2k} \cdot \mu(\bar{f})$ for $1 \le i \le l$ and $d_{\bar{f}}(S_i, S_j) \ge 2\epsilon := \frac{2\Delta\delta}{ch}$ for $i \ne j \in [l]$ by the third property of $P'_1, \ldots, P'_m$.

Now, we apply the smooth localization procedure in Lemma 5.16 on each S_i with $\epsilon = \frac{\Delta\delta}{ch}$ to obtain an embedding $\bar{f}_i : V \to \mathbb{R}^h$ for $1 \le i \le l$. First, we check that $\bar{f}_1, \ldots, \bar{f}_l$ are disjointly supported. This follows from $d_{\bar{f}}(S_i, S_j) \ge 2\epsilon$ for $i \ne j$ and the second property in Lemma 5.16. Second, since $\mu(S_i) \ge \frac{1}{2k} \cdot \mu(\bar{f})$ and $\bar{f}_i(v) = \bar{f}(v)$ for $v \in S_i$ by the first property in Lemma 5.16, it follows that $\mu(\bar{f}_i) = \mu(S_i) \ge \frac{1}{2k} \cdot \mu(\bar{f})$. Finally, by the third property in Lemma 5.16, it follows that

$$\mathcal{E}(\bar{f}_i) = \sum_{v \in V} \pi(v) \max_{u:u \to v} \left\| \bar{f}_i(u) - \bar{f}_i(v) \right\|^2 \le \left(1 + \frac{2ch}{\Delta\delta}\right)^2 \sum_{v \in V} \pi(v) \max_{u:u \to v} \left\| \bar{f}(u) - \bar{f}(v) \right\|^2 \lesssim \left(1 + \frac{h}{\Delta\delta}\right)^2 \mathcal{E}(\bar{f}).$$

Therefore, we conclude that $\bar{f}_1, \ldots, \bar{f}_l$ satisfy all the properties stated in Lemma 5.11.

5.4 Cheeger Rounding

The fourth step is to apply the results in Section 3 on $\bar{f}_1, \ldots, \bar{f}_l$ from Lemma 5.11 to obtain disjoint subsets with small vertex expansion.

Lemma 5.17 (Cheeger Rounding). *Let $G = (V, E)$ be an undirected graph with maximum degree d and π be a probability distribution π on V. Given an orientation $\overrightarrow{E}$ and an embedding $\bar{f} : V \to \mathbb{R}^h$, there exists a set $S \subseteq \mathrm{supp}(\bar{f})$ with*

$$\psi(S)^2 \lesssim \min\{h, \log d\} \cdot \frac{\mathcal{E}(\bar{f})}{\mu(\bar{f})}.$$

Proof. Given $\overrightarrow{E}$ and $\bar{f}$, we apply the Gaussian projection step in Proposition 3.14 to obtain a 1-dimensional embedding $x : V \to \mathbb{R}$ with $\mathcal{E}(x)/\mu(x) \lesssim \log d \cdot \mathcal{E}(\bar{f})/\mu(\bar{f})$. Alternatively, if $h \leq \log d$, we can choose the best coordinate from $\bar{f}$ to obtain a 1-dimensional embedding $x : V \to \mathbb{R}$ with $\mathcal{E}(x)/\mu(x) \leq h \cdot \mathcal{E}(\bar{f})/\mu(\bar{f})$ as was done in [OZ22]. So we have a 1-dimensional embedding x with $\mathcal{E}(x)/\mu(x) \lesssim \min\{h, \log d\} \cdot \mathcal{E}(\bar{f})/\mu(\bar{f})$.

Then, we apply the threshold rounding step in Proposition 3.19 on x to obtain a set $S \subseteq \mathrm{supp}(x) \subseteq \mathrm{supp}(\bar{f})$ with $\overrightarrow{\psi}(S)^2 \lesssim \mathcal{E}(x)/\mu(x) \leq \min\{h, \log d\} \cdot \mathcal{E}(\bar{f})/\mu(\bar{f})$. Finally, we apply the postprocessing step in Lemma 3.20 to obtain a set $S' \subseteq S$ with $\psi(S') \leq 2\overrightarrow{\psi}(S)$ satisfying the requirements of this lemma. $\qquad\square$

We are ready to put together the steps to prove the hard direction of the higher-order Cheeger inequality for vertex expansion.

Theorem 5.18 (Hard Direction for Multiway Vertex Expansion). *Let $G = (V, E)$ be an undirected graph with maximum degree d and π be a probability distribution π on V. For any $2 \leq k \leq n$ and $0 \leq \epsilon < 1$, let $\xi := \max\{\epsilon, \frac{1}{2k}\}$, it holds that*

$$\psi_{(1-\epsilon)k}(G) \lesssim \frac{\log k}{\xi^4} \cdot \sqrt{\log d \cdot \sigma_k^*(G)}.$$

Proof. The first step is to compute an optimal solution (f^*, g^*) to the dual program in Proposition 5.2 with objective value $\kappa(G) = \sigma_k^*(G)$. Then, we use Lemma 5.4 to obtain a solution $f : V \to \mathbb{R}^n$ to the directed program $\overrightarrow{\kappa}(G)$ with energy $\mathcal{E}(f) \leq 2\sigma_k^*(G)$ and $\mu(f) = k$. As f satisfies the sub-isotropy condition $\sum_u \pi(u)f(u)f(u)^T \preccurlyeq I_n$ in $\overrightarrow{\kappa}(G)$, we know from Proposition 5.13 that f is $(\Delta, 1/(k(1-\Delta^2)))$-spreading for any $\Delta \in (0,1)$ of our choice.

The second step is to apply the Gaussian projection algorithm in Lemma 5.10 on f with $\eta = 1/(k(1-\Delta^2))$ to obtain $\bar{f} : V \to \mathbb{R}^h$ with

$$h \lesssim \frac{1}{\Delta^2} \log\left(\frac{1}{\Delta\eta}\right) = \frac{1}{\Delta^2} \log\left(\frac{k(1-\Delta^2)}{\Delta}\right)$$

such that

$$\mathcal{E}(\bar{f}) \lesssim \left(1 + \frac{\log d}{h}\right) \cdot \mathcal{E}(f) \quad \text{and} \quad \mu(\bar{f}) \gtrsim \mu(f) \quad \text{and} \quad \bar{f} \text{ is } \left(\frac{\Delta}{4}, \frac{1}{k(1-\Delta)}\right)\text{-spreading}.$$

The third step is to apply the spectral partitioning algorithm in Lemma 5.11 on f. Let $l := (1-\epsilon)k$ be the target number of output sets. By setting

$$\Delta = \min\left\{\frac{1}{2}, \frac{2(k-l)+1}{2(k+l)-3}\right\} \quad \text{and} \quad \delta = \frac{2(k-l)+1}{4k},$$

we can check that the conditions of Lemma 5.11 are satisfied, and so we can construct functions $\bar{f}_1, \ldots, \bar{f}_l : V \to \mathbb{R}^h$ with disjoint support, such that for each $1 \le i \le l$ it holds that

$$\mathcal{E}(\bar{f}_i) \lesssim \left(1 + \frac{h}{\Delta\delta}\right)^2 \cdot \mathcal{E}(\bar{f}) \quad \text{and} \quad \mu(\bar{f}_i) \gtrsim \frac{1}{k}\mu(\bar{f}).$$

The fourth step is to apply Lemma 5.17 to $\bar{f}_1, \ldots \bar{f}_l$ to obtain disjoint subsets $S_1, \ldots, S_l$, such that for every $1 \le i \le l$,

$$
\begin{aligned}
\psi(S_i)^2 &\lesssim \min\{h, \log d\} \cdot \frac{\mathcal{E}(\bar{f}_i)}{\mu(\bar{f}_i)} \\
&\lesssim \min\{h, \log d\} \cdot k \cdot \left(1 + \frac{h}{\Delta\delta}\right)^2 \cdot \frac{\mathcal{E}(\bar{f})}{\mu(\bar{f})} \\
&\lesssim \min\{h, \log d\} \cdot \left(1 + \frac{\log d}{h}\right) \cdot \left(1 + \frac{h}{\Delta\delta}\right)^2 \cdot k \cdot \frac{\mathcal{E}(f)}{\mu(f)} \\
&\lesssim \log d \cdot \left(1 + \frac{h}{\Delta\delta}\right)^2 \cdot \sigma_k^*(G) \\
&\lesssim \log d \cdot \frac{1}{\Delta^6\delta^2} \cdot \log^2\left(\frac{k(1-\Delta^2)}{\Delta}\right) \cdot \sigma_k^*(G),
\end{aligned}
$$

where the fourth inequality uses that $\min\{h, \log d\} \cdot \left(1 + \frac{\log d}{h}\right) \le 2\log d$ and $\mu(f) = k$, and the inequality uses the fact that $\frac{h}{\Delta\delta} \ge 1$. This implies that

$$\psi_l(G) \lesssim \frac{1}{\Delta^3\delta} \log\left(\frac{k}{\Delta}\right) \cdot \sqrt{\log d \cdot \sigma_k^*(G)}.$$

Finally, we plug in $l = (1-\epsilon)k$ and consider two cases. In the case when $\epsilon \ge \frac{1}{2k}$, we see that $\Delta = \Theta(\epsilon)$ and $\delta = \Theta(\epsilon)$, and so $\psi_l(G) \le \frac{1}{\epsilon^4} \cdot \log k \cdot \sqrt{\log d \cdot \sigma_k^*(G)}$. In the case when $\epsilon < \frac{1}{2k}$, we simply set $l = k$ and see that $\Delta = \Theta(1/k)$ and $\delta = \Theta(1/k)$ and so $\psi_l(G) \le k^4 \cdot \log k \cdot \sqrt{\log d \cdot \sigma_k^*(G)}$. Combining the two cases proves the theorem. $\square$

We prove the following easy direction in Appendix C. Note that it is about $\lambda_k^*(G)$ instead of $\sigma_k^*(G)$.

Lemma 5.19 (Easy Direction for Multiway Vertex Expansion). *For any undirected graph $G = (V, E)$ and any probability distribution π on V, $\lambda_k^*(G) \le 2\psi_k(G)$ for any $k \ge 2$.*

Combining Lemma 5.19 and Theorem 5.18, we conclude this section with the following higher-order Cheeger inequality for vertex expansion that implies Theorem 1.10.

Theorem 5.20 (Higher-Order Cheeger Inequality for Vertex Expansion)**.** *For any undirected graph* $G = (V, E)$ *with maximum degree* d *and any probability distribution* π *on* V,

$$\frac{\sigma_k^*(G)}{k} \leq \lambda_k^*(G) \lesssim \psi_k(G) \lesssim k^4 \log k \sqrt{\log d \cdot \sigma_k^*(G)} \leq k^{\frac{9}{2}} \log k \sqrt{\log d \cdot \lambda_k^*(G)}$$

Furthermore, for any $1 > \epsilon \geq \frac{1}{2k}$,

$$\psi_{(1-\epsilon)k}(G) \lesssim \frac{1}{\epsilon^4} \log k \sqrt{\log d \cdot \sigma_k^*(G)} \leq \frac{1}{\epsilon^4} \log k \sqrt{k \log d \cdot \lambda_k^*(G)}.$$

6 Improved Cheeger Inequality for Vertex Expansion

The goal of this section is to prove Theorem 1.11. The proof in [KLLOT13] has two main steps. The first step is to prove that if the second eigenfunction is close to a k-step function, then the approximation guarantee of threshold rounding is improved. The second step is to prove that if λ_k is large for a small k, then the second eigenfunction is close to a k-step function.

We follow the a similar plan to prove an improved version of the Cheeger inequality for $\gamma^{(1)}(G)$ in Theorem 3.15, where in the second step we replace λ_k by $\sigma_k^*(G)$. First we begin with the definition of a k-step function.

Definition 6.1 (k-Step Function and Approximation)**.** *Let* $G = (V, E)$ *be an undirected graph and* π *be a probability distribution on* V. *Given* $y : V \to \mathbb{R}$ *and* $1 \leq k \leq n$, *we call* y *a* k-step function *if the number of distinct values in* $\{y(u)\}_{u \in V}$ *is at most* k.

Given $x : V \to \mathbb{R}$, *we say* y *is a* k-step ϵ-approximation *to* x *if* y *is a* k-step function and $\|x - y\|_\pi \leq \epsilon$, *where* $\|z\|_\pi^2 := \sum_{v \in V} \pi(v) z(v)^2$ *for a vector* $z : V \to \mathbb{R}$.

