Spectral Properties of Hypergraph Laplacian and Approximation Algorithms *

T-H. Hubert Chan† Anand Louis‡ Zhihao Gavin Tang† Chenzi Zhang†

Abstract

The celebrated Cheeger's Inequality [AM85, Alo86] establishes a bound on the edge expansion of a graph via its spectrum. This inequality is central to a rich spectral theory of graphs, based on studying the eigenvalues and eigenvectors of the adjacency matrix (and other related matrices) of graphs. It has remained open to define a suitable spectral model for hypergraphs whose spectra can be used to estimate various combinatorial properties of the hypergraph.

In this paper we introduce a new hypergraph Laplacian operator generalizing the Laplacian matrix of graphs. In particular, the operator is induced by a diffusion process on the hypergraph, such that within each hyperedge, measure flows from vertices having maximum weighted measure to those having minimum. Since the operator is non-linear, we have to exploit other properties of the diffusion process to recover a spectral property concerning the "second eigenvalue" of the resulting Laplacian. Moreover, we show that higher order spectral properties cannot hold in general using the current framework.

We consider a stochastic diffusion process, in which each vertex also experiences Brownian noise from outside the system. We show a relationship between the second eigenvalue and the convergence behavior of the process.

We show that various hypergraph parameters like multi-way expansion and diameter can be bounded using this operator's spectral properties. Since higher order spectral properties do not hold for the Laplacian operator, we instead use the concept of procedural minimizers to consider higher order Cheeger-like inequalities. For any $k \in \mathbb{N}$, we give a polynomial time algorithm to compute an $O(\log r)$-approximation to the k-th procedural minimizer, where r is the maximum cardinality of a hyperedge. We show that this approximation factor is optimal under the SSE hypothesis (introduced by [RS10]) for constant values of k.

Moreover, using the factor preserving reduction from vertex expansion in graphs to hypergraph expansion, we show that all our results for hypergraphs extend to vertex expansion in graphs.

---

*A preliminary version of this paper appeared in STOC 2015 [Lou15] and the current paper is the result of a merge with [CTZ15].

†Department of Computer Science, the University of Hong Kong.

‡Princeton University. Supported by the Simons Collaboration on Algorithms and Geometry. Part of the work was done while the author was a student at Georgia Tech and supported by Santosh Vempala's NSF award CCF-1217793.

Contents

1 Introduction

There is a rich spectral theory of graphs, based on studying the eigenvalues and eigenvectors of the adjacency and other related matrices of graphs [AM85, Alo86, AC88, ABS10, LRTV11, LRTV12, LOT12]. We refer the reader to [Chu97, MT06] for a comprehensive survey on Spectral Graph Theory. A fundamental graph parameter is its expansion or conductance defined for a graph $G = (V, E)$ as:

$$\phi_G := \min_{S \subset V} \frac{|\partial S|}{\min\left\{\mathsf{vol}(S), \mathsf{vol}(\bar{S})\right\}},$$

where by $\mathsf{vol}(S)$ we denote the sum of degrees of the vertices in S, and ∂S is the set of edges in the cut induced by S. Cheeger's inequality [AM85, Alo86], a central inequality in Spectral Graph Theory, establishes a bound on expansion via the spectrum of the graph:

$$\frac{\lambda_2}{2} \leq \phi_G \leq \sqrt{2\lambda_2},$$

where λ_2 is the second smallest eigenvalue of the normalized Laplacian matrix $\mathcal{L}_G := \mathsf{W}^{-1/2}(\mathsf{W} - A)\mathsf{W}^{-1/2}$, and A is the adjacency matrix of the graph and W is the diagonal matrix whose (i, i)-th entry is the degree of vertex i. This theorem and its many (minor) variants have played a major role in the design of algorithms as well as in understanding the limits of computation [SJ89, SS96, Din07, ARV09, ABS10]. We refer the reader to [HLW06] for a comprehensive survey.

Edge expansion can be generalized to edge-weighted hypergraphs. In a hypergraph $H = (V, E)$, an edge $e \in E$ is a non-empty subset of V. The edges have non-negative weights indicated by $w : E \to \mathbb{R}_+$. We say that H is an r-graph (or r-uniform) if every edge contains exactly r vertices. (Hence, a normal graph is a 2-graph.) Each vertex $v \in V$ has weight $w_v := \sum_{e \in E: v \in e} w_e$. A subset S of vertices has weight $w(S) := \sum_{v \in S} w_v$, and the edges it cuts is $\partial S := \{e \in E : e \text{ intersects both } S \text{ and } V \setminus S\}$. The *edge expansion* of $S \subset V$ is defined as $\phi(S) := \frac{w(\partial S)}{w(S)}$. The expansion of H is defined as:

$$\phi_H := \min_{\emptyset \subsetneq S \subsetneq V} \max\{\phi(S), \phi(V \setminus S)\}. \tag{1.1}$$

It has remained open to define a spectral model of hypergraphs, whose spectra can be used to estimate hypergraph parameters. Hypergraph expansion and related hypergraph partitioning problems are of immense practical importance, having applications in parallel and distributed computing [CA99], VLSI circuit design and computer architecture [KAKS99, GGLP00], scientific computing [DBH$^+$06] and other areas. Inspite of this, hypergraph expansion problems haven't been studied as well as their graph counterparts (see Section 1.1 for a brief survey). Spectral graph partitioning algorithms are widely used in practice for their efficiency and the high quality of solutions that they often provide [BS94, HL95]. Besides being of natural theoretical interest, a spectral theory of hypergraphs might also be relevant for practical applications.

The various spectral models for hypergraphs considered in the literature haven't been without shortcomings. An important reason for this is that there is no canonical matrix representation of hypergraphs. For an r-uniform hypergraph $H = (V, E)$ on the vertex set V and having edge set $E \subseteq \binom{V}{r}$, one can define the canonical r-tensor form A as follows:

$$A_{(i_1, \ldots, i_r)} := \begin{cases} 1 & \{i_1, \ldots, i_r\} \in E \\ 0 & \text{otherwise} \end{cases}.$$

This tensor form and its minor variants have been explored in the literature (see Section 1.1 for a brief survey), but have not been understood very well. Optimizing over tensors is NP-hard [HL13]; even getting good approximations might be intractable [BV09]. Moreover, the spectral properties of tensors seem to be unrelated to combinatorial properties of hypergraphs (See Appendix A).

Another way to study a hypergraph, say $H = (V, E)$, is to replace each hyperedge $e \in E$ by a complete 2-graph or a low degree expander on the vertices of e to obtain a 2-graph $G = (V, E')$. If we let r denote the size of the largest hyperedge in E, then it is easy to see that the combinatorial properties of G and H, like min-cut, sparsest-cut, among others, could be separated by a factor of $\Omega(r)$. Therefore, this approach will not be useful when r is *large*.

In general, one cannot hope to have a linear operator for hypergraphs whose spectra captures hypergraph expansion in a Cheeger-like manner. This is because the existence of such an operator will imply the existence of a polynomial time algorithm obtaining a $\mathcal{O}\left(\sqrt{\mathsf{OPT}}\right)$ bound on hypergraph expansion, but we rule this out by giving a lower bound of $\Omega(\sqrt{\mathsf{OPT}\,\log r})$ for computing hypergraph expansion, where r is the size of the largest hyperedge (Theorem 3.18).

Our main contribution is the definition of a new Laplacian operator for hypergraphs, obtained by generalizing the random-walk operator on graphs. Our operator does not require the hypergraph to be uniform (i.e. does not require all the hyperedges to have the same size). We describe this operator in Section 4 (see also Figure 3.1). We present our main results about this hypergraph operator in Section 4 and Section 6. Most of our results are independent of r (the size of the hyperedges), some of our bounds have a logarithmic dependence on r, and none of our bounds have a polynomial dependence on r. All our bounds are generalizations of the corresponding bounds for 2-graphs.

1.1 Related Work

Freidman and Wigderson [FW95] studied the canonical tensors of hypergraphs. They bounded the second eigenvalue of such tensors for hypergraphs drawn randomly from various distributions and showed their connections to randomness dispersers. Rodriguez [Rod09] studied the eigenvalues of a graph obtained by replacing each hyperedge by a clique (Note that this step incurs a loss of $\mathcal{O}(r^2)$, where r is the size of the hyperedge). Cooper and Dutle [CD12] studied the roots of the characteristic polynomial of hypergraphs and related it to its chromatic number. [HQ13, HQ14] also studied the canonical tensor form of the hypergraph and related its eigenvectors to some configured components of that hypergraph. Lenz and Mubayi [LM12, LM15, LM13] related the eigenvector corresponding to the second largest eigenvalue of the canonical tensor to hypergraph quasi-randomness. Chung [Chu93] defined a notion of Laplacian for hypergraphs and studied the relationship between its eigenvalues and a very different notion of hypergraph cuts and homologies. [PRT12, PR12, Par13, KKL14, SKM14] studied the relation of simplicial complexes with rather different notions of Laplacian forms, and considered isoperimetric inequalities, homologies and mixing times. Ene and Nguyen [EN14] studied the hypergraph multiway partition problem (generalizing the graph multiway partition problem) and gave a $\frac{4}{3}$-approximation algorithm for 3-uniform hypergraphs. Concurrent to this work, [LM14b] gave approximation algorithms for hypergraph expansion, and more generally, hypergraph small set expansion; they gave an $\tilde{\mathcal{O}}\left(k\sqrt{\log n}\right)$-approximation algorithm and an $\tilde{\mathcal{O}}\left(k\sqrt{\mathsf{OPT}\,\log r}\right)$ approximation bound for the problem of computing the set of vertices of size at most $|V|/k$ in a hypergraph $H = (V, E)$, having the least expansion.

Bobkov, Houdré and Tetali [BHT00] defined a Poincairé-type functional graph parameter called λ_∞ and showed that it relates to the vertex expansion of a graph in a Cheeger-like manner, i.e. it satisfies $\frac{\lambda_\infty}{2} \leq \phi^{\mathsf{V}} = \mathcal{O}\left(\sqrt{\lambda_\infty}\right)$ where ϕ^{V} is the vertex expansion of the graph (see Section 3.4 for the definition of vertex expansion of a graph). [LRV13] gave an $\mathcal{O}\left(\sqrt{\mathsf{OPT}\log d}\right)$-approximation bound for computing the vertex expansion in graphs having the largest vertex degree d. Feige *et al.* [FHL08] gave an $\mathcal{O}\left(\sqrt{\log n}\right)$-approximation algorithm for computing the vertex expansion of graphs (having arbitrary vertex degrees).

Peres *et al.* [PSSW09] study a "tug of war" Laplacian operator on graphs that is similar to our hypergraph heat operator and use it to prove that every bounded real-valued Lipschitz function F on a subset Y of a length space X admits a unique absolutely minimal extension to X. Subsequently a variant of this operator was used for analyzing the rate of convergence of local dynamics in bargaining networks [CDP10]. [LRTV11, LRTV12, LOT12, LM14a] study higher eigenvalues of graph Laplacians and relate them to graph multi-partitioning parameters (see Section 3.3).

2 Notation

Recall that we consider an edge-weighted hypergraph $H = (V, E, w)$, where V is the vertex set, E is the set of hyperedges and $w : E \to \mathbb{R}_+$ gives the edge weights. We let $n := |V|$ and $m := |E|$. The weight of a vertex $v \in V$ is $w_v := \sum_{e \in E: v \in e} w(e)$. Without loss of generality, we assume that all vertices have positive weights, since any vertex with zero weight can be removed. We use $\mathbb{R}^V$ to denote the set of column vectors. Given $f \in \mathbb{R}^V$, we use f_u or $f(u)$ (if we need to use the subscript to distinguish between different vectors) to indicate the coordinate corresponding to $u \in V$. We use A^T to denote the transpose of a matrix A. For a positive integer s, we denote $[s] := \{1, 2, \ldots, s\}$.

We let I denote the identity matrix and $\mathsf{W} \in \mathbb{R}^{n \times n}$ denote the diagonal matrix whose (i, i)-th entry is w_i. We use $r_{\min} := \min_{e \in E} |e|$ to denote the size of the smallest hyperedge and use $r_{\max} := \max_{e \in E} |e|$ to denote the size of the largest hyperedge. Since, most of our bounds will only need $r_{\max}$, we use $r := r_{\max}$ for brevity. We say that a hypergraph is *regular* if all its vertices have the same degree. We say that a hypergraph is *uniform* if all its hyperedges have the same cardinality. Recall that the expansion ϕ_H of a hypergraph H is defined in (1.1). We drop the subscript whenever the hypergraph is clear from the context.

Hop-Diameter. A list of edges $e_1, \ldots, e_l$ such that $e_i \cap e_{i+1} \neq \emptyset$ for $i \in [l-1]$ is referred as a *path*. The length of a path is the number of edges in it. We say that a path $e_1, \ldots, e_l$ connects two vertices $u, v \in V$ if $u \in e_1$ and $v \in e_l$. We say that the hypergraph is *connected* if for each pair of vertices $u, v \in V$, there exists a path connecting them. The *hop-diameter* of a hypergraph, denoted by $\mathsf{diam}(H)$, is the smallest value $l \in \mathbb{N}$, such that each pair of vertices $u, v \in V$ have a path of length at most l connecting them.

For an $x \in \mathbb{R}$, we define $x^+ := \max\{x, 0\}$ and $x^- := \max\{-x, 0\}$. For a vector u, we use $\|u\| := \|u\|_2$ to denote its Euclidean norm; if $\|u\| \neq 0$, we define $\tilde{u} := \frac{u}{\|u\|}$. We use $\mathbf{1} \in \mathbb{R}^V$ to denote the vector having 1 in every coordinate. For a vector $x \in \mathbb{R}^V$, we define its support as the set of coordinates at which x is non-zero, i.e. $\mathsf{supp}(x) := \{i : x_i \neq 0\}$. We use $\mathbb{I}[\cdot]$ to denote the indicator variable, i.e. $\mathbb{I}[\mathcal{E}]$ is equal to 1 if event $\mathcal{E}$ occurs, and is equal to 0 otherwise. We use $\chi_S \in \mathbb{R}^V$ to denote the indicator vector of the set $S \subset V$, i.e.

$$\chi_S(v) = \begin{cases} 1 & v \in S \\ 0 & \text{otherwise} \end{cases} .$$

In classical spectral graph theory, the edge expansion is related to the *discrepancy ratio*, which is defined as

$$\mathsf{D}_w(f) := \frac{\sum_{e \in E} w_e \max_{u,v \in e} (f_u - f_v)^2}{\sum_{u \in V} w_u f_u^2},$$

for each non-zero vector $f \in \mathbb{R}^V$. Note that $0 \leq \mathsf{D}_w(f) \leq 2$, where the upper bound can be achieved, say, by a complete bipartite graph with f having 1's on one side and -1's on the other side. Observe that if $f = \chi_S$ is the indicator vector for a subset $S \subset V$, then $\mathsf{D}_w(f) = \phi(S)$. In this paper, we use three isomorphic spaces described as follows. As we shall see, sometimes it is more convenient to use one space to describe the results.

Weighted Space. This is the space associated with the discrepancy ratio D_w to consider edge expansion. For $f, g \in \mathbb{R}^V$, the inner product is defined as $\langle f, g \rangle_w := f^\mathsf{T} \mathsf{W} g$, and the associated norm is $\|f\|_w :=$

$\sqrt{\langle f, f \rangle_w}$. We use $f \perp_w g$ to denote $\langle f, g \rangle_w = 0$.

Normalized Space. Given $f \in \mathbb{R}^V$ in the weighted space, the corresponding vector in the normalized space is $x := \mathsf{W}^{\frac{1}{2}} f$. The normalized discrepancy ratio is $\mathcal{D}(x) := \mathsf{D}_w(\mathsf{W}^{-\frac{1}{2}} x) = \mathsf{D}_w(f)$.

In the normalized space, the usual ℓ_2 inner product and norm are used. Observe that if x and y are the corresponding normalized vectors for f and g in the weighted space, then $\langle x, y \rangle = \langle f, g \rangle_w$.

A well-known result [Chu97] is that the *normalized Laplacian* for a 2-graph can be defined as $\mathcal{L} := \mathsf{I} - \mathsf{W}^{-\frac{1}{2}} A \mathsf{W}^{-\frac{1}{2}}$ (where A is the symmetric matrix giving the edge weights) such that $\mathcal{D}(x)$ coincides with the *Rayleigh quotient* of the Laplacian defined as follows:

$$\mathcal{R}(x) := \frac{\langle x, \mathcal{L}x \rangle}{\langle x, x \rangle}.$$

Measure Space. This is the space associated with the diffusion process that we shall define later. Given f in the weighted space, the corresponding vector in the measure space is given by $\varphi := \mathsf{W} f$. Observe that a vector in the measure space can have negative coordinates. We do not consider inner product explicitly in this space, and so there is no special notation for it. However, we use the ℓ_1-norm, which is not induced by an inner product. For vectors $\varphi_i = \mathsf{W}^{\frac{1}{2}} x_i$, we have

$$\sqrt{w_{\min}} \cdot \|x_1 - x_2\|_2 \leq \|\varphi_1 - \varphi_2\|_1 \leq \sqrt{w(V)} \cdot \|x_1 - x_2\|_2,$$

where the upper bound comes from the Cauchy-Schwarz inequality.

In the diffusion process, we consider how φ will move in the future. Hence, unless otherwise stated, all derivatives considered are actually right-hand-derivatives $\frac{d\varphi(t)}{dt} := \lim_{\Delta t \to 0^+} \frac{\varphi(t+\Delta t) - \varphi(t)}{\Delta t}$.

Transformation between Different Spaces. We use the Roman letter f for vectors in the weighted space, x for vectors in the normalized space, and Greek letter φ for vectors in the measure space. Observe that an operator defined on one space induces operators on the other two spaces. For instance, if L is an operator defined on the measure space, then $\mathsf{L}_w := \mathsf{W}^{-1} \mathsf{L} \mathsf{W}$ is the corresponding operator on the weighted space and $\mathcal{L} := \mathsf{W}^{-\frac{1}{2}} \mathsf{L} \mathsf{W}^{\frac{1}{2}}$ is the one on the normalized space. Moreover, all three operators have the same eigenvalues. Recall that the Rayleigh quotients are defined as $\mathsf{R}_w(f) := \frac{\langle f, \mathsf{L}_w f \rangle_w}{\langle f, f \rangle_w}$ and $\mathcal{R}(x) := \frac{\langle x, \mathcal{L}x \rangle}{\langle x, x \rangle}$. For $\mathsf{W}^{\frac{1}{2}} f = x$, we have $\mathsf{R}_w(f) = \mathcal{R}(x)$.

Given a set S of vectors in the normalized space, Π_S is the orthogonal projection operator onto the subspace spanned by S. The orthogonal projection operator Π_S^w can also be defined for the weighted space.

3 Overview of Results

A major contribution of this paper is to define a hypergraph Laplacian operator $\mathcal{L}$ whose spectral properties are related to the expansion properties of the underlying hypergraph.

3.1 Laplacian and Diffusion Process

In order to gain insights on how to define the Laplacian for hypergraphs, we first illustrate that the Laplacian for 2-graphs can be related to a diffusion process. Suppose edge weights w of a 2-graph are given by the (symmetric) matrix A.

Suppose $\varphi \in \mathbb{R}^V$ is some measure on the vertices, which, for instance, can represent a probability distribution on the vertices. A random walk on the graph can be characterized by the transition matrix $\mathsf{M} := A \mathsf{W}^{-1}$. Observe that each column of M sums to 1, because we apply M to the column vector φ to get the distribution $\mathsf{M}\varphi$ after one step of the random walk.

We wish to define a continuous diffusion process. Observe that, at this moment, the measure vector φ is moving in the direction of $\mathsf{M}\varphi - \varphi = (\mathsf{M} - \mathsf{I})\varphi$. Therefore, if we define an operator $\mathsf{L} := \mathsf{I} - \mathsf{M}$ on the measure space, we have the differential equation $\frac{d\varphi}{dt} = -\mathsf{L}\varphi$.

Using the transformation into the weighted space $f = \mathsf{W}^{-1}\varphi$ and the normalized space $x = \mathsf{W}^{-\frac{1}{2}}\varphi$, we can define the corresponding operators $\mathsf{L}_w := \mathsf{W}^{-1}\mathsf{L}\mathsf{W} = \mathsf{I} - \mathsf{W}^{-1}A$ and $\mathcal{L} := \mathsf{W}^{-\frac{1}{2}}\mathsf{L}\mathsf{W}^{\frac{1}{2}} = \mathsf{I} - \mathsf{W}^{-\frac{1}{2}}A\mathsf{W}^{-\frac{1}{2}}$, which is exactly the normalized Laplacian for 2-graphs. In the literature, the (weighted) Laplacian is defined as $\mathsf{W} - A$, which is $\mathsf{W}\mathsf{L}_w$ in our notation. Hence, to avoid confusion, we only consider the normalized Laplacian in this paper.

Interpreting the Diffusion Process. In the above diffusion process, we consider more carefully the rate of change for the measure at a certain vertex u:

$$\frac{d\varphi_u}{dt} = \sum_{v:\{u,v\}\in E} w_{uv}(f_v - f_u), \tag{3.1}$$

where $f = \mathsf{W}^{-1}\varphi$ is the weighted measure. Observe that for a stationary distribution of the random walk, the measure at a vertex u should be proportional to its (weighted) degree w_u. Hence, given an edge $e = \{u, v\}$, equation (3.1) indicates that there should be a contribution of measure flowing from the vertex with higher f value to the vertex with smaller f value. Moreover, this contribution has rate given by $c_e := w_e \cdot |f_u - f_v|$.

Generalizing Diffusion Rule to Hypergraphs. Suppose in a hypergraph $H = (V, E, w)$ the vertices have measure $\varphi \in \mathbb{R}^V$ (corresponding to $f = \mathsf{W}^{-1}\varphi$). For $e \in E$, we define $I_e(f) \subseteq e$ as the vertices u in e whose $f_u = \frac{\varphi_u}{w_u}$ values are minimum, $S_e(f) \subseteq e$ as those whose corresponding values are maximum, and $\Delta_e(f) := \max_{u,v\in E}(f_u - f_v)$ as the discrepancy within edge e. Then, inspired from the case of 2-graphs, the diffusion process should satisfy the following rules.

(R1) When the measure distribution is at state φ (where $f = \mathsf{W}^{-1}\varphi$), there can be a positive rate of measure flow from u to v due to edge $e \in E$ only if $u \in S_e(f)$ and $v \in I_e(f)$.

(R2) For every edge $e \in E$, the total rate of measure flow **due to** e from vertices in $S_e(f)$ to $I_e(f)$ is $c_e := w_e \cdot \Delta_e(f)$.

We shall later elaborate how the rate c_e of flow due to edge e is distributed among the pairs in $S_e(f) \times I_e(f)$. Figure 3.1 summarizes this framework.

Given a hypergraph $H = (V, E, w)$, we define the (normalized) Laplacian operator as follows. Suppose $x \in \mathbb{R}^V$ is in the normalized space with the corresponding $\varphi := \mathsf{W}^{\frac{1}{2}} x$ in the measure space and $f := \mathsf{W}^{-1}\varphi$ in the weighted space.

1. For each hyperedge $e \in E$, let $I_e(f) \subseteq e$ be the set of vertices u in e whose $f_u = \frac{\varphi_u}{w_u}$ values are minimum and $S_e(f) \subseteq e$ be the set of vertices in e whose corresponding values are maximum. Let $\Delta_e(f) := \max_{u,v \in E}(f_u - f_v)$.
2. *Weight Distribution.* For each $e \in E$, the weight w_e is "somehow" distributed among pairs in $S_e(f) \times I_e(f)$ satisfying (R1) and (R2). Observe that if $I_e = S_e$, then $\Delta_e = 0$, and it does not matter how the weight w_e is distributed.

   For each $(u, v) \in S_e(f) \times I_e(f)$, there exists $a^e_{uv} = a^e_{uv}(f)$ such that $\sum_{(u,v) \in S_e \times I_e} a^e_{uv} = w_e$, and the rate of flow from u to v (due to e) is $a^e_{uv} \cdot \Delta_e$.

   For ease of notation, we let $a^e_{uv} = a^e_{vu}$. Moreover, for other pairs $\{u', v'\}$ that do not receive any weight from e, we let $a^e_{u'v'} = 0$.
3. The distribution of hyperedge weights induces a symmetric matrix A_f as follows. For $u \neq v$, $A_f(u, v) = a_{uv} := \sum_{e \in E} a^e_{uv}(f)$; the diagonal entries are chosen such that entries in the row corresponding to vertex u sum to w_u. Observe that A_f depends on φ because $f = \mathsf{W}^{-1}\varphi$.

Then, the operator $\mathsf{L}(\varphi) := (\mathsf{I} - A_f \mathsf{W}^{-1})\varphi$ is defined on the measure space, and the diffusion process is described by $\frac{d\varphi}{dt} = -\mathsf{L}\varphi$.
This induces the (normalized) Laplacian $\mathcal{L} := \mathsf{W}^{-\frac{1}{2}} \mathsf{L} \mathsf{W}^{\frac{1}{2}}$, and the operator $\mathsf{L}_w := \mathsf{W}^{-1}\mathsf{L}\mathsf{W}$ on the weighted space.

Figure 3.1: Defining Laplacian via Diffusion Framework

How to distribute the weight w_e in Step (2) in Figure 3.1? In order to satisfy rule (R1), it turns out that the weight cannot be distributed arbitrarily. We show that the following straightforward approaches will not work.

- *Assign the weight w_e to just one pair $(u, v) \in S_e \times I_e$.* For the case $|S_e| \geq 2$, after infinitesimal time, among vertices in S_e, only φ_u (and f_u) will decrease due to e. This means u will no longer be in S_e after infinitesimal time, and we will have to pick another vertex in S_e immediately. However, we will run into the same problem again if we try to pick another vertex from S_e, and the diffusion process cannot continue.
- *Distribute the weight w_e evenly among pairs in $S_e \times I_e$.*[1] In Example B.3, there is an edge $e_5 = \{a, b, c\}$ such that the vertex in $I_{e_5} = \{c\}$ receives measure from the vertices in $S_{e_5} = \{a, b\}$. However, vertex b also gives some measure to vertex d because of the edge $e_2 = \{b, d\}$. In the example, all vertices have the same weight. Now, if w_{e_5} is distributed evenly among $\{a, c\}$ and $\{b, c\}$, then the measure of a decreases more slowly than that of b because b loses extra measure due to e_2. Hence, after infinitesimal time, b will no longer be in S_{e_5}. This means that the measure of b should not have been decreased at all due to e_5, contradicting the choice of distributing w_{e_5} evenly.

What properties should the Laplacian operator have? Even though the weight distribution in Step 2 does not satisfy rule (R1), some operator could still be defined. The issue is whether such an operator would have any desirable properties. In particular, the spectral properties of the Laplacian should have be related to the expansion properties of the hypergraph. Recall that the normalized discrepancy ratio

[1]Through personal communication, Jingcheng Liu and Alistair Sinclair have informed us that they also noticed that distributing the weight of a hyperedge uniformly will not work, and discovered independently a similar method for resolving ties.

$\mathcal{D}(x)$ is defined for non-zero $x \in \mathbb{R}^V$, and is related to hypergraph edge expansion.

Definition 3.1 (Procedural Minimizers) *Define $x_1 := \mathsf{W}^{\frac{1}{2}}\vec{\mathbf{1}}$, where $\vec{\mathbf{1}} \in \mathbb{R}^V$ is the all-ones vector; $\gamma_1 := \mathcal{D}(x_1) = 0$. Suppose $\{(x_i, \gamma_i)\}_{i \in [k-1]}$ have been constructed. Define $\gamma_k := \min\{\mathcal{D}(x) : \vec{\mathbf{0}} \neq x \perp \{x_i : i \in [k-1]\}\}$, and x_k to be any such minimizer that attains $\gamma_k = \mathcal{D}(x_k)$.*

Properties of Laplacian in 2-graphs. For the case of 2-graphs, it is known that the discrepancy ratio $\mathcal{D}(x)$ coincides with the Rayleigh quotient $\mathcal{R}(x) := \frac{\langle x, \mathcal{L}x \rangle}{\langle x, x \rangle}$ of the normalized Laplacian $\mathcal{L}$, which can be interpreted as a symmetric matrix. Hence, it follows that the sequence $\{\gamma_i\}$ obtained by the procedural minimizers also gives the eigenvalues of $\mathcal{L}$. Observe that for a 2-graph, the sequence $\{\gamma_i\}$ is uniquely defined, even though the minimizers $\{x_i\}$ might not be unique (even modulo scalar multiple) in the case of repeated eigenvalues. On the other hand, for hypergraphs, γ_2 is uniquely defined, but we shall see in Example B.1 that γ_3 could depend on the choice of minimizer x_2.

Theorem 3.2 (Diffusion Process and Laplacian) *Given an edge-weighted hypergraph, a diffusion process satisfying rules (R1) and (R2) can be defined and uniquely induces a normalized Laplacian $\mathcal{L}$ (that is not necessarily linear) on the normalized space having the following properties.*

1. *For all $\vec{\mathbf{0}} \neq x \in \mathbb{R}^V$, the Rayleigh quotient $\frac{\langle x, \mathcal{L}x \rangle}{\langle x, x \rangle}$ coincides with the discrepancy ratio $\mathcal{D}(x)$. This implies that all eigenvalues of $\mathcal{L}$ are non-negative.*
2. *There is an operator $\mathsf{L} := \mathsf{W}^{\frac{1}{2}} \mathcal{L} \mathsf{W}^{-\frac{1}{2}}$ on the measure space such that the diffusion process can be described by the differential equation $\frac{d\varphi}{dt} = -\mathsf{L}\varphi$.*
3. *Any procedural minimizer x_2 attaining $\gamma_2 := \min_{\vec{\mathbf{0}} \neq x \perp \mathsf{W}^{\frac{1}{2}}\vec{\mathbf{1}}} \mathcal{D}(x)$ satisfies $\mathcal{L}x_2 = \gamma_2 x_2$.*

However, there exists a hypergraph (Example B.4) such that for all procedural minimizers $\{x_1, x_2\}$, any procedural minimizer x_3 attaining $\gamma_3 := \min_{\vec{\mathbf{0}} \neq x \perp \{x_1, x_2\}} \mathcal{D}(x)$ is not an eigenvector of $\Pi_{\{x_1, x_2\}^\perp} \mathcal{L}$.

The first three statements are proved in Lemmas 4.2, 4.8 and Theorem 4.1. Example B.4 suggests that the current approach cannot be generalized to consider higher order eigenvalues of the Laplacian $\mathcal{L}$, since any diffusion process satisfying rules (R1) and (R2) uniquely determines the Laplacian $\mathcal{L}$.

We remark that for hypergraphs, the Laplacian $\mathcal{L}$ is non-linear. In general, non-linear operators can have more or fewer than n eigenvalues. Theorem 3.2 implies that apart from $x_1 = \mathsf{W}^{\frac{1}{2}}\vec{\mathbf{1}}$, the Laplacian has another eigenvector x_2, which is a procedural minimizer attaining γ_2. It is not clear if $\mathcal{L}$ has any other eigenvalues. We leave as an open problem the task of investigating if other eigenvalues exist.

Diffusion Process and Steepest Descent. We can interpret the above diffusion process in terms of deepest descent with respect to the following quadratic potential function on the weighted space:

$$\mathsf{Q}_w(f) := \frac{1}{2} \sum_{e \in E} w_e \max_{u,v \in e} (f_u - f_v)^2.$$

Specifically, we can imagine a diffusion process in which the motion is leading to a decrease in the potential function. For 2-graphs, one can check that in fact we have $\frac{df}{dt} = -\mathsf{W}^{-1} \nabla_f \mathsf{Q}_w(f)$. Hence, we could try to define $\mathsf{L}_w f$ as $\mathsf{W}^{-1} \nabla_f \mathsf{Q}_w(f)$. Indeed, Lemma 4.10 confirms that our diffusion process implies that $\frac{d}{dt} \mathsf{Q}_w(f) = -\|\mathsf{L}_w f\|_w^2$. However, because of the maximum operator in the definition of $\mathsf{Q}_w(\cdot)$, one eventually has to consider the issue of resolving ties in order to give a meaningful definition of $\nabla_f \mathsf{Q}_w(f)$.

Comparison to other operators. One could ask if there can be a "better" operator? A natural operator that one would be tempted to try is the *averaging* operator, which corresponds to a diffusion process that attempts to transfer measure between *all* vertices in a hyperedge to approach the stationary distribution. However, for each hyperedge $e \in E$, the averaging operator will yield information about $\mathbb{E}_{i,j \in e} (f_i - f_j)^2$, instead of $\max_{i,j \in e} (f_i - f_j)^2$ that is related to edge expansion. In particular, the averaging operator will have a gap of factor $\Omega(r)$ between the hypergraph expansion and the square root of its second smallest eigenvalue.

3.2 Diffusion Processes

The diffusion process described in Figure 3.1 is given by the differential equation $\frac{d\varphi}{dt} = -\mathsf{L}\varphi$, where $\varphi \in \mathbb{R}^V$ is in the measure space. The diffusion process is deterministic, and no measure enters or leaves the system. We believe that it will be of independent interest to consider the case when each vertex can experience independent noise from outside the system, for instance, in risk management applications [Mer69, Mer71]. Since the diffusion process is continuous in nature, we consider Brownian noise.

For some $\eta \geq 0$, we assume that the noise experienced by each vertex u follows the Brownian motion whose rate of variance is ηw_u. Then, the measure $\Phi_t \in \mathbb{R}^V$ of the system is an Itō process defined by the stochastic differential equation $d\Phi_t = -\mathsf{L}\Phi_t\, dt + \sqrt{\eta} \cdot \mathsf{W}^{\frac{1}{2}}\, dB_t$. For $\eta = 0$, this reduces to the deterministic diffusion process in a closed system.

We consider the transformation into the normalized space $X_t := \mathsf{W}^{-\frac{1}{2}}\Phi_t$, and obtain the corresponding equation $dX_t = -\mathcal{L}X_t\, dt + \sqrt{\eta}\, dB_t$, where $\mathcal{L}$ is the normalized Laplacian. Observe that the random noise in the normalized space is spherically symmetric.

Convergence Metric. Given a measure vector $\varphi \in \mathbb{R}^V$, denote $\varphi^* := \frac{\langle \vec{\mathbf{1}}, \varphi \rangle}{w(V)} \cdot \mathsf{W}\vec{\mathbf{1}}$, which is the corresponding *stationary* measure vector obtained by distributing the total measure $\sum_{u \in V} \varphi_u = \langle \vec{\mathbf{1}}, \varphi \rangle$ among the vertices such that each vertex u receives an amount proportional to its weight w_u.

For the normalized vector $x = \mathsf{W}^{-\frac{1}{2}}\varphi$, observe that $x^* := \mathsf{W}^{-\frac{1}{2}}\varphi^* = \frac{\langle \vec{\mathbf{1}}, \varphi \rangle}{w(V)} \cdot \mathsf{W}^{\frac{1}{2}}\vec{\mathbf{1}}$ is the projection of x into the subspace spanned by $x_1 := \mathsf{W}^{\frac{1}{2}}\vec{\mathbf{1}}$. We denote by Π the orthogonal projection operator into the subspace orthogonal to x_1.

Hence, given $x = \mathsf{W}^{-\frac{1}{2}}\varphi$, we have $x = x^* + \Pi x$, where x^* is the stationary component and Πx is the transient component. Moreover, $\varphi - \varphi^* = \mathsf{W}^{\frac{1}{2}}\Pi x$.

We derive a relationship between γ_2 and the system's convergence behavior.

Theorem 3.3 (Convergence and Spectral Gap) *Suppose $\gamma_2 = \min_{0 \neq x \perp x_1} \mathcal{R}(x)$. Then, in the stochastic diffusion process described above, for each $t \geq 0$, the random variable $\|\Pi X_t\|_2$ is stochastically dominated by $\|\widehat{X}_t\|_2$, where $\widehat{X}_t$ has distribution $e^{-\gamma_2 t}\Pi X_0 + \sqrt{\frac{\eta}{2\gamma_2} \cdot (1 - e^{-2\gamma_2 t})} \cdot N(0,1)^V$, and $N(0,1)^V$ is the standard n-dimensional Guassian distribution with independent coordinates.*

Mixing Time for Deterministic Diffusion Process. For the special case $\eta = 0$, one can consider an initial probability measure $\varphi_0 \in \mathbb{R}_+^V$ such that $\langle \vec{\mathbf{1}}, \varphi_0 \rangle = 1$. We denote the the stationary distribution $\varphi^* := \frac{1}{w(V)} \cdot \mathsf{W}\vec{\mathbf{1}}$. For $\delta > 0$, the *mixing time* $\mathsf{t}_\delta^{\mathsf{mix}}(\varphi_0)$ is the smallest time τ such that for all $t \geq \tau$, $\|\varphi_t - \varphi^*\|_1 \leq \delta$.

Theorem 3.4 (Upper Bound for Mixing Time) *Consider the deterministic diffusion process with some initial probability measure $\varphi_0 \in \mathbb{R}_+^V$. Then, for all $\delta > 0$, the mixing time $\mathsf{t}_\delta^{\mathsf{mix}}(\varphi_0) \leq \frac{1}{\gamma_2} \log \frac{1}{\delta\sqrt{\varphi^*_{\min}}}$, where $\varphi^*_{\min} := \min_{u \in V} \varphi^*(u)$.*

Observe that for a regular hypergraph (i.e., w_u is the same for all $u \in V$), Theorem 3.4 says that the mixing time can be $\mathcal{O}(\log n)$. We believe that this fact might have applications in counting/sampling problems on hypergraphs à la MCMC (Markov chain monte carlo) algorithms on graphs.

Towards Local Clustering Algorithms for Hypergraphs We believe that the hypergraph diffusion process has applications in computing combinatorial properties and sampling problems in hypergraphs. As a concrete example, we show that the diffusion process can be useful towards computing sets of vertices having small expansion. We show that if the diffusion process mixes slowly, then the hypergraph must contain a set of vertices having small expansion. This is analogous to the corresponding fact for graphs, and can be used as a tool to certify an upper bound on hypergraph expansion.

Theorem 3.5 *Given a hypergraph $H = (V, E, w)$ and a probability distribution $\varphi_0 : V \to [0, 1]$, let φ_t denote the probability distribution at time t according to the diffusion process (Figure 3.1) and φ^* be the stationary distribution.*

Let $\delta > 0$. Suppose initially $\|\varphi_0 - \varphi^*\|_1 > \delta$ and for some time $T > 0$, $\|\varphi_T - \varphi^*\|_1 > \delta$. Then, there exists a set $S \subset V$ such that $\varphi^*(S) \le \frac{1}{2}$ and

$$\phi(S) \le \mathcal{O}\left(\frac{1}{T} \ln \frac{\|\varphi_0 - \varphi^*\|_1}{\sqrt{\varphi^*_{\min}} \cdot \delta}\right).$$

As in the case of graphs, this upper bound might be better than the guarantee obtained using an SDP relaxation (3.22) in certain settings.

One could ask if the converse of the statement of Theorem 3.5 is true, i.e., if the hypergraph $H = (V, E, w)$ has a "sparse cut", then is there a polynomial time computable probability distribution $\varphi_0 : V \to [0, 1]$ such that the diffusion process initialized with this φ_0 mixes "slowly"? Theorem 3.6 shows that there exists such a distribution φ_0, but it is not known if such a distribution can be computed in polynomial time. We leave this as an open problem.

Theorem 3.6 (Lower bound on Mixing Time) *Given a hypergraph $H = (V, E, w)$, there exists a probability measure φ_0 on V such that $\|\varphi_0 - \varphi^*\|_1 \ge \frac{1}{2}$, and for small enough δ,*

$$\mathsf{t}^{\mathsf{mix}}_{\delta}(\varphi_0) = \Omega\left(\frac{1}{\gamma_2} \ln \frac{\varphi^*_{\min}}{\delta}\right).$$

See Theorem 5.5 for the formal statement of Theorem 3.6. We view the condition in Theorem 3.6 that the starting distribution φ_0 satisfy $\|\varphi_0 - \varphi^*\|_1 \ge \frac{1}{2}$ as the analogue of a random walk in a graph starting from some vertex.

Discretized Diffusion Operator and Hypergraph Diameter A well known fact about regular 2-graphs is that the diameter of a graph G is at most $\mathcal{O}(\log n / (\log(1/(1 - \gamma_2))))$.

We define a discretized diffusion operator $\mathsf{M} := \mathsf{I} - \frac{1}{2}\mathsf{L}$ on the measure space in Section 5.4, and use it to prove an upper bound on the hop-diameter of a hypergraph.

Theorem 3.7 *Given a hypergraph $H = (V, E, w)$, its hop-diameter $\mathsf{diam}(H)$ is at most $\mathcal{O}\left(\frac{\log N_w}{\gamma_2}\right)$, where $N_w := \max_{u \in V} \frac{w(V)}{w_u}$.*

3.3 Cheeger Inequalities

We generalize the Cheeger's inequality [AM85, Alo86] to hypergraphs.

Theorem 3.8 (Hypergraph Cheeger Inequalities) *Given an edge-weighted hypergraph H, its expansion ϕ_H is defined as in (1.1). Then, we have the following:*

$$\frac{\gamma_2}{2} \le \phi_H \le 2\sqrt{\gamma_2}.$$

However, to consider higher-order Cheeger inequalities for hypergraphs, at this moment, we cannot use the spectral properties of the Laplacian $\mathcal{L}$. Moroever, the sequence $\{\gamma_i\}$ generated by procedural minimizers might not be unique. We consider the following parameters.

Orthogonal Minimaximizers. Define $\xi_k := \min_{x_1,\dots x_k} \max_{i \in [k]} \mathcal{D}(x_i)$ and $\zeta_k := \min_{x_1,\dots x_k} \max\{\mathcal{D}(x) : x \in \mathrm{span}\{x_1, \dots x_k\}\}$, where the minimum is over k non-zero mutually orthogonal vectors $x_1, \dots, x_k$ in the normalized space. (All involved minimum and maximum can be attained because $\mathcal{D}$ is continuous and all vectors could be chosen from the surface of a unit ball, which is compact.)

For 2-graphs, the three parameters $\xi_k = \gamma_k = \zeta_k$ coincide with the eigenvalues of the normalized Laplacian $\mathcal{L}$. Indeed, most proofs in the literature concerning expansion and Cheeger inequalities (e.g., [LOT12, KLL$^+$13]) just need to use the underlying properties of γ_k, ξ_k and ζ_k with respect to the discrepancy ratio, without explicitly using the spectral properties of the Laplacian. However, the three parameters can be related to one another in the following lemma, whose proof is in Section 6.2.

Lemma 3.9 (Comparing Discrepancy Minimizers) *Suppose $\{\gamma_k\}$ is some sequence produced by the procedural minimizers. For each $k \geq 1$, $\xi_k \leq \gamma_k \leq \zeta_k \leq k\xi_k$. In particular, $\gamma_2 = \zeta_2$, but it is possible that $\xi_2 < \gamma_2$.*

Given a parameter $k \in \mathbb{N}$, the multi-way small-set expansion problem asks to compute k disjoint sets $S_1, S_2, \ldots, S_k$ that all have small expansion. This problem has a close connection with the Unique Games Conjecture [RS10, ABS10]. In recent works, higher eigenvalues of Laplacians were used to bound small-set expansion in 2-graphs [LRTV12, LOT12]. In particular, the following result is achieved.

Fact 3.10 (Higher-Order Cheeger Inequalities for 2-Graphs) *There exists an absolute constant $c > 0$ such that for any 2-graph $G = (V, E, w)$ and any integer $k < |V|$, there exist $\Theta(k)$ non-empty disjoint sets $S_1, \ldots, S_{\lfloor ck \rfloor} \subset V$ such that*

$$\max_{i \in [ck]} \phi(S_i) = \mathcal{O}\left(\sqrt{\gamma_k \log k}\right) .$$

Moreover, for any k disjoint non-empty sets $S_1, \ldots, S_k \subset V$

$$\max_{i \in [k]} \phi(S_i) \geq \frac{\gamma_k}{2} .$$

We prove the following generalizations to hypergraphs (see Theorems 6.6 and 6.14 for formal statements).

Theorem 3.11 (Small Set Expansion) *Given hypergraph $H = (V, E, w)$ and parameter $k < |V|$, suppose $f_1, f_2, \ldots, f_k$ are k orthonormal vectors in the weighted space such that $\max_{s \in [k]} \mathsf{D}_w(f_s) \leq \xi$. Then, there exists a set $S \subset V$ such that $|S| = \mathcal{O}(|V|/k)$ satisfying*

$$\phi(S) = \mathcal{O}\left(k \log k \log\log k \cdot \sqrt{\log r} \cdot \sqrt{\xi}\right) ,$$

where r is the size of the largest hyperedge in E.

Theorem 3.12 (Higher-Order Cheeger Inequalities for Hypergraphs) *There exist absolute constants $c > 0$ such that the following holds. Given a hypergraph $H = (V, E, w)$ and any integer $k < |V|$, suppose $f_1, f_2, \ldots, f_k$ are k orthonormal vectors in the weighted space such that $\max_{s \in [k]} \mathsf{D}_w(f_s) \leq \xi$. Then, there exists $\Theta(k)$ non-empty disjoint sets $S_1, \ldots, S_{\lfloor ck \rfloor} \subset V$ such that*

$$\max_{i \in [ck]} \phi(S_i) = \mathcal{O}\left(k^2 \log k \log\log k \cdot \sqrt{\log r} \cdot \sqrt{\xi}\right) .$$

Moreover, for any k disjoint non-empty sets $S_1, \ldots, S_k \subset V$

$$\max_{i \in [k]} \phi(S_i) \geq \frac{\zeta_k}{2} .$$

3.4 Hardness via Vertex Expansion in 2-Graphs

Given a graph $G = (V, E, w)$ having maximum vertex degree d and a set $S \subset V$, its internal boundary $N^{\mathsf{in}}(S)$, and external boundary $N^{\mathsf{out}}(S)$ is defined as follows:

$N^{\mathsf{in}}(S) := \big\{ v \in S : \exists u \in \bar{S} \text{such that } \{u, v\} \in E \big\}$ and

$N^{\mathsf{out}}(S) := \big\{ v \in \bar{S} : \exists u \in S \text{such that } \{u, v\} \in E \big\}$.

The vertex expansion $\phi^{\mathsf{V}}(S)$ of a set S is defined as

$$\phi^{\mathsf{V}}(S) := \frac{\left|N^{\mathsf{in}}(S)\right| + \left|N^{\mathsf{out}}(S)\right|}{|S|} .$$

Vertex expansion is a fundamental graph parameter that has applications both as an algorithmic primitive and as a tool for proving communication lower bounds [LT80, Lei80, BTL84, AK95, SM00].

Bobkov *et al.* [BHT00] defined a Poincairé-type functional graph parameter as follows. Given an undirected graph $G = (V, E)$, denote $v \sim u$ if $\{v, u\} \in E$, and define

$$\lambda_\infty := \min_{f \in \mathbb{R}^V} \frac{\sum_{u \in V} \max_{v \sim u} \left(f_u - f_v\right)^2}{\sum_{u \in V} f_u^2 - \frac{1}{n} \left(\sum_{u \in V} f_u\right)^2} .$$

Observe that the expression to be minimized does not change if the same constant is added to every coordinate. Hence, without loss of generality, we can assume that the above minimization is taken over all non-zero vectors such that $f \perp \vec{\mathbf{1}}$. Therefore, we can equivalently write

$$\lambda_\infty = \min_{0 \neq f \perp \vec{\mathbf{1}}} \mathsf{D}^V(f), \tag{3.2}$$

where $\mathsf{D}^V(\cdot)$ is the discrepancy ratio for vertex expansion:

$$\mathsf{D}^V(f) := \frac{\sum_{u \in V} \max_{v \sim u} \left(f_u - f_v\right)^2}{\sum_{u \in V} f_u^2} .$$

If χ_S is the characteristic vector of the susbet S of vertices, then it follows that $\phi^{\mathsf{V}}(S) = \mathsf{D}^V(\chi_S)$. We can see that there are many similarities with edge expansion, and indeed a Cheeger-type Inequality for vertex expansion in graphs was proved in [BHT00].

Fact 3.13 ([BHT00]) *For an un-weighted graph $G = (V, E)$,*

$$\frac{\lambda_\infty}{2} \leq \phi_G^{\mathsf{V}} \leq \sqrt{2\lambda_\infty} .$$

Given the similarities between vertex expansion in 2-graphs and hyperedge expansion, one could imagine that a diffusion process can be defined with respect to vertex expansion in order to construct a similar Laplacian operator, which would have λ_∞ as an eigenvalue. However, instead of repeating the whole argument and analysis, we remark that there is a well known reduction from vertex expansion in 2-graphs to hyperedge expansion.

Reduction 3.14
Input*: Undirected* 2*-graph* $G = (V, E)$*.*
Output*:We construct hypergraph* $H = (V, E')$ *as follows. For every vertex* $v \in V$*, we add the (unit-weighted) hyperedge* $\{v\} \cup N^{\mathsf{out}}(\{v\})$ *to* E'*.*

Fact 3.15 ([LM14b]) *Given a graph $G = (V, E, w)$ of maximum degree d and minimum degree $c_1 d$ (for some constant c_1), the hypergraph $H = (V, E')$ obtained from Reduction 3.14 has hyperedges of cardinality at most $d + 1$ and,*

$$c_1 \phi_H(S) \leq \frac{1}{d+1} \cdot \phi_G^{\mathsf{V}}(S) \leq \phi_H(S) \qquad \forall S \subset V .$$

Remark 3.16 *The dependence on the degree in Fact 3.15 is only because vertex expansion and hypergraph expansion are normalized differently. The vertex expansion of a set S is defined as the number of vertices in the boundary of S divided by the cardinality of S, whereas the hypergraph expansion of a set S is defined as the number hyperedges crossing S divided by the sum of the degrees of the vertices in S.*

Using Fact 3.15, we can apply our results for hypergraph edge expansion to vertex expansion in d-regular 2-graphs. In particular, we relate λ_∞ with the parameter γ_2 associated with the hypergraph achieved in Reduction 3.14.

Theorem 3.17 *Let $G = (V, E)$ be a undirected d-regular 2-graph with parameter λ_∞, and let $H = (V, E')$ be the hypergraph obtained in Reduction 3.14 having parameter γ_2. Then,*

$$\frac{\gamma_2}{4} \leq \frac{\lambda_\infty}{d} \leq \gamma_2 .$$

The computation of λ_∞ is not known to be tractable. For graphs having maximum vertex degree d, [LRV13] gave a $\mathcal{O}(\log d)$-approximation algorithm for computing λ_∞, and showed that there exists an absolute constant C such that is SSE-hard to get better than a $C \log d$-approximation to λ_∞. Indeed, such a hardness result implies that the hyperedge expansion and the spectral gap γ_2 cannot be efficiently approximated. See Section 7 for a definition of SSE hypothesis. Specifically, we show the following hardness results for computing hyperedge expansion (see Theorem 7.3) and γ_2 (see Theorem 7.4).

Theorem 3.18 (Informal Statement) *Given a hypergraph H, it is* SSE*-hard to get better than an $\mathcal{O}\left(\sqrt{\phi_H \cdot \frac{\log r}{r}}\right)$ bound on hypergraph expansion in polynomial time. (Note that this is non-trivial only when $\phi_H \leq \frac{\log r}{r}$.)*

Theorem 3.19 (Informal Statement) *When $\gamma_2 \leq \frac{1}{r}$, it is* SSE*-hard to output a number $\widehat{\gamma}$ in polynomial time such that $\gamma_2 \leq \widehat{\gamma} = \mathcal{O}(\gamma_2 \log r)$.*

3.5 Approximation Algorithms

We do not know how to efficiently find orthonormal vectors $f_1, f_2, \ldots, f_k$ in the weighted space that attain ξ_k. In view of Theorems 3.11 and 3.12, we consider approximation algorithms to find k such vectors to minimize $\max_{i \in [k]} \mathsf{D}_w(f_i)$.

Approximate Procedural Minimizers. Our approximation algorithms are based on the following result on finding approximate procedural minimizers.

Theorem 3.20 *Suppose for $k \geq 2$, $\{f_i\}_{i \in [k-1]}$ is a set of orthonormal vectors in the weighted space, and define $\gamma := \min\{\mathsf{D}_w(f) : \vec{\mathbf{0}} \neq f \perp_w \{f_i : i \in [k-1]\}\}$. Then, there is a randomized procedure that produces a non-zero vector f that is orthogonal to $\{f_i\}_{i \in [k-1]}$ in polynomial time, such that with high probability, $\mathsf{D}_w(f) = \mathcal{O}(\gamma \log r)$, where r is the size of the largest hyperedge.*

Using the procedure in Theorem 3.20 as a subroutine for generating procedural minimizers, we can show that the resulting vectors provide an $\mathcal{O}(k \log r)$-approximation to ξ_k.

Theorem 3.21 (Approximating ξ_k) *There exists a randomized polynomial time algorithm that, given a hypergraph $H = (V, E, w)$ and a parameter $k < |V|$, outputs k orthonormal vectors $f_1, \ldots, f_k$ in the weighted space such that with high probability, for each $i \in [k]$,*

$$\mathsf{D}_w(f_i) \leq \mathcal{O}\left(i \log r \cdot \xi_i\right).$$

Algorithmic Applications. Applying Theorem 3.21, we readily have approximation algorithms for the problems in Theorems 3.8, 3.11 and 3.12.

Corollary 3.22 (Hyperedge Expansion) *There exists a randomized polynomial time algorithm that given a hypergraph $H = (V, E, w)$, outputs a set $S \subset V$ such that $\phi(S) = \mathcal{O}\left(\sqrt{\phi_H \log r}\right)$ with high probability, where r is the size of the largest hyperedge in E.*

We note that Corollary 3.22 also follows directly from [LM14b].

Many theoretical and practical applications require multiplicative approximation guarantees for hypergraph sparsest cut. In a seminal work, Arora, Rao and Vazirani [ARV09] gave a $\mathcal{O}\left(\sqrt{\log n}\right)$-approximation algorithm for the (uniform) sparsest cut problem in graphs. [LM14b] gave a $\mathcal{O}\left(\sqrt{\log n}\right)$-approximation algorithm for hypergraph expansion.