The following is the precise statement of the first step for $\gamma^{(1)}(G)$, which informally says that if there is a good k-step approximation to an optimal solution to $\gamma^{(1)}(G)$ then the performance of threshold rounding is better than that in Theorem 3.15.

Proposition 6.2 (Rounding k-Step Approximation)**.** *Let* $G = (V, E)$ *be an undirected graph and* π *be a probability distribution on* V. *For any feasible solution* (f, g) *to the* $\gamma^{(1)}(G)$ *program with objective value* γ_f *and any* k-step function $y_f : V \to \mathbb{R}$ *approximating* f,

$$\psi(G) \lesssim k \cdot \gamma_f + k \, \|f - y_f\|_\pi \, \sqrt{\gamma_f}.$$

Our second step is to prove that if $\sigma_k^*(G)$ in Proposition 5.2 is large for a small k, then there is a good k-step approximation to a good solution to $\gamma^{(1)}(G)$.

Proposition 6.3 (Constructing k-Step Approximation)**.** *Let* $G = (V, E)$ *be an undirected graph and* π *be a probability distribution on* V. *For any feasible solution* (f, g) *to the* $\gamma^{(1)}(G)$ *program with objective value* γ_f, *there exists a* k-step function y *with*

$$\|f - y\|_\pi^2 \lesssim \frac{k \cdot \gamma_f}{\sigma_k^*(G)}.$$

Assuming Proposition 6.2 and Proposition 6.3, we prove an exact analog of the improved Cheeger's inequality in [KLLOT13] for vertex expansion, with $\sigma_k^*(G)/k$ playing the role of $\lambda_k^*(G)$.

Theorem 6.4 (Improved Cheeger Inequality for Vertex Expansion). *For any undirected graph $G = (V, E)$ and any probability distribution π on V and any $k \geq 2$,*

$$\gamma^{(1)}(G) \lesssim \psi(G) \lesssim k \cdot \gamma^{(1)}(G) \cdot \sqrt{\frac{k}{\sigma_k^*(G)}}.$$

Proof. The easy direction is proved in Lemma 3.16. For the hard direction, let (f^*, g^*) be an optimal solution to the $\gamma^{(1)}(G)$ program in Definition 3.2 with objective value γ^*, and σ^* be the optimal value of the $\sigma_k^*(G)$ program in Proposition 5.2. By Proposition 6.3, there exists a k-step function $y : V \to \mathbb{R}$ with

$$\|f^* - y\|_\pi^2 \lesssim \frac{k \cdot \gamma^*}{\sigma^*}.$$

Applying Proposition 6.2 with y, it follows that

$$\psi(G) \lesssim k \cdot \left(\gamma^* + \|f^* - y\|_\pi \cdot \sqrt{\gamma^*}\right) \lesssim k \cdot \gamma^* \left(1 + \sqrt{\frac{k}{\sigma^*}}\right) \lesssim k \cdot \gamma^* \cdot \sqrt{\frac{k}{\sigma^*}}.$$

$\square$

Note that Theorem 1.11 follows immediately from Theorem 6.4.

Proof of Theorem 1.11.

$$\lambda_2^*(G) = \gamma(G) \leq \gamma^{(1)}(G) \lesssim \psi(G) \lesssim k \cdot \gamma^{(1)}(G) \cdot \sqrt{\frac{k}{\sigma_k^*(G)}} \lesssim k \cdot \log d \cdot \gamma(G) \cdot \sqrt{\frac{k}{\lambda_k^*(G)}},$$

where we use $\gamma(G) = \lambda_2^*(G)$ in Proposition 3.1 and $\gamma(G) \leq \gamma^{(1)}(G) \lesssim \log d \cdot \gamma(G)$ in Proposition 3.14 and $\sigma_k^*(G) \geq \lambda_k^*(G)$ in Proposition 5.2. $\square$

Remark 6.5 (Tight Examples). *We remark that Theorem 6.4 is tight. The loss in Theorem 1.11 is because of the factor $\log d$ loss in the dimension reduction step and the factor k loss in the transition from $\lambda_k^*(G)$ to $\sigma_k^*(G)$.*

As an example, let G be an n-cycle, where n is odd, and π the uniform distribution. Suppose $k \ll n$. Then $\sigma_k^*(G) = \sigma_k(G)$, because the only possible "reweighting" is the one with equal edge weight. Since $\psi(G) = \Theta(1/n), \gamma^{(1)}(G) = O(1/n^2)$ (choose $f : V \to \mathbb{R}$ that maps vertex $l \in [n]$ to point $C(1 - \frac{4}{n} \cdot \min(l, n - l))$, where $C = \Theta(1)$ is a normalizing factor, and $g(l) \equiv \frac{8C^2}{n^2}$), and

$$\sigma_k(G) = \sum_{l=0}^{k-1} \left(1 - \cos\left(\frac{2\pi l}{n}\right)\right) = \Theta\left(\frac{k^3}{n^2}\right),$$

one can verify that in this case the hard direction of Theorem 6.4 is tight.

We prove Proposition 6.2 and Proposition 6.3 in the following two subsections.

6.1 Rounding k-Step Approximation

We prove Proposition 6.2 in this subsection. First, we do some preprocessing on the solution (f, g) as in Section 3. Then, as in [KLLOT13], the main step is to use a modified probability distribution on the thresholds based on the k-step approximation y to analyze the threshold rounding algorithm.

Given a feasible solution (f, g) to the $\gamma^{(1)}(G)$ program with objective value γ_f, we use Lemma 3.12 and Lemma 3.13 to obtain a solution f and $\overrightarrow{E}$ to the directed program $\overrightarrow{\gamma}^{(1)}(G)$ in Definition 3.11 with objective value at most $2\gamma_f$. Then we apply the truncation step in Lemma 3.17 on f and $\overrightarrow{E}$ to obtain a solution x and $\overrightarrow{E}$ with $x \geq 0$ and $\pi(\text{supp}(x)) \leq 1/2$ and

$$\sum_{v \in V} \pi(v) \max_{u:uv \in \overrightarrow{E}} (x(u) - x(v))^2 \leq 8\gamma_f \quad \text{and} \quad \sum_{v \in V} \pi(v)x(v)^2 = 1.$$

Let y_f be the k-step approximation of f in the assumption of Proposition 6.2. Note that the way we construct x from f consists of shifting, truncating, and scaling by a factor of at most 2. So, applying the same transformations to y_f will still give us a k-step function y with $\|x - y\|_\pi \leq 2\|f - y_f\|_\pi$. Henceforth, we work with x and its k-step approximation y.

The main step is to prove that applying the threshold rounding algorithm on x will find a set $S \subseteq V$ with small directed vertex expansion $\overrightarrow{\psi}(S)$ as defined in Definition 3.18. For the analysis, we take the k-step approximation y into consideration and give higher weight to a threshold t if t is far away from the function values in y.

The following weighting scheme is from [KLLOT13]. Let $x_{\max} := \max_{v \in V} x(v)$. Suppose the k-step function y takes values $0 \leq y_1 \leq y_2 \leq \cdots \leq y_k$. For $t \in [0, x_{\max}]$, define $\rho(t) := \min_{1 \leq i \leq k} |t - y_i|$. In words, $\rho(t)$ is the distance from t to the closest value of the k-step function y. We sample $t \in [0, x_{\max}]$ with probability proportional to $\rho(t)$. That is, for $0 \leq a < b \leq x_{\max}$,

$$\Pr\left[t \in [a, b]\right] = \frac{1}{Z} \int_a^b \rho(t)\, dt,$$

where $Z := \int_0^{x_{\max}} \rho(t)dt$ is the normalizing factor.

We use the distribution described above to analyze the threshold rounding algorithm. Let $S_t := \{v \in V \mid x(v) > t\}$ be a threshold set. By a standard averaging argument,

$$\min_{t \in [0, x_{\max}]} \overrightarrow{\psi}(S_t) \leq \frac{\mathbb{E}_t[\pi(\overrightarrow{\partial}S_t)]}{\mathbb{E}_t[\pi(S_t)]}$$

The denominator is

$$\mathbb{E}_t[\pi(S_t)] \;=\; \frac{1}{Z} \cdot \sum_{v \in V} \pi(v) \cdot \int_0^{x(v)} \rho(t)\, dt \;\gtrsim\; \frac{1}{kZ} \sum_{v \in V} \pi(v)x(v)^2 = \frac{1}{kZ},$$

where the inequality can be seen as follows: For any $v \in V$, let $k' \leq k$ be the largest index so that $0 \leq y_1 \leq \cdots \leq y_{k'} \leq x(v)$ and let $z_0 := 0$, $z_i := y_i$ for $1 \leq i \leq k'$ and $z_{k'+1} := x(v)$, then

$$\int_0^{x(v)} \rho(t)\, dt = \sum_{i=0}^{k'} \int_{z_i}^{z_{i+1}} \rho(t)\, dt \geq \frac{1}{4}\sum_{i=0}^{k'}(z_{i+1} - z_i)^2 \geq \frac{1}{4(k'+1)}\left(\sum_{i=0}^{k'}(z_{i+1} - z_i)\right)^2 \gtrsim \frac{1}{k}x(v)^2,$$

where the first inequality is by simple calculus and the second inequality is by Cauchy-Schwarz.

The numerator is

$$
\begin{aligned}
\mathbb{E}_t\left[\pi\big(\overrightarrow{\partial} S_t\big)\right] &= \sum_v \pi(v) \cdot \Pr_t[\text{ there exists an edge } uv \in \overrightarrow{E} \text{ with } uv \in \delta(S_t)] \\
&= \sum_v \pi(v) \cdot \max_{u:u\to v} \Pr_t\left[uv \in \delta(S_t)\right] \\
&= \sum_v \pi(v) \cdot \max_{u:u\to v} \frac{1}{Z} \cdot \left| \int_{x(v)}^{x(u)} \rho(t)\, dt \right| \\
&\le \frac{1}{Z} \sum_v \pi(v) \cdot \max_{u:u\to v} \left| \int_{x(v)}^{x(u)} \Big(\rho\big(x(v)\big) + |t - x(v)| \Big)\, dt \right| \\
&= \frac{1}{Z} \sum_v \pi(v) \cdot \max_{u:u\to v} \left[|x(u) - x(v)| \cdot \rho\big(x(v)\big) + \frac{1}{2}\big(x(u) - x(v)\big)^2 \right] \\
&\le \frac{1}{Z} \cdot \left(4\gamma_f + \sum_v \pi(v) \cdot \rho\big(x(v)\big) \cdot \max_{u:u\to v} |x(u) - x(v)| \right) \\
&\le \frac{1}{Z} \cdot \left(4\gamma_f + \sqrt{\left(\sum_v \pi(v) \cdot \rho\big(x(v)\big)^2\right) \cdot \left(\sum_v \pi(v) \max_{u:u\to v}\big(x(u) - x(v)\big)^2\right)} \right) \\
&\le \frac{1}{Z} \cdot \left(4\gamma_f + \sqrt{\|x - y\|_\pi^2 \cdot 8\gamma_f} \right) \\
&\lesssim \frac{1}{Z} \cdot \left(\gamma_f + \|x - y\|_\pi \cdot \sqrt{\gamma_f} \right),
\end{aligned}
$$

where the first inequality is because $\rho(t)$ is a 1-Lipschitz function, the third inequality is by Cauchy-Schwarz, and the fourth inequality is because $\rho(x(v)) \le |x(v) - y(v)|$ by the definition of ρ.

Combining the bounds on the denominator and the numerator,

$$
\min_{t\in[0,x_{\max}]} \overrightarrow{\psi}(S_t) \le \frac{\mathbb{E}_t\big[\pi\big(\overrightarrow{\partial} S_t\big)\big]}{\mathbb{E}_t\big[\pi(S_t)\big]} \lesssim k\gamma_f + k\,\|x - y\|_\pi \cdot \sqrt{\gamma_f} \lesssim k\gamma_f + k\,\|f - y_f\|_\pi \cdot \sqrt{\gamma_f}.
$$

Since $\pi(\mathrm{supp}(x)) \le \frac{1}{2}$, the output set S has $\pi(S) \le \frac{1}{2}$. Finally, we apply the postprocessing step in Lemma 3.20 to obtain a set $S' \subseteq S$ with $\psi(S') \le 2\overrightarrow{\psi}(S)$. We conclude that

$$
\psi(G) \le \psi(S') \lesssim \min_t \overrightarrow{\psi}(S_t) \lesssim \gamma_f + k\,\|f - y_f\|_\pi \cdot \sqrt{\gamma_f}.
$$

6.2 Constructing k-Step Approximation

We prove Proposition 6.3 in this subsection. The high level plan is similar to that in [KLLOT13]. Given a feasible solution (f, g) to the $\gamma^{(1)}(G)$ program, we aim to construct a good k-step approximation y of f using a simple procedure. If we fail to do so, then we show that f can be used to construct a good k-dimensional solution $\bar{f} = (\bar{f}_1, \bar{f}_2, \ldots, \bar{f}_k)$ to the $\sigma_k^*(G)$ program, contradicting

43

the value of $\sigma_k^*(G)$ is large. Therefore, the simple process must succeed to find a good k-step approximation y of f.