Corollary 3.23 (Small Set Expansion) *There exists a randomized polynomial time algorithm that given hypergraph $H = (V, E, w)$ and parameter $k < |V|$, produces a set $S \subset V$ such that with high probability, $|S| = \mathcal{O}\left(\frac{n}{k}\right)$ and*

$$\phi(S) = \mathcal{O}\left(k^{1.5} \log k \log \log k \cdot \log r \cdot \sqrt{\xi_k}\right),$$

where r is the size of the largest hyperedge in E.

In contrast, a polynomial-time algorithm is given in [LM14b] that returns a subset S with size $\mathcal{O}\left(\frac{n}{k}\right)$ whose expansion is at most $\mathcal{O}\left(k \log k \log \log k \cdot \sqrt{\log n}\right)$ times the smallest expansion over all vertex sets of size at most $\frac{n}{k}$.

Corollary 3.24 (Multi-way Hyperedge Expansion) *There exist absolute constants $c, c' > 0$ such that the following holds. There exists a randomized polynomial time algorithm that given hypergraph $H = (V, E, w)$ and parameter $k < |V|$, produces $\Theta(k)$ non-empty disjoint sets $S_1, \ldots, S_{\lfloor ck \rfloor} \subset V$ such that with high probability,*

$$\max_{i \in [ck]} \phi(S_i) = \mathcal{O}\left(k^{2.5} \log k \log \log k \cdot \log r \cdot \sqrt{\xi_k}\right).$$

In contrast, for 2-graphs, a polynomial-time bi-criteria approximation algorithm [LM14a] outputs $(1 - \epsilon)k$ disjoint subsets such that each subset has expansion at most $O_\epsilon(\sqrt{\log n \log k})$ times the optimal value.

3.6 Sparsest Cut with General Demands

An instance of the problem consists of a hypergraph $H = (V, E, w)$ with edge weights w and a collection $T = \{(\{s_i, t_i\}, D_i) : i \in [k]\}$ of demand pairs, where each pair $\{s_i, t_i\}$ has demand D_i. For a subset $S \subset V$, its expansion with respect to T is

$$\Phi(S) := \frac{w(\partial S)}{\sum_{i \in [k]} D_i \left|\chi_S(s_i) - \chi_S(t_i)\right|}.$$

The goal is to find S to minimize $\Phi(S)$. We denote $\Phi_H := \min_{S \subset V} \Phi(S)$.

Arora, Lee and Naor [ALN08] gave a $\mathcal{O}\left(\sqrt{\log k} \log \log k\right)$-approximation algorithm for the sparsest cut in 2-graphs with general demands. We give a similar bound for the sparsest cut in hypergraphs with general demands.

Theorem 3.25 *There exists a randomized polynomial time algorithm that given an instance of the hypergraph Sparsest Cut problem with hypergraph $H = (V, E, w)$ and k demand pairs in $T = \{(\{s_i, t_i\}, D_i) : i \in [k]\}$, outputs a set $S \subset V$ such that with high probability,*

$$\Phi(S) \leq \mathcal{O}\left(\sqrt{\log k \log r} \log \log k\right) \Phi_H,$$

where $r = \max_{e \in E} |e|$.

3.6.1 Discussion

We stress that none of our bounds have a polynomial dependence on r, the size of the largest hyperedge (Theorem 3.11 has a dependence on $\tilde{\mathcal{O}}(\min\{r, k\})$). In many of the practical applications, the typical instances have $r = \Theta(n^\alpha)$ for some $\alpha = \Omega(1)$; in such cases having bounds of $\mathsf{poly}(r)$ would not be of any practical utility. All our results generalize the corresponding results for 2-graphs.

3.7 Organization

We formally define the diffusion process and our Laplacian operator in Section 4. We prove the existence of a non-trivial eigenvalue for the Laplacian operator in Theorem 4.1.

In Section 5, we define the stochastic diffusion process, and prove our bounds on the mixing time (Theorem 3.4 and Theorem 3.6). We define a discrete diffusion operator and give a bound on the hypergraph diameter (Theorem 3.7) in Section 5.4.

In Section 6, we prove the basic hypergraph Cheeger inequality (Theorem 3.8) and also the higher-order variants (Theorem 3.11 and Theorem 3.12).

In Section 7, we explore the relationship between hyperedge expansion and vertex expansion in 2-graphs. Using hardness results for vertex expansion, we prove our hardness results for computing hypergraph eigenvalues (Theorem 3.19) and for hypergraph expansion (Theorem 3.18).

In Section 8, we give our approximation algorithm for procedural minimizers (Theorem 3.20). We present our algorithm for sparsest cut with general demands (Theorem 3.25) in Section 9.

4 Defining Diffusion Process and Laplacian for Hypergraphs

A classical result in spectral graph theory is that for a 2-graph whose edge weights are given by the adjacency matrix A, the parameter $\gamma_2 := \min_{\vec{0} \neq x \perp \mathsf{W}^{\frac{1}{2}} \vec{1}} \mathcal{D}(x)$ is an eigenvalue of the normalized Laplacian $\mathcal{L} := \mathsf{I} - \mathsf{W}^{-\frac{1}{2}} A \mathsf{W}^{-\frac{1}{2}}$, where a corresponding minimizer x_2 is an eigenvector of $\mathcal{L}$. Observe that γ_2 is also an eigenvector on the operator $\mathsf{L}_w := \mathsf{I} - \mathsf{W}^{-1} A$ induced on the weighted space. However, in the literature, the (weighted) Laplacian is defined as $\mathsf{W} - A$, which is $\mathsf{W}\mathsf{L}_w$ in our notation. Hence, to avoid confusion, we only consider the normalized Laplacian in this paper.

In this section, we generalize the result to hypergraphs. Observe that any result for the normalized space has an equivalent counterpart in the weighted space, and vice versa.

Theorem 4.1 (Eigenvalue of Hypergraph Laplacian) *For a hypergraph with edge weights w, there exists a normalized Laplacian $\mathcal{L}$ such that the normalized discrepancy ratio $\mathcal{D}(x)$ coincides with the corresponding Rayleigh quotient $\mathcal{R}(x)$. Moreover, the parameter $\gamma_2 := \min_{\vec{0} \neq x \perp \mathsf{W}^{\frac{1}{2}} \vec{1}} \mathcal{D}(x)$ is an eigenvalue of $\mathcal{L}$, where any minimizer x_2 is a corresponding eigenvector.*

However, we show in Example B.4 that the above result for our Laplacian does not hold for γ_3.

Intuition from Random Walk and Diffusion Process. We further elaborate the intuition described in Section 3.1. Given a 2-graph whose edge weights w are given by the (symmetric) matrix A, we illustrate the relationship between the Laplacian and a diffusion process in an underlying measure space, in order to gain insights on how to define the Laplacian for hypergraphs.

Suppose $\varphi \in \mathbb{R}^V$ is some measure on the vertices, which, for instance, can represent a probability distribution on the vertices. A random walk on the graph can be characterized by the transition matrix $\mathsf{M} := A\mathsf{W}^{-1}$. Observe that each column of M sums to 1, because we apply M to the column vector φ to get the distribution $\mathsf{M}\varphi$ after one step of the random walk.

We wish to define a continuous diffusion process. Observe that, at this moment, the measure vector φ is moving in the direction of $\mathsf{M}\varphi - \varphi = (\mathsf{M} - \mathsf{I})\varphi$. Therefore, if we define an operator $\mathsf{L} := \mathsf{I} - \mathsf{M}$ on the measure space, we have the differential equation $\frac{d\varphi}{dt} = -\mathsf{L}\varphi$.

To be mathematically precise, we are considering how φ will move in the future. Hence, unless otherwise stated, all derivatives considered are actually right-hand-derivatives $\frac{d\varphi(t)}{dt} := \lim_{\Delta t \to 0^+} \frac{\varphi(t+\Delta t)-\varphi(t)}{\Delta t}$.

Using the transformation into the weighted space $f = \mathsf{W}^{-1}\varphi$ and the normalized space $x = \mathsf{W}^{-\frac{1}{2}}\varphi$, we can define the corresponding operators $\mathsf{L}_w := \mathsf{W}^{-1}\mathsf{L}\mathsf{W} = \mathsf{I} - \mathsf{W}^{-1}A$ and $\mathcal{L} := \mathsf{W}^{-\frac{1}{2}}\mathsf{L}\mathsf{W}^{\frac{1}{2}} = \mathsf{I} - \mathsf{W}^{-\frac{1}{2}}A\mathsf{W}^{-\frac{1}{2}}$, which is exactly the normalized Laplacian for 2-graphs.

Generalizing the Diffusion Rule from 2-Graphs to Hypergraphs. We consider more carefully the rate of change for the measure at a certain vertex u: $\frac{d\varphi_u}{dt} = \sum_{v:\{u,v\}\in E} w_{uv}(f_v - f_u)$, where $f = \mathsf{W}^{-1}\varphi$ is the weighted measure. Observe that for a stationary distribution of the random walk, the measure at a vertex u should be proportional to its (weighted) degree w_u. Hence, given an edge $e = \{u, v\}$, by comparing the values f_u and f_v, measure should move from the vertex with higher f value to the vertex with smaller f value, at the rate given by $c_e := w_e \cdot |f_u - f_v|$.

To generalize this to a hypergraph $H = (V, E)$, for $e \in E$ and measure φ (corresponding to $f = \mathsf{W}^{-1}\varphi$), we define $I_e(f) \subseteq e$ as the vertices u in e whose $f_u = \frac{\varphi_u}{w_u}$ are minimum, $S_e(f) \subseteq e$ as those whose corresponding values are maximum, and $\Delta_e(f) := \max_{u,v\in E}(f_u - f_v)$ as the discrepancy within edge e. Then, the diffusion process obeys the following rules.

(R1) When the measure distribution is at state φ (where $f = \mathsf{W}^{-1}\varphi$), there can be a positive rate of measure flow from u to v due to edge $e \in E$ only if $u \in S_e(f)$ and $v \in I_e(f)$.

(R2) For every edge $e \in E$, the total rate of measure flow **due to** e from vertices in $S_e(f)$ to $I_e(f)$ is $c_e := w_e \cdot \Delta_e(f)$. In other words, the weight w_e is distributed among $(u, v) \in S_e(f) \times I_e(f)$ such that for each such (u, v), there exists $a^e_{uv} = a^e_{uv}(f)$ such that $\sum_{(u,v)\in S_e \times I_e} a^e_{uv} = w_e$, and the rate of flow from u to v (due to e) is $a^e_{uv} \cdot \Delta_e$. (For ease of notation, we write $a^e_{uv} = a^e_{vu}$.) Observe that if $I_e = S_e$, then $\Delta_e = 0$ and it does not matter how the weight w_e is distributed.

Observe that the distribution of hyperedge weights will induce a symmetric matrix A_f such that for $u \neq v$, $A_f(u, v) = a_{uv} := \sum_{e\in E} a^e_{uv}(f)$, and the diagonal entries are chosen such that entries in the row corresponding to vertex u sum to w_u. Then, the operator $\mathsf{L}(\varphi) := (\mathsf{I} - A_f\mathsf{W}^{-1})\varphi$ is defined on the measure space to obtain the differential equation $\frac{d\varphi}{dt} = -\mathsf{L}\varphi$. As in the case for 2-graph, we show in Lemma 4.2 that the corresponding operator L_w on the weighted space and the normalized Laplacian $\mathcal{L}$ are induced such that $\mathsf{D}_w(f) = \mathsf{R}_w(f)$ and $\mathcal{D}(x) = \mathcal{R}(x)$, which hold no matter how the weight w_e of hyperedge e is distributed among edges in $S_e(f) \times I_e(f)$.

Lemma 4.2 (Rayleigh Quotient Coincides with Discrepancy Ratio) *Suppose L_w on the weighted space is defined such that rules (R1) and (R2) are obeyed. Then, the Rayleigh quotient associated with L_w satisfies that for any f in the weighted space, $\mathsf{R}_w(f) = \mathsf{D}_w(f)$. By considering the isomorphic normalized space, we have for each x, $\mathcal{R}(x) = \mathcal{D}(x)$.*

Proof: It suffices to show that $\langle f, \mathsf{L}_w f\rangle_w = \sum_{e\in E} w_e \max_{u,v\in e}(f_u - f_v)^2$.

Recall that $\varphi = \mathsf{W}f$, and $\mathsf{L}_w = \mathsf{I} - \mathsf{W}^{-1}A_f$, where A_f is chosen as above to satisfy rules (R1) and (R2).

Hence, it follows that

$\langle f, \mathsf{L}_w f\rangle_w = f^\mathsf{T}(\mathsf{W} - A_f)f = \sum_{uv\in\binom{V}{2}} a_{uv}(f_u - f_v)^2$

$= \sum_{uv\in\binom{V}{2}} \sum_{e\in E:\{uv,vu\}\cap S_e\times I_e\neq\emptyset} a^e_{uv}(f_u - f_v)^2 = \sum_{e\in E} w_e \max_{u,v\in e}(f_u - f_v)^2$, as required. ■

4.1 Defining Diffusion Process to Construct Laplacian

Recall that $\varphi \in \mathbb{R}^V$ is the measure vector, where each coordinate contains the "measure" being dispersed. Observe that we consider a closed system here, and hence $\langle \vec{\mathbf{1}}, \varphi \rangle$ remains invariant. To facilitate the analysis, we also consider the weighted measure $f := \mathsf{W}^{-1}\varphi$.

Our goal is to define a diffusion process that obeys rules (R1) and (R2). Then, the operator on the measure space is given by $\mathsf{L}\varphi := -\frac{d\varphi}{dt}$. By observing that the weighted space is achieved by the transformation $f = \mathsf{W}^{-1}\varphi$, the operator on the weighted space is given by $\mathsf{L}_w f := -\frac{df}{dt}$.

In Figure 4.1, we give a procedure that takes $f \in \mathbb{R}^V$ and returns $r = \frac{df}{dt} \in \mathbb{R}^V$. This defines $\mathsf{L}_w f = -r$, and the Laplacian is induced $\mathcal{L} := \mathsf{W}^{\frac{1}{2}} \mathsf{L}_w \mathsf{W}^{-\frac{1}{2}}$ on the normalized space $x = \mathsf{W}^{\frac{1}{2}} f$.

Suppose we have the measure vector $\varphi \in \mathbb{R}^V$ and the corresponding weighted vector $f = \mathsf{W}^{-1}\varphi$. Observe that even though we call φ a measure vector, φ can still have negative coordinates. We shall construct a vector $r \in \mathbb{R}^V$ that is supposed to be $\frac{df}{dt}$. For $u \in V$ and $e \in E$, let $\rho_u(e)$ be the rate of change of the measure φ_u due to edge e. Then, $\rho_u := \sum_{e \in E} \rho_u(e)$ gives the rate of change of φ_u.

We show that r and ρ must satisfy certain constraints because of rules (R1) and (R2). Then, it suffices to show that there exists a unique $r \in \mathbb{R}^V$ that satisfies all the constraints.

First, since $\frac{df}{dt} = \mathsf{W}^{-1}\frac{d\varphi}{dt}$, we have for each vertex $u \in V$, $r_u = \frac{\rho_u}{w_u}$.

Rule (R1) implies the following constraint:

for $u \in V$ and $e \in E$, $\rho_u(e) < 0$ only if $u \in S_e(f)$, and $\rho_u(e) > 0$ only if $u \in I_e(f)$.

Rule (R2) implies the following constraint:

for each $e \in E$, we have $\sum_{u \in I_e(f)} \rho_u(e) = -\sum_{u \in S_e(f)} \rho_u(e) = w_e \cdot \Delta_e(f)$.

Construction of A_f. Observe that for each $e \in E$, once all the $\rho_u(e)$'s are determined, the weight w_e can be distributed among edges in $S_e \times I_e$ by considering a simple flow problem on the complete bipartite graph, where each $u \in S_e$ is a source with supply $-\frac{\rho_u(e)}{\Delta_e}$, and each $v \in I_e$ is a sink with demand $\frac{\rho_v(e)}{\Delta_e}$. Then, from any feasible flow, we can set a^e_{uv} to be the flow along the edge $(u, v) \in S_e \times I_e$.

Infinitesimal Considerations. In the previous discussion, we argue that if a vertex u is losing measure due to edge e, then it should remain in S_e for infinitesimal time, which holds only if the rate of change of f_u is the maximum among vertices in S_e. A similar condition should hold if the vertex u is gaining measure due to edge e. This translates to the following constraints.

Rule (R3) First-Order Derivative Constraints:

- If $\rho_u(e) < 0$, then $r_u \geq r_v$ for all $v \in S_e$.
- If $\rho_u(e) > 0$, then $r_u \leq r_v$ for all $v \in I_e$.

We remark that rule (R3) is only a necessary condition in order for the diffusion process to satisfy rule (R1). Even though A_f might not be unique, we shall show that these rules are sufficient to define a unique $r \in \mathbb{R}^V$, which is returned by the procedure in Figure 4.1.

Moreover, observe that if $f = \alpha g$ for some $\alpha > 0$, then in the above flow problem to determine the symmetric matrix, we can still have $A_f = A_g$. Hence, even though the resulting $\mathsf{L}_w(f) := (\mathsf{I} - \mathsf{W}^{-1}A_f)f$ might not be linear, we still have $\mathsf{L}_w(\alpha g) = \alpha \mathsf{L}_w(g)$.

Given a hypergraph $H = (V, E, w)$ and a vector $f \in \mathbb{R}^V$ in the weighted space, we describe a procedure to return $r \in \mathbb{R}^V$ that is supposed to be $r = \frac{df}{dt}$ in the diffusion process.

1. Define an equivalence relation on V such that u and v are in the same equivalence class *iff* $f_u = f_v$.
2. We consider each such equivalence class $U \subset V$ and define the r values for vertices in U. Denote $E_U := \{e \in E : \exists u \in U, u \in I_e \cup S_e\}$.

   Recall that $c_e := w_e \cdot \max_{u,v \in E}(f_u - f_v)$. For $F \subset E$, denote $c(F) := \sum_{e \in F} c_e$.

   For $X \subset U$, define $I_X := \{e \in E_U : I_e \subseteq X\}$ and $S_X := \{e \in E_U : S_e \cap X \neq \emptyset\}$.

   Denote $C(X) := c(I_X) - c(S_X)$ and $\delta(X) := \frac{C(X)}{w(X)}$.
3. Find any $P \subset U$ such that $\delta(P)$ is maximized.

   For all $u \in P$, set $r_u := \delta(P)$.
4. Recursively, find the r values for the remaining points $U' := U \setminus P$ using $E_{U'} := E_U \setminus (I_P \cup S_P)$.

Figure 4.1: Determining the Vector $r = \frac{df}{dt}$

Uniqueness of Procedure. In step (3) of Figure 4.1, there could be more than one choice of P to maximize $\delta(P)$. In Section 4.2, we give an efficient algorithm to find such a P. Moreover, we shall show that the procedure will return the same $r \in \mathbb{R}^V$ no matter what choice the algorithm makes. In Lemma 4.8, we prove that rules (R1)-(R3) imply that $\frac{df}{dt}$ must equal to such an r.

4.2 A Densest Subset Problem

In step (3) of Figure 4.1, we are solving the following variant of the densest subset problem restricted to some set U of vertices, with multi-sets $I := \{e \cap U : e \in E, I_e(f) \cap U \neq \emptyset\}$ and $S := \{e \cap U : e \in E, S_e(f) \cap U \neq \emptyset\}$.

Definition 4.3 (Densest Subset Problem) *The input is a hypergraph $H_U = (U, I \cup S)$, where we allow multi-hyperedges in $I \cup S$. Each $v \in U$ has weight $w_v > 0$, and each $e \in I \cup S$ has value $c_e > 0$.*

For $X \subset U$, define $I_X := \{e \in I : e \subset X\}$ and $S_X := \{e \in S : e \cap X \neq \emptyset\}$.

The output is a non-empty $P \subset U$ such that $\delta(P) := \frac{c(I_P) - c(S_P)}{w(P)}$ is maximized, and we call such P a densest subset.

We use an LP similar to the one given by Charikar [Cha00] used for the basic densest subset problem.

$$
\begin{array}{lll}
\text{maximize} & c(x) := \sum_{e \in I} c_e x_e - \sum_{e \in S} c_e x_e & \\
\text{subject to} & \sum_{v \in U} w_v y_v = 1 & \\
 & x_e \leq y_v & \forall e \in I, v \in e \\
 & x_e \geq y_v & \forall e \in S, v \in e \\
 & y_v, x_e \geq 0 & \forall v \in U, e \in I \cup S
\end{array}
$$

We analyze this LP using a similar approach given in [BBC+15]. Given a subset $P \subset U$, we define the following feasible solution $z^P = (x^P, y^P)$.

$$
x_e^P = \begin{cases} \frac{1}{w(P)} & \text{if } e \in I_P \cup S_P \\ 0 & \text{otherwise.} \end{cases}
$$

$$y_v^P = \begin{cases} \frac{1}{w(P)} & \text{if } v \in P \\ 0 & \text{otherwise.} \end{cases}$$

Feasibility of z^P can be verified easily and it can be checked that the objective value is $c(x^P) = \delta(P)$.

Given a feasible solution $z = (x, y)$, we say that a non-empty P is a *level set* of z if there exists $r > 0$ such that $P = \{v \in U : y_v \geq r\}$.

The following lemma has a proof similar to [BBC$^+$15, Lemma 4.1].

Lemma 4.4 *Suppose $z^* = (x^*, y^*)$ is an optimal (fractional) solution of the LP. Then, every (non-empty) level set P of z^* is a densest set and $\delta(P) = c(x^*)$.*

Proof: Suppose $z^* = (x^*, y^*)$ is an optimal solution. We prove the result by induction on the number k of level sets of z^*, which is also the number of distinct non-zero values found in the coordinates of y^*. For the base case when $k = 1$, z^* only has one level set $P = \mathsf{supp}(y^*)$. Because $\sum_{v \in U} w_v y_v^* = 1$, it follows that we must have $z^* = z^P$, and hence P must be a densest set and the result holds for $k = 1$.

For the inductive step, suppose y^* has $k \geq 2$ non-zero distinct values in its coordinates. Let $P := \mathsf{supp}(y^*)$ and $\alpha := \min\{y_v^* : v \in P\}$. Observe P is a level set of z^* and $\alpha \cdot w(P) \leq \sum_{v \in U} w_v y_v^* = 1$. Moreover, observe that if $x_e^* > 0$, then $x_e^* \geq \alpha$.

Define $\widehat{z} = (\widehat{x}, \widehat{y})$ as follows.

$$\widehat{x}_e = \begin{cases} \frac{x_e^* - \alpha}{1 - \alpha \cdot w(P)} & \text{if } x_e^* > 0 \\ 0 & \text{otherwise.} \end{cases}$$

$$\widehat{y}_v = \begin{cases} \frac{y_v^* - \alpha}{1 - \alpha \cdot w(P)} & \text{if } v \in P \\ 0 & \text{otherwise.} \end{cases}$$

Hence, $z^* = \alpha \cdot w(P) \cdot z^S + (1 - \alpha \cdot w(P))\widehat{z}$, and the number of level sets of $\widehat{z}$ is exactly $k - 1$. In particular, the level sets of z^* are P together with those of $\widehat{z}$.

Hence, to complete the inductive step, it suffices to show that $\widehat{z}$ is a feasible solution to the LP. To see why this is enough, observe that the objective function is linear, $c(x^*) = \alpha \cdot w(P) \cdot c(x^P) + (1 - \alpha \cdot w(P)) \cdot c(\widehat{x})$. Hence, if both z^P and $\widehat{z}$ are feasible, then both must be optimal. Then, the inductive hypothesis on $\widehat{z}$ can be used to finish the inductive step.

Hence, it remains to check the feasibility of $\widehat{z}$.

First, $\sum_{v \in U} w_v \widehat{y}_v = \sum_{v \in P} w_v \frac{y_v^* - \alpha}{1 - \alpha \cdot w(P)} = 1$.

Observe in the objective value, we want to increase x_e for $e \in I$ and decrease x_e for $e \in S$. Hence, the optimality of z^* implies that

$$x_e^* = \begin{cases} \min_{v \in e} y_v^* & \text{if } e \in I \\ \max_{v \in e} y_v^* & \text{if } e \in S. \end{cases}$$

For $x_e^* = 0$, then $\widehat{x}_e = 0$ and the corresponding inequality is satisfied.

Otherwise, $x_e^* \geq \alpha$, we have

$$\widehat{x}_e = \frac{x_e^* - \alpha}{1 - \alpha \cdot w(P)} = \begin{cases} \frac{\min_{v \in e} y_v^* - \alpha}{1 - \alpha \cdot w(P)} = \min_{v \in e} \widehat{y}_v & \text{if } e \in I \\ \frac{\max_{v \in e} y_v^* - \alpha}{1 - \alpha \cdot w(P)} = \max_{v \in e} \widehat{y}_v & \text{if } e \in S. \end{cases}$$

Therefore, $\widehat{z}$ is feasible and this completes the inductive step. ■

Given two densest subsets P_1 and P_2, it follows that that $\frac{z^{P_1} + z^{P_2}}{2}$ is an optimal LP solution. Hence, by considering its level sets, Lemma 4.4 implies the following corollary.

Corollary 4.5 (Properties of Densest Subsets) *1. Suppose P_1 and P_2 are both densest subsets. Then, $P_1 \cup P_2$ is also a densest subset. Moreover, if $P_1 \cap P_2$ is non-empty, then it is also a densest subset.*

2. *The maximal densest subset is unique and contains all densest subsets.*

The next two lemmas show that the procedure defined in Figure 4.1 will return the same $r \in \mathbb{R}^V$, no matter which densest subset is returned in step (3). Lemma 4.6 implies that if P is a maximal densest subset in the given instance, then the procedure will assign r values to the vertices in P first and each $v \in P$ will receive $r_v := \delta(P)$.

Lemma 4.6 (Remaining Instance) *Suppose in an instance $(U, I \cup S)$ with density function δ, some densest subset X is found, and the remaining instance $(U', I' \cup S')$ is defined with $U' := U \setminus X$, $I' := \{e \cap U' : e \in I \setminus I_X\}$, $S' := \{e \cap U' : e \in S \setminus S_X\}$ and the corresponding density function δ'. Then, for any $Y \subset U'$, $\delta'(Y) \leq \delta(X)$, where equality holds* iff $\delta(X \cup Y) = \delta(X)$.

Proof: Denote $\delta_M := \delta(X) = \frac{c(I_X) - c(S_X)}{w(X)}$.

Observe that $c(I'_Y) = c(I_{X \cup Y}) - c(I_X)$ and $c(S'_Y) = c(S_{X \cup Y}) - c(S_X)$.

Hence, we have $\delta'(Y) = \frac{c(I'_Y) - c(S'_Y)}{w(Y)} = \frac{\delta(X \cup Y) \cdot w(X \cup Y) - \delta_M \cdot w(X)}{w(X \cup Y) - w(X)}$.

Therefore, for each $\bowtie \in \{<, =, >\}$, we have $\delta'(Y) \bowtie \delta_M$ *iff* $\delta(X \cup Y) \bowtie \delta_M$.

We next see how this implies the lemma. For $\bowtie$ being "$>$", we know $\delta'(Y) > \delta(X)$ is impossible, because this implies that $\delta(X \cup Y) > \delta(X)$, violating the assumption that X is a densest subset.