Suppose our k-step function y takes values $y_1 \leq y_2 \leq \cdots \leq y_k$. For convenience $y_0 := -\infty$. We use these values to define k disjoint subsets $S_1, \ldots, S_k \subseteq V$ where $S_i := \{v \in V \mid y_{i-1} < f(v) \leq y_i\}$, and define functions $\bar{f}_1, \ldots, \bar{f}_k$ supported on $S_1, \ldots, S_k$ respectively where

$$\bar{f}_i(v) := \begin{cases} \min \{|y_i - f(v)|, |f(v) - y_{i-1}|\}, & \text{if } f(v) \in (y_{i-1}, y_i) \\ 0 & \text{otherwise.} \end{cases}$$

The role of $\bar{f}_i$ is to measure how well the two threshold values y_{i-1} and y_i approximate the values of the vertices in S_i in f. We would like to choose $y_1 \leq y_2 \leq \cdots \leq y_k$ so that each $\|\bar{f}_i\|_\pi^2$ is small and $S_1 \cup \cdots \cup S_k = V$. As we will show, this would imply that there exists a k-step function y with threshold values $y_1 \leq y_2 \leq \cdots \leq y_k$ so that $\|f - y\|_\pi = \sum_{i=1}^k \|\bar{f}_i\|_\pi^2$ is small, and thus y is a good k-step approximation to f.

Consider the following simple procedure to choose the threshold values $y_1 \leq y_2 \leq \ldots \leq y_k$ of the k-step function. Let μ be a parameter to be determined later. We would like to guarantee that $\|\bar{f}_i\|_\pi^2 \leq \mu$ for $1 \leq i \leq k$ and $S_1 \cup S_2 \cup \cdots \cup S_k = V$ such that all vertices are covered by these functions $\bar{f}_1, \ldots, \bar{f}_k$. We choose y_i successively as follows: given $y_0, \ldots, y_{i-1}$, set y_i to be the smallest number such that $\|\bar{f}_i\|_\pi^2 = \mu$. If the smallest y_i does not exist, then we set $y_i = \ldots = y_k := \max_{v \in V} f(v)$ and call the procedure a "success". Otherwise, if we set all $y_1 \leq y_2 \leq \ldots \leq y_k$ and $S_1 \cup S_2 \cup \cdots S_k \subset V$ such that not all vertices are covered by $\bar{f}_1, \ldots, \bar{f}_k$, then we call the procedure a "failure".

Now we set

$$\mu = \frac{2\gamma_f}{\sigma_k^*(G)}.$$

When the procedure succeeds, then $\|\bar{f}_i\|_\pi^2 \leq \mu$ for $1 \leq i \leq k$ and $S_1 \cup S_2 \cup \cdots \cup S_k = V$. We would like to show that the there exists a k-step function y with threshold values $y_1 \leq y_2 \leq \ldots \leq y_k$ such that $\|y - f\|_\pi^2 \leq k\mu$. Define $y : V \to \mathbb{R}$ to be

$$y(v) := \mathrm{argmin}_{\alpha \in \{y_1, \ldots, y_k\}} |\alpha - f(v)|,$$

such that each vertex v is assigned to its closest threshold value. Then

$$\|f - y\|_\pi^2 = \sum_{v \in V} \pi(v) \cdot |f(v) - y(v)|^2 = \sum_{i=1}^k \sum_{v \in S_i} \pi(v) \bar{f}_i(v)^2 = \sum_{i=1}^k \|\bar{f}_i\|_\pi^2 \leq k\mu \lesssim \frac{k\gamma_f}{\sigma_k^*(G)},$$

where the second equality uses that $S_1 \cup S_2 \cup \cdots \cup S_k = V$ and the first inequality uses that each $\|\bar{f}_i\|_\pi^2 \leq \mu$. Thus y is a k-step function that satisfies the statement of the proposition.

To complete the proof, we would like to show that the procedure always succeeds. When the procedure fails, then there exist $y_1 < y_2 < \cdots < y_k$ such that $\|\bar{f}_i\|_\pi^2 = \mu$ for $1 \leq i \leq k$. We will construct from $\bar{f}_1, \ldots, \bar{f}_k$ a solution $(\bar{f}, \bar{g})$ to the $\sigma_k^*(G)$ program with objective value less than $\sigma_k^*(G)$, thus arriving at a contradiction. Define $\bar{f} : V \to \mathbb{R}^n$ and $\bar{g} : V \to \mathbb{R}$ as follows:

$$\bar{f}(v) := \left(\frac{\bar{f}_1(v)}{\sqrt{\mu}}, \ldots, \frac{\bar{f}_k(v)}{\sqrt{\mu}}, 0, \ldots, 0 \right)^T \quad \text{and} \quad \bar{g}(v) := \frac{1}{\mu} g(v).$$

We will check that $(\bar{f}, \bar{g})$ is a feasible solution to the $\sigma_k^*(G)$ program in Proposition 5.2. For the sub-isotropy condition, note that each $\bar{f}(v)$ has at most one nonzero entry, and

$$\sum_{v \in V} \pi(v) \bar{f}(v) \bar{f}(v)^T = \operatorname{diag}\left(\frac{1}{\mu} \sum_{v \in S_1} \pi(u) \bar{f}_1(u)^2, \frac{1}{\mu} \sum_{v \in S_2} \pi(v) \bar{f}_2(u)^2, \ldots, \frac{1}{\mu} \sum_{v \in S_k} \pi(v) \bar{f}_k(v)^2, 0, \ldots, 0\right)$$

$$= \frac{1}{\mu} \operatorname{diag}\left(\|\bar{f}_1\|_\pi^2, \|\bar{f}_2\|_\pi^2, \ldots, \|\bar{f}_k\|_\pi^2, 0, \ldots, 0\right) = \operatorname{diag}(1, 1, \ldots, 1, 0, \ldots, 0) \preccurlyeq I_n.$$

The mass constraint is satisfied as

$$\sum_{v \in V} \pi(v) \|\bar{f}(v)\|^2 = \operatorname{tr}\left(\sum_{u \in V} \pi(u) \bar{f}(u) \bar{f}(u)^T\right) = \operatorname{tr}\left(\operatorname{diag}\left(1, 1, \ldots, 1, 0, \ldots, 0\right)\right) = k.$$

For the constraint on each edge $uv \in E$,

$$\left\|\bar{f}(u) - \bar{f}(v)\right\|^2 = \frac{1}{\mu} \sum_{i=1}^k \left(\bar{f}_i(u) - \bar{f}_i(v)\right)^2 \leq \frac{1}{\mu}\left(f(u) - f(v)\right)^2 \leq \frac{1}{\mu}\left(g(u) + g(v)\right) = \bar{g}(u) + \bar{g}(v),$$

where for the first inequality we consider two cases: (i) suppose $u \in S_i$ and $v \in S_j$ for $i = j$, then $\sum_{l=1}^k \left(\bar{f}_l(u) - \bar{f}_l(v)\right)^2 = \left(\bar{f}_i(u) - \bar{f}_i(v)\right)^2 \leq \left(f(u) - f(v)\right)^2$, and (ii) suppose $u \in S_i$ and $v \in S_j$ for $i \neq j$, then $\sum_{l=1}^k \left(\bar{f}_l(u) - \bar{f}_l(v)\right)^2 = \left(\bar{f}_i(u) - \bar{f}_i(v)\right)^2 + \left(\bar{f}_j(u) - \bar{f}_j(v)\right)^2 \leq \left(f(u) - f(v)\right)^2$ since $\left|\bar{f}_i(u) - \bar{f}_i(v)\right| + \left|\bar{f}_j(u) - \bar{f}_j(v)\right| \leq \left|f(u) - f(v)\right|$.

Therefore, $(\bar{f}, \bar{g})$ is a feasible solution to the $\sigma_k^*(G)$ program, and its objective value is

$$\sum_{v \in V} \pi(v) \bar{g}(v) = \frac{1}{\mu} \sum_{v \in V} \pi(v) g(v) = \frac{1}{\mu} \cdot \gamma_f = \frac{\sigma_k^*(G)}{2} < \sigma_k^*(G),$$

a contradiction to the definition of $\sigma_k^*(G)$.

To conclude, the procedure must succeed and return a k-step function y with $\|f - y\|_\pi^2 \lesssim k\gamma_f/\sigma_k^*(G)$.

7 Vertex Expansion of $0/1$-Polytopes

The goal of this section is to present a construction of $0/1$-polytopes with poor vertex expansion as described in Theorem 1.12, which has implications about sampling from the uniform distribution as described in Section 1.1.5.

A $0/1$-polytope is defined by a subset of vertices in the boolean hypercube $\{0, 1\}^n$. Our examples are based on the following simple probabilistic construction.

Definition 7.1 (Probabilistic Construction). *Let n be an even number and $k < n/2$. For a binary string $x \in \{0, 1\}^n$, denote its 1-norm by $|x| := \sum_{i=1}^n |x_i|$. The set of vertices of our constructed polytope is the union of three subsets:*

1. *A left part $L := \{x \in \{0, 1\}^n \mid |x| = k\}$ consists of all binary strings with k ones.*

45

2. *A right part $R := \{x \in \{0,1\}^n \mid |x| = n - k\}$ consists of all binary strings with $n - k$ ones.*

3. *A middle part $M \subset \{x \in \{0,1\}^n \mid |x| = n/2\}$ consists of $\Theta(4^k n^2)$ number of uniformly random binary strings with $n/2$ ones.*

The graph $G_Q = (V, E)$ of a polytope Q is defined as the 1-skeleton of the polytope Q. Our plan is to prove that for a random polytope Q constructed in Definition 7.1, the middle part M "blocks" all the edges between L and R in G_Q with constant probability. This would imply that $\partial L \subseteq M$ and $\partial R \subseteq M$, and thus $\psi(L), \psi(R) \lesssim 4^k n^2 / n^k$ are very small.

The organization of this section is as follows. First, in Section 7.1, we provide a sufficient condition for two binary strings x, y to have no edge in G_Q, using geometric arguments. Then, in Section 7.2, we outline the main probabilistic argument to prove Theorem 1.12, by using a union bound over the set of linear threshold functions. Finally, we show Corollary 1.13 in Section 7.3. We defer all the proofs in Section 7.1 and Section 7.2 to Appendix D.

As mentioned in Remark 1.14 in the introduction, Gillmann [Gil07] has constructed similar examples of 0/1-polytopes with poor vertex expansion. Following the same simple argument as in Section 7.3, one obtains analogous lower bounds on the fastest mixing time of these polytopes. We remark that the construction in [Gil07] is similar to our construction, but the proofs are different and so we present our proofs even though the results follow from [Gil07].

7.1 A Sufficient Condition for Edge Blocking

Let Q be a 0/1-polytope and $G_Q = (V, E)$ be its graph/1-skeleton. For two binary strings $x, y \in \{0,1\}^n$, if xy is an edge in G_Q, then there is a separating hyperplane l with $l(x), l(y) \geq 0$ while $l(z) < 0$ for all other binary strings z in the 0/1-polytope Q.

In the construction of Q in Definition 7.1, if $x \in L$ and $y \in R$ then $\frac{1}{2}(x + y)$ has 1-norm equal to $n/2$. If xy is an edge in G_Q, then there is a separating hyperplane l with $l\left(\frac{1}{2}(x + y)\right) \geq 0$ while $l(z) < 0$ for all other binary strings z in the middle part M. So, if we could establish that $\frac{1}{2}(x + y)$ is in the convex hull $\text{conv}(M)$ of M for all $x \in L$ and $y \in R$, then there are no edges between L and R in the graph G_Q. This is the sufficient condition that we will formalize.

In the analysis, we use the following definitions to group the pairs of vertices $x \in L$, $y \in R$ based on their common patterns.

Definition 7.2 (Patterns). *For $n \in \mathbb{N}$, a pattern is an element $p \in \{0, 1, \vee\}^n$, where 0, 1, and $\vee$ are regarded as symbols.*

The support of a pattern $p \in \{0, 1, \vee\}^n$ is defined as $\text{supp}(p) := \{i \in [n] \mid p_i \neq \vee\}$. We also define $\text{supp}_0(p) := \{i \in [n] \mid p_i = 0\}$ and $\text{supp}_1(p) := \{i \in [n] \mid p_i = 1\}$.

Given two binary strings $x, y \in \{0,1\}^n$, their common pattern $p^{(x,y)} \in \{0, 1, \vee\}^n$ is defined as

$$
p_i^{(x,y)} = \begin{cases} 0, & \text{if } x_i = y_i = 0 \\ 1, & \text{if } x_i = y_i = 1 \\ \vee, & \text{if } x_i \neq y_i. \end{cases}
$$

46

Given a pattern $p \in \{0, 1, \vee\}^n$ and a binary string $x \in \{0, 1\}^n$, x is said to match p if and only if $p_i \neq \vee$ implies $p_i = x_i$.

For each pattern p, we consider a potential separating hyperplane of the following specific form, which will be convenient for the probabilistic analysis

Definition 7.3 (Consistent Affine Function). *Let $p \in \{0, 1, \vee\}^n$ be a pattern. An affine function $l : (u_1, u_2, \ldots, u_n) \in \mathbb{R}^n \mapsto \beta + \sum_i \alpha_i u_i$ is called p-consistent if*

$$\alpha_i = 0 \text{ for } i \in \text{supp}(p) \qquad \text{and} \qquad \beta + \frac{1}{2} \sum_{i : i \notin \text{supp}(p)} \alpha_i = 0.$$

If p is the common pattern of x and y, then l being p-consistent implies $l\left(\frac{1}{2}(x + y)\right) = 0$.

We formulate the sufficient condition described above for the middle part M blocking the edge xy for $x \in L$ and $y \in R$ using the definitions that we have developed. The proof is deferred to Appendix D.

Lemma 7.4 (Blocking One Edge). *Let $Q = L \cup M \cup R$ be a 0/1-polytope from Definition 7.1. Let $x \in L$, $y \in R$ and $p = p^{(x,y)} \in \{0, 1, \vee\}^n$ be the common pattern of x and y. If for any p-consistent affine function l there exists a point $z \in M$ matching the pattern p and satisfying $l(z) \geq 0$, then there is no edge connecting x and y in the graph of Q.*

The following is a sufficient condition for the middle part M blocking all the edges between L and R, by considering all possible common patterns of an $x \in L$ and a $y \in R$.

Lemma 7.5 (Blocking All Edges). *Let $Q = L \cup M \cup R$ be a 0/1-polytope from Definition 7.1. Suppose for every pattern $p \in \{0, 1, \vee\}^n$ with $|\text{supp}_0(p)| = |\text{supp}_1(p)| \leq k$ and for any p-consistent affine function $l : \mathbb{R}^n \to \mathbb{R}$, there is $z \in M$ matching the pattern p with $l(z) \geq 0$. Then there are no edges between L and R in the graph of Q.*

7.2 Probabilistic Analysis

Our plan is to use the sufficient condition in Lemma 7.5 to prove that a random M with not many points can block all the edges between L and R. To this end, we prepare with two simple lemmas about the probability that $z \in M$ satisfying $l(z) \geq 0$ and matching a particular pattern p with $|\text{supp}_0(p)| = |\text{supp}_1(p)| \leq k$.

The geometric intuition of the first lemma is simple: when we restrict a p-consistent affine function l on the coordinates in $[n] \setminus \text{supp}(p)$, then l is an "unbiased" hyperplane that goes through the point $\frac{1}{2} \cdot \vec{1}$ on $[n] \setminus \text{supp}(p)$ (because of the second condition in Definition 7.3), and thus a random vertex in M matching the pattern p lies on the non-negative side of l with probability at least $1/2$.

Lemma 7.6. *Let $Q = L \cup M \cup R$ be a 0/1-polytope from Definition 7.1. Let p be the common pattern of $x \in L$ and $y \in R$, and l be a p-consistent affine function. Let Z be the uniform distribution on $\{z \in \{0, 1\}^n : |z| = \frac{n}{2}\}$. Then,*

$$\Pr_{z \sim Z} \left[l(z) \geq 0 \mid z \text{ matches pattern } p \right] \geq \frac{1}{2}.$$

The second lemma gives a lower bound on the probability that a random point $z \sim Z$ matches a pattern p with $|\operatorname{supp}_0(p)| = |\operatorname{supp}_1(p)| \leq k$.

Lemma 7.7. *Let $p \in \{0, 1, \vee\}^n$ be a pattern with $|supp_0(p)| = |supp_1(p)| = s \leq k$, and let Z be the uniform distribution on $\{z \in \{0,1\}^n : |z| = n/2\}$. Then*

$$\Pr_{z \sim Z} [z \text{ matches pattern } p] \gtrsim 4^{-s}.$$

With the above two lemmas, we can show that for any pattern p with $|\operatorname{supp}_0(p)| = |\operatorname{supp}_1(p)| = s \leq k$ and any p-consistent affine function l, the probability that a random point in $\{z \in \{0,1\}^n : |z| = n/2\}$ matches the pattern p and satisfies $l(z) \geq 0$ with probability not too small. So, by adding enough number of random points in the middle part M, such a point z exists in M with high probability for a fixed p and l. Then, we would like to use a union bound over p and l to prove that there will be no edges between L and R in the graph of the polytope with constant probability.

One technical issue of this approach is that there are infinitely many affine functions $l : \mathbb{R}^n \to \mathbb{R}$. Note, however, that we only care about the values of l on the hypercube vertices. This reduces the number of different functions to 2^{2^n}. Indeed, we only care about whether $l(z) \geq 0$ for $z \in \{0,1\}^n$ for Lemma 7.5. Therefore, we only need to apply a union bound over the set of linear threshold functions over the boolean hypercube, which further reduces the number of different such functions to 2^{n^2}.

Proposition 7.8 ([Cov65]). *The number of linear threshold functions on $\{0,1\}^n$ is at most 2^{n^2}. A linear threshold function on $\{0,1\}^n$ is a function of the form $\tau : \{0,1\}^n \to \{0,1\}$, where*

$$\tau(u_1, \ldots, u_n) = \begin{cases} 1, & \text{if } \beta + \sum_i \alpha_i u_i \geq 0; \\ 0, & \text{if } \beta + \sum_i \alpha_i u_i < 0, \end{cases}$$

for some $\alpha_1, \ldots, \alpha_n, \beta \in \mathbb{R}$.

The complete proof of Theorem 1.12 can be found in Appendix D.

7.3 Mixing Time

Corollary 1.13 follows from Theorem 1.12 and the easy direction of Cheeger's inequality for vertex expansion in Theorem 1.3.

Proof of Corollary 1.13. Let $Q = L \cup M \cup R$ be a 0/1-polytope from Definition 7.1. The number of vertices of Q is $|L| + |M| + |R| \lesssim \binom{n}{k} \leq (en/k)^k$. By Theorem 1.12, the vertex expansion of the graph G_Q is $\psi(G_Q) \lesssim (4k)^k/n^{k-2}$. Therefore, by the easy direction in Theorem 1.3, the mixing time of any reversible chain $P \in \mathbb{R}^{|V| \times |V|}$ on G_Q with stationary distribution $\pi = \frac{1}{|V|}\vec{1}$ is at least

$$\tau^*(G) \gtrsim \frac{1}{\psi(G_Q)} \gtrsim \frac{n^{k-2}}{(4k)^k} \gtrsim \left(\frac{en}{k}\right)^{k-2} \gtrsim |V|^{1-\frac{2}{k}},$$

where the second last inequality is by the assumption that k is a constant and so only the exponent of n matters, and the last inequality is by $|V| \leq (en/k)^k$ explained above. $\square$

The implication of Corollary 1.13 has been discussed in Section 1.1.5 and we won't repeat here.

8 Tight Example to Cheeger's Inequality for Vertex Expansion

The goal of this section is to construct a family of tight examples to Theorem 1.4 when $\pi = \vec{1}/|V|$. The graphs constructed will have non-constant maximum degree d and satisfy

$$\frac{\psi(G)^2}{\log d} \asymp \lambda_2^*(G).$$

These examples are suggested to us by Shayan Oveis Gharan.

The graphs are realized as proximity graphs on $\S^{k-1}$. We employ the following notations: μ denotes the normalized Lebesgue measure on $\S^{k-1}$. $\Delta(\cdot, \cdot)$ denotes the geodesic distance on $\S^{k-1}$. For $\theta \in \S^{k-1}$ and $r > 0$, $B(\theta, r)$ denotes the set of points of distance less than r from θ. $Cap(r)$ denotes a generic spherical cap on $\S^{k-1}$ of radius r. For $S \subseteq \S^{k-1}$, we use S_r or $S + Cap(r)$ to denote the set of points of distance less than r from S.

Definition 8.1 (Spherical Proximity Graph). *Given k, a (γ, δ)-spherical proximity graph in k dimensions is a graph $G_k = (V, E)$ that can be constructed as follows:*

1. *Partition $\S^{k-1}$ into n cells: $S_1, S_2, \ldots, S_n$, such that each cell has diameter at most γ and measure between $\epsilon := \mu(Cap(\gamma/4))$ and 2ϵ. Take a point $x_i \in S_i$ for each $i \in [n]$.*

2. *Set $V = \{x_i\}_{i \in [n]}$ and $E = \{(x_i, x_j) \in V \times V : \Delta(x_i, x_j) < \delta\}$.*

We shall show that, for suitable choices of γ and δ, the resulting graph will be a vertex expander with $\lambda_2^* = O(1/\log d)$.

Theorem 8.2 (Tight Example of Theorem 1.4). *Given any $k \in \mathbb{N}$, if we set $\gamma = c_1/\sqrt{k}, \delta = c_2/\sqrt{k}$ for constants $0 < c_1 < c_2$, then for any (γ, δ)-spherical proximity graph G_k:*

- *the maximum degree d of G_k is $2^{O(k)}$;*

- *$\psi(G_k) = \Theta(1)$; and*

- *$\lambda_2^*(G_k) = O(1/\log d)$.*

The rest of the section is organized as follows. First, we show that the construction of the spherical proximity graph is indeed possible. Then, we prove the degree and expansion bounds, by relating them to volume ratios on the sphere. Finally, we prove the bound on reweighted eigenvalue, using the spherical embedding of the graph. All the proofs in this section are deferred to Appendix E.

8.1 The Graph Construction

The following proposition is modified from [GM12, Lemma 8.3.22] and shows that we can always construct a spherical proximity graph.

Proposition 8.3 (Constructing Spherical Proximity Graph). *For every k and $\gamma > 0$, there exists n such that $\S^{k-1}$ can be partitioned into cells $S_1, \ldots, S_n$ with $\operatorname{diam}(S_i) \le \gamma$ and $\mu(S_i) \in [\epsilon, 2\epsilon]$ for all $i \in [n]$, where $\epsilon := \mu(Cap(\gamma/4))$. Therefore, $n = \Theta(\mu(Cap(\gamma/4))^{-1})$.*

8.2 Degree and Expansion Bounds

Now, we establish bounds on the maximum degree and vertex expansion of G_k. We do so by connecting the quantities with the continuous notion of volume.

Recall our choice of parameters $\gamma = c_1/\sqrt{k}$ and $\delta = c_2/\sqrt{k}$, where $0 < c_1 < c_2$ are constants.

Lemma 8.4 (Degree, Expansion, and Volume). *Let $G = (V, E)$ be the graph constructed above.*

- *The maximum degree of G is at most*

$$\frac{\mu(Cap(\delta + \gamma))}{\mu(Cap(\gamma/4))}.$$

- *For any $T \subseteq V$ with $|T| \leq |V|/2$, if we let $S_T := \cup_{i \in T} S_i$, then*

$$\psi(T) \geq \frac{\mu(S_T + Cap(\delta - \gamma)) - \mu(S_T)}{2\mu(S_T)}.$$

After relating the graph degree and vertex expansion to volumes of portions of sphere, we shall now use results from high-dimensional geometry to obtain bounds on the maximum degree and vertex expansion.

Proposition 8.5 (Degree Bound). *For constants $0 < c_1 < c_2$, it holds that*

$$\frac{\mu(Cap(\delta + \gamma))}{\mu(Cap(\gamma/4))} \leq 2^{O(k)}.$$

Proposition 8.6 (Small-Volume Expansion of the Sphere). *For constants $0 < c_1 < c_2$ with $c_2 - c_1$ sufficiently large, it holds that for all $T \subseteq V$ with $|T| \leq |V|/2$,*

$$\frac{\mu(S_T + Cap(\delta - \gamma)) - \mu(S_T)}{\mu(S_T)} \geq \Omega(1).$$

Combining Lemma 8.4, Proposition 8.5, and Proposition 8.6 gives the first two parts of Theorem 8.2.