For $\bowtie$ being "$=$", this gives $\delta'(Y) = \delta(X)$ *iff* $\delta(X \cup Y) = \delta(X)$, as required. ■

Corollary 4.7 (Procedure in Figure 4.1 is well-defined.) *The procedure defined in Figure 4.1 will return the same $r \in \mathbb{R}^V$, no matter which densest subset is returned in step (3). In particular, if P is the (unique) maximal densest subset in the given instance, then the procedure will assign r values to the vertices in P first and each $v \in P$ will receive $r_v := \delta(P)$. Moreover, after P is removed from the instance, the maximum density in the remaining instance is strictly less than $\delta(P)$.*

4.3 Densest Subset Procedure Defines Laplacian

We next show that rules (R1) to (R3) imply that in the diffusion process, $\frac{df}{dt}$ must equal to the vector $r \in \mathbb{R}^V$ returned by the procedure described in Figure 4.1.

We denote $r_S(e) := \max_{u \in S_e} r_u$ and $r_I(e) := \min_{u \in I_e} r_u$.

Lemma 4.8 (Defining Laplacian from Diffusion Process) *Given a measure vector $\varphi \in \mathbb{R}^V$ (and the corresponding $f = \mathsf{W}^{-1}\varphi$ in the weighted space), rules (R1) to (R3) uniquely determine $r = \frac{df}{dt} \in \mathbb{R}^V$ (and $\rho = \mathsf{W}r$), which can be found by the procedure described in Figure 4.1. This defines the operators $\mathsf{L}_w f := -r$ and $\mathsf{L}\varphi := -\mathsf{W}r$. The normalized Laplacian is also induced $\mathcal{L} := \mathsf{W}^{-\frac{1}{2}}\mathsf{L}\mathsf{W}^{\frac{1}{2}}$.*

Moreover, $\sum_{e \in E} c_e(r_I(e) - r_S(e)) = \sum_{u \in V} \rho_u r_u = \|r\|_w^2$.

Proof: As in Figure 4.1, we consider each equivalence class U, where all vertices in a class have the same f values.

For each such equivalence class $U \subset V$, define $I_U := \{e \in E : \exists u \in U, u \in I_e\}$ and $S_U := \{e \in E : \exists u \in U, u \in S_e\}$. Notice that each e is in exactly one such I's and one such S's.

As remarked in Section 4.1, for each $e \in E$, once all $\rho_u(e)$ is defined for all $u \in S_e \cup I_e$, it is simple to determine a^e_{uv} for $(u, v) \in S_e \times I_e$ by considering a flow problem on the bipartite graph $S_e \times I_e$. The "uniqueness" part of the proof will show that $r = \frac{df}{dt}$ must be some unique value, and the "existence" part of the proof shows that this r can determine the $\rho_u(e)$'s.

Considering Each Equivalence Class U. We can consider each equivalence class U independently by analyzing r_u and $\rho_u(e)$ for $u \in U$ and $e \in I_U \cup S_U$ that satisfy rules (R1) to (R3).

Proof of Uniqueness. We next show that rules (R1) to (R3) imply that r must take a unique value that can be found by the procedure in Figure 4.1.

For each $e \in I_U \cup S_U$, recall that $c_e := w_e \cdot \Delta_e(f)$, which is the rate of flow due to e into U (if $e \in I_U$) or out of U (if $e \in S_U$). For $F \subseteq I_U \cup S_U$, denote $c(F) := \sum_{e \in F} c_e$.

Suppose T is the set of vertices that have the maximum r values within the equivalence class, i.e., for all $u \in T$, $r_u = \max_{v \in U} r_v$. Observe that to satisfy rule (R3), for $e \in I_U$, there is positive rate c_e of measure flow into T due to e *iff* $I_e \subseteq T$; otherwise, the entire rate c_e will flow into $U \setminus T$. On the other hand, for $e \in S_U$, if $S_e \cap T \neq \emptyset$, then there is a rate c_e of flow out of T due to e; otherwise, the rate c_e flows out of $U \setminus T$.

Based on this observation, we define for $X \subset U$, $I_X := \{e \in I_U : I_e \subseteq X\}$ and $S_X := \{e \in S_U : S_e \cap X \neq \emptyset\}$. Note that these definitions are consistent with I_U and S_U. We denote $C(X) := c(I_X) - c(S_X)$.

To detect which vertices in U should have the largest r values, we define $\delta(X) := \frac{C(X)}{w(X)}$, which, loosely speaking, is the average weighted (with respect to W) measure rate going into vertices in X. Observe that if r is feasible, then the definition of T implies that for all $v \in T$, $r_v = \delta(T)$.

Corollary 4.7 implies that the procedure in Figure 4.1 will find the unique maximal densest subset P with $\delta_M := \delta(P)$.

We next show that $T = P$. Observe that for all edges $e \in I_P$ have $I_e \subset P$, and hence, there must be at least rate of $c(I_P)$ going into P; similarly, there is at most rate of $c(S_P)$ going out of P. Hence, we have $\sum_{u \in P} w_u r_u \geq c(I_P) - c(S_P) = w(P) \cdot \delta(P)$. Therefore, there exists $u \in P$ such that $\delta(P) \leq r_u \leq \delta(T)$, where the last inequality holds because every vertex $v \in T$ is supposed to have the maximum rate $r_v = \delta(T)$. This implies that $\delta(T) = \delta_M$, $T \subseteq P$ and the maximum r value is $\delta_M = \delta(T) = \delta(P)$. Therefore, the above inequality becomes $w(P) \cdot \delta_M \geq \sum_{u \in P} w_u r_u \geq w(P) \cdot \delta(P)$, which means equality actually holds. This implies that every vertex $u \in P$ has the maximum rate $r_u = \delta_M$, and so $T = P$.

Recursive Argument. Hence, it follows that the set T can be uniquely identified in Figure 4.1 as the set of vertices have maximum r values, which is also the unique maximal densest subset. Then, the uniqueness argument can be applied recursively for the smaller instance with $U' := U \setminus T$, $I_{U'} := I_U \setminus I_T$, $S_{U'} := S_U \setminus S_T$.

Proof of Existence. We show that once T is identified in Figure 4.1, it is possible to assign for each $v \in T$ and edge e where $v \in I_e \cup S_e$, the values $\rho_v(e)$ such that $\delta_M = r_v = \sum_e \rho_v(e)$.

Consider an arbitrary configuration ρ in which edge $e \in I_T$ supplies a rate of c_e to vertices in T, and each edge $e \in S_T$ demands a rate of c_e from vertices in T. Each vertex $v \in T$ is supposed to gather a net rate of $w_v \cdot \delta_M$, where any deviation is known as the *surplus* or *deficit*.

Given configuration ρ, define a directed graph G_ρ with vertices in T such that there is an arc (u, v) if non-zero measure rate can be transferred from u to v. This can happen in one of two ways: (i) there exists $e \in I_T$ containing both u and v such that $\rho_u(e) > 0$, or (ii) there exists $e \in S_T$ containing both u and v such that $\rho_v(e) < 0$.

Hence, if there is a directed path from a vertex u with non-zero surplus to a vertex v with non-zero deficit, then the surplus at vertex u (and the deficit at vertex v) can be decreased.

We argue that a configuration ρ with minimum surplus must have zero surplus. (Observe that the minimum can be achieved because ρ comes from a compact set.) Otherwise, suppose there is at least one vertex with positive surplus, and let T' be all the vertices that are reachable from some vertex with positive surplus in the directed graph G_ρ. Hence, it follows that for all $e \notin I_{T'}$, for all $v \in T'$, $\rho_v(e) = 0$, and for all $e \in S_{T'}$, for all $u \notin T'$, $\rho_u(e) = 0$. This means that the rate going into T' is $c(I_{T'})$ and all comes from $I_{T'}$, and the rate going out of T' is $c(S_{T'})$. Since no vertex in T' has a deficit and at least one has positive surplus, it follows that $\delta(T') > \delta_M$, which is a contradiction.

After we have shown that a configuration ρ with zero surplus exists, it can be found by a standard flow problem, in which each $e \in I_T$ has supply c_e, each $v \in T$ has demand $w_v \cdot \delta_M$, and each $e \in S_T$ has demand c_e. Moreover, in the flow network, there is a directed edge (e, v) if $v \in I_e$ and (v, e) if $v \in S_e$. Suppose in a feasible solution, there is a flow with magnitude θ along a directed edge. If the flow is in the direction (e, v), then $\rho_v(e) = \theta$; otherwise, if it is in the direction (v, e), then $\rho_v(e) = -\theta$.

Recursive Application. The feasibility argument can be applied recursively to the smaller instance defined on $(U', I_{U'}, S_{U'})$ with the corresponding density function δ'. Indeed, Corollary 4.7 implies that that $\delta'_M := \max_{\emptyset \neq Q \subset U'} \delta'(Q) < \delta_M$.

Claim. $\sum_{e \in E} c_e(r_I(e) - r_S(e)) = \sum_{u \in V} \rho_u r_u$.

Consider T defined above with $\delta_M = \delta(T) = r_u$ for $u \in T$.

Observe that $\sum_{u \in T} \rho_u r_u = (c(I_T) - c(S_T)) \cdot \delta_M = \sum_{e \in I_T} c_e \cdot r_I(e) - \sum_{e \in S_T} c_e \cdot r_S(e)$, where the last equality is due to rule (R3).

Observe that every $u \in V$ will be in exactly one such T, and every $e \in E$ will be accounted for exactly once in each of I_T and S_T, ranging over all T's. Hence, summing over all T's gives the result. ■

Comment on the Robustness of Diffusion Process. Recall that in Section 3.1, we mention that if the weight distribution is not carefully designed in Figure 3.1, then the diffusion process cannot actually continue. The following lemma implies that our diffusion process resulting from the procedure in Figure 4.1 will be robust.

Lemma 4.9 *In the diffusion process resulting from Figure 4.1 with the differential equation* $\frac{df}{dt} = -\mathsf{L}_w f$, *at any time* t_0, *there exists some* $\epsilon > 0$ *such that* $\frac{df}{dt}$ *is continuous in* $(t_0, t_0 + \epsilon)$.

Proof: Observe that as long as the equivalence classes induced by f do not change, then each of them act as a super vertex, and hence the diffusion process goes smoothly.

At the very instant that equivalence classes merge into some U, Figure 4.1 is actually used to determine whether the vertices will stay together in the next moment.

An equivalence class can be split in two ways. The first case is that the equivalence class U is peeled off layer by layer in the recursive manner described above, because they receive different r values. In particular, the (unique) maximal densest subset T is such a layer.

The second case is more subtle, because it is possible that vertices within T could be split in the next moment. For instance, there could be a proper subset $X \subsetneq T$ whose r values might be marginally larger than the rest after infinitesimal time.

The potential issue is that if the vertices in X go on their own, then the vertices X and also the vertices in $T \setminus X$ might experience a sudden jump in their rate r, thereby nullifying the "work" performed in Figure 4.1

Fortunately, this cannot happen, because if the set X could go on its own, it must be the case that $\delta_M = \delta(T) = \delta(X)$. Corollary 4.7 states that in this case, after X is separated on its own, then in the remaining instance, we must still have $\delta'(T \setminus X) = \delta_M$. Hence, the behavior of the remaining vertices is still consistent with the r value produced in Figure 4.1, and the r value cannot suddenly jump.

Hence, we can conclude that if equivalence classes merge or split at time t_0, there exists some $\epsilon > 0$ such that $\frac{df}{dt}$ is continuous in $(t_0, t_0 + \epsilon)$, until the next time equivalence classes merge or split. ■

4.4 Spectral Properties of Laplacian

We next consider the spectral properties of the normalized Laplacian $\mathcal{L}$ induced by the diffusion process defined in Section 4.1.

Lemma 4.10 (First-Order Derivatives) *Consider the diffusion process satisfying rules (R1) to (R3) on the measure space with $\varphi \in \mathbb{R}^V$, which corresponds to $f = \mathsf{W}^{-1}\varphi$ in the weighted space. Suppose L_w is the induced operator on the weighted space such that $\frac{df}{dt} = -\mathsf{L}_w f$. Then, we have the following derivatives.*

1. $\frac{d\|f\|_w^2}{dt} = -2\langle f, \mathsf{L}_w f\rangle_w$.
2. $\frac{d\langle f, \mathsf{L}_w f\rangle_w}{dt} = -2\|\mathsf{L}_w f\|_w^2$.
3. *Suppose $\mathsf{R}_w(f)$ is the Rayleigh quotient with respect to the operator L_w on the weighted space. Then, for $f \neq 0$, $\frac{d\mathsf{R}_w(f)}{dt} = -\frac{2}{\|f\|_w^4} \cdot (\|f\|_w^2 \cdot \|\mathsf{L}_w f\|_w^2 - \langle f, \mathsf{L}_w f\rangle_w^2) \leq 0$, by the Cauchy-Schwarz inequality on the $\langle \cdot, \cdot \rangle_w$ inner product, where equality holds* iff $\mathsf{L}_w f \in \mathrm{span}(f)$.

Proof: For the first statement, $\frac{d\|f\|_w^2}{dt} = 2\langle f, \frac{df}{dt}\rangle_w = -2\langle f, \mathsf{L}_w f\rangle_w$.

For the second statement, recall from Lemma 4.2 that $\langle f, \mathsf{L}_w f\rangle_w = \sum_{e \in E} w_e \max_{u,v \in e}(f_u - f_v)^2$. Moreover, recall also that $c_e = w_e \cdot \max_{u,v \in e}(f_u - f_v)$. Recall that $r = \frac{df}{dt}$, $r_S(e) = \max_{u \in S_e} r_u$ and $r_I(e) = \min_{u \in I_e} r_u$.

Hence, by the Envelope Theorem, $\frac{d\langle f, \mathsf{L}_w f\rangle_w}{dt} = 2\sum_{e \in E} c_e \cdot (r_S(e) - r_I(e))$. From Lemma 4.8, this equals $-2\|r\|_w^2 = -2\|\mathsf{L}_w f\|_w^2$.

Finally, for the third statement, we have

$\frac{d}{dt}\frac{\langle f, \mathsf{L}_w f\rangle_w}{\langle f, f\rangle_w} = \frac{1}{\|f\|_w^4}(\|f\|_w^2 \cdot \frac{d\langle f, \mathsf{L}_w f\rangle_w}{dt} - \langle f, \mathsf{L}_w f\rangle_w \cdot \frac{d\|f\|_w^2}{dt}) = -\frac{2}{\|f\|_w^4} \cdot (\|f\|_w^2 \cdot \|\mathsf{L}_w f\|_w^2 - \langle f, \mathsf{L}_w f\rangle_w^2)$, where the last equality follows from the first two statements. ■

We next prove some properties of the normalized Laplacian $\mathcal{L}$ with respect to orthogonal projection in the normalized space.

Lemma 4.11 (Laplacian and Orthogonal Projection) *Suppose $\mathcal{L}$ is the normalized Laplacian defined in Lemma 4.8. Moreover, denote $x_1 := \mathsf{W}^{\frac{1}{2}}\vec{\mathbf{1}}$, and let Π denote the orthogonal projection into the subspace that is orthogonal to x_1. Then, for all x, we have the following:*

1. $\mathcal{L}(x) \perp x_1$,
2. $\langle x, \mathcal{L}x\rangle = \langle \Pi x, \mathcal{L}\Pi x\rangle$.
3. *For all real numbers α and β, $\mathcal{L}(\alpha x_1 + \beta x) = \beta\mathcal{L}(x)$.*

Proof: For the first statement, observe that since the diffusion process is defined on a closed system, the total measure given by $\sum_{u \in V} \varphi_u$ does not change. Therefore, $0 = \langle \vec{\mathbf{1}}, \frac{d\varphi}{dt}\rangle = \langle \mathsf{W}^{\frac{1}{2}}\vec{\mathbf{1}}, \frac{dx}{dt}\rangle$, which implies that $\mathcal{L}x = -\frac{dx}{dt} \perp x_1$.

For the second statement, observe that from Lemma 4.2, we have:

$\langle x, \mathcal{L}x\rangle = \sum_{e \in E} w_e \max_{u,v \in e}(\frac{x_u}{\sqrt{w_u}} - \frac{x_v}{\sqrt{x_v}})^2 = \langle (x + \alpha x_1), \mathcal{L}(x + \alpha x_1)\rangle$, where the last equality holds for all real numbers α. It suffices to observe that $\Pi x = x + \alpha x_1$, for some suitable real α.

For the third statement, it is more convenient to consider transformation into the weighted space $f = \mathsf{W}^{-\frac{1}{2}}x$. It suffices to show that $\mathsf{L}_w(\alpha\vec{\mathbf{1}} + \beta f) = \beta\mathsf{L}_w(f)$. This follows immediately because in the definition of the diffusion process, it can be easily checked that $\Delta_e(\alpha\vec{\mathbf{1}} + \beta f) = \beta\Delta_e(f)$. ■

Proof of Theorem 4.1: Suppose $\mathcal{L}$ is the normalized Laplacian induced by the diffusion process in Lemma 4.8. Let $\gamma_2 := \min_{\vec{\mathbf{0}} \neq x \perp \mathsf{W}^{\frac{1}{2}}\vec{\mathbf{1}}} \mathcal{R}(x)$ be attained by some minimizer x_2. We use the isomorphism between the three spaces: $\mathsf{W}^{-\frac{1}{2}}\varphi = x = \mathsf{W}^{\frac{1}{2}}f$.

The third statement of Lemma 4.10 can be formulated in terms of the normalized space, which states that $\frac{d\mathcal{R}(x)}{dt} \leq 0$, where equality holds *iff* $\mathcal{L}x \in \mathrm{span}(x)$.

We claim that $\frac{d\mathcal{R}(x_2)}{dt} = 0$. Otherwise, suppose $\frac{d\mathcal{R}(x_2)}{dt} < 0$. From Lemma 4.11, we have $\frac{dx}{dt} = -\mathcal{L}x \perp \mathsf{W}^{\frac{1}{2}}\vec{\mathbf{1}}$. Hence, it follows that at this moment, the current normalized vector is at position x_2, and is moving towards the direction given by $x' := \frac{dx}{dt}|_{x=x_2}$ such that $x' \perp \mathsf{W}^{\frac{1}{2}}\vec{\mathbf{1}}$, and $\frac{d\mathcal{R}(x)}{dt}|_{x=x_2} < 0$. Therefore, for sufficiently small $\epsilon > 0$, it follows that $x_2' := x_2 + \epsilon x'$ is a non-zero vector that is perpendicular to $\mathsf{W}^{\frac{1}{2}}\vec{\mathbf{1}}$ and $\mathcal{R}(x_2') < \mathcal{R}(x_2) = \gamma_2$, contradicting the definition of x_2.

Hence, it follows that $\frac{d\mathcal{R}(x_2)}{dt} = 0$, which implies that $\mathcal{L}x_2 \in \mathrm{span}(x_2)$. Since $\gamma_2 = \mathcal{R}(x_2) = \frac{\langle x_2, \mathcal{L}x_2 \rangle}{\langle x_2, x_2 \rangle}$, it follows that $\mathcal{L}x_2 = \gamma_2 x_2$, as required. ■

5 Diffusion Processes

In Section 4, we define a diffusion process in a closed system with respect to a hypergraph according to the equation $\frac{d\varphi}{dt} = -\mathsf{L}\varphi$, where $\varphi \in \mathbb{R}^V$ is the measure vector, and L is the corresponding operator on the measure space. In this section, we consider related diffusion processes. In the stochastic diffusion process, on the top of the diffusion process, each vertex is subject to independent Brownian noise. We also consider a discretized diffusion operator, which we use to analyze the hop-diameter of a hypergraph.

5.1 Stochastic Diffusion Process

We analyze the process using Itō calculus, and the reader can refer to the textbook by Øksendal [Øks14] for relevant background.

Randomness Model. We consider the standard multi-dimensional Wiener process $\{B_t \in \mathbb{R}^V : t \geq 0\}$ with independent Brownian motion on each coordinate. Suppose the variance of the Brownian motion experienced by each vertex is proportional to its weight. To be precise, there exists $\eta \geq 0$ such that for each vertex $u \in V$, the Brownian noise introduced to u till time t is $\sqrt{\eta w_u} \cdot B_t(u)$, whose variance is $\eta w_u t$. It follows that the net amount of measure added to the system till time t is $\sum_{u \in V} \sqrt{\eta w_u} \cdot B_t(u)$, which has normal distribution $N(0, \eta t \cdot w(V))$. Observe that the special case for $\eta = 0$ is just the diffusion process in a closed system.

This random model induces an Itō process on the measure space given by the following stochastic differential equation:

$$d\Phi_t = -\mathsf{L}\Phi_t \, dt + \sqrt{\eta} \cdot \mathsf{W}^{\frac{1}{2}} \, dB_t,$$

with some initial measure Φ_0

By the transformation into the normalized space $x := \mathsf{W}^{-\frac{1}{2}}\varphi$, we consider the corresponding stochastic differential equation in the normalized space:

$$dX_t = -\mathcal{L}X_t \, dt + \sqrt{\eta} \, dB_t,$$

where $\mathcal{L}$ is the normalized Laplacian from Lemma 4.8. Observe that the random noise in the normalized space is spherically symmetric.

Convergence Metric. Given a measure vector $\varphi \in \mathbb{R}^V$, denote $\varphi^* := \frac{\langle \vec{\mathbf{1}}, \varphi \rangle}{w(V)} \cdot \mathsf{W}\vec{\mathbf{1}}$, which is the measure vector obtained by distributing the total measure $\sum_{u \in V} \varphi_u = \langle \vec{\mathbf{1}}, \varphi \rangle$ among the vertices such that each vertex u receives an amount proportional to its weight w_u.

For the normalized vector $x = \mathsf{W}^{-\frac{1}{2}}\varphi$, observe that $x^* := \mathsf{W}^{-\frac{1}{2}}\varphi^* = \frac{\langle \vec{\mathbf{1}}, \varphi \rangle}{w(V)} \cdot \mathsf{W}^{\frac{1}{2}}\vec{\mathbf{1}}$ is the projection of x into the subspace spanned by $x_1 := \mathsf{W}^{\frac{1}{2}}\vec{\mathbf{1}}$. We denote by Π the orthogonal projection operator into the subspace orthogonal to x_1.

Hence, to analyze how far the measure is from being stationary, we consider the vector $\Phi_t - \Phi_t^*$, whose ℓ_1-norm is $\|\Phi_t - \Phi_t^*\|_1 \leq \sqrt{w(V)} \cdot \|\Pi X_t\|_2$. As random noise is constantly delivered to the system, we cannot hope to argue that these random quantities approach zero as t tends to infinity. However, we can show that these random variables are stochastically dominated by distributions with bounded mean and variance as t tends to infinity. The following lemma states that a larger value of γ_2 implies that the measure is closer to being stationary.

Lemma 5.1 (Stochastic Dominance) *Suppose $\gamma_2 = \min_{0 \neq x \perp x_1} \mathcal{R}(x)$. Then, in the stochastic diffusion process described above, for each $t \geq 0$, the random variable $\|\Pi X_t\|_2$ is stochastically dominated by $\|\widehat{X}_t\|_2$, where $\widehat{X}_t$ has distribution $e^{-\gamma_2 t}\Pi X_0 + \sqrt{\frac{\eta}{2\gamma_2} \cdot (1 - e^{-2\gamma_2 t})} \cdot N(0,1)^V$, and $N(0,1)^V$ is the standard n-dimensional Guassian distribution with independent coordinates.*

Proof: Consider the function $h : \mathbb{R}^V \to \mathbb{R}$ given by $h(x) := \|\Pi x\|_2^2 = \|x - x^*\|_2^2$, where $x^* := \frac{\langle x_1, x \rangle}{w(V)} \cdot x_1$ and $x_1 := \mathsf{W}^{\frac{1}{2}}\vec{1}$. Then, one can check that the gradient is $\nabla h(x) = 2\,\Pi x$, and the Hessian is $\nabla^2 h(x) = 2(\mathsf{I} - \frac{1}{w(V)} \cdot \mathsf{W}^{\frac{1}{2}} J \mathsf{W}^{\frac{1}{2}})$, where J is the matrix where every entry is 1.

Define the Itō process $Y_t := h(X_t) = \langle \Pi X_t, \Pi X_t \rangle$. By the Itō's lemma, we have

$dY_t = \langle \nabla h(X_t), dX_t \rangle + \frac{1}{2}(dX_t)^{\mathsf{T}} \nabla^2 h(X_t)\,(dX_t)$.

To simplify the above expression, we make the substitution $dX_t = -\mathcal{L}X_t\,dt + \sqrt{\eta}\,dB_t$. From Lemma 4.11, we have for all x, $\mathcal{L}x \perp x_1$ and $\langle x, \mathcal{L}x \rangle = \langle \Pi x, \mathcal{L}\Pi x \rangle$.

Moreover, the convention for the product of differentials is $0 = dt \cdot dt = dt \cdot dB_t(u) = dB_t(u) \cdot dB_t(v)$ for $u \neq v$, and $dB_t(u) \cdot dB_t(u) = dt$. Hence, only the diagonal entries of the Hessian are relevant.

We have $dY_t = -2\langle \Pi X_t, \mathcal{L}\Pi X_t \rangle\,dt + \eta \sum_{u \in V}(1 - \frac{w_u}{w(V)})\,dt + 2\sqrt{\eta} \cdot \langle \Pi X_t, dB_t \rangle$. Observing that $\Pi X_t \perp x_1$, from the definition of γ_2, we have $\langle \Pi X_t, \mathcal{L}\Pi X_t \rangle \geq \gamma_2 \cdot \langle \Pi X_t, \Pi X_t \rangle$. Hence, we have the following inequality: $dY_t \leq -2\gamma_2 Y_t\,dt + \eta n\,dt + 2\sqrt{\eta} \cdot \langle \Pi X_t, dB_t \rangle$.

We next define another Itō process $\widehat{Y}_t := \langle \widehat{X}_t, \widehat{X}_t \rangle$ with initial value $\widehat{X}_0 := \Pi X_0$ and stochastic differential equation: $d\widehat{Y}_t = -2\gamma_2 \widehat{Y}_t\,dt + \eta n\,dt + 2\sqrt{\eta} \cdot \langle \widehat{X}_t, d\widehat{B}_t \rangle$.

We briefly explain why Y_t is stochastically dominated by $\widehat{Y}_t$ by using a simple coupling argument. If $Y_t < \widehat{Y}_t$, then we can choose dB_t and $d\widehat{B}_t$ to be independent. If $Y_t = \widehat{Y}_t$, observe that $\langle \Pi X_t, dB_t \rangle$ and $\langle \widehat{X}_t, d\widehat{B}_t \rangle$ have the same distribution, because both dB_t and $d\widehat{B}_t$ are spherically symmetric. Hence, in this case, we can choose a coupling between dB_t and $d\widehat{B}_t$ such that $\langle \Pi X_t, dB_t \rangle = \langle \widehat{X}_t, d\widehat{B}_t \rangle$.

Using Itō's lemma, one can verify that the above stochastic differential equation can be derived from the following equation involving $\widehat{X}_t$: $d\widehat{X}_t = -\gamma_2 \widehat{X}_t\,dt + \sqrt{\eta}\,d\widehat{B}_t$.

Because $d\widehat{B}_t$ has independent coordinates, it follows that the equation can be solved independently for each vertex u. Again, using the Itō lemma, one can verify that $d(e^{\gamma_2 t} X_t) = \sqrt{\eta} \cdot e^{\gamma_2 t}\,d\widehat{B}_t$. Therefore, we have the solution $\widehat{X}_t = e^{-\gamma_2 t}\widehat{X}_0 + \sqrt{\eta} \cdot e^{-\gamma_2 t} \int_0^t e^{\gamma_2 s}\,d\widehat{B}_s$, which has the same distribution as:

$e^{-\gamma_2 t}\widehat{X}_0 + \sqrt{\frac{\eta}{2\gamma_2} \cdot (1 - e^{-2\gamma_2 t})} \cdot N(0,1)^V$, as required. ■

Corollary 5.2 (Convergence and Laplacian) *In the stochastic diffusion process, as t tends to infinity, $\|\Phi_t - \Phi_t^*\|_1^2$ is stochastically dominated by $\frac{\eta \cdot w(V)}{2\gamma_2} \cdot \chi^2(n)$, where $\chi^2(n)$ is the chi-squared distribution with n degrees of freedom. Hence, $\lim_{t \to \infty} E[\|\Phi_t - \Phi_t^*\|_1] \leq \sqrt{\frac{\eta n \cdot w(V)}{2\gamma_2}}$.*

Remark. Observe that the total measure introduced into the system is $\sum_{u \in V} \sqrt{\eta w_u} \cdot B_t(u)$, which has standard deviation $\sqrt{\eta t \cdot w(V)}$. Hence, as t increases, the "error rate" is at most $\sqrt{\frac{n}{2\gamma_2 t}}$.

Proof: Observe that, as t tends to infinity, $\widehat{Y}_t = \|\widehat{X}_t\|_2^2$ converges to the distribution $\frac{\eta}{2\gamma_2} \cdot \chi^2(n)$, where $\chi^2(n)$ is the chi-squared distribution with n degrees of freedom (having mean n and standard deviation $\sqrt{2n}$).

Finally, observing that $\|\Phi_t - \Phi_t^*\|_1^2 \leq w(V) \cdot \|\Pi X_t\|_2^2$, it follows that as t tends to infinity, $\|\Phi_t - \Phi_t^*\|_1^2$ is stochastically dominated by the distribution $\frac{\eta \cdot w(V)}{2\gamma_2} \cdot \chi^2(n)$, which has mean $\frac{\eta n \cdot w(V)}{2\gamma_2}$ and standard deviation $\frac{\eta\sqrt{n}\cdot w(V)}{\sqrt{2}\gamma_2}$. ∎

Corollary 5.3 (Upper Bound for Mixing Time for $\eta = 0$) *Consider the deterministic diffusion process with $\eta = 0$, and some initial probability measure $\varphi_0 \in \mathbb{R}_+^V$ such that $\langle \vec{\mathbf{1}}, \varphi_0 \rangle = 1$. Denote $\varphi^* := \frac{1}{w(V)} \cdot \mathsf{W}\vec{\mathbf{1}}$, and $\varphi^*_{\min} := \min_{u \in V} \varphi^*(u)$. Then, for any $\delta > 0$ and $t \geq \frac{1}{\gamma_2} \log \frac{1}{\delta\sqrt{\varphi^*_{\min}}}$, we have $\|\Phi_t - \varphi^*\|_1 \leq \delta$.*

Proof: In the deterministic process with $\eta = 0$, stochastic dominance becomes $\|\Pi X_t\|_2 \leq e^{\gamma_2 t} \cdot \|\Pi X_0\|_2$.

Relating the norms, we have $\|\Phi_t - \varphi^*\|_1 \leq \sqrt{w(V)} \cdot \|\Pi X_t\|_2 \leq \sqrt{w(V)} \cdot e^{-\gamma_2 t} \cdot \|\Pi X_0\|_2$.

Observe that $\|\Pi X_0\|_2^2 \leq \langle X_0, X_0 \rangle = \langle \varphi_0, \mathsf{W}^{-1}\varphi_0 \rangle \leq \frac{1}{\min_u w_u}$.

Hence, it follows that $\|\Phi_t - \varphi^*\|_1 \leq \frac{1}{\sqrt{\varphi^*_{\min}}} \cdot e^{-\gamma_2 t}$, which is at most δ, for $t \geq \frac{1}{\gamma_2} \log \frac{1}{\delta\sqrt{\varphi^*_{\min}}}$. ∎

5.2 Bottlenecks for the Hypergraph Diffusion Process

In this section we prove that if the hypergraph diffusion process mixes slowly, then it must have a set of vertices having small expansion (Theorem 3.5).

Theorem 5.4 (Restatement of Theorem 3.5) *Given a hypergraph $H = (V, E, w)$ and a probability distribution $\varphi_0 : V \to [0, 1]$, let φ_t denote the probability distribution at time t according to the diffusion process (Figure 3.1) and φ^* be the stationary distribution.*

Let $\delta > 0$. Suppose initially $\|\varphi_0 - \varphi^*\|_1 > \delta$ and for some time $T > 0$, $\|\varphi_T - \varphi^*\|_1 > \delta$. Then, there exists a set $S \subset V$ such that $\varphi^*(S) \leq \frac{1}{2}$ and

$$\phi(S) \leq \mathcal{O}\left(\frac{1}{T} \ln \frac{\|\varphi_0 - \varphi^*\|_1}{\sqrt{\varphi^*_{\min}} \cdot \delta}\right).$$

Proof: We consider the transformation $x_t := \mathsf{W}^{-\frac{1}{2}}\varphi_t$. We denote by Π the orthogonal projection operator into the subspace orthogonal to $x_1 := \mathsf{W}^{\frac{1}{2}}\vec{\mathbf{1}}$. Consider the projection $\widehat{x}_t := \Pi x_t$ onto the subspace orthogonal to x_1. Denote $x^* := \mathsf{W}^{-\frac{1}{2}}\varphi^* = \frac{1}{w(V)} \cdot \mathsf{W}^{\frac{1}{2}}\vec{\mathbf{1}}$, which is the projection of x_0 into the subspace spanned by $x_1 := \mathsf{W}^{\frac{1}{2}}\vec{\mathbf{1}}$.

Observe that $x_t = x^* + \widehat{x}_t$, where x^* is the stationary component and $\widehat{x}_t$ is the transient component. Moreover, $\varphi_t - \varphi^* = \mathsf{W}^{\frac{1}{2}}\widehat{x}_t$.

The diffusion process on the measure space induces the differential equation on $\widehat{x}_t$ as follows:

$\frac{d\widehat{x}_t}{dt} = -\mathcal{L}\widehat{x}_t$.

By expressing Lemma 4.10 (1) in terms of the normalized space, we have

$\frac{d\|\widehat{x}_t\|^2}{dt} = -2\mathcal{R}(\widehat{x}_t) \cdot \|\widehat{x}_t\|^2$.

Integrating from $t = 0$ to T and simplifying, we have

$\ln \frac{\|\widehat{x}_0\|}{\|\widehat{x}_T\|} = \int_0^T \mathcal{R}(\widehat{x}_t) dt \geq T \cdot \mathcal{R}(\widehat{x}_T)$,

where the last inequality holds because $\mathcal{R}(\widehat{x}_t)$ is decreasing according to Lemma 4.10 (3).

Since the norms are related by $\sqrt{w_{\min}} \cdot \|x\|_2 \leq \|\varphi\|_1 \leq \sqrt{w(V)} \cdot \|x\|_2$, we have

$\mathcal{R}(\widehat{x}_T) \leq \frac{1}{T} \ln \frac{\|\widehat{x}_0\|}{\|\widehat{x}_T\|} \leq \frac{1}{T} \ln(\frac{1}{\sqrt{\varphi^*_{\min}}} \cdot \frac{\|\varphi_0 - \varphi^*\|_1}{\|\varphi_T - \varphi^*\|_1}) \leq \frac{1}{T} \ln \frac{\|\varphi_0 - \varphi^*\|_1}{\sqrt{\varphi^*_{\min}} \cdot \delta}$.

Finally, observing that $\widehat{x}_T \perp x_1$, Proposition 6.2 implies that there exists a set $S \subset V$ such that $\varphi^*(S) \leq \frac{1}{2}$, and $\phi(S) \leq \mathcal{O}\left(\sqrt{\mathcal{R}(\widehat{x}_T)}\right) \leq \mathcal{O}\left(\frac{1}{T} \ln \frac{\|\varphi_0 - \varphi^*\|_1}{\sqrt{\varphi^*_{\min}} \cdot \delta}\right)$. ■

5.3 Lower Bounds on Mixing Time

Next we prove Theorem 3.6.

Theorem 5.5 (Formal statement of Theorem 3.6) *Given a hypergraph $H = (V, E, w)$, suppose there exists a vector $y \perp x_1$ in the normalized space such that $\mathcal{R}(y) \leq \gamma$. Then, there exists an initial probability distribution $\varphi_0 \in \mathbb{R}_+^V$ in the measure space such that $\|\varphi_0 - \varphi^*\|_1 \geq \frac{1}{2}$. Moreover, for any $\delta > 0$ and $t \leq \frac{1}{4\gamma} \ln \frac{\sqrt{\varphi^*_{\min}}}{2\delta}$, at time t of the diffusion process, we have*
$\|\varphi_t - \varphi^*\|_1 \geq \delta$.

We consider the diffusion process from the perspective of the normalized space. Recall that $x_1 := \mathsf{W}^{\frac{1}{2}}\vec{\mathbf{1}}$ is an eigenvector of the normalized Laplacian $\mathcal{L}$ with eigenvalue 0. From Lemma 4.11 (1), $\mathcal{L}(x) \perp x_1$ for all $x \in \mathbb{R}^V$. Therefore, the diffusion process has no effect on the subspace spanned by x_1, and we can focus on its orthogonal space.

Lemma 5.6 *Suppose $y \in \mathbb{R}^V$ is a non-zero vector in the normalized space such that $y \perp x_1$ and $\mathcal{R}(y) = \gamma$. If we start the diffusion process with $y_0 := y$, then after time $t \geq 0$, we have $\|y_t\|_2 \geq e^{-\gamma t} \cdot \|y_0\|_2$.*

Proof: By Lemma 4.10 (1) interpreted for the normalized space, we have
$\frac{d\|y_t\|^2}{dt} = -2\mathcal{R}(y_t) \cdot \|y_t\|^2 \geq -2\gamma \cdot \|y_t\|^2$, where the last inequality holds because from Lemma 4.10 (3), $t \mapsto \mathcal{R}(y_t)$ is a decreasing funtion, which implies that $\mathcal{R}(y_t) \leq \mathcal{R}(y_0) = \gamma$.

Integrating the above gives
$\|y_t\|^2 \geq e^{-2\gamma t} \cdot \|y_0\|^2$. ■

The next lemma shows that given a vector in the normalized space that is orthogonal to x_1, a corresponding probability distribution in the measure space that has large distance from the stationary distribution $\varphi^* := \frac{\mathsf{W}\vec{\mathbf{1}}}{w(V)}$ can be constructed.

Lemma 5.7 *Suppose $y \in \mathbb{R}^V$ is a non-zero vector in the normalized space such that $y \perp x_1$ and $\mathcal{R}(y) = \gamma$. Then, there exists $\widehat{y} \perp x_1$ such that $\mathcal{R}(\widehat{y}) \leq 4\gamma$ and $\varphi_0 := \varphi^* + \mathsf{W}^{\frac{1}{2}}\widehat{y}$ is a probability distribution (i.e., $\varphi_0 \geq 0$), and $\left\|\mathsf{W}^{\frac{1}{2}}\widehat{y}\right\|_1 \geq \frac{1}{2}$.*

Proof: One could try to consider $\varphi^* + \mathsf{W}^{\frac{1}{2}}(\alpha y)$ for some $\alpha \in \mathbb{R}$, but the issue is that to ensure that every coordinate is non-negative, the scalar α might need to have very small magnitude, leading to a very small $\left\|\mathsf{W}^{\frac{1}{2}}(\alpha y)\right\|_1$.

We construct the desired vector in several steps. We first consider $z := y + cx_1$ for an appropriate scalar $c \in \mathbb{R}$ such that both $w(\mathsf{supp}(z^+))$ and $w(\mathsf{supp}(z^-))$ are at most $\frac{1}{2} \cdot w(V)$, where z^+ is obtained from z by keeping only the positive coordinates, and z^- is obtained similarly from the negative coordinates. Observe that we have $z = z^+ + z^-$.

We use Π to denote the projection operator into the space orthogonal to x_1 in the normalized space. Then, we have $y = \Pi z = \Pi z^+ + \Pi z^-$. Without loss of generality, by replacing z with $-z$, we can assume that $\|\Pi z^+\| \geq \frac{1}{2}\|y\|$.

Observe that $\langle \Pi z^+, \mathcal{L}\Pi z^+\rangle = \langle z^+, \mathcal{L}z^+\rangle \leq \langle z, \mathcal{L}z\rangle = \langle y, \mathcal{L}y\rangle$,

where the middle inequality follows because $\langle z, \mathcal{L}z\rangle = \sum_{e \in E} w_e \max_{u,v \in e}(\frac{z_u}{\sqrt{w_u}} - \frac{z_v}{\sqrt{w_v}})^2$.

Hence, we have $\mathcal{R}(\Pi z^+) \leq 4\mathcal{R}(y)$, and we consider an appropriate scaled vector $\widehat{y} := \Pi\widehat{z}$, where $\widehat{z} = cz^+$ for some $c > 0$ such that $\langle \vec{\mathbf{1}}, \mathsf{W}^{\frac{1}{2}}\widehat{z}\rangle = 1$.

Hence, it follows that $\widehat{y} = \widehat{z} - \frac{\langle \mathsf{W}^{\frac{1}{2}}\vec{\mathbf{1}}, \widehat{z}\rangle}{w(V)} \cdot \mathsf{W}^{\frac{1}{2}}\vec{\mathbf{1}}$, which implies that $\mathsf{W}^{\frac{1}{2}}\widehat{y} = \mathsf{W}^{\frac{1}{2}}\widehat{z} - \varphi^*$.

Therefore, we have $\varphi_0 := \varphi^* + \mathsf{W}^{\frac{1}{2}}\widehat{y} = \mathsf{W}^{\frac{1}{2}}\widehat{z} \geq 0$.

Moreover, $\left\|\mathsf{W}^{\frac{1}{2}}\widehat{y}\right\|_1 \geq \langle \vec{\mathbf{1}}, \mathsf{W}^{\frac{1}{2}}\widehat{z}\rangle - \frac{w(\mathsf{supp}(z^+))}{w(V)} + \frac{w(\mathsf{supp}(z^-))}{w(V)} \geq \frac{1}{2}$, where the last inequality follows from $w(\mathsf{supp}(z^+)) \leq \frac{1}{2}w(V)$. ■

Proof of Theorem 5.5: Using Lemma 5.7, we can construct $\widehat{y}$ from y such that $\widehat{y} \perp x_1$ and $\mathcal{R}(\widehat{y}) \leq 4\gamma$.

Then, we can define the initial probability distribution $\varphi_0 := \varphi^* + \mathsf{W}^{\frac{1}{2}}\widehat{y}$ in the measure space with the corresponding $y_0 := \widehat{y}$ vector in the normalized subspace orthogonal to x_1.

By Lemma 5.6, at time t of the difffusion process, we have $\|y_t\|_2 \geq e^{-4\gamma t} \cdot \|y_0\|_2$.

Relating the norms of the measure space and the normalized space, we have

$\|\varphi_t - \varphi^*\|_1 \geq \sqrt{w_{\min}} \cdot \|y_t\|_2 \geq \sqrt{w_{\min}} \cdot e^{-4\gamma t} \cdot \|y_0\|_2 \geq \sqrt{\varphi^*_{\min}} \cdot e^{-4\gamma t} \cdot \|\varphi_0 - \varphi^*\|_1 \geq \sqrt{\varphi^*_{\min}} \cdot e^{-4\gamma t} \cdot \frac{1}{2}$.

Hence, for $t \leq \frac{1}{4\gamma} \ln \frac{\sqrt{\varphi^*_{\min}}}{2\delta}$, we have $\|\varphi_t - \varphi^*\|_1 \geq \delta$, as required. ■

Remark 5.8 *Observe that we do not know how to efficiently find $x_2 \perp x_1$ to attain $\mathcal{R}(x_2) = \gamma_2$. However, the approximation algorithm in Theorem 8.2 allows us to efficiently compute some y such that $\mathcal{R}(y) \leq O(\log r) \cdot \gamma_2$.*

Hence, we can compute a probability distribution φ_0 in polynomial time such

$$\|\varphi_0 - \varphi^*\|_1 \geq \frac{1}{2} \qquad \textit{and} \qquad \mathsf{t}^{\mathsf{mix}}_{\delta}(\varphi_0) \geq \Omega\left(\frac{1}{\gamma_2 \log r} \log \frac{\varphi^*_{\min}}{\delta}\right).$$

5.4 Hypergraph Diameter

In this section we prove Theorem 3.7.

Theorem 5.9 (Restatement of Theorem 3.7) *Given a hypergraph $H = (V, E, w)$, its hop-diameter is*

$$\mathsf{diam}(H) = \mathcal{O}\left(\frac{\log N_w}{\gamma_2}\right),$$

where $N_w := \max_{u \in V} \frac{w(V)}{w_u}$ and γ_2 is the eigenvalue of the normalized Laplacian as defined in Theorem 4.1.

We start by defining the notion of discretized diffusion operator.

Definition 5.10 (Discretized Diffusion Operator) *Recalling that a diffusion process in the measure space is defined in Section 4.1 by $\frac{d\varphi}{dt} = -\mathsf{L}\varphi$, we define a discretized diffusion operator on the measure space by $\mathsf{M} := \mathsf{I} - \frac{1}{2} \cdot \mathsf{L}$.*

Moreover, using the isomorphism between the measure space and the normalized space, we define the corresponding operator on the normalized space $\mathcal{M} := \mathsf{I} - \frac{1}{2} \cdot \mathcal{L}$.

When we consider the diffusion process, it is more convenient to think in terms of the measure space. However, the normalized space is more convenient for considering orthogonality.

Next, we bound the norm of the discretized diffusion operator.

Lemma 5.11 *For a vector x in the normalized space such that $x \perp x_1 := \mathsf{W}^{\frac{1}{2}}\vec{\mathbf{1}}$, we have $\|\mathcal{M}x\|_2 \leq \sqrt{1 - \frac{\gamma_2}{2}} \cdot \|x\|_2$.*

Proof: Fix $x \perp x_1 := \mathsf{W}^{\frac{1}{2}}\vec{\mathbf{1}}$. Observe that $\mathcal{M}x = \widehat{M}x$ for some symmetric matrix $\widehat{M} := \mathsf{I} - \frac{1}{2} \cdot \widehat{L}$, where the matrix $\widehat{L}$ depends on x and has the form $\widehat{L} := \mathsf{I} - \mathsf{W}^{-\frac{1}{2}}\widehat{A}\mathsf{W}^{-\frac{1}{2}}$. The precise definition of $\widehat{A}$ (depending on x) is given in Section 4.1, but the important property is that $\widehat{A}$ is a non-negative symmetric matrix such that sum of entries in row u is w_u.

Standard spectral graph theory and linear algebra state that $\mathbb{R}^V$ has a basis consisting of orthonormal eigenvectors $\{v_1, v_2, \ldots, v_n\}$ of $\widehat{L}$, whose eigenvalues are in $[0, 2]$. Hence, the matrix $\widehat{M}$ has the same eigenvectors; suppose the eigenvalue of v_i is $\lambda_i \in [0, 1]$.

We write $x := \sum_{i=1}^n c_i v_i$ for some real c_i's. Then, we have $\|\mathcal{M}x\|_2^2 = \sum_i \lambda_i^2 c_i^2 \leq \sum_i \lambda_i c_i^2 = \langle x, \mathcal{M}x\rangle = \langle x, x\rangle - \frac{1}{2}\langle x, \mathcal{L}x\rangle \leq (1 - \frac{\gamma_2}{2}) \|x\|_2^2$,

where the last inequality follows from $\langle x, \mathcal{L}x\rangle \geq \gamma_2 \|x\|_2^2$, because of the definition of γ_2 and $x \perp x_1$.

Hence, the result follows. ■

Proof of Theorem 3.7: The high level idea is based on the following observation. Suppose S is the support of a non-negative vector φ in the measure space. Then, applying the discretized diffusion operator M to φ has the effect of spreading the measure on S to vertices that are within one hop from S, where two vertices u and v are within one hop from each other if there is an edge e that contains both u and v.

Therefore, to prove that a hypergraph has hop-diameter at most l, it suffices to show that, starting from a measure vector φ whose support consists of only one vertex, applying the operator M to φ for l times spreads the support to all vertices. Since we consider orthogonal projection, it will be more convenient to perform the calculation in the normalized space.

Given a vertex $u \in V$, denote $\chi_u \in \mathbb{R}^V$ as the corresponding characteristic unit vector in the normalized space. The goal is to show that if l is large enough, then for all vertices u and v, we have $\langle \chi_u, \mathcal{M}^l(\chi_v)\rangle > 0$.

We use Π to denote the projection operator into the subspace that is orthogonal to $x_1 := \mathsf{W}^{\frac{1}{2}}\vec{\mathbf{1}}$. Then, we have $\chi_u = \frac{\sqrt{w_u}}{w(V)} \cdot x_1 + \Pi\chi_u$.

Lemma 4.11 implies that for all x, $\mathcal{M}(x) \perp x_1$, and for all real α, $\mathcal{M}(\alpha x_1 + x) = \alpha x_1 + \mathcal{M}(x)$.

Hence, we have $\langle \chi_u, \mathcal{M}^l \chi_v\rangle = \frac{\sqrt{w_u w_v}}{w(V)} + \langle \Pi\chi_u, \mathcal{M}^l(\Pi\chi_v)\rangle$. Observe that the first term $\frac{\sqrt{w_u w_v}}{w(V)} \geq \frac{1}{N_w}$, where $N_w := \max_{u \in V} \frac{w(V)}{w_u}$.

For the second term, we have $\langle \Pi\chi_u, \mathcal{M}^l(\Pi\chi_v)\rangle \leq \|\Pi\chi_u\|_2 \cdot \left\|\mathcal{M}^l(\Pi\chi_v)\right\|_2 \leq (1 - \frac{\gamma_2}{2})^{l/2}$, where the first inequality follows from Cauchy-Schwartz and the second inequality follows from applying Lemma 5.11 for l times.

Hence, for l larger than $\frac{2\log N_w}{\log \frac{1}{1-\frac{\gamma_2}{2}}} = \mathcal{O}\left(\frac{\log N_w}{\gamma_2}\right)$, we have $\langle \Pi\chi_u, \mathcal{M}^l(\Pi\chi_v)\rangle > 0$, as required.

■

6 Cheeger Inequalities for Hypergraphs

In this section, we generalize the Cheeger inequalities to hypergraphs. For the basic version, we relate the expansion of a hypergraph with the eigenvalue γ_2 of the Laplacian $\mathcal{L}$ defined in Section 4. However, at the moment, we cannot exploit the higher order spectral properties of $\mathcal{L}$. Instead, we achieve higher order Cheeger inequalities in terms of the orthogonal minimaximizers defined in Section 3.3.

6.1 Basic Cheeger Inequalities for Hypergraphs

We prove the basic Cheeger inequalities for hypergraphs.

Theorem 6.1 (Restatement of Theorem 3.8) *Given an edge-weighted hypergraph H, we have:*

$$\frac{\gamma_2}{2} \leq \phi_H \leq \gamma_2 + 2\sqrt{\frac{\gamma_2}{r_{\min}}} \leq 2\sqrt{\gamma_2},$$

where ϕ_H is the hypergraph expansion and γ_2 is the eignenvalue of $\mathcal{L}$ as in Theorem 4.1.

Towards proving this theorem, we first show that a *good* line-embedding of the hypergraph suffices to upper bound the expansion.

Proposition 6.1 *Let $H = (V, E, w)$ be a hypergraph with edge weights $w : E \to \mathbb{R}^+$ and let $f \in \mathbb{R}_+^V$ be a non-zero vector. Then, there exists a set $S \subseteq \mathsf{supp}(f)$ such that*

$$\phi(S) \leq \frac{\sum_{e \in E} w_e \max_{u,v \in e} |f_u - f_v|}{\sum_u w_u f_u}.$$

Proof: The proof is similar to the proof of the corresponding statement for vertex expansion in graphs [LRV13]. Observe that in the result, the upper bound on the right hand side does not change if f is multiplied by a positive scalar. Hence, we can assume, without loss of generality, that $f \in [0,1]^V$.

We define a family of functions $\{F_r : [0,1] \to \{0,1\}\}_{r \in [0,1]}$ as follows.

$$F_r(x) = \begin{cases} 1 & x \geq r \\ 0 & \text{otherwise} \end{cases}.$$

For $r \geq 0$ and a vector $f \in [0,1]^V$, we consider the induced vector $F_r(f) \in \{0,1\}^V$, whose coordinate corresponding to v is $F_r(f_v)$. Let S_r denote the support of the vector $F_r(f)$. For any $a \in [0,1]$ we have

$$\int_0^1 F_r(a) \, \mathsf{dr} = a\,. \tag{6.1}$$

Now, observe that if $a - b \geq 0$, then $F_r(a) - F_r(b) \geq 0, \forall r \in [0,1]$; similarly, if $a - b \leq 0$ then $F_r(a) - F_r(b) \leq 0, \forall r \in [0,1]$. Therefore,

$$\int_0^1 |F_r(a) - F_r(b)| \, \mathsf{dr} = \left| \int_0^1 F_r(a)\mathsf{dr} - \int_0^1 F_r(b)\mathsf{dr} \right| = |a - b|\ . \tag{6.2}$$

Also, for a hyperedge e, if $u = \arg\max_{u \in e} f_u$ and $v = \arg\min_{u \in e} f_u$, then

$$|F_r(f_u) - F_r(f_v)| \geq |F_r(f_{u'}) - F_r(f_{v'})|\,, \quad \forall r \in [0,1] \text{ and } \forall u', v' \in e\,. \tag{6.3}$$

Therefore, we have

$$\begin{aligned} \frac{\int_0^1 \sum_e w_e \max_{u,v \in e} |F_r(f_u) - F_r(f_v)| \, \mathsf{dr}}{\int_0^1 \sum_u w_u F_r(f_u) \mathsf{dr}} &= \frac{\sum_e w_e \max_{u,v \in e} \int_0^1 |F_r(f_u) - F_r(f_v)| \, \mathsf{dr}}{\int_0^1 \sum_u w_u F_r(f_u) \mathsf{dr}} && \text{(Using 6.3)} \\ &= \frac{\sum_e w_e \max_{u,v \in e} \left| \int_0^1 F_r(f_u)\mathsf{dr} - \int_0^1 F_r(f_v)\mathsf{dr} \right|}{\sum_u w_u \int_0^1 F_r(f_u)\mathsf{dr}} && \text{(Using 6.2)} \\ &= \frac{\sum_e w_e \max_{u,v \in e} |f_u - f_v|}{\sum_u w_u f_u}. && \text{(Using 6.1)} \end{aligned}$$

Therefore, there exists $r' \in [0,1]$ such that

$$\frac{\sum_e w_e \max_{u,v \in e} |F_{r'}(f_u) - F_{r'}(f_v)|}{\sum_u w_u F_{r'}(f_u)} \leq \frac{\sum_e w_e \max_{u,v \in e} |f_u - f_v|}{\sum_u w_u f_u}\,.$$

Since $F_{r'}(\cdot)$ takes value in $\{0,1\}$, we have

$$\frac{\sum_e w_e \max_{u,v\in e} |F_{r'}(f_u) - F_{r'}(f_v)|}{\sum_{u\in V} w_u F_{r'}(f_u)} = \frac{\sum_e w_e \cdot \mathbb{I}\left[e \text{ is cut by } S_{r'}\right]}{\sum_{u\in S_{r'}} w_u} = \phi(S_{r'})\,.$$

Therefore,

$$\phi(S_{r'}) \le \frac{\sum_e w_e \max_{u,v\in e} |f_u - f_v|}{\sum_u w_u f_u} \qquad \text{and} \qquad S_{r'} \subseteq \mathsf{supp}(f)\,.$$

■

Proposition 6.2 *Given an edge-weighted hypergraph $H = (V, E, w)$ and a non-zero vector $f \in \mathbb{R}^V$ such that $f \perp_w \vec{\mathbf{1}}$, there exists a set $S \subset V$ such that $w(S) \le \frac{w(V)}{2}$ and*

$$\phi(S) \le \mathsf{D}_w(f) + 2\sqrt{\frac{\mathsf{D}_w(f)}{r_{\min}}},$$

where $\mathsf{D}_w(f) = \frac{\sum_{e\in E} w_e \max_{u,v\in e} (f_u - f_v)^2}{\sum_{u\in V} w_u f_u^2}$ and $r_{\min} = \min_{e\in E} |e|$.