8.3 Reweighted Eigenvalue Bound

Finally, we bound $\lambda_2^*(G)$. On a high level, if a graph can be embedded in a unit sphere via $i \mapsto v_i$, such that the centre of mass is the origin and that each edge ij satisfies $\|v_i - v_j\| \leq \delta$, then the embedding certifies that $\lambda_2^*(G) \leq \delta^2$. This is basically the proof. The details can be found in Appendix E.

Proposition 8.7 (Reweighted Eigenvalue Bound). *For constants $0 < c_1 < c_2$, it holds that*

$$\lambda_2^*(G_n) \lesssim \frac{1}{k} \lesssim \frac{1}{\log d}.$$

Proposition 8.7 gives the last part of Theorem 8.2. Thus, the proof of Theorem 8.2 is complete.

9   Concluding Remarks

We present a new spectral theory which relates (i) reweighted eignevalues, (ii) vertex expansion and (iii) fastest mixing time. This is analogous to the classical spectral theory which relates (i) eigenvalues, (ii) edge conductance and (iii) mixing time. This spectral approach for vertex expansion has the advantage that most existing results and proofs for edge conductances and eigenvalues have a close analog for vertex expansion and reweighted eigenvalues with almost tight bounds. We do not intend to be exhaustive in this paper, and we fully expect that other results relating eigenvalues and edge conductances also have an analog for vertex expansion using reweighted eigenvalues.

To conclude, we believe that our work provides an interesting spectral theory for vertex expansion, as the formulations have the natural interpretation as reweighted eigenvalues and also have close connections to other important problems such as fastest mixing time and the reweighting conjectures in approximation algorithms. We also believe that this approach can be extended further for hypergraph edge expansion.

Acknowledgements

A   Deferred Proofs for Weighted Vertex Expansion

Proof of Lemma 3.16. First we consider the case that there is an optimizer $S \subseteq V$ to $\psi(G)$, with $0 < \pi(S) \le 1/2$ and $\pi(\partial S)/\pi(S) = \psi(G)$. For $a, b \in \mathbb{R}$ to be chosen below, we define a solution f, g to $\gamma^{(1)}(G)$ in Definition 3.2 as follows:

$$f(v) := \begin{cases} a, & \text{if } v \in S; \\ b, & \text{if } v \notin S, \end{cases} \qquad g(v) := \begin{cases} (a-b)^2, & \text{if } v \in \partial S; \\ 0, & \text{otherwise.} \end{cases}$$

Then the constraints $g(u) + g(v) \ge (f(u) - f(v))^2$ for $uv \in E$ are satisfied by construction. Note that we can always choose $a, b \neq 0$ that satisfy the two equations $\pi(S) \cdot a + (1 - \pi(S)) \cdot b = 0$ and $\pi(S) \cdot a^2 + (1 - \pi(S)) \cdot b^2 = 1$ simultaneously. This choice ensures that the two constraints $\sum_{v \in V} \pi(v) f(v) = 0$ and $\sum_{v \in V} \pi(v) f(v)^2 = 1$ are satisfied, and so f, g as defined is a feasible solution to $\gamma^{(1)}(G)$. The objective value is

$$\sum_{v \in V} \pi(v) g(v) = \pi(\partial S) \cdot (a-b)^2 = \psi(G) \cdot \pi(S) \cdot (a-b)^2 \le 2\psi(G) \cdot \pi(S) \cdot (a^2 + b^2) \le 2\psi(G),$$

where the last inequality uses that $\pi(S) \le 1/2$ and so $\pi(S) \cdot (a^2 + b^2) \le \pi(S) \cdot a^2 + (1 - \pi(S)) \cdot b^2 = 1$. This proves $\gamma^{(1)}(G) \le 2\psi(G)$ in this case.

The other case is when $\psi(G) = 1$. We will show that $\gamma^{(1)}(G) \le 2$ and this would imply that $\gamma^{(1)}(G) \le 2\psi(G)$. Let v be a vertex with $0 < \pi(v) \le 1/2$, which must exist as long as the graph has at least two vertices. We define a solution f, g to $\gamma^{(1)}(G)$ as above with $S = \{v\}$. Following the same arguments, f, g is a feasible solution to $\gamma^{(1)}(G)$ with objective value

$$\sum_{v \in V} \pi(v) g(v) = \pi(v)(a-b)^2 \le 2\pi(v)(a^2 + b^2) \le 2(\pi(v)a^2 + (1 - \pi(v))b^2) = 2.$$

This proves $\gamma^{(1)} \le 2\psi(G)$ in the other case when $\psi(G) = 1$. $\qquad \square$

Proof of Lemma 3.17. First we shift y so that both $\pi(\{v \in V : y(v) > 0\})$ and $\pi(\{v \in V : y(v) < 0\}|)$ are at most $1/2$, which is always possible thanks to the constraint $\sum_{v \in V} \pi(v) y(v) = 0$. By shifting, the objective value of the solution does not change. Let y^+ and y^- be the solutions with $y^+(v) := \max\{y(v), 0\}$ and $y^-(v) := \min\{y(v), 0\}$ for $v \in V$. Note that the objective values for y^+ and y^- are at most that for y, as $|y^+(u) - y^+(v)| \le |y(u) - y(v)|$ and $|y^-(u) - y^-(v)| \le |y(u) - y(v)|$. By Lemma A.1, either $\sum_{v \in V} \pi(v) y^+(v)^2$ or $\sum_{v \in V} \pi(v) y^-(v)^2$ is at least $1/4$.

We define x to be y^+ if $\sum_{v \in V} \pi(v) y^+(v)^2 \ge \sum_{v \in V} \pi(v) y^-(v)^2$, or otherwise we define x to be $-y^-$. Then $x \ge 0$ and $\pi(\text{supp}(x)) \le 1/2$. The numerator is at most the objective value of y which is $\overrightarrow{\gamma}^{(1)}(G)$, and the denominator is at least $1/4$ by Lemma A.1 and so the lemma follows. $\qquad \square$

Lemma A.1. *Given a finite set X, a vector $f : X \to \mathbb{R}$, a distribution π on X. Let*

$$\text{Var}_\pi(f) := \sum_{i \in X} \pi(i) f(i)^2 - \left(\sum_{i \in X} \pi(i) f(i) \right)^2$$

be the π-variance of f. Let $f^+ := \max(f, 0)$ and $f^- := \max(-f, 0)$ be the positive and negative parts of f respectively (so that $f = f^+ - f^-$). Then,

$$2\big(\operatorname{Var}_\pi(f^+) + \operatorname{Var}_\pi(f^-)\big) \geq \operatorname{Var}_\pi(f).$$

Proof. Let $S^+ \subseteq X := \{i \in X : f(i) \geq 0\}$ and $S^- \subseteq X := \{i \in X : f(i) < 0\}$. Suppose $\pi(S^+), \pi(S^-) > 0$; otherwise either $f^+ = f$ or $f^- = f$, and there is nothing to prove. Then

$$
\begin{aligned}
\operatorname{Var}_\pi(f^+) &= \sum_{i \in S^+} \pi(i) f(i)^2 - \left(\sum_{i \in S^+} \pi(i) f(i) \right)^2 \\
&= \pi(S^+) \sum_{i \in S^+} \frac{\pi(i)}{\pi(S^+)} f(i)^2 - \pi(S^+)^2 \left(\sum_{i \in S^+} \frac{\pi(i)}{\pi(S^+)} f(i) \right)^2 \\
&= \pi(S^+)\pi(S^-) \sum_{i \in S^+} \frac{\pi(i)}{\pi(S^+)} f(i)^2 + \pi(S^+)^2 \left[\sum_{i \in S^+} \frac{\pi(i)}{\pi(S^+)} f(i)^2 - \left(\sum_{i \in S^+} \frac{\pi(i)}{\pi(S^+)} f(i) \right)^2 \right]
\end{aligned}
$$

Similarly,

$$
\operatorname{Var}_\pi(f^-) = \pi(S^-)\pi(S^+) \sum_{i \in S^-} \frac{\pi(i)}{\pi(S^-)} f(i)^2 + \pi(S^-)^2 \left[\sum_{i \in S^-} \frac{\pi(i)}{\pi(S^-)} f(i)^2 - \left(\sum_{i \in S^-} \frac{\pi(i)}{\pi(S^-)} f(i) \right)^2 \right]
$$

Therefore,

$$
\begin{aligned}
&\operatorname{Var}_\pi(f) \\
&= \sum_{i \in X} \pi(i) f(i)^2 - \left(\sum_{i \in X} \pi(i) f(i) \right)^2 \\
&= \pi(S^+) \sum_{i \in S^+} \frac{\pi(i)}{\pi(S^+)} f(i)^2 + \pi(S^-) \sum_{i \in S^-} \frac{\pi(i)}{\pi(S^-)} f(i)^2 \\
&\quad - \left(\pi(S^+) \sum_{i \in S^+} \frac{\pi(i)}{\pi(S^+)} f(i) + \pi(S^-) \sum_{i \in S^-} \frac{\pi(i)}{\pi(S^-)} f(i) \right)^2 \\
&= \operatorname{Var}_\pi(f^+) + \operatorname{Var}_\pi(f^-) - 2\pi(S^+)\pi(S^-) \left(\sum_{i \in S^+} \frac{\pi(i)}{\pi(S^+)} f(i) \right) \left(\sum_{i \in S^-} \frac{\pi(i)}{\pi(S^-)} f(i) \right) \\
&\leq \operatorname{Var}_\pi(f^+) + \operatorname{Var}_\pi(f^-) + \pi(S^+)\pi(S^-) \left[\left(\sum_{i \in S^+} \frac{\pi(i)}{\pi(S^+)} f(i) \right)^2 + \left(\sum_{i \in S^-} \frac{\pi(i)}{\pi(S^-)} f(i) \right)^2 \right] \\
&\leq \operatorname{Var}_\pi(f^+) + \operatorname{Var}_\pi(f^-) + \pi(S^+)\pi(S^-) \left[\left(\sum_{i \in S^+} \frac{\pi(i)}{\pi(S^+)} f(i)^2 \right) + \left(\sum_{i \in S^-} \frac{\pi(i)}{\pi(S^-)} f(i)^2 \right) \right] \\
&\leq 2\big(\operatorname{Var}_\pi(f^+) + \operatorname{Var}_\pi(f^-) \big),
\end{aligned}
$$

$\square$

Proof of Theorem 3.15. The easy direction $\gamma^{(1)}(G) \lesssim \psi(G)$ is proved in Lemma 3.16. For the hard direction, given a solution y and $\overrightarrow{E}$ to $\overrightarrow{\gamma}^{(1)}(G)$, we apply the truncation step in Lemma 3.17 to obtain $x \geq 0$ with $\pi(\mathrm{supp}(x)) \leq 1/2$ and $\left(\sum_{v \in V} \pi(v) \max_{u:u \to v}(x(u) - x(v))^2\right)/\left(\sum_{v \in V} \pi(v) x(v)^2\right) \leq 4\overrightarrow{\gamma}^{(1)}(G)$. Then, we apply the threshold rounding step in Proposition 3.19 on x to obtain a set S with $S \subseteq \mathrm{supp}(x)$ and $\overrightarrow{\psi}(S) \lesssim \sqrt{\overrightarrow{\gamma}^{(1)}(G)}$. If $\overrightarrow{\psi}(S) \geq 1/2$, then it implies that $\overrightarrow{\gamma}^{(1)}(G) = \Omega(1)$, and so the inequality $\psi(G)^2 \lesssim \overrightarrow{\gamma}^{(1)}(G)$ holds trivially as $\psi(G) \leq 1$ by definition. Otherwise, if $\overrightarrow{\psi}(S) < 1/2$, we apply the postprocessing step in Lemma 3.20 on S to obtain $S' \subseteq S$ with $\psi(S') \leq 2\overrightarrow{\psi}(S)$. Therefore, S' is a set with $\pi(S') \leq \pi(S) \leq \pi(\mathrm{supp}(x)) \leq 1/2$ and $\psi(S') \leq 2\overrightarrow{\psi}(S) \lesssim \sqrt{\overrightarrow{\gamma}^{(1)}(G)}$. Thus we conclude the hard direction that $\psi(G) \leq \psi(S') \lesssim \sqrt{\overrightarrow{\gamma}^{(1)}(G)} \lesssim \sqrt{\gamma^{(1)}(G)}$. $\square$

B Deferred Proofs for Bipartite Vertex Expansion

Proof of Lemma 4.8. First we consider the case that there is an optimizer $S \subseteq V$ to $\psi_B(G)$, with bipartition $S = (S_1, S_2)$ and $\pi(\partial S)/\pi(S) = \psi_B(G)$. Define $f : V \to \mathbb{R}$ and $g : V \to \mathbb{R}$ as follows:

- $f(u) = 1$ if $u \in S_1$, and $f(u) = -1$ if $u \in S_2$, and $f(u) = 0$ if $u \notin S$;

- $g(u) = 1$ if $u \in \partial S$, and $g(u) = 0$ otherwise.

For each edge $uv \in E$, we claim that $g(u) + g(v) \geq (f(u) + f(v))^2$. Note that u and v cannot both belong to S_1 or both belong to S_2. One can check that the constraint is satisfied in all the remaining cases: (1) $u \in S_1$ and $v \in S_2$ (or vice versa), (2) $u \in S$ and $v \notin S$ (or vice versa), and (3) $u, v \notin S$. So, (f, g) is a feasible solution to the $\nu^{(1)}(G)$ program, and the objective value is

$$\frac{\sum_{v \in V} \pi(v) g(v)}{\sum_{v \in V} \pi(v) f(v)^2} = \frac{\sum_{v \in \partial S} \pi(v)}{\sum_{v \in S_1 \cup S_2} \pi(v)} = \frac{\pi(\partial S)}{\pi(S)} = \psi_B(S).$$

This implies that $\nu^{(1)}(G) \leq 2\psi_B(S)$ in this case. The other case is when $\psi_B(G) = 1$. Choosing the feasible solution $f \equiv 1$ and $g \equiv 2$ to the $\nu^{(1)}(G)$ program shows that $\nu^{(1)}(G) \leq 2$. This implies that $\nu^{(1)}(G) \leq 2\psi_B(G)$ in the other case. $\square$

Proof of Theorem 4.4. The easy direction $\nu^{(1)}(G) \lesssim \psi_B(G)$ is proved in Lemma 4.8. For the hard direction, given a solution x and $\overrightarrow{E}$ to $\overrightarrow{\nu}^{(1)}(G)$, we apply the threshold rounding step in Proposition 4.10 on x to obtain two disjoint sets $S_1, S_2 \subseteq V$ with $\overrightarrow{\psi}(S_1, S_2) \lesssim \sqrt{\overrightarrow{\nu}^{(1)}(G)}$. If $\overrightarrow{\psi}(S_1, S_2) \geq 1/2$, then it implies that $\overrightarrow{\nu}^{(1)}(G) = \Omega(1)$, and so the inequality $\psi_B(G)^2 \lesssim \overrightarrow{\nu}^{(1)}(G)$ holds trivially as $\psi_B(G) \leq 1$ by definition. Otherwise, if $\overrightarrow{\psi}(S_1, S_2) < 1/2$, we apply the postprocessing step in Lemma 4.11 on (S_1, S_2) to obtain (S_1', S_2') so that $S_1' \cup S_2'$ is an induced bipartite graph in G and $\psi(S_1' \cup S_2') \leq 2\overrightarrow{\psi}(S_1, S_2)$. Thus we conclude the hard direction that $\psi_B(G) \leq \psi_B(S_1' \cup S_2') \leq 2\overrightarrow{\psi}(S_1, S_2) \lesssim \sqrt{\overrightarrow{\nu}^{(1)}(G)}$. $\square$

C Deferred Proofs for Multiway Vertex Expansion

Proof of Proposition 5.9. By the Laurent-Massart bound of χ-squared distribution (c.f. Lemma 1 of [LM00]), for any $\delta > 0$ and $i \leq d$,

$$\Pr\left[Y_i - 1 \geq 2\sqrt{\delta/m} + 2(\delta/m)\right] \leq e^{-\delta}.$$

Since

$$1 + 2\sqrt{\delta/m} + 2(\delta/m) \leq 4 + 4(\delta/m)$$

it follows that

$$\Pr[Y_i \geq 4 + 4(\delta/m)] \leq e^{-\delta}.$$

Recall that $Y = \max_{i \leq d} Y_i$. Taking union bound over the Y_i's,

$$\Pr[Y \geq 4 + 4(\delta/m)] \leq d \cdot e^{-\delta}.$$

With this probability tail bound, the expectation of Y can be bounded as follows:

$$
\begin{aligned}
\mathbb{E}[Y] &= \int_0^\infty \Pr[Y \geq t]\, dt \;\leq\; 4 + \int_0^\infty \Pr[Y \geq 4 + t]\, dt \;\leq\; 4 + \int_0^\infty \min\left\{1, d \cdot e^{-mt/4}\right\} dt \\
&= 4 + \frac{4\log d}{m} + \int_{\frac{4\log d}{m}}^\infty d \cdot e^{-mt/4}\, dt \;=\; 4 + \frac{4\log d}{m} - \frac{4d}{m} \cdot e^{-mt/4}\Big|_{\frac{4\log d}{m}}^\infty \;=\; 4 + \frac{4\log d}{m} + \frac{4}{m}.
\end{aligned}
$$

$\square$

Proof of Lemma 5.19. If $\psi_k(G) \geq 1$, then the lemma holds trivially as $\lambda_k^*(G) \leq 2$. Henceforth, we assume $\psi_k(G) < 1$, and there are nonempty disjoint subsets $S_1, \ldots, S_k \subseteq V$ with $\max_{1 \leq i \leq k} \psi(S_i) = \psi_k(G)$.

Using the notation in Proposition 5.2, the λ_k^* program in Definition 1.9 can be written as

$$
\begin{aligned}
\lambda_k^*(G) \;:=\; \max_{Q \geq 0} \; \min_{f_1,\ldots,f_k : V \to \mathbb{R}} \; \max_{1 \leq i \leq k} \quad & f_i^T(\Pi - Q)f_i \\
\text{subject to} \quad & \sum_v \pi(v) f_i(v)^2 = 1 && \forall 1 \leq i \leq k \\
& \sum_v \pi(v) f_i(v) f_j(v) = 0 && \forall 1 \leq i \neq j \leq k \\
& Q(u,v) = 0 && \forall uv \notin E \\
& \sum_{v \in V} Q(u,v) = \pi(u) && \forall u \in V \\
& Q(u,v) = Q(v,u) && \forall uv \in E.
\end{aligned}
$$

Each $f_i^T(\Pi - Q)f_i$ can be written as

$$\sum_{uv \in E} Q(u,v)\big(f_i(u) - f_i(v)\big)^2 = \sum_{uv \in E} \pi(u)P(u,v)\big(f_i(u) - f_i(v)\big)^2.$$

We set

$$f_i(u) := \begin{cases} \frac{1}{\sqrt{\pi(S_i)}}, & \text{if } u \in S_i \\ 0, & \text{otherwise.} \end{cases}$$

We can check that the constraints on f_i are satisfied. Moreover, for any P satisfying the constraints,

$$\sum_{uv \in E} \pi(u)P(u,v)\big(f_i(u) - f_i(v)\big)^2 \;=\; \sum_{v \in S_i, u \in \partial(S_i)} \pi(u)P(u,v)\frac{1}{\pi(S_i)}$$

$$\leq \; \frac{1}{\pi(S_i)} \sum_{v \in V, u \in \partial(S_i)} \pi(u)P(u,v) \;=\; \frac{\pi(\partial(S_i))}{\pi(S_i)} \;=\; \psi(S_i).$$

So, $\max_{1 \leq i \leq k} f_i^T(\Pi - \Pi P)f_i \leq \max_{1 \leq i \leq k} \psi(S_i)$. Taking maximum over P gives $\lambda_k^*(G) \leq \psi_k(G)$.
$\square$

D Deferred Proofs for 0/1-Polytopes with Poor Vertex Expansion

Proof of Lemma 7.4. The plan is to prove that the stated conditions imply $\frac{1}{2}(x + y) \in \operatorname{conv}(M)$, which would then immediately imply that there is no edge connecting x and y in Q. We will prove the contrapositive: if $\frac{1}{2}(x + y) \notin \operatorname{conv}(M)$, then there is a p-consistent affine function l such that $l(z) < 0$ for all $z \in M$ matching the pattern p.

Denote $w := \frac{1}{2}(x + y)$. As $w \notin \operatorname{conv}(M)$, by Proposition 2.6, there is an affine function $l' : (u_1, \ldots, u_n) \mapsto \beta' + \sum_i \alpha_i' u_i$ such that $l'(w) = 0$ and $l'(z) < 0$ for all $z \in M$. We would like to modify l' to obtain an affine function $l : (u_1, \ldots, u_n) \mapsto \beta + \sum_i \alpha_i u_i$ such that (i) $l(w) = 0$, (ii) $\alpha_i = 0$ for $i \in \operatorname{supp}(p)$, and (iii) $l(z) < 0$ for all $z \in M$ matching the pattern p.

Note that, by the definition of common pattern, for $i \in \operatorname{supp}(p)$, either $x_i = y_i = 1$ or $x_i = y_i = 0$, and so $w_i \in \{0, 1\}$. Also, for any $z \in \{0, 1\}^n$ that matches the pattern p, we must have $z_i = w_i$ for $i \in \operatorname{supp}(p)$. So, for any such z,

$$l'(z) = \beta' + \sum_{i=1}^n \alpha_i' z_i = \left(\beta' + \sum_{i \in \operatorname{supp}(p)} \alpha_i' z_i\right) + \sum_{i \notin \operatorname{supp}(p)} \alpha_i' z_i = \left(\beta' + \sum_{i \in \operatorname{supp}(p)} \alpha_i' w_i\right) + \sum_{i \notin \operatorname{supp}(p)} \alpha_i' z_i.$$

Hence, if we set $\beta := \beta' + \sum_{i \in \operatorname{supp}(p)} \alpha_i' w_i$, and $\alpha_i = \alpha_i'$ for $i \notin \operatorname{supp}(p)$ and $\alpha_i = 0$ for $i \in \operatorname{supp}(p)$, then the affine function $l : (u_1, u_2, \ldots, u_n) \in \mathbb{R}^n \mapsto \beta + \sum_i \alpha_i u_i$ satisfies $l(z) = l'(z)$ for any z that matches the pattern p. Therefore, l is an affine function that satisfies the three properties that (i) $l(w) = l'(w) = 0$, (ii) $\alpha_i = 0$ for $i \in \operatorname{supp}(p)$, and (iii) $l(z) = l'(z) < 0$ for $z \in M$ matching the pattern p.
$\square$

Proof of Lemma 7.5. Note that for $x \in L$ and $y \in R$ in our construction in Definition 7.1, their common pattern p satisfies $|\operatorname{supp}_0(p)| = |\operatorname{supp}_1(p)| \leq k$. The lemma follows by applying Lemma 7.4 on all possible such patterns.
$\square$

Proof of Lemma 7.6. Let $z \in \{0,1\}^n$ be such that $|z| = \frac{n}{2}$ and z matching the pattern p. Let $z^\tau \in \{0,1\}^n$ be its "opposite point" formed by toggling the coordinates of z_i for $i \notin \text{supp}(p)$ and leaving other coordinates unchanged. Note that the lemma follows from the following two facts: (i) $|z^\tau| = \frac{n}{2}$ and z^τ matches the pattern p, and (ii) $l(z) + l(z^\tau) = 0$.

For the first fact, z^τ matches the pattern p because $z_i^\tau = z_i$ for $i \in \text{supp}(p)$. And $|z^\tau| = \frac{n}{2}$ because $|z| = \frac{n}{2}$ and there are the same number of zeroes and ones in p, the latter being a consequence of $|x| + |y| = k + (n-k) = n$.