Proof: Let $g = f + c\vec{\mathbf{1}}$ for an appropriate $c \in \mathbb{R}$ such that both $w(\mathsf{supp}(g^+))$ and $w(\mathsf{supp}(g^-))$ are at most $\frac{w(V)}{2}$. For instance, sort the coordinates of f such that $f(v_1) \le f(v_2) \le \cdots \le f(v_n)$ and pick $c = f(v_i)$, where i is the smallest index such that $\sum_{j=1}^{i} w(v_j) \ge \frac{w(V)}{2}$.

Since $f \perp_w \vec{\mathbf{1}}$, it follows that $\langle g, \vec{\mathbf{1}}\rangle_w = c\langle \vec{\mathbf{1}}, \vec{\mathbf{1}}\rangle_w$. Hence, we have $\langle f, f\rangle_w = \langle g, g\rangle_w - 2c\langle g, \vec{\mathbf{1}}\rangle_w + c^2\langle \vec{\mathbf{1}}, \vec{\mathbf{1}}\rangle_w = \langle g, g\rangle_w - c^2\langle \vec{\mathbf{1}}, \vec{\mathbf{1}}\rangle_w \le \langle g, g\rangle_w$.

Therefore, we have

$$\mathsf{D}_w(f) = \frac{\sum_{e\in E} w_e \max_{u,v\in e} (g_u - g_v)^2}{\langle f, f\rangle_w} \ge \frac{\sum_{e\in E} w_e \max_{u,v\in e} (g_u - g_v)^2}{\langle g, g\rangle_w} = \mathsf{D}_w(g)\,.$$

For any $a, b \in R$, we have

$$(a^+ - b^+)^2 + (a^- - b^-)^2 \le (a - b)^2.$$

Therefore, we have

$$\begin{aligned}
\mathsf{D}_w(f) &\ge \mathsf{D}_w(g) = \frac{\sum_{e\in E} w_e \max_{u,v\in e} (g_u - g_v)^2}{\sum_u w_u g_u^2} \\
&\ge \frac{\left(\sum_{e\in E} w_e \max_{u,v\in e} (g_u^+ - g_v^+)^2\right) + \left(\sum_{e\in E} w_e \max_{u,v\in e} (g_u^- - g_v^-)^2\right)}{\sum_u w_u (g_u^+)^2 + \sum_u w_u (g_u^-)^2} \\
&\ge \min\left\{\frac{\sum_{e\in E} w_e \max_{u,v\in e} (g_u^+ - g_v^+)^2}{\sum_u w_u (g_u^+)^2}, \frac{\sum_{e\in E} w_e \max_{u,v\in e} (g_u^- - g_v^-)^2}{\sum_u w_u (g_u^-)^2}\right\} \\
&= \min\left\{\mathsf{D}_w(g^+), \mathsf{D}_w(g^-)\right\}.
\end{aligned}$$

Let $h \in \{g^+, g^-\}$ be the vector corresponding the minimum in the previous inequality. Then, we have

$$\begin{aligned}
\sum_{e \in E} w_e \max_{u,v \in e} \left| h_u^2 - h_v^2 \right| &= \sum_{e \in E} w_e \max_{u,v \in e} \left| h_u - h_v \right| (h_u + h_v) \\
&= \sum_{e \in E} w_e \max_{u,v \in e} (h_u - h_v)^2 + 2 \sum_{e \in E} w_e \min_{u \in e} h_u \max_{u,v \in e} \left| h_u - h_v \right| \\
&\leq \sum_{e \in E} w_e \max_{u,v \in e} (h_u - h_v)^2 + 2 \sqrt{\sum_{e \in E} w_e \max_{u,v \in e} (h_u - h_v)^2} \sqrt{\sum_{e \in E} w_e \cdot \frac{\sum_{u \in e} h_u^2}{r_{\min}}} \\
&= \sum_{e \in E} w_e \max_{u,v \in e} (h_u - h_v)^2 + 2 \sqrt{\sum_{e \in E} w_e \max_{u,v \in e} (h_u - h_v)^2} \sqrt{\frac{\sum_{u \in V} w_u h_u^2}{r_{\min}}},
\end{aligned}$$

where the inequality follows from the Cauchy-Schwarz's Inequality.

Using $\mathsf{D}_w(h) \leq \mathsf{D}_w(f)$,

$$\frac{\sum_{e \in E} w_e \max_{u,v \in e} \left| h_u^2 - h_v^2 \right|}{\sum_u w_u h_u^2} \leq \mathsf{D}_w(h) + 2 \sqrt{\frac{\mathsf{D}_w(h)}{r_{\min}}} \leq \mathsf{D}_w(f) + 2 \sqrt{\frac{\mathsf{D}_w(f)}{r_{\min}}}.$$

Invoking Proposition 6.1 with vector h^2, we get that there exists a set $S \subset \mathsf{supp}\,(h)$ such that

$$\phi(S) \leq \mathsf{D}_w(f) + 2 \sqrt{\frac{\mathsf{D}_w(f)}{r_{\min}}} \qquad \text{and} \qquad w(S) \leq w(\mathsf{supp}\,(h)) \leq \frac{w(V)}{2}.$$

■

The "hypergraph orthogonal separators" construction due to [LM14b] can also be used to prove Proposition 6.2, albeit with a much larger absolute constant in the bound on the expansion of the set S.

We are now ready to prove Theorem 6.1.

Proof of Theorem 6.1 (and 3.8):

1. Let $S \subset V$ be any set such that $w(S) \leq \frac{w(V)}{2}$, and let $g \in \{0,1\}^V$ be the indicator vector of S. Let f be the component of g orthogonal to $\vec{\mathbf{1}}$ (in the weighted space). Then, $g = f + c\vec{\mathbf{1}}$, where $c = \frac{\langle g, \vec{\mathbf{1}} \rangle_w}{\langle \vec{\mathbf{1}}, \vec{\mathbf{1}} \rangle_w} = \frac{w(S)}{w(V)}$.

   Moreover, as in the proof of Proposition 6.2, we have $\langle f, f \rangle_w = \langle g, g \rangle_w - c^2 \langle \vec{\mathbf{1}}, \vec{\mathbf{1}} \rangle_w = w(S) \cdot (1 - \frac{w(S)}{w(V)}) \geq \frac{w(S)}{2}$.

   Then, since $g \neq \vec{\mathbf{1}}$, we have $0 \neq f \perp_w \vec{\mathbf{1}}$ and so we have

$$\begin{aligned}
\gamma_2 \leq \mathsf{D}_w(f) &= \frac{\sum_e w_e \max_{u,v \in e} (g_u - g_v)^2}{\langle f, f \rangle_w} \\
&\leq \frac{w(\partial S)}{w(S)/2} = 2\phi(S).
\end{aligned}$$

   Since the choice of the set S was arbitrary, we have $\frac{\gamma_2}{2} \leq \phi_H$.

2. Invoking Proposition 6.2 with the minimizer h_2 such that $\gamma_2 = \mathsf{D}_w(h_2)$, we get that $\phi_H \leq \gamma_2 + 2\sqrt{\frac{\gamma_2}{r_{\min}}}$.

   For $\gamma_2 \leq \frac{1}{4}$, we observe that $r_{\min} \geq 2$ and have $\phi_H \leq (\frac{1}{2} + \sqrt{2}) \cdot \sqrt{\gamma_2} \leq 2\sqrt{\gamma_2}$; for $\gamma_2 > \frac{1}{4}$, observe that we have $\phi_H \leq 1 \leq 2\sqrt{\gamma_2}$.

We remark that the constant 2 in the upper bound can be improved slightly by optimizing the threshold for γ_2 in the above case analysis, and further considering cases whether $r_{\min} = 2$ or $r_{\min} \geq 3$.

■

6.2 Higher Order Orthogonal Minimaximizers

As mentioned in Section 3.3, we do not yet know about higher order spectral properties of the Laplacian $\mathcal{L}$. Hence, to achieve results like higher order Cheeger-like inequalities, we consider the notion of orthogonal minimaximizers with respect to the discrepancy ratio.

In Section 3.3, the parameters ξ_k and ζ_k are defined in terms of the normalized space. We can equivalently define them in terms of the weighted space as $\xi_k := \min_{f_1,\ldots,f_k} \max_{i\in[k]} \mathsf{D}_w(f_i)$ and $\zeta_k := \min_{f_1,\ldots,f_k} \max\{\mathsf{D}_w(f) : f \in \mathrm{span}\{f_1,\ldots,f_k\}\}$, where the minimum is over k non-zero mutually orthogonal vectors $f_1, f_2, \ldots, f_k$ in the weighted space. The proofs shall work with either the normalized or the weighted space, depending on which is more convenient.

We do not know an efficient method to find k orthonormal vectors that achieve ξ_k or ζ_k. In Section 8, we describe how approximations of these vectors can be obtained.

We prove Lemma 3.9 that compares the parameters γ_k, ξ_k and ζ_k by the following claims.

Claim 6.2 *For $k \geq 1$, $\xi_k \leq \gamma_k$.*

Proof: Suppose the procedure produces $\{\gamma_i : i \in [k]\}$, which is attained by orthonormal vectors $X_k := \{x_i : i \in [k]\}$ in the normalized space. Observe that $\max_{i\in[k]} \mathcal{D}(x_i) = \mathcal{D}(x_k) = \gamma_k$, since x_k could have been a candidate in the minimum for defining γ_i because $x_k \perp x_j$, for all $j \in [k-1]$.

Since X_k is a candidate for taking the minimum over sets of k orthonormal vectors in the definition of ξ_k, it follows that $\xi_k \leq \gamma_k$. ■

Claim 6.3 *For $k \geq 1$, $\gamma_k \leq \zeta_k$.*

Proof: For $k = 1$, $\gamma_1 = \zeta_1 = 0$.

For $k > 1$, suppose the $\{\gamma_i : i \in [k-1]\}$ have already been constructed with the corresponding orthonormal minimizers $X_{k-1} := \{x_i : i \in [k-1]\}$.

Let $Y_k := \{y_i : i \in [k]\}$ be an arbitrary set of k orthonormal vectors. Since the subspace orthogonal to X_{k-1} has rank $n - k + 1$ and the span of Y_k has rank k, there must be a non-zero $y \in \mathrm{span}(Y_k) \cap X_{k-1}^{\perp}$. Hence, it follows that $\gamma_k = \min_{\vec{0} \neq x \in X_{k-1}^{\perp}} \mathcal{D}(x) \leq \max_{y\in\mathrm{span}(Y_k)} \mathcal{D}(y)$. Since this holds for any set Y_k of k orthonormal vectors, the result follows. ■

Claim 6.4 *Given any k orthogonal vectors $\{f_i : i \in [k]\}$ in the weighted space. We have,*

$$\zeta_k \leq k \max_{i\in[k]} \mathsf{D}_w(f_i).$$

Moreover, if the f_i's have disjoint support, we have

$$\zeta_k \leq 2 \max_{i\in[k]} \mathsf{D}_w(f_i).$$

Proof: Here it will be convenient to consider the equivalent discrepancy ratios for the weighted space.

It suffices to show that for any $h \in \mathrm{span}(\{f_i : i \in [k]\})$, $\mathsf{D}_w(h) \leq k \max_{i\in[k]} \mathsf{D}_w(f_i)$.

Suppose for some scalars α_i's, $h = \sum_{i\in[k]} \alpha_i f_i$.

For $u, v \in V$ we have

$$\begin{aligned}(h(u) - h(v))^2 &= (\sum_{i \in [k]} \alpha_i (f_i(u) - f_i(v)))^2 \\ &\leq k \sum_{i \in [k]} \alpha_i^2 (f_i(u) - f_i(v))^2,\end{aligned}$$

where the last inequality follows from Cauchy-Schwarz inequality. In the case f_i's have disjoint support, we have

$$(h(u) - h(v))^2 \leq 2 \sum_{i \in [k]} \alpha_i^2 (f_i(u) - f_i(v))^2.$$

For each $e \in E$ we have

$$\begin{aligned}\max_{u,v \in e} (h(u) - h(v))^2 &\leq \max_{u,v \in e} k \sum_{i \in [k]} \alpha_i^2 (f_i(u) - f_i(v))^2 \\ &\leq k \sum_{i \in [k]} \alpha_i^2 \max_{u,v \in e} (f_i(u) - f_i(v))^2.\end{aligned}$$

Therefore, we have

$$\begin{aligned}\mathsf{D}_w(h) &= \frac{\sum_e w_e \max_{u,v \in e} (h(u) - h(v))^2}{\sum_{u \in V} w_u h(u)^2} \\ &\leq \frac{k \sum_{i \in [k]} \alpha_i^2 \sum_e w_e \max_{u,v \in e} (f_i(u) - f_i(v))^2}{\sum_{i \in [k]} \alpha_i^2 \sum_{u \in V} w_u f_i(u)^2} \\ &\leq k \max_{i \in [k]} \mathsf{D}_w(f_i),\end{aligned}$$

as required. ■

Claim 6.5 *We have* $\gamma_2 = \zeta_2$.

Proof: From Claim 6.3, we already have $\gamma_2 \leq \zeta_2$. Hence, it suffices to show the other direction. We shall consider the discrepancy ratio for the weighted space.

Suppose $f \perp_w \mathbf{1}$ attains $\mathsf{D}_w(f) = \gamma_2$. Then, we have

$$\begin{aligned}\zeta_2 &\leq \max_{g = af + b\mathbf{1}} \frac{\sum_{e \in E} w_e \max_{u,v \in e} (g_u - g_v)^2}{\sum_{v \in V} w_v g_v^2} \\ &= \max_{g = af + b\mathbf{1}} \frac{\sum_{e \in E} w_e \max_{u,v \in e} a^2 (f_u - f_v)^2}{\sum_{v \in V} w_v (af_v + b)^2} \\ &= \max_{g = af + b\mathbf{1}} \frac{\sum_{e \in E} w_e \max_{u,v \in e} a^2 (f_u - f_v)^2}{\sum_{v \in V} w_v (a^2 f_v^2 + b^2) + 2ab \sum_{v \in V} w_v f_v} \\ &\leq \max_{g = af + b\mathbf{1}} \frac{\sum_{e \in E} w_e \max_{u,v \in e} a^2 (f_u - f_v)^2}{\sum_{v \in V} a^2 w_v f_v^2} = \gamma_2.\end{aligned}$$

■

6.3 Small Set Expansion

Even though we do not have an efficient method to generate k orthonormal vectors that attain ξ_k. As a warm up, we show that an approximation can still give us a bound on the expansion of a set of size at most $O(\frac{n}{k})$.

Theorem 6.6 (Formal Statement of 3.11) *Suppose $H = (V, E, w)$ is a hypergraph, and $f_1, f_2, \ldots, f_k$ are k orthonormal vectors in the weighted space such that $\max_{s\in[k]} \mathsf{D}_w(f_s) \leq \xi$. Then, a random set $S \subset V$ can be constructed in polynomial time such that with $\Omega(1)$ probability, $|S| \leq \frac{24|V|}{k}$ and*

$$\phi(S) \leq C \min\{\sqrt{r \log k}, k \log k \log\log k \cdot \sqrt{\log r}\} \cdot \sqrt{\xi},$$

where C is an absolute constant and r is the size of the largest hyperedge in E.

Our proof is achieved by a randomized polynomial time Algorithm 1 that computes a set S satisfying the conditions of the theorem, given vectors whose discrepancy ratios are at most ξ. We will use the following *orthogonal separator* [LM14b] subroutine. We say that a set S *cuts* another set e, if there exist $u, v \in e$ such that $u \in S$ and $v \notin S$.

Fact 6.7 (Orthogonal Separator [LM14b]) *There exists a randomized polynomial time algorithm that, given a set of unit vectors $\{\bar{u}\}_{u\in V}$, parameters $\beta \in (0,1)$ and $\tau \in \mathbb{Z}^+$, outputs a random set $\widehat{S} \subset \{\bar{u}\}_{u\in V}$ such that for some absolute constant c_1 and $\alpha = \Theta(\frac{1}{\tau})$, we have the following.*

1. *For every $\bar{u}$, $\Pr[\bar{u} \in \widehat{S}] = \alpha$.*
2. *For every $\bar{u}, \bar{v}$ such that $\langle \bar{u}, \bar{v}\rangle \leq \beta$,*
$$\Pr[\bar{u} \in \widehat{S} \text{ and } \bar{v} \in \widehat{S}] \leq \frac{\alpha}{\tau}\,.$$
3. *For any $e \subset \{\bar{u}\}_{u\in V}$*
$$\Pr[e \text{ is “cut” by } \widehat{S}] \leq \frac{c_1}{\sqrt{1-\beta}} \cdot \alpha\tau \log\tau \log\log\tau \sqrt{\log |e|} \cdot \max_{\bar{u},\bar{v}\in e} \|\bar{u} - \bar{v}\|\,.$$

Remark 6.8 *We remark that the vectors do not have to satisfy the ℓ_2^2-constraints in this version of orthogonal separators [LM14b].*

Algorithm 1 Small Set Expansion

---

1. **Spectral Embedding**. Let $f_1, \ldots, f_k$ be orthonormal vectors in the **weighted space** such that $\max_{s\in[k]} \mathsf{D}_w(f_s) \leq \xi$. We map a vertex $i \in V$ to a vector $u_i \in \mathbb{R}^k$ defined as follows. For $i \in V$ and $s \in [k]$,
$$u_i(s) = f_s(i)\,.$$
In other words, we map the vertex u to the vector formed by taking the coordinate corresponding to vertex u from $f_1, \ldots, f_k$. We consider the Euclidean ℓ_2 norm in $\mathbb{R}^k$.
2. **Normalization**. For every $i \in V$, let $\tilde{u}_i = \frac{u_i}{\|u_i\|}$.
3. **Random Projection**. Using Fact 6.7 (orthogonal separator), sample a random set $\widehat{S}$ from the set of vectors $\{\tilde{u}_i\}_{i\in V}$ with $\beta = 99/100$ and $\tau = k$, and define the vector $X \in \mathbb{R}^V$ as follows.
$$X_i := \begin{cases} \|u_i\|^2 & \text{if } \tilde{u}_i \in \widehat{S} \\ 0 & \text{otherwise} \end{cases}.$$
4. **Sweep Cut**. Sort the coordinates of the vector X in decreasing order and output the prefix having the least expansion (See Proposition 6.1).

We first prove some basic facts about the spectral embedding (Lemma 6.9), where the analogous facts for graphs are well known.

Lemma 6.9 (Spectral embedding) *We have the following.*

1.
$$\frac{\sum_{e\in E} w_e \max_{i,j\in e} \|u_i - u_j\|^2}{\sum_{i\in V} w_i \|u_i\|^2} \leq \max_{s\in[k]} \mathsf{D}_w(f_s)\,.$$

2.
$$\sum_{i\in V} w_i \|u_i\|^2 = k\,.$$

3.
$$\sum_{i\in V} w_i \langle u_j, u_i\rangle^2 = \|u_j\|^2\,, \qquad \forall j \in V\,.$$

4.
$$\sum_{e\in E} w_e \max_{i\in e} \|u_i\| \cdot \max_{i,j\in e} \|u_i - u_j\| \leq k \cdot \sqrt{\max_{s\in[k]} \mathsf{D}_w(f_s)}.$$

Proof:

1. For the first statement, we have
$\frac{\sum_{e\in E} w_e \max_{i,j\in e}\|u_i-u_j\|^2}{\sum_{i\in V} w_i\|u_i\|^2} = \frac{\sum_{e\in E} w_e \max_{i,j\in e}\sum_{s\in[k]}(f_s(i)-f_s(j))^2}{\sum_{i\in V} w_i \sum_{s\in[k]} f_s(i)^2}$
$\leq \frac{\sum_{s\in[k]}\sum_{e\in E} w_e \max_{i,j\in e}(f_s(i)-f_s(j))^2}{\sum_{s\in[k]}\sum_{i\in V} w_i f_s(i)^2} \leq \max_{s\in[k]} \mathsf{D}_w(f_s).$

2. The second statement follows because each f_s has norm 1 in the weighted space.

3. For the third statement,

$$\begin{aligned}\sum_{i\in V} w_i \langle u_j, u_i\rangle^2 &= \sum_{i\in V} w_i \left(\sum_{s\in[k]} f_s(j)f_s(i)\right)^2\\ &= \sum_{i\in V} w_i \sum_{s,t\in[k]} f_s(j)f_s(i)f_t(j)f_t(i)\\ &= \sum_{s,t\in[k]} f_s(j)f_t(j) \sum_{i\in V} w_i f_s(i)f_t(i)\\ &= \sum_{s,t\in[k]} f_s(j)f_t(j)\cdot\langle f_s, f_t\rangle_w\\ &= \sum_{s,t\in[k]} f_s(j)f_t(j)\cdot\mathbb{I}\left[s=t\right]\\ &= \sum_{s\in[k]} u_j(s)^2\\ &= \|u_j\|^2\,.\end{aligned}$$

4. For the fourth statement, using the Cauchy-Schwarz inequality, we have

$$\sum_{e\in E} w_e \max_{i\in e}\|u_i\|\cdot\max_{i,j\in e}\|u_i-u_j\| \le \sqrt{\sum_{e\in E} w_e \max_{i\in e}\|u_i\|^2}\cdot\sqrt{\sum_{e\in E} w_e \max_{i,j\in e}\|u_i-u_j\|^2}$$
$$= \sum_{e\in E} w_e \max_{i\in e}\|u_i\|^2\cdot\sqrt{\frac{\sum_{e\in E} w_e \max_{i,j\in e}\|u_i-u_j\|^2}{\sum_{e\in E} w_e \max_{i\in e}\|u_i\|^2}}$$
$$\le \sum_{e\in E} w_e \max_{i\in e}\|u_i\|^2\cdot\sqrt{\max_{s\in[k]} \mathsf{D}_w(f_s)},$$

where the last inequality follows from the first statement.

To finish with the proof, observe that

$\sum_{e\in E} w_e \max_{i\in e}\|u_i\|^2 \le \sum_{i\in V} w_i\|u_i\|^2 = k$, where the last equality follows from the second statement.

■

We denote $D := \frac{\tau}{\sqrt{1-\beta}}\cdot\log\tau\log\log\tau\cdot\sqrt{\log\tau}$.

Main Analysis To prove that Algorithm 1 outputs a set which meets the requirements of Theorem 6.6, we will show that the vector X meets the requirements of Proposition 6.1. We prove an upper bound on the numerator $\sum_{e\in E} w_e \max_{i,j\in e}|X_i - X_j|$ in Lemma 6.11 and a lower bound on the denominator $\sum_{i\in V} w_i X_i$ in Lemma 6.13. We first show a technical lemma.

Lemma 6.10 *For any non-zero vectors u and v, $\|\tilde{u}-\tilde{v}\| \le 2\frac{\|u-v\|}{\sqrt{\|u\|^2+\|v\|^2}}$.*

Proof: Denote $a := \|u\|$, $b := \|v\|$ and $\theta := \langle\tilde{u},\tilde{v}\rangle$. Then, we have

$$\|\tilde{u}-\tilde{v}\|^2\left(\|u\|^2+\|v\|^2\right) = (2-2\theta)(a^2+b^2)$$
$$\le 4(a^2-2ab\theta+b^2) = 4\|u-v\|^2,$$

where the inequality is equivalent to $(1+\theta)(a^2+b^2)-4ab\theta \ge 0$.

To see why this is true, consider the function $h(\theta) := (1+\theta)(a^2+b^2)-4ab\theta$ for $\theta\in[-1,1]$. Since $h'(\theta)$ is independent of θ, h is either monotonically increasing or decreasing. Hence, to show that h is non-negative, it suffices to check that both $h(-1)$ and $h(1)$ are non-negative. ■

Lemma 6.11 *We have $\mathbb{E}[\sum_{e\in E} w_e \max_{i,j\in e}|X_i - X_j|] \le O(D)\cdot\sqrt{\xi}$.*

Proof: For an edge $e\in E$ we have

$$\mathbb{E}[\max_{i,j\in e}|X_i-X_j|] \le \max_{i,j\in e}\left|\|u_i\|^2-\|u_j\|^2\right|\cdot\Pr[\tilde{u}_i\in\widehat{S}\ \forall i\in e] + \max_{i\in e}\|u_i\|^2\cdot\Pr[e \text{ is cut by } \widehat{S}]. \quad (6.4)$$

By Fact 6.7 (1), the probability in the first term is at most $\Theta(\frac{1}{k})$. Hence, the first term is at most

$$\frac{\Theta(1)}{k}\cdot\max_{i,j\in e}\left|\|u_i\|^2-\|u_j\|^2\right| \le \frac{\Theta(1)}{k}\cdot\max_{i,j\in e}\|u_i-u_j\|\cdot\|u_i+u_j\| \le \frac{\Theta(1)}{k}\cdot\max_{i,j\in e}\|u_i-u_j\|\max_{i\in e}\|u_i\|\ . \quad (6.5)$$

To bound the second term in (6.4), we divide the edge set E into E_1 and E_2 as follows.

$$E_1 := \left\{ e \in E : \max_{i,j \in e} \frac{\|u_i\|^2}{\|u_j\|^2} \leq 2 \right\} \quad \text{and} \quad E_2 := \left\{ e \in E : \max_{i,j \in e} \frac{\|u_i\|^2}{\|u_j\|^2} > 2 \right\} .$$

E_1 is the set of those edges whose vertices have roughly equal lengths and E_2 is the set of those edges whose vertices have large disparity in lengths.

Claim 6.12 *Suppose E_1 and E_2 are as defined above. Then, the following holds.*

(a) For $e \in E_1$, we have
$\Pr[e$ *is cut by* $\widehat{S}] \leq O(\alpha D) \cdot \frac{\max_{i,j \in e} \|u_i - u_j\|}{\max_{i \in e} \|u_i\|}$.

(b) For $e \in E_2$, we have $\max_{i \in e} \|u_i\|^2 \leq 4 \max_{i \in e} \|u_i\| \max_{i,j \in e} \|u_i - u_j\|$.

Proof: We prove the two statements.

(a) For $e \in E_1$, using Lemma 6.10 and Fact 6.7, the probability that e is cut by $\widehat{S}$ is at most

$$O(\alpha D) \cdot \max_{i,j \in e} \frac{\|u_i - u_j\|}{\sqrt{\|u_i\|^2 + \|u_j\|^2}} \leq O(\alpha D) \cdot \frac{\max_{i,j \in e} \|u_i - u_j\|}{\max_{i \in e} \|u_i\|},$$

where the inequality follows because $e \in E_1$.

(b) Fix any $e \in E_2$, and suppose the vertices in $e = [r]$ are labeled such that $\|u_1\| \geq \|u_2\| \geq \ldots \geq \|u_r\|$. Then, from the definition of E_2, we have

$$\frac{\|u_1\|^2}{\|u_r\|^2} > 2 .$$

Hence, $\max_{i,j \in e} \|u_i - u_j\| \geq \|u_1 - u_r\| \geq (1 - \frac{1}{\sqrt{2}}) \cdot \|u_1\|$. Therefore, $\max_{i \in e} \|u_i\|^2 \leq 4 \max_{i \in e} \|u_i\| \max_{i,j \in e} \|u_i - u$

■

For a hyperedge $e \in E_1$, using Claim 6.12 (a), the second term in (6.4) is at most
$\frac{O(D)}{k} \max_{i \in e} \|u_i\| \max_{i,j \in e} \|u_i - u_j\|$.

For $e \in E_2$, in the second term of (6.4), we can just upperbound the probability trivially by $1 \leq \frac{O(D)}{k}$, and use Claim 6.12 (b) to conclude that the second term is also at most
$\frac{O(D)}{k} \max_{i \in e} \|u_i\| \max_{i,j \in e} \|u_i - u_j\|$.

Hence, inequality (6.4) becomes:

$$\mathbb{E}[\max_{i,j \in e} |X_i - X_j|] \leq \frac{O(D)}{k} \cdot \max_{i \in e} \|u_i\| \max_{i,j \in e} \|u_i - u_j\| .$$

Summing over all hyperedges $e \in E$, we have

$$\mathbb{E}[\sum_{e \in E} w_e \max_{i,j \in e} |X_i - X_j|] \leq \frac{O(D)}{k} \cdot \sum_{e \in E} w_e \max_{i \in e} \|u_i\| \cdot \max_{i,j \in e} \|u_i - u_j\|$$
$$\leq O(D) \cdot \sqrt{\xi},$$

where the last inequality follows from Lemma 6.9 (4). ■

Lemma 6.13 *We have*

$$\Pr[\sum_{i\in V} w_i X_i > \frac{1}{2}] \geq \frac{1}{12}\,.$$

Proof: We denote $Y := \sum_{i\in V} w_i X_i$. We first compute $\mathbb{E}[Y]$ as follows.

$$\begin{aligned}\mathbb{E}[Y] &= \sum_{i\in V} w_i \|u_i\|^2 \Pr[\tilde{u} \in \widehat{S}] \\ &= \sum_{i\in V} w_i \|u_i\|^2 \cdot \alpha && \text{(From Fact 6.7 (1))} \\ &= k\alpha && \text{(Using Lemma 6.9 (2))}\,.\end{aligned}$$

Next we give an upper bound of $\mathbb{E}[Y^2]$.

$$\begin{aligned}\mathbb{E}[Y^2] &= \sum_{i,j\in V} w_i w_j \|u_i\|^2 \|u_j\|^2 \Pr[\tilde{u}_i, \tilde{u}_j \in \widehat{S}] \\ &\leq \sum_{\substack{i,j:\\ \langle \tilde{u}_i,\tilde{u}_j\rangle \leq \beta}} w_i w_j \|u_i\|^2 \|u_j\|^2 \Pr[\tilde{u}_i, \tilde{u}_j \in \widehat{S}] + \sum_{\substack{i,j:\\ \langle \tilde{u}_i,\tilde{u}_j\rangle > \beta}} w_i w_j \|u_i\|^2 \|u_j\|^2 \Pr[\tilde{u}_i, \tilde{u}_j \in \widehat{S}].\end{aligned}$$

We use Fact 6.7 (2) to bound the first term, and use the trivial bound of $\frac{1}{k}$ (Fact 6.7 (1)) to bound $\Pr[\tilde{u}_i, \tilde{u}_j \in S]$ in the second term. Therefore,

$$\begin{aligned}\mathbb{E}[Y^2] &\leq \sum_{\substack{i,j:\\ \langle \tilde{u}_i,\tilde{u}_j\rangle \leq \beta}} w_i w_j \|u_i\|^2 \|u_j\|^2 \cdot \frac{\alpha}{k} + \sum_{\substack{i,j:\\ \langle \tilde{u}_i,\tilde{u}_j\rangle > \beta}} w_i w_j \|u_i\|^2 \|u_j\|^2 \cdot \frac{\langle \tilde{u}_i, \tilde{u}_j\rangle^2}{\beta^2} \cdot \alpha \\ &\leq \sum_{i,j} w_i w_j \left(\frac{\alpha \|u_i\|^2 \|u_j\|^2}{k} + \frac{\alpha}{\beta^2} \langle u_i, u_j\rangle^2 \right) \\ &= \frac{\alpha}{k} \left(\sum_i w_i \|u_i\|^2 \right)^2 + \frac{\alpha}{\beta^2} \sum_{i,j} w_i w_j \langle u_i, u_j\rangle^2 \\ &= \frac{\alpha}{k} \cdot k^2 + \frac{\alpha}{\beta^2} \cdot k = \alpha k (1 + \frac{1}{\beta^2}) \leq 3k\alpha. && \text{(Using Lemma 6.9)}\end{aligned}$$

Since Y is a non-negative random variable, we get using the Paley-Zygmund inequality that

$$\Pr[Y \geq \frac{1}{2}\mathbb{E}[Y]] \geq \left(\frac{1}{2}\right)^2 \frac{\mathbb{E}[Y]^2}{\mathbb{E}[Y^2]} = \frac{1}{4} \cdot \frac{1}{3} = \frac{1}{12}\,.$$

This finishes the proof of the lemma. ■

We are now ready to finish the proof of Theorem 6.6.

Proof of Theorem 6.6:

(1) We first show that Algorithm 1 gives $S \subset V$ such that $|S| = O(\frac{n}{k})$ and $\phi(S) = O(k \log k \log\log k \cdot \sqrt{\xi \log r})$.

By the definition of Algorithm 1,

$$\mathbb{E}[|\mathsf{supp}(X)|] = \frac{n}{k}\,.$$

Therefore, by Markov's inequality,

$$\Pr[|\mathsf{supp}(X)| \leq \frac{24n}{k}] \geq 1 - \frac{1}{24}\,. \tag{6.6}$$

Using Markov's inequality and Lemma 6.11, for some large enough constant $C_1 > 0$,

$$\Pr[\sum_{e \in E} w_e \max_{u,v \in e} |X_u - X_v| \leq C_1 D \cdot \sqrt{\xi}] \geq 1 - \frac{1}{48}. \tag{6.7}$$

Therefore, using a union bound over (6.6), (6.7) and Lemma 6.11, we get that with probability at least $\frac{1}{48}$, the following happens.

(1) $\frac{\sum_{e \in E} w_e \max_{i,j \in e} |X_i - X_j|}{\sum_{i \in V} w_i X_i} \leq O(D) \cdot \sqrt{\xi}$, and

(2) $|\mathsf{supp}(X)| \leq \frac{24n}{k}$.

When these two events happen, from Proposition 6.1, Algorithm 1 outputs a set S such that $\phi(S) \leq O(D) \cdot \sqrt{\xi}$ and $|S| \leq |\mathsf{supp}(X)| = O(\frac{n}{k})$, as required.

(2) We next show that algorithmic version [LRTV12, LOT12] of Fact 3.10 for 2-graphs can give us $S \subset V$ such that $|S| = O(\frac{n}{k})$ and $\phi(S) = O(\sqrt{r\xi \log k})$.

Given edge-weighted hypergraph $H = (V, E, w)$, we define an edge-weighted 2-graph $G = (V, E')$ as follows. For each $e \in E$, where $r_e = |e|$, add a complete graph on e with each pair having weight $\frac{w_e}{r_e - 1}$. Observe that eventually a pair $\{u, v\}$ in G has weight derived from all $e \in E$ such that both u and v are in e. In this construction, each vertex u has the same weight in H and G.

We first relate the discrepancy ratios of the two graphs by showing that $\mathsf{D}_w^G(f) \leq \frac{r}{2} \cdot \mathsf{D}_w^H(f)$. Since the denominators are the same, we compare the contribution of each hyperedge $e \in E$ to the numerators. For $e \in E$ with $r_e = |e|$, its contribution to the numerator of $\mathsf{D}_w^G(f)$ is $\frac{w_e}{r_e - 1} \sum_{\{u,v\} \in \binom{e}{2}} (f_u - f_v)^2 \leq w_e \cdot \frac{r_e}{2} \cdot \max_{u,v \in e} (f_u - f_v)^2$, which is $\frac{r_e}{2}$ times the contribution of e to the numerator of $\mathsf{D}_w^H(f)$.

Hence, Fact 3.10 for 2-graphs implies that given vectors orthogonal vectors $f_1, f_2, \ldots, f_k$ in the weighted space (where $\max_{i \in [k]} \mathsf{D}_w^G(f_i) \leq \frac{r\xi}{2}$), there is a procedure to return S such that $|S| = O(\frac{n}{k})$ and $\phi^G(S) = O(\sqrt{r\xi \log k})$.

Therefore, it suffices to prove that $\phi^H(S) \leq \phi^G(S)$. Again, the denominators involved are the same. Hence, we compare the numerators. For each hyperedge $e \in \partial S$, suppose $r_e = |e|$ and $a_e = |e \cap S|$, where $0 < a_e < r_e$. Then, the contribution of e to the numerator of $\phi^G(S)$ is $\frac{w_e}{r_e - 1} \cdot a_e(r_e - a_e) \geq w_e$, which is exactly the contribution of e to the numerator of $\phi^H(S)$. Hence, the result follows.

■

6.4 Higher Order Cheeger Inequalities for Hypergraphs

In this section, we achieve an algorithm that, given k orthonormal vectors $f_1, f_2, \ldots, f_k$ in the weighted space such that $\max_{s \in [k]} \mathsf{D}_w(f_s) \leq \xi$, returns $\Theta(k)$ non-empty disjoint subsets with small expansion.

Theorem 6.14 (Restatement of Theorem 3.12) *Suppose $H = (V, E, w)$ is a hypergraph. Then, we have the following.*

(a) Suppose $f_1, f_2, \ldots, f_k$ are k orthonormal vectors in the weighted space such that $\max_{s \in [k]} \mathsf{D}_w(f_s) \leq \xi$. There is a randomized procedure that runs in polynomial time such that for every $\epsilon \geq \frac{1}{k}$, with $\Omega(1)$ probability, returns $\lfloor (1-\epsilon)k \rfloor$ non-empty disjoint sets $S_1, \ldots, S_{\lfloor (1-\epsilon)k \rfloor} \subset V$ such that

$$\max_{i \in [\lfloor (1-\epsilon)k \rfloor]} \phi(S_i) = \mathcal{O}\left(\frac{k^2}{\epsilon^{1.5}} \log \frac{k}{\epsilon} \log\log \frac{k}{\epsilon} \sqrt{\log r} \cdot \sqrt{\xi} \right).$$

(b) For any k disjoint non-empty sets $S_1, \ldots, S_k \subset V$

$$\max_{i \in [k]} \phi(S_i) \geq \frac{\zeta_k}{2},$$

where ζ_k is defined in Section 3.3.

Proof of Theorem 6.14 (b):

For an arbitrary collection of k disjoint non-empty sets $\{S_l\}_l$, let f_l be the corresponding indicator function S_l. Then, the vectors f_l's have disjoint support, and by Claim 6.4, we have

$$\frac{\zeta_k}{2} \le \max_{l \in [k]} \mathsf{D}_w(f_l) = \max_{l \in [k]} \phi(S_l).$$

■

For statement (a), the proof is similar to Section 6.3, and we also have a similar sampling algorithm.

Algorithm 2 Sample algorithm

1: Suppose $f_1, \ldots, f_k$ are orthonormal vectors in the weighted space such that $\max_{s \in [k]} \mathsf{D}_w(f_s) \le \xi$. We map a vertex $i \in V$ to a vector $u_i \in \mathbb{R}^k$ defined as follows. For $i \in V$ and $s \in [k]$,

$$u_i(s) = f_s(i)\,.$$

2: For each $i \in V$, normalize $\tilde{u}_i \leftarrow \frac{u_i}{\|u_i\|}$.

3: Using Fact 6.7 (orthogonal separator), sample $T := \frac{2 \log 4n}{\alpha}$ independent subsets $S_1, \ldots, S_T \subset V$ with the set of vectors $\{\tilde{u}_i\}_{i \in V}$, $\beta = 1 - \frac{\epsilon}{72}$ and $\tau = \frac{16k}{\epsilon}$.

4: Define measure $\mu(S) := \sum_{i \in S} w_i \|u_i\|^2$.
For each $l \in [T]$, define S'_l as follows:

$$S'_l = \begin{cases} S_l & \text{if } \mu(S_l) \le 1 + \frac{\epsilon}{4}; \\ \emptyset & \text{otherwise.} \end{cases}$$

5: For each $l \in [T]$, let $S''_l = S'_l \backslash (\cup_{j \in [l-1]} S'_j)$.

6: Arbitrarily merge sets from $\{S''_l\}$ to form sets having μ-measure in $[\frac{1}{4}, 1 + \frac{\epsilon}{4}]$ (while discarding sets with total measure at most $\frac{1}{4}$). We name the resulting sets to be $B = \{B_1, \ldots, B_t\}$.

7: For each $j \in [t]$, set $\hat{B}_j = \{i \in B_j : \|u_i\|^2 \ge r_j\}$, where r_j is chosen to minimize $\phi(\hat{B}_j)$.

8: Output the non-empty sets $\hat{B}_j$ with the smallest expansion $\phi(\hat{B}_j)$, for $j \in [t]$.

Forming Disjoint Subsets. The algorithm first uses orthogonal separator to generate subsets S_l's independently. If the μ-measure of a subset is larger than $1 + \frac{\epsilon}{4}$, then it is discarded. We first show that with high probability, each vertex is contained in some subset that is not discarded.

Lemma 6.15 (Similar to [LM14a, Lemma 2.5]) *For every vertex $i \in V$, and $l \in [T]$, we have*

$$\Pr[i \in S'_l] \ge \frac{\alpha}{2}.$$

Proof: Recall that we sample S_l using Fact 6.7 with $\beta = 1 - \frac{\epsilon}{72}$ and $\tau = \frac{16k}{\epsilon}$.

Fix $i \in V$. If $i \in S_l$, then $i \in S'_l$ unless $\mu(S_l) > 1 + \frac{\epsilon}{4}$. Hence , we only need to show that $\Pr[\mu(S_l) > 1 + \frac{\epsilon}{4} | i \in S_l] \le \frac{1}{2}$.

Define the sets V_1 and V_2 as follows

$$V_1 = \{j \in V : \langle \tilde{u}_i, \tilde{u}_j \rangle > \beta\}$$

and

$$V_2 = \{j \in V : \langle \tilde{u}_i, \tilde{u}_j \rangle \le \beta\}.$$

We next give an upper bound for $\mu(V_1)$. From Fact 6.9 (3), we have

$$1 = \sum_{j \in V} w_j \|u_j\|^2 \langle \tilde{u}_i, \tilde{u}_j \rangle^2 \geq \beta^2 \sum_{j \in V_1} w_j \|u_j\|^2 = \beta^2 \cdot \mu(V_1).$$

Hence, $\mu(V_1) \leq \beta^{-2} \leq 1 + \frac{\epsilon}{8}$.

For any $j \in V_2$, we have $\langle \tilde{u}_i, \tilde{u}_j \rangle \leq \beta$. Hence, by Fact 6.7 (2) of orthogonal separators,

$$\Pr[j \in S_l | i \in S_l] \leq \frac{1}{\tau}.$$

Therefore,

$$E[\mu(S_l \cap V_2) | i \in S_l] \leq \frac{\mu(V_2)}{\tau} \leq \frac{\mu(V)}{\tau} = \frac{\epsilon}{16},$$

where the equality holds because $\mu(V) = k$ and $\tau = \frac{16k}{\epsilon}$.

By Markov's inequality, $\Pr[\mu(S_l \cap V_2) \geq \frac{\epsilon}{8} | i \in S_l] \leq \frac{1}{2}$.

Since $\mu(S_l) = \mu(S_l \cap V_1) + \mu(S_l \cap V_2)$, we get

$\Pr[\mu(S_l) > 1 + \frac{\epsilon}{4} | i \in S_l] \leq \Pr[\mu(S_l \cap V_2) \geq \frac{\epsilon}{8} | i \in S_l] \leq \frac{1}{2}$, as required. ■

Lemma 6.16 *With probability at least $\frac{3}{4}$, every vertex is contained in at least one S'_l. Moreover, when this happens, Algorithm 2 returns at least $t \geq \lfloor k(1-\epsilon) \rfloor$ non-empty disjoint subsets.*

Proof: From Lemma 6.15, the probability that a vertex is not included in S'_l for all $l \in [T]$ is at most $(1 - \frac{\alpha}{2})^T \leq \exp(-\frac{\alpha T}{2}) \leq \frac{1}{4n}$. Hence, by the union bound, the probability that there exists a vertex not included in at least one S'_l is at most $\frac{1}{4}$.

When every vertex is included in some S'_l, then the total μ-measure of the S''_l's is exactly $\mu(V) = k$. Since we merge the S''_l's to form subsets of μ-measure in the range $[\frac{1}{4}, 1 + \frac{\epsilon}{4}]$, at most a measure of $\frac{1}{4}$ will be discarded.

Hence, the number of subsets formed is at least $t \geq \frac{k - \frac{1}{4}}{1 + \frac{\epsilon}{4}} \geq (1 - \epsilon)k$, where the last inequality holds because $\frac{1}{k} \leq \epsilon < 1$. ■

Bounding Expansion. After we have shown that the algorithm returns enough number of subsets (each of which having μ-measure at least $\frac{1}{4}$), it remains to show that their expansion is small. In addition to measure μ, we also consider measure

$\nu(S) := \sum_{e \subset S} w_e \max_{i,j \in e}(\|u_i\|^2 - \|u_j\|^2) + \sum_{e \in \partial S} w_e \max_{i \in S \cap e} \|u_i\|^2$.

The next lemma shows that there is a non-empty subset of S having expansion at most $\frac{\nu(S)}{\mu(S)}$.

Lemma 6.17 *Suppose S is a subset of V. For $r \geq 0$, denote $S_r := \{i \in S : \|u_i\|^2 \geq r\}$. Then, there exists $r > 0$ such that $S_r \neq \emptyset$ and $\phi(S_r) \leq \frac{\nu(S)}{\mu(S)}$.*

Proof: Suppose r is sampled uniformly from the interval $(0, M)$, where $M := \max_{i \in S} \|u_i\|^2$. Observe that for $r \in (0, M)$, S_r is non-empty.

Then, it follows that an edge e can be in ∂S_r only if $e \subset S$ or $e \in \partial S$.

For $e \subset S$, $e \in \partial S_r$ *iff* there exists $i, j \in e$ such that $\|u_i\|^2 < r \leq \|u_j\|^2$.

On the other hand, if $e \in \partial S$, then $e \in \partial S_r$ *iff* $r \leq \max_{i \in S \cap e} \|u_i\|^2$.

Hence, $\mathbb{E}[w(\partial S_r)] = \frac{\nu(S)}{M}$.

Similarly, $i \in S$ is in S_r *iff* $r \leq \|u_i\|^2$. Hence, $\mathbb{E}[w(S_r)] = \frac{\mu(S)}{M}$.

Therefore, there exists $M > \rho > 0$ such that $\phi(S_\rho) = \frac{w(\partial S_\rho)}{w(S_\rho)} \le \frac{\mathbb{E}[w(\partial S_r)]}{\mathbb{E}[w(S_r)]} = \frac{\nu(S)}{\mu(S)}$. ■

In view of Lemma 6.17, it suffices to show that the algorithm generates subsets with small ν-measure.

Lemma 6.18 *Algorithm 2 produces subsets B_j's such that*

$$\mathbb{E}[\max_{l\in[t]} \nu(B_l)] \le O(D) \cdot k\sqrt{\xi_k},$$

where $D = \frac{\tau}{\sqrt{1-\beta}} \cdot \log\tau \log\log\tau\sqrt{\log r}$, and $r = \max_{e\in E}|e|$.

Proof:

Let $E_{cut} := \cup_{l\in[t]}\partial B_l$ be the set of edges cut by $B_1, \ldots, B_t$. Then, for all $l \in [t]$,
$\nu(B_l) \le \sum_{e\in E_{cut}} w_e \max_{i\in e}\|u_i\|^2 + \sum_{e\in E} w_e \max_{i,j\in e}(\|u_i\|^2 - \|u_j\|^2)$.
Hence, $\max_{l\in[t]} \nu(B_j)$ also has the same upper bound. Taking expectation, we have

$$\mathbb{E}[\max_{l\in[t]} \nu(B_j)] \le \mathbb{E}[\sum_{e\in E_{cut}} w_e \max_{i\in e}\|u_i\|^2] + \sum_{e\in E} w_e \max_{i,j\in e}(\|u_i\|^2 - \|u_j\|^2). \tag{6.8}$$

The second term in (6.8) is

$$\begin{aligned}\sum_{e\in E} w_e \max_{i,j\in e}(\|u_i\|^2 - \|u_j\|^2) &\le \sum_{e\in E} w_e \max_{i,j\in e}\|u_i - u_j\| \cdot \|u_i + u_j\| \\ &\le 2\sum_{e\in E} w_e \max_{i,j\in e}\|u_i - u_j\| \max_{i\in e}\|u_i\|.\end{aligned}$$

To bound the first term in (6.8), we divide the edge set E into two parts E_1 and E_2 as follows

$$E_1 = \{e \in E : \max_{i,j\in e}\frac{\|u_i\|^2}{\|u_j\|^2} \le 2\} \quad \text{and} \quad E_2 = \{e \in E : \max_{i,j\in e}\frac{\|u_i\|^2}{\|u_j\|^2} > 2\}.$$

The first term in (6.8) is

$$\mathbb{E}[\sum_{e\in E_{cut}} w_e \max_{i\in e}\|u_i\|^2] \le \sum_{e\in E_1} \Pr[e \in \cup_{l\in[t]}\partial B_l] \cdot w_e \max_{i\in e}\|u_i\|^2 + \sum_{e\in E_2} w_e \max_{i\in e}\|u_i\|^2. \tag{6.9}$$

We next bound the contribution from edges in E_1. Fix an edge $e \in E_1$. Recall that for $l \in [T]$, the set S_l is generated independently by the orthogonal separator (Lemma 6.7). For $l \in [T]$, we define $\mathcal{E}_l$ to be the event that for $l' \in [l-1]$, $S_{l'} \cap e = \emptyset$ and $e \in \partial S_l$.

Observe that $e \in \cup_{l\in[t]}\partial B_l$ implies that there exists $l \in [T]$ such that the event $\mathcal{E}_l$ happens. Next, if $\widehat{S}$ is sampled from the orthogonal separator in Lemma 6.7, then Lemma 6.15 implies that $\Pr[\widehat{S} \cap e = \emptyset] \le 1 - \frac{\alpha}{2}$, and Claim 6.12 (a) states that

$\Pr[e \in \partial\widehat{S}] \le O(\alpha D) \cdot \frac{\max_{i,j\in e}\|u_i - u_j\|}{\max_{i\in e}\|u_i\|}$.

Therefore, we have

$$\begin{aligned}\Pr[e \in \cup_{l\in[t]} \partial B_l] &\le \sum_{l\in[T]} \Pr[\mathcal{E}_l] \\ &\le \sum_{l\in[T]} (1-\frac{\alpha}{2})^{l-1} \cdot \Pr[e \in \partial \widehat{S}] \\ &\le \frac{2}{\alpha} \cdot \Pr[e \in \partial \widehat{S}] \\ &\le O(D) \cdot \frac{\max_{i,j\in e} \|u_i - u_j\|}{\max_{i\in e} \|u_i\|}.\end{aligned}$$

Hence, the first term in (6.9) is
$\sum_{e\in E_1} \Pr[e \in \cup_{l\in[t]} \partial B_l] \cdot w_e \max_{i\in e} \|u_i\|^2 \le \sum_{e\in E_1} w_e \max_{i,j\in e} \|u_i - u_j\| \cdot \max_{i\in e} \|u_i\|$.
For $e \in E_2$, Claim 6.12 (b) implies that the second term in (6.9) is
$\sum_{e\in E_2} w_e \max_{i\in e} \|u_i\|^2 \le \sum_{e\in E_2} 4 w_e \max_{i\in e} \|u_i\| \max_{i,j\in e} \|u_i - u_j\|$.
Therefore, it follows that

$$\begin{aligned}\mathbb{E}[\max_{l\in[t]} \nu(B_l)] &= O(D) \cdot \sum_{e\in E} w_e \max_{i\in e} \|u_i\| \max_{i,j\in e} \|u_i - u_j\| \\ &\le O(D) \cdot k \sqrt{\max_{s\in[k]} \mathsf{D}_w(f_s)} \\ &\le O(D) \cdot k \cdot \sqrt{\xi},\end{aligned}$$

where the second to last inequality comes from Lemma 6.9 (4). ■

Proof of Theorem 6.6 (a): We run Algorithm 2. By Lemma 6.16, with probability at least $\frac{3}{4}$, it produces at least $t \ge (1-\epsilon)k$ subsets $B_1, \ldots, B_t$, each of which has μ-measure at least $\frac{1}{4}$.

Using Markov's inequality and Lemma 6.18, with probability at least $\frac{3}{4}$, we have $\max_{l\in[t]} \nu(B_l) \le 4\mathbb{E}[\max_{l\in[t]} \nu(B_l)] = O(Dk) \cdot \sqrt{\xi}$.

By union bound, with probability at least $\frac{1}{2}$, the algorithm produces at least $t \ge (1-\epsilon)k$ disjoint subsets B_l, each of which satisfies $\nu(B_l) = O(Dk) \cdot \sqrt{\xi}$ and $\mu(B_l) \ge \frac{1}{4}$.

Hence, Lemma 6.17 implies that each such B_l contains a non-empty subset $\hat{B}_l$ such that $\phi(\hat{B}_j) \le \frac{\nu(B_l)}{\mu(B_l)} = O(Dk) \cdot \sqrt{\xi}$, as required.

■

7 Vertex Expansion in 2-Graphs and Hardness

As mentioned in Section 3.4, vertex expansion in 2-graphs is closely related to hyperedge expansion. Indeed, Reduction 3.14 implies that vertex expansion in d-regular graphs can be reduced to hyperedge expansion. We show that this reduction also relates the parameter λ_∞ (see (3.2)) defined by Bobkov *et al.* [BHT00] with the parameter γ_2 associated with the Laplacian we define (in Section 4.1) for hypergraphs.

Theorem 7.1 (Restatement of Theorem 3.17) *Let $G = (V, E)$ be a undirected d-regular 2-graph with parameter λ_∞, and let $H = (V, E')$ be the hypergraph obtained in Reduction 3.14 having parameter γ_2. Then,*

$$\frac{\gamma_2}{4} \le \frac{\lambda_\infty}{d} \le \gamma_2 \,.$$

Proof: Using Theorem 4.1 for hypergraphs, the parameter γ_2 of H can be reformulated in terms of the weighted space as:

$$\gamma_2 = \min_{f \perp \mathbf{1}} \frac{\sum_{u \in V} \max_{i,j \in (\{u\} \cup N(u))} (f_i - f_j)^2}{d \sum_{u \in V} f_u^2} .$$

Therefore, it follows that $\frac{\lambda_\infty}{d} \leq \gamma_2$.

Next, using $(x+y)^2 \leq 4 \max \left\{ x^2, y^2 \right\}$ for any $x, y \in \mathbb{R}$, we get

$$\gamma_2 = \min_{f \perp \mathbf{1}} \frac{\sum_{u \in V} \max_{i,j \in (\{u\} \cup N(u))} (f_i - f_u + f_u - f_j)^2}{d \sum_{u \in V} f_u^2} \leq \min_{f \perp \mathbf{1}} \frac{\sum_{u \in V} 4 \max_{v \sim u} (f_v - f_u)^2}{d \sum_{u \in V} f_u^2} = \frac{4\lambda_\infty}{d} .$$

■

7.1 Hardness via the Small-Set Expansion Hypothesis

We state the Small-Set Expansion Hypothesis proposed by Raghavendra and Steurer [RS10].

Hypothesis 1 (Small-Set Expansion (SSE) Hypothesis) *For every constant $\eta > 0$, there exists sufficiently small $\delta > 0$ such that, given a graph G (with unit edge weights), it is NP-hard to distinguish the following two cases:*

YES: *there exists a vertex set S with $\delta \leq \frac{|S|}{n} \leq 10\delta$ and edge expansion $\phi(S) \leq \eta$,*

NO: *all vertex sets S with $\delta \leq \frac{|S|}{n} \leq 10\delta$ have expansion $\phi(S) \geq 1 - \eta$.*

Small-Set Expansion Hypothesis Apart from being a natural optimization problem, the small-set expansion problem is closely tied to the Unique Games Conjecture. Recent work by Raghavendra-Steurer [RS10] established the reduction from the small-set expansion problem to the well known Unique Games problem, thereby showing that Small-Set Expansion Hypothesis implies the Unique Games Conjecture. We refer the reader to [RST12] for a comprehensive discussion on the implications of Small-Set Expansion Hypothesis. We shall use the following hardness result for vertex expansion based on Small-Set Expansion Hypothesis.

Fact 7.2 ([LRV13]) *For every $\eta > 0$, there exists an absolute constant C_1 such that $\forall \varepsilon > 0$ it is SSE-hard to distinguish between the following two cases for a given graph $G = (V, E, w)$ with maximum degree $d \geq 100/\varepsilon$ and minimum degree $c_1 d$ (for some absolute constant c_1).*

YES *: There exists a set $S \subset V$ of size $|S| \leq |V|/2$ such that*

$$\phi^{\mathsf{V}}(S) \leq \varepsilon$$

NO *: For all sets $S \subset V$,*

$$\phi^{\mathsf{V}}(S) \geq \min \left\{ 10^{-10}, C_1 \sqrt{\varepsilon \log d} \right\} - \eta.$$

Reduction 3.14 implies that vertex expansion in 2-graphs is closely related to hyperedge expansion. Therefore, the hardness of vertex expansion as stated in Fact 7.2 should imply the hardness of hyperedge expansion. We formalize this intuition in the following theorems.

Theorem 7.3 (Formal statement of 3.18) *For every $\eta > 0$, there exists an absolute constant C such that for all $\widehat{\varepsilon} > 0$ it is SSE-hard to distinguish between the following two cases for a given hypergraph $H = (V, E, w)$ with maximum hyperedge size r such that $\widehat{\varepsilon} r \log r \in [\eta^2, c_2]$ (for some absolute constant c_2) and $r_{\min} \geq c_1 r$ (for some absolute constant c_1).*

YES : *There exists a set $S \subset V$ such that*
$$\phi_H(S) \leq \widehat{\varepsilon}$$

NO : *For all sets $S \subset V$,*
$$\phi_H(S) \geq C\sqrt{\widehat{\varepsilon} \cdot \frac{\log r}{r}}.$$

Proof: Given an undirected graph G with maximum degree d and minimum degree $\Omega(d)$ as in Fact 7.2, we apply Reduction 3.14 to obtain a hypergraph H with maximum edge cardinality $r = d + 1$. Then, Fact 3.15 implies that for any subset S of vertices, $c_i \cdot \phi_H(S) \leq \frac{\phi^{\mathsf{V}}_G(S)}{d+1} \leq \phi_H(S)$.

Fix some small enough $\eta > 0$ and corresponding $C_1 > 0$ as in Fact 7.2. Let $\varepsilon > \frac{100}{d+1} = \frac{100}{r}$.

Under the YES case of vertex expansion in Fact 7.2, there is some subset S such that $|S| \leq \frac{|V|}{2}$ and $\phi^{\mathsf{V}}_G(S) \leq \varepsilon$. This implies that $\phi_H(S) \leq \frac{\varepsilon}{c_1 r}$, and we denote $\widehat{\varepsilon} := \frac{\varepsilon}{c_1 r} > \frac{100}{c_1 r^2}$.

Under the NO case of vertex expansion in Fact 7.2, we have the fact that any $S \subset V$ has vertex expansion $\phi^{\mathsf{V}}_G(S) \geq \min\left\{10^{-10}, C_1\sqrt{\varepsilon \log d}\right\} - \eta$.

This implies that for some constant C' depending on C_1 and c_1,

$\phi_H(S) \geq \frac{\phi^{\mathsf{V}}_G(S)}{r} \geq \min\left\{\frac{10^{-10}}{r}, C'\sqrt{\widehat{\varepsilon} \cdot \frac{\log r}{r}}\right\} - \frac{\eta}{r}$.

Observe that this lower bound is non-trivial under the case

$\frac{10^{-10}}{r} \geq C'\sqrt{\widehat{\varepsilon} \cdot \frac{\log r}{r}} \geq 2 \cdot \frac{\eta}{r}$, which is equivalent to $\widehat{\varepsilon} r \log r \in [\eta^2, c_2]$, for some constant c_2 depending on C_1 and c_1. Hence, under this case, we have $\phi_H(S) \geq \frac{C'}{2} \cdot \sqrt{\widehat{\varepsilon} \cdot \frac{\log r}{r}}$.

Hence, the SSE-hardness in Fact 7.2 finishes the proof. ∎

Theorem 7.4 (Formal statement of 3.19) *For every $\eta > 0$, there exists an absolute constant C such that $\forall \overline{\varepsilon} > 0$ it is* SSE*-hard to distinguish between the following two cases for a given hypergraph $H = (V, E, w)$ with maximum hyperedge size r such that $\overline{\varepsilon} r \log r \in [\eta^2, c_2]$ (for some absolute constant c_2), $r_{\min} \geq c_1 r$ (for some absolute constant c_1) and $\gamma_2 \leq \frac{1}{r}$ where γ_2 is the parameter associated with H as in Theorem 6.1.*

YES : $\gamma_2 \leq \overline{\varepsilon}$.

NO : $\gamma_2 \geq C\overline{\varepsilon}\log r$.

Proof: We shall use the hardness result in Theroem 7.3, and the Cheeger inequality for hyeprgraphs in Theorem 6.1 and Proposition 6.2.

Given a hypergraph H, we have

$\frac{\gamma_2}{2} \leq \phi_H \leq \gamma_2 + 2\sqrt{\frac{\gamma_2}{r_{\min}}} \leq O(\sqrt{\frac{\gamma_2}{r}})$, where the last inequality follow because $r_{\min} = \Omega(r)$ and $\gamma_2 \leq \frac{1}{r}$.

Hence, the YES case in Theorem 7.3 implies that $\gamma_2 \leq 2\widehat{\varepsilon}$.

The NO case in Theorem 7.3 implies that $\gamma_2 = \Omega(\widehat{\varepsilon} \log r)$.

Therefore, the hardness result in Theorem 7.3 finishes the proof. ∎

8 Polynomial Time Approximation Algorithm for Procedural Minimizers

Observe the procedures in Section 6 take k orthonormal vectors $f_1, f_2, \ldots, f_k$ in the weighted space such that $\max_{i\in[k]} \mathsf{D}_w(f_i)$ is small. However, we do not know of an efficient algorithm to generate such k vectors to attain the minimum ξ_k. In this section, we consider an approximation algorithm to produce these vectors.

Theorem 8.1 (Restatement of Theorem 3.21) *There exists a randomized polynomial time algorithm that, given a hypergraph $H = (V, E, w)$ and a parameter $k < |V|$, outputs k orthonormal vectors $f_1, \ldots, f_k$ in the weighted space such that with high probability, for each $i \in [k]$,*

$$\mathsf{D}_w(f_i) \leq \mathcal{O}\left(i \log r \cdot \xi_i\right).$$

Observe that Theorem 8.1 gives a way to generate k orthonormal vectors in the weighted space such that the maximum discrepancy ratio $\mathsf{D}_w(\cdot)$ is at most $k \log r \cdot \xi_k$. Hence, these vectors can be used as inputs for the procedures in Theorem 6.1 (more precisely, we use an approximate f_2 in Proposition 6.1), Theorems 6.6 and 6.14 to give approximation algorithms as described in Corollaries 3.22, 3.23 and 3.24.

The approximate algorithm in Theorem 8.1 achieves the k vectors by starting with $f_1 \in \operatorname{span}(\vec{\mathbf{1}})$, and repeatedly using the algorithm in the following theorem to generate approximate procedural minimizers.

Theorem 8.2 (Restatement of Theorem 3.20) *Suppose for $k \geq 2$, $\{f_i\}_{i\in[k-1]}$ is a set of orthonormal vectors in the weighted space, and define $\gamma := \min\{\mathsf{D}_w(f) : \vec{\mathbf{0}} \neq f \perp_w \{f_i : i \in [k-1]\}\}$. Then, there is a randomized procedure that produces a non-zero vector f that is orthogonal to $\{f_i\}_{i\in[k-1]}$ in polynomial time, such that with high probability, $\mathsf{D}_w(f) = \mathcal{O}\left(\gamma \log r\right)$, where r is the size of the largest hyperedge.*

Proof of Theorem 8.1: On a high level, we start with $f_1 := \frac{\vec{\mathbf{1}}}{\|\vec{\mathbf{1}}\|_w}$. For $1 < i \leq k$, assuming that orthonormal vectors $\{f_l : l \in [i-1]\}$ are already constructed, we apply Theorem 8.2 to generate f_i. Hence, it suffices to show that $\mathsf{D}_w(f_i) \leq \mathcal{O}\left(i \log r \cdot \xi_i\right)$.

We prove that if $\xi := \min\{\mathsf{D}_w(f) : \vec{\mathbf{0}} \neq f \perp_w \{f_l : l \in [i-1]\}\}$, then $\xi \leq i \cdot \xi_i$. Hence, Theorem 8.2 implies that $\mathsf{D}_w(f_i) \leq \mathcal{O}\left(\xi \log r\right) \leq \mathcal{O}\left(i \log r \cdot \xi_i\right)$.

Therefore, it remains to show $\xi \leq i \cdot \xi_i$. Suppose $g_1, g_2, \ldots, g_i$ are orthonormal vectors in the weighted space that attain ζ_i (which is defined in Section 6.2).

Since $\operatorname{span}(\{g_1, g_2, \ldots, g_i\})$ has dimension i, there exists non-zero $g \in \operatorname{span}(\{g_1, g_2, \ldots, g_i\})$ such that $g \perp_w \{f_1, f_2, \ldots, f_{i-1}\}$. By the definition of ζ_i, we have $\mathsf{D}_w(g) \leq \zeta_i \leq i\xi_i$, where the last inequality follows from Claim 6.4. Hence, we have $\xi \leq i\xi_i$, as required. ■

We next give an SDP relaxation (8.3) and a rounding algorithm (Algorithm 3) to prove Theorem 8.2.

8.1 An SDP Relaxation to Approximate Procedural Minimizers: Proof of Theorem 8.2

We present SDP 8.3 to compute a vector in the weighted space that is orthogonal to $f_1, \ldots, f_{k-1}$ having the least discrepancy ratio $\mathsf{D}_w(\cdot)$. In the SDP, for each $u \in V$, the vector $\vec{\mathbf{g_u}}$ represents the u-th coordinate of the vector $f \in \mathbb{R}^V$ that we try to compute. The objective function of the SDP and equation (8.1) seek to minimize the discrepancy ratio $\mathsf{D}_w(\cdot)$. We shall see that equation (8.2) ensures that after rounding, the resulting vector f is orthogonal to $f_1, \ldots, f_{k-1}$ in the weighted space and achieves $\mathcal{O}\left(\log r\right)$-approximation with constant probability.

SDP 8.3

$$\mathsf{SDPval} := \min \sum_{e \in E} w_e \max_{u,v \in e} \left\| \vec{\mathbf{g_u}} - \vec{\mathbf{g_v}} \right\|^2$$

subject to

$$\sum_{v \in V} w_v \left\| \vec{\mathbf{g_v}} \right\|^2 = 1 \tag{8.1}$$

$$\sum_{v \in V} w_v f_i(v)\, \vec{\mathbf{g_v}} = \vec{\mathbf{0}} \qquad \forall i \in [k-1] \tag{8.2}$$

Algorithm 3 Rounding Algorithm for Computing Eigenvalues

1: Solve SDP 8.3 to generate vectors $\vec{\mathbf{g_v}} \in \mathbb{R}^n$ for $v \in V$.
2: Sample a random Gaussian vector $\vec{\mathbf{z}} \sim \mathcal{N}(0,1)^n$. For $v \in V$, set $f(v) := \langle \vec{\mathbf{g_v}}, \vec{\mathbf{z}} \rangle$.
3: Output f.

Lemma 8.4 (Feasibility) *With probability 1, Algorithm 3 outputs a non-zero vector f such that $f \perp_w \{f_1, f_2, \ldots, f_{k-1}\}$.*

Proof: Because of equation (8.1), there exists $v \in V$ such that $\vec{\mathbf{g_v}} \neq \vec{\mathbf{0}}$. Hence, when $\vec{\mathbf{z}}$ is sampled from $\mathcal{N}(0,1)^n$, the probability that $f(v) := \langle \vec{\mathbf{z}}, \vec{\mathbf{g_v}} \rangle$ is non-zero is 1.

For any $i \in [k-1]$, we use equation 8.2 to achieve:

$$\langle f, f_i \rangle_w = \sum_{v \in V} w_v \left\langle \vec{\mathbf{g_v}}, \vec{\mathbf{z}} \right\rangle f_i(v) = \left\langle \sum_{v \in V} w_v f_i(v) \vec{\mathbf{g_v}}, \vec{\mathbf{z}} \right\rangle = 0\,.$$

■

Lemma 8.5 (Approximation Ratio) *With probability at least $\frac{1}{24}$, Algorithm 3 outputs a vector f such that* $\mathsf{D}_w(f) \leq 384 \log r \cdot \mathsf{SDPval}$.

Proof: To give an upper bound on $\mathsf{D}_w(f_k)$, we prove an upper bound on the numerator and a lower bound on the denominator in the definition of $\mathsf{D}_w(\cdot)$.

For the numerator, we have

$$\begin{aligned} \mathbb{E}\left[\sum_{e \in E} w(e) \max_{u,v \in e} (f(u) - f(v))^2\right] &= \sum_{e \in E} w(e) \cdot \mathbb{E}\left[\max_{u,v \in e} (f(u) - f(v))^2\right] \\ &\leq 8 \log r \sum_{e \in E} w(e) \max_{u,v \in e} \left\| \vec{\mathbf{g_u}} - \vec{\mathbf{g_v}} \right\|^2 \quad \text{(Using Fact 8.6)} \\ &= 8 \log r \cdot \mathsf{SDPval}, \end{aligned}$$

where the inequality follows from Fact 8.6 in the following manner. For each $e \in E$, observe that the $\max_{u,v \in e}$ is over a set of cardinality $\binom{r}{2} \leq \frac{r^2}{2}$. Moreover for $u, v \in e$, $f(u) - f(v) = \langle \vec{\mathbf{g_u}} - \vec{\mathbf{g_v}}, \vec{\mathbf{z}} \rangle$ is a normal distribution with variance $\left\| \vec{\mathbf{g_u}} - \vec{\mathbf{g_v}} \right\|^2$ and mean 0. Hence, Fact 8.6 implies that $\mathbb{E}\left[\max_{u,v \in e} (f(u) - f(v))^2\right] \leq 8 \log r \cdot \max_{u,v \in e} \left\| \vec{\mathbf{g_u}} - \vec{\mathbf{g_v}} \right\|^2$.

Therefore, by Markov's Inequality,

$$\Pr\left[\sum_{e \in E} w(e) \max_{u,v \in e} (f(u) - f(v))^2 \leq 192 \log r \cdot \mathsf{SDPval}\right] \geq 1 - \frac{1}{24}\,. \tag{8.3}$$

For the denominator, using linearity of expectation, we get

$$\mathbb{E}\left[\sum_{v\in V} w_v f(v)^2\right] = \sum_{v\in V} w_v \mathbb{E}\left[\langle \vec{\mathbf{g}}_{\mathbf{v}}, \vec{\mathbf{z}}\rangle^2\right] = \sum_{v\in V} w_v \left\|\vec{\mathbf{g}}_{\mathbf{v}}\right\|^2 = 1 \qquad \text{(Using Equation 8.1).}$$

Now applying Fact 8.7 to the denominator we conclude

$$\Pr\left[\sum_{v\in V} w_v f(v)^2 \geq \frac{1}{2}\right] \geq \frac{1}{12}\,. \tag{8.4}$$

Using the union bound on Inequality (8.3) and Inequality (8.4) we get that

$$\Pr\left[\mathsf{D}_w(f) \leq 384\log r \;\cdot \mathsf{SDPval}\right] \geq \frac{1}{24}.$$

■

Fact 8.6 (Variant of Massart's Lemma) *Suppose $Y_1, Y_2, \ldots, Y_d$ are normal random variables that are not necessarily independent. For each $i \in [d]$, suppose $\mathbb{E}[Y_i] = 0$ and $\mathbb{E}\left[Y_i^2\right] = \sigma_i^2$. Denote $\sigma := \max_{i\in[d]} \sigma_i$. Then, we have*

1. $\mathbb{E}\left[\max_{i\in[d]} Y_i^2\right] \leq 4\sigma^2 \ln d$, *and*
2. $\mathbb{E}\left[\max_{i\in[d]} |Y_i|\right] \leq 2\sigma \cdot \sqrt{\ln d}$.

Proof: For $i \in [d]$, we write $Y_i = \sigma_i Z_i$, where Z_i has the standard normal distribution $\mathcal{N}(0,1)$. Observe that for any real numbers $x_1, x_2, \ldots, x_d$, for any positive integer p, we have $\max_{i\in[d]} x_i^2 \leq (\sum_{i\in[d]} x_i^{2p})^{\frac{1}{p}}$. Hence, we have

$$\begin{aligned}
\mathbb{E}\left[\max_{i\in[d]} Y_i^2\right] &\leq \mathbb{E}\left[\left(\sum_{i\in[d]} Y_i^{2p}\right)^{\frac{1}{p}}\right] \leq \left(\mathbb{E}\left[\sum_{i\in[d]} Y_i^{2p}\right]\right)^{\frac{1}{p}} \quad \text{(by Jensen's Inequality, because } t \mapsto t^{\frac{1}{p}} \text{ is concave)} \\
&\leq \sigma^2\left(\mathbb{E}\left[\sum_{i\in[d]} Z_i^{2p}\right]\right)^{\frac{1}{p}} = \sigma^2\left(\sum_{i\in[d]} \frac{(2p)!}{(p)!2^p}\right)^{\frac{1}{p}} \quad \text{(For } Z_i \sim \mathcal{N}(0,1), \mathbb{E}\left[Z_i^{2p}\right] = \frac{(2p)!}{(p)!2^p}\text{)} \\
&\leq \sigma^2 p d^{\frac{1}{p}}. \quad \text{(using } \frac{(2p)!}{p!} \leq (2p)^p \text{)}
\end{aligned}$$

Picking $p = \lceil \log d \rceil$ gives the first result $\mathbb{E}\left[\max_{i\in[d]} Y_i^2\right] \leq 4\sigma^2 \log d$. Moreover, the inequality $\mathbb{E}[|Y|] \leq \sqrt{\mathbb{E}[Y^2]}$ immediately gives the second result. ■

Fact 8.7 *Let $Y_1, \ldots, Y_m$ be normal random variables (that are not necessarily independent) having mean 0 such that $\mathbb{E}\left[\sum_i Y_i^2\right] = 1$ then*

$$\Pr\left[\sum_i Y_i^2 \geq \frac{1}{2}\right] \geq \frac{1}{12}\,.$$

Proof: We will bound the second moment of the random variable $R := \sum_i Y_i^2$ as follows.

$$
\begin{aligned}
\mathbb{E}\left[R^2\right] &= \sum_{i,j} \mathbb{E}\left[Y_i^2 Y_j^2\right] \\
&\le \sum_{i,j} \left(\mathbb{E}\left[Y_i^4\right]\right)^{\frac{1}{2}} \left(\mathbb{E}\left[Y_j^4\right]\right)^{\frac{1}{2}} && \text{(Using Cauchy-Schwarz Inequality)} \\
&= \sum_{i,j} 3\,\mathbb{E}\left[Y_i^2\right]\mathbb{E}\left[Y_j^2\right] && \text{(Using } \mathbb{E}\left[Z^4\right] = 3\left(\mathbb{E}\left[Z^2\right]\right)^2 \text{ for Gaussian } Z\text{)} \\
&= 3\left(\sum_i \mathbb{E}\left[Y_i^2\right]\right)^2 = 3.
\end{aligned}
$$

By the Paley-Zygmund inequality,

$$
\Pr\left[R \ge \frac{1}{2}\cdot\mathbb{E}\left[R\right]\right] \ge \left(\frac{1}{2}\right)^2 \cdot \frac{\mathbb{E}\left[R\right]^2}{\mathbb{E}\left[R^2\right]} \ge \frac{1}{12}\,.
$$

■

9 Sparsest Cut with General Demands

In this section, we study the Sparsest Cut with General Demands problem (defined in Section 3.6) and give an approximation algorithm for it (Theorem 3.25).

Theorem 9.1 (Restatement of Theorem 3.25) *There exists a randomized polynomial time algorithm that given an instance of the hypergraph Sparsest Cut problem with hypergraph $H = (V, E, w)$ and k demand pairs in $T = \{(\{s_i, t_i\}, D_i) : i \in [k]\}$, outputs a set $S \subset V$ such that with high probability,*

$$
\Phi(S) \le \mathcal{O}\left(\sqrt{\log k \log r}\,\log\log k\right)\Phi_H,
$$

where $r = \max_{e\in E}|e|$.

Proof: We prove this theorem by giving an SDP relaxation for this problem (SDP 9.2) and a rounding algorithm for it (Algorithm 4). We introduce a variable $\bar{u}$ for each vertex $u \in V$. Ideally, we would want all vectors $\bar{u}$ to be in the set $\{0,1\}$ so that we can identify the cut, in which case $\max_{u,v\in e}\|\bar{u}-\bar{v}\|^2$ will indicate whether the edge e is cut or not. Therefore, our objective function will be $\sum_{e\in E} w(e)\max_{u,v\in e}\|\bar{u}-\bar{v}\|^2$. Next, we add (9.1) as a scaling constraint. Finally, we add ℓ_2^2-triangle inequality constraints (9.2) between all triplets of vertices, as all integral solutions of the relaxation will trivially satisfy this. Therefore, SDP 9.2 is a relaxation of the problem and its objective value is at most Φ_H.

SDP 9.2

$$
\min \sum_{e\in E} w_e \max_{u,v\in e}\|\bar{u}-\bar{v}\|^2
$$

subject to

$$
\sum_{i\in[k]} D_i \cdot \|\bar{s}_i - \bar{t}_i\|^2 = 1 \tag{9.1}
$$

$$
\|\bar{u}-\bar{v}\|^2 + \|\bar{v}-\bar{w}\|^2 \ge \|\bar{u}-\bar{w}\|^2 \qquad \forall u,v,w \in V \tag{9.2}
$$

Our main ingredient is the following result due to [ALN08].

Fact 9.3 ([ALN08]) *Let (V,d) be an arbitrary metric space, and let $U \subset V$ be any k-point subset. If the space (V,d) is a metric of the negative type, then there exists a 1-Lipschitz map $f : V \to \ell_2$ such that the map $f|_U : U \to \ell_2$ has distortion $\mathcal{O}\left(\sqrt{\log k}\,\log\log k\right)$.*

Algorithm 4 Rounding Algorithm for Sparsest Cut

1: Solve SDP 9.2.
2: Compute the map $f : (V, \ell_2^2) \to \mathbb{R}^n$ using Fact 9.3, with U being the set of vertices that appear in the demand pairs in T.
3: Sample $\vec{\mathbf{z}} \sim \mathcal{N}(0,1)^n$ and define $x \in \mathbb{R}^V$ such that $x(v) := \langle \vec{\mathbf{z}}, f(v) \rangle$ for each $v \in V$.
4: Arrange the vertices of V as $v_1, \ldots, v_n$ such that $x(v_j) \leq x(v_{j+1})$ for each $1 \leq j \leq n-1$. Output the sparsest cut of the form

$$(\{v_1, \ldots, v_i\}, \{v_{i+1}, \ldots, v_n\}) .$$

Without loss of generality, we may assume that the map f is such that $f|_U$ has the least distortion (on vertices in demand pairs) among all 1-Lipschitz maps $f : (V, \ell_2^2) \to \ell_2$ ([ALN08] gives a polynomial time algorithm to compute such a map.) For the sake of brevity, let $\Lambda = \mathcal{O}\left(\sqrt{\log k} \log \log k\right)$ denote the distortion factor guaranteed in Fact 9.3. Since SDP 9.2 is a relaxation of Φ_H, we also get that objective value of the SDP is at most Φ_H. Suppose $x \in \mathbb{R}^V$ is the vector produced by the rounding algorithm.

We next analyze the following quantity. The numerator is related to the objective function, and the denominator is related to the expression in (9.1):

$$\varphi(x) := \frac{\sum_{e \in E} w_e \max_{u,v \in e} |x(u) - x(v)|}{\sum_{i \in [k]} D_i \cdot |x(s_i) - x(t_i)|}. \tag{9.3}$$

The following analysis is similar to the proof of Lemma 8.5.

For each edge e, obsever that for $u, v \in e$, $x_u - x_v$ is a random variable having normal distribution with mean 0 and variance $\|f(u) - f(v)\|^2$. Hence, using Fact 8.6 (2), we get

$$\mathbb{E}\left[\max_{u,v \in e} |x(u) - x(v)|\right] \leq 4\sqrt{\log r} \max_{u,v \in e} \|f(u) - f(v)\| \leq 4\sqrt{\log r} \max_{u,v \in e} \|\bar{u} - \bar{v}\|^2 ,$$

where the last inequality follows because $f : (V, \ell_2^2) \to \ell_2$ is 1-Lipschitz.

The expectation of the numerator of (9.3) is

$\mathbb{E}\left[\sum_{e \in E} w_e \max_{u,v \in e} |x(u) - x(v)|\right] \leq 4\sqrt{\log r} \cdot \Phi_H$.

Using Markov's inequality