For the second fact, as l is p-consistent,

$$l(z) + l(z^\tau) = \left(\beta + \sum_{i \notin \text{supp}(p)} \alpha_i z_i\right) + \left(\beta + \sum_{i \notin \text{supp}(p)} \alpha_i z_i^\tau\right) = 2\left(\beta + \frac{1}{2}\sum_{i \notin \text{supp}(p)} \alpha_i\right) = 0,$$

where the last equality is from the second condition in Definition 7.3. $\qquad\square$

Proof of Lemma 7.7. The number of points $z \in \{0,1\}^n$ with $|z| = \frac{n}{2}$ is $\binom{n}{n/2}$, whereas the number of such points that matches pattern p is $\binom{n-2s}{n/2-s}$. Therefore,

$$\Pr_{z \sim Z}[z \text{ matches pattern } p] = \binom{n-2s}{n/2-s} \Big/ \binom{n}{n/2} \gtrsim \left(\sqrt{\frac{2}{\pi(n-2s)}} \cdot 2^{n-2s}\right) \Big/ \left(\sqrt{\frac{2}{\pi n}} \cdot 2^n\right) \geq 4^{-s},$$

where we used Stirling's approximation $n! \approx_n \sqrt{2\pi n}(n/e)^n$. $\qquad\square$

Proof of Theorem 1.12. Let Q be a 0/1-polytope from Definition 7.1. We would like to apply Lemma 7.5 to prove the theorem.

Let Z be the uniform distribution on $\{z \in \{0,1\}^n : |z| = n/2\}$. Combining Lemma 7.6 and Lemma 7.7, it follows that for any pattern p with $|\text{supp}_0(p)| = |\text{supp}_1(p)| = s \leq k$ and any p-consistent affine function l,

$$\Pr_{z \sim Z}[l(z) \geq 0 \text{ and } z \text{ matches pattern } p] \geq c \cdot 4^{-s}$$

for some universal constant $c > 0$. Therefore, if we take independent samples $z_1, \ldots, z_m \sim Z$ and set $M := \{z_1, \ldots, z_m\}$ where m is a value to be determined later, then for any pattern p with $|\text{supp}_0(p)| = |\text{supp}_1(p)| = s \leq k$ and any p-consistent affine function l,

$$\Pr[\nexists z \in M \text{ with } l(z) \geq 0 \text{ and } z \text{ matching pattern } p] \leq (1 - c \cdot 4^{-s})^m \leq (1 - c \cdot 4^{-k})^m.$$

To apply a union bound, we upper bound the numbers of different such p and l. The number of patterns p with $|\text{supp}_0(p)| = |\text{supp}_1(p)| = s \leq k$ is

$$\sum_{s=0}^{k} \binom{n}{2s} \cdot \binom{2s}{s} \leq (k+1) \cdot n^{2k}.$$

The number of p-consistent affine functions l with different sign patterns on the boolean hypercube $\{0,1\}^n$ is upper bounded by the number of affine threshold functions on $\{0,1\}^n$, which is at most

2^{n^2} by Proposition 7.8. Combining the two estimates and the above probability bound, the failure probability is

$$\Pr[\exists p \; \exists l \text{ s.t. } \nexists z \in M \text{ with } l(z) \geq 0 \text{ and } z \text{ matching pattern } p] \leq (1 - c \cdot 4^{-k})^m \cdot (k+1) \cdot n^{2k} \cdot 2^{n^2}.$$

Setting

$$m = \frac{4^k}{c} \cdot (1 + \log(k+1) + 2k \log n + n^2 \log 2) \lesssim 4^k n^2,$$

the failure probability is at most

$$
\begin{aligned}
& (1 - c \cdot 4^{-k})^m \cdot (k+1) \cdot n^{2k} \cdot 2^{n^2} \\
\leq \quad & \exp(-c \cdot 4^{-k} \cdot m) \cdot \exp(\log(k+1) + 2k \log n + n^2 \log 2) \\
\leq \quad & \exp(-(1 + \log(k+1) + 2k \log n + n^2 \log 2)) \cdot \exp(\log(k+1) + 2k \log n + n^2 \log 2) \\
= \quad & e^{-1}.
\end{aligned}
$$

Therefore, by Lemma 7.5, we conclude that there exists $M \subseteq \{z \in \{0,1\}^n : |z| = n/2\}$ with $|M| \lesssim 4^k n^2$ such that there are no edges between L and R in the graph of Q.

Finally, as $|L| = \binom{n}{k} \geq (n/k)^k$ and $\partial L \subseteq M$, it follows that

$$\psi(L) \leq \frac{|M|}{|L|} \lesssim \frac{4^k n^2}{(n/k)^k} = \frac{(4k)^k}{n^{k-2}}.$$

$\square$

E   Deferred Proofs for Spherical Proximity Graph

Proof of Proposition 8.3.   The construction largely follows that of [GM12, Lemma 8.3.22]. Given n and $\gamma > 0$, we iteratively choose points $y_1, y_2, \ldots, y_m \in \S^{k-1}$, such that each new point y_{i+1} has distance at least $\gamma/2$ from $y_1, \ldots, y_i$. We stop when it is no longer possible to choose a point that is $(\gamma/2)$-far from all existing points.

We now let $S'_1, S'_2, \ldots, S'_m$ be the cells of the Voronoi diagram of $y_1, y_2, \ldots, y_m$. That is, for any $x \in \S^{k-1}$, $x \in S'_i$ iff $\Delta(x, y_i) = \min_{j \in [m]} \Delta(x, y_j)$. Note that cell S'_i contains $B(y_i, \gamma/4)$ and is contained in $B(y_i, \gamma/2)$. Therefore, the measure of each S'_i is at least $\epsilon = \mu(Cap(\gamma/4))$ and the diamater of each S'_i is at most γ. By further subdividing the cells (evenly) until each cell has measure $\leq 2\epsilon$, we obtain $S_1, S_2, \ldots, S_n$, such that the measure of each S_i is between ϵ and 2ϵ, and $\mathrm{diam}(S_i) \leq \gamma$ for each $i \in [n]$. We can choose the points $x_i \in S_i$ arbitrarily. $\square$

Proof of Lemma 8.4.   First, we prove the bound on the maximum degree. For any vertex $i \in V$, its degree is equal to the number of points x_j that are δ-close to x_i. We can count the number of such points using volume estimation. If a point x_j is within distance δ from x_i, then the entire cell

S_j is within distance $\delta + \gamma$ from x_i. This contributes at least $\epsilon = \mu(Cap(\gamma/4))$ total measure to $B(x_i, \delta + \gamma)$. The total measure of these cells is at most $\mu(B(x_i, \delta + \gamma))$, so

$$|\partial(i)| \cdot \mu(Cap(\gamma/4)) \leq \mu(B(x_i, \delta + \gamma)).$$

Rearranging gives the desired upper bound on $|\partial(i)|$, and thus on the maximum degree.

Next, we prove the bound on the vertex expansion. Given any $T \subseteq V$, we wish to lower bound $|\partial(T)|$. For $j \in V$, if the cell S_j is completely contained in $S_T + Cap(\delta)$, then $x_j \in T \cup \partial(T)$. If we take the union of all such cells, then the set will contain $S_T + Cap(\delta - \gamma)$:

$$\cup\{S_j : S_j \subseteq S_T + Cap(\delta)\} \supseteq S_T + Cap(\delta - \gamma).$$

This is because, for any point $u \in S_T + Cap(\delta - \gamma)$, the cell that it is in will be contained in

$$\{u\} + Cap(\gamma) \subseteq S_T + Cap(\delta - \gamma) + Cap(\gamma) = S_T + Cap(\delta).$$

It follows that

$$\cup\{S_j : j \notin T \text{ and } S_j \subseteq S_T + Cap(\delta)\} \supseteq (S_T + Cap(\delta - \gamma)) \setminus S_T.$$

Combining previous observations, and since each cell has measure at most 2ϵ,

$$
\begin{aligned}
|\partial(T)| = |(T \cup \partial(T)) \setminus T| &\geq |\{S_j : j \notin T \text{ and } S_j \subseteq S_T + Cap(\delta)\}| \\
&\geq \frac{\mu(S_T + Cap(\delta - \gamma)) - \mu(S_T)}{2\epsilon}.
\end{aligned}
$$

We are done after substituting this and $|T| \leq \mu(S_T)/\epsilon$ into $\psi(T) = |\partial(T)|/|T|$. $\qquad \square$

Proof of Proposition 8.5. We will use the following formula for spherical cap volume:

$$\mu(Cap(x)) = \frac{\int_0^x \sin^{k-2} \theta \, d\theta}{\int_0^\pi \sin^{k-2} \theta \, d\theta}.$$

This comes from the formula for unnormalized spherical cap volume [Li11]:

$$\mu_0(Cap(x)) = \frac{2\pi^{(k-1)/2}}{\Gamma((k-1)/2)} \cdot \int_0^x \sin^{k-2} \theta \, d\theta.$$

We shall use the approximation

$$\sin^k \theta \approx_k \left(\theta - \frac{\theta^3}{3!}\right)^k$$

for $0 \leq \theta \leq O(1/\sqrt{k})$. The third-degree approximation is sufficient because $(\theta - \theta^3/3! + O(\theta^5))^k \approx_k (\theta - \theta^3/3!)^k$ for θ in this range.

Then, for $x = O(1/\sqrt{k})$,

$$\int_0^x \sin^k \theta \, d\theta \quad \approx_k \quad \int_0^x \left(\theta - \frac{\theta^3}{3!}\right)^k d\theta$$

$$\approx_k \quad \int_0^x \left(\theta - \frac{\theta^3}{3!}\right)^k \cdot (1 - \theta^2/2) \, d\theta \quad (\because (1 - \theta^2/2) \text{ is close to } 1)$$

$$\overset{y:=\theta-\theta^3/6}{=} \quad \int_0^{x-x^3/6} y^k \, dy \quad (\because y(\theta) := \theta - \theta^3/6 \text{ is increasing for } \theta \in [0, x])$$

$$= \quad \frac{(x - x^3/6)^{k+1}}{k+1}.$$

Therefore,

$$\frac{\mu(Cap(\delta + \gamma))}{\mu(Cap(\gamma/4))} \quad = \quad \frac{\int_0^{\delta+\gamma} \sin^{k-2} \theta \, d\theta}{\int_0^{\gamma/4} \sin^{k-2} \theta \, d\theta}$$

$$\approx_k \quad \left(\frac{\delta + \gamma}{\gamma/4}\right)^{k-1} \cdot \left(\frac{1 - (\delta + \gamma)^2/6}{1 - (\gamma/4)^2/6}\right)^{k-1}$$

$$= \quad \left(\frac{4(c_1 + c_2)}{c_1}\right)^{k-1} \cdot \left(\frac{1 - (c_1 + c_2)/6k}{1 - c_1/96k}\right)^{k-1}$$

$$\lesssim \quad \left(\frac{4(c_1 + c_2)}{c_1}\right)^{k}$$

$$\leq \quad 2^{O(k)}.$$

$\square$

Proof of Proposition 8.6.   By isoperimetric results on the sphere, the worst case is when the set S_T is a spherical cap. Therefore, it suffices to prove that, for any $\tau \in (0, \pi/2]$,

$$\mu(Cap(\tau + (\delta - \gamma))) \geq (1 + \Omega(1)) \cdot \mu(Cap(\tau)).$$

Let $c := c_2 - c_1 > 0$. This is equivalent to

$$\int_0^{\tau + c/\sqrt{k}} \sin^{k-2} \theta \, d\theta \geq (1 + \Omega(1)) \int_0^\tau \sin^{k-2} \theta \, d\theta.$$

For technical reasons, we first deal with the case where τ is close to $\pi/2$. For $\tau \geq \pi/2 - c/\sqrt{k}$, the result follows from well-known upper bounds on spherical cap volume. For example, we may use the upper bound in [Tko12]:

$$\mu(Cap(\pi/2 - \theta)) \leq e^{-k \sin^2 \theta/2}, \qquad \theta \in [0, \pi/2)$$

and the fact that $\sin\theta \approx_k \theta$ for $\theta = O(1/\sqrt{k})$. Then, if c is such that $\exp\left[-k\sin^2\left(\frac{c}{2\sqrt{k}}\right)/2\right] \leq 1/3$,

$$\frac{\mu(Cap(\tau + c/\sqrt{k}))}{\mu(Cap(\tau))} \geq \min\left(\frac{\mu(Cap(\pi/2))}{\mu(Cap(\pi/2 - c/2\sqrt{k}))}, \frac{\mu(Cap(\pi/2 + c/2\sqrt{k}))}{\mu(Cap(\pi/2))}\right) \geq \frac{4}{3}.$$

Therefore, we may assume that $\tau < \pi/2 - c/\sqrt{k}$.

We will actually prove the following relation: for all $x \in [0, \pi/2 - c/\sqrt{k}]$,

$$\int_x^{x+c/\sqrt{k}} \sin^{k-2}\theta\, d\theta \geq (1 + \Omega(1)) \int_{x-c/\sqrt{k}}^x \sin^{k-2}\theta\, d\theta.$$

If this relation is proven to be true, then by writing

$$\int_0^{\tau+c/\sqrt{k}} \sin^{k-2}\theta\, d\theta = \left(\int_\tau^{\tau+c/\sqrt{k}} + \int_{\tau-c/\sqrt{k}}^\tau + \cdots + \int_0^{\tau-tc/\sqrt{k}}\right)\sin^{k-2}\theta\, d\theta, \quad \left(t := \left\lfloor\frac{\tau\sqrt{k}}{c}\right\rfloor\right),$$

and applying the above relation to each term on RHS, we obtain the desired result.

In order to prove the relation, we show that, for all $x \in [0, \pi/2 - c/\sqrt{k}]$,

$$\sin^{k-2}(x + c/\sqrt{k}) \geq (1 + \Omega(1)) \cdot \sin^{k-2}x.$$

that is, we wish to show that the function

$$f(x) := \frac{\sin^{k-2}(x + c/\sqrt{k})}{\sin^{k-2}(x)}$$

is at least $1 + \Omega(1)$ for $x \in [0, \pi/2 - c/\sqrt{k}]$.

Check that, for $k \geq 3$, $f(x)$ is decreasing by differentiating $\sin(x + c/\sqrt{k})/\sin(x)$. It remains to compute

$$
\begin{aligned}
f(\pi/2 - c/\sqrt{k}) &= \frac{\sin^{k-2}(\pi/2)}{\sin^{k-2}(\pi/2 - c/\sqrt{k})} \\
&= [\cos(c/\sqrt{k})]^{-(k-2)} \\
&= \left(1 - \frac{c^2}{2k}\right)^{-(k-2)} + o(1) \qquad \text{(by Taylor expansion)} \\
&= \exp(c^2/2) + o_k(1).
\end{aligned}
$$

Therefore, for sufficiently large k, $f(x) \geq 1 + \Omega(1)$, where the constant depends on c. $\qquad\square$

Proof of Proposition 8.7. We wish to construct test vectors $\{v_i\}_{i \in [n]}$, such that $\sum_{i \in [n]} v_i = \vec{0}$ and

$$\frac{\sum_{(i,j) \in E} P(i,j) \|v_i - v_j\|^2}{\sum_{i \in [n]} \|v_i\|^2}$$

is small for any doubly stochastic reweighting P of the graph G_n.

We claim that setting $v_i := x_i - z$ works, where $z := \frac{1}{n} \sum_{j \in [n]} x_j$. By construction, $\sum_{i \in [n]} v_i = \vec{0}$. We next show that

$$\sum_{i \in [n]} \|v_i\|^2 \geq \Omega(n).$$

Since $\sum_{i \in [n]} v_i = \vec{0}$, note that

$$\sum_{i \in [n]} \|v_i\|^2 = \frac{1}{2n} \sum_{i,j \in [n]} \|v_i - v_j\|^2 = \frac{1}{2n} \sum_{i,j \in [n]} \|x_i - x_j\|^2.$$

It then suffices to show that

$$\sum_{j \in [n]} \|x_i - x_j\|^2 \geq \Omega(n)$$

for all $i \in [n]$. This follows from the fact that at least $\frac{n}{3}$ of the points x_j are of geodesic distance at least $\frac{\pi}{2} - \gamma \geq \Omega(1)$ from x_i, and so are of Euclidean distance $\Omega(1)$ from x_i.

Therefore, for any doubly stochastic reweighting P,

$$\lambda_2(P) \leq \frac{\sum_{(i,j) \in E} P(i,j) \|v_i - v_j\|^2}{\sum_{i \in V} \|v_i\|^2} \leq \frac{\sum_{i \in [n]} \sum_{j:(i,j) \in E} P(i,j) \cdot \delta^2}{\Omega(n)} = O(\delta^2).$$

Note that, in the second inequality, we used the fact that straight-line distance in $\mathbb{R}^k$ is less than or equal to geodesic distance in $\S^{k-1}$.

We conclude that $\lambda_2^*(G) \lesssim \delta^2 \lesssim 1/k$. Combining with the degree bound in Proposition 8.5,

$$\lambda_2^*(G) \lesssim \frac{1}{k} \lesssim \frac{1}{\log d}.$$

$\square$

References

[Alo86] Noga Alon. *Eigenvalues and expanders.* Combinatorica, 6, 83–96, 1986. 7, 12

[AM85] Noga Alon, Vitali Milman. λ_1, *isoperimetric inequalities for graphs, and superconcentrators.* Journal of Combinatorial Theory, Series B, 38(1), 73–88, 1985. 7, 12

[ALOV19] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, Cynthia Vinzant. *Log-concave polynomials II: high-dimensional walks and an FPRAS for counting bases of a matroid.* In Proceedings of the 51th Annual Symposium on Theory of Computing (STOC), 1–12, 2019. 5

[ABS10] Sanjeev Arora, Boaz Barak, David Steurer. *Subexponential algorithms for unique games and related problems.* In Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS), 563–572, 2010.

[AG11] Sanjeev Arora, Rong Ge. *New tools for graph coloring.* In Proceedings of the 14th International Workshop on Approximation Algorithms for Combinatorial Optimization (APPROX), 1–12, 2011. 2, 4

[ARV09] Sanjeev Arora, Satish Rao, Umesh Vazirani. *Expander flows, geometric embeddings and graph partitioning.* Journal of the ACM, 56(2):1–37, 2009. 7

[Blu94] Avrim Blum. *New approximation algorithms for graph coloring.* Journal of the ACM, 41:470–516, 1994. 4

[BHT00] S. Bobkov, C. Houdré, Prasad Tetali. λ_∞, *vertex isoperimetry and concentration.* Combinatorica 20(2):153–172, 2000. 3, 7, 8, 9, 16, 17, 20

[BDPX09] Stephen Boyd, Persi Diaconis, Pablo Parrilo, Lin Xiao. *Fastest mixing time on graphs with symmetries.* SIAM Journal on Optimization, 20(2):792–819, 2009. 1

[BDSX06] Stephen Boyd, Persi Diaconis, J. Sun, Lin Xiao. *Fastest mixing Markov chain on a path.* American Mathematical Monthly, 113(1):70–74, 2006. 1

[BDX04] Stephen Boyd, Persi Diaconis, Lin Xiao. *Fastest mixing markov chain on a graph.* SIAM review, 46(4):667–689, 2004. 1, 12, 13, 15

[CLTZ18] T.-H. Hubert Chan, Anand Louis, Zhihao Gavin Tang, Chenzi Zhang. *Spectral properties of hypergraph Laplacian and approximation algorithms.* Journal of the ACM, 65(3):1–48, 2018. 5, 8

[CTWZ17] T.-H. Hubert Chan, Zhihao Gavin Tang, Xiaowei Wu, Chenzi Zhang. *Diffusion operator and spectral analysis for directed hypergraph Laplacian.* CoRR abs/1711.01560, 2017. 8

[Che70] Jeff Cheeger. *A lower bound for the smallest eigenvalue of the Laplacian.* Problems in Analysis, Princeton University Press, 195–199, 1970. 12

[CA15] O. Cihan, M. Akar. *Fastest mixing reversible Markov chains on graphs with degree proportional stationary distributions.* Transactions on Automatic Control, 60(1):227–232, 2015. 1

[Cov65] Thomas M. Cover. *Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition.* IEEE Transactions on Electronic Computers, 3:326–334, 1965. 48

[FLT20] Majid Farhadi, Anand Louis, Prasad Tetali. *On the complexity of λ_∞, vertex expansion, and spread constant of trees.* In arXiv:2003.05582, 2020. 7, 16

[FM92] Tomás Feder, Milena Mihail. *Balanced matroids.* In Proceedings of the 24th Annual Symposium on Theory of Computing (STOC), 26–38, 1992. 5

[FHL08] Uriel Feige, MohammadTaghi Hajiaghayi, James R. Lee. *Improved approximation algorithms for minimum weight vertex separators.* SIAM Journal on Computing, 38(2):629–657, 2008. 7

[FK13] J.A. Fill, Jeff Kahn. *Comparision inequalities and fastest mixing Markov chains.* Annals of Applied Probability, 23(5):1778–1816, 2013. 1

[GM12] Bernd Gärtner and Jiri Matousek. *Approximation algorithms and semidefinite programming.* Springer, 2012. 49, 58

[Gil07] Rafael Gillmann. *0/1-polytopes: typical and extremal properties.* Ph.D. thesis, Technische Universität Berlin, 2007. 6, 46

[GKL03] Anupam Gupta, Robert Krauthgamer, James R. Lee. *Bounded geometries, fractals, and low-distortion embeddings.* In Proceedings of the 44th Annual Symposium on Foundations of Computer Science (FOCS), 534–543, 2003. 36

[JPV22] Vishesh Jain, Huy Tuan Pham, and Thuy-Duong Vuong. *Dimension reduction for maximum matchings and the Fastest Mixing Markov Chain.* In arXiv:2203.03858, 2022. 10

[Kai04] Volker Kaibel. *On the expansion of graphs of 0/1-polytopes.* The Sharpest Cut. MPS-SAIM series on optimization 4, 199–216, 2004. 5

[KR03] Volker Kaibel, Anja Remshagen. *On the graph-density of random 0/1-polytopes.* In Proceedings of the 6th International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM), 318–328, 2003. 14

[KLLOT13] Tsz Chiu Kwok, Lap Chi Lau, Yin Tat Lee, Shayan Oveis Gharan, Luca Trevisan. *Improved Cheeger's inequality: Analysis of spectral partitioning algorithms through higher order spectral gap.* In Proceedings of the 45th Annual Symposium on Theory of Computing (STOC), 11–20, 2013. 5, 9, 40, 41, 42, 43

[LM00] Beatrice Laurent, Pascal Massart. *Adaptive estimation of a quadratic functional by model selection.* Annals of Statistics, 1302–1338, 2000. 55

[LOT12] James R. Lee, Shayan Oveis Gharan, Luca Trevisan. *Multi-way spectral partitioning and higher-order Cheeger inequalities*. In Proceedings of the 44th Annual Symposium on Theory of Computing (STOC), 1117–1130, 2012. 4, 9, 29, 33, 35, 36

[LP17] David Levin and Yuval Peres. *Markov chains and mixing times*. American Mathematical Society, Volume 107, 2017. 11

[Li11] Shengqiao Li. *Concise formulas for the area and volume of a hyperspherical cap*. Asian Journal of Mathematics and Statistics, Volume 4, No. 1, 66–70, 2011. 59

[Lou15] Anand Louis. *Hypergraph Markov operators, eigenvalues and approximation algorthms*. In Proceedings of the 47th Annual Symposium on Theory of Computing (STOC), 713–722, 2015. 8

[LRTV12] Anand Louis, Prasad Raghavendra, Prasad Tetali, Santosh Vempala. *Many sparse cuts via higher eigenvalues*. In Proceedings of the 44th Annual ACM Symposium on Theory of Computing (STOC), 1131–1140, 2012. 4

[LRV13] Anand Louis, Prasad Raghavendra, Santosh Vempala. *The complexity of approximating vertex expansion*. In Proceedings of the 54th IEEE Annual Symposium on Foundations of Computer Secience (FOCS), 360–369, 2013. 2, 3, 7, 8, 9, 14, 16, 17, 19, 24

[OZ22] Sam Olesker-Taylor, Luca Zanetti. *Geometric bounds on the fastest mixing Markov chain*. In the 13th Innovations in Theoretical Computer Science Conference (ITCS 2022). 1, 2, 3, 9, 10, 14, 15, 17, 19, 20, 22, 38

[RS10] Prasad Raghavendra, David Steurer. *Graph expansion and the unique games conjecture*. In Proceedings of the 42nd Annual ACM Symposium on Theory of Computing (STOC), 755–764, 2010. 2

[Roc05] Sébastien Roch. *Bounding fastest mixing*. Electron. Commun. Probab. 10:282–296, 2005. 1, 9, 13, 14, 23, 29, 31

[Sim95] Stephen Simons. *Minimax theorems and their proofs*. Minimax and applications, 1–23, 1995. 13

[Ste10] David Steurer. *On the complexity of unique games and graph expansion*. Ph.D. thesis, Princeton University, 2010. 2

[Tan84] R.M. Tanner. *Explicit construction of concentrators from generalized N-gons*. SIAM J. Algebraic Discrete Methods 5: 287–294, 1984. 7

[Tko12] Tomasz Tkocz. *An upper bound for spherical caps*. The American Mathematical Monthly, Volume 119, No. 7, 606–607, 2012. 60

[Tre09] Luca Trevisan. *Max cut and the smallest eigenvalue*. In Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC), 263–272, 2009. 3, 9, 25, 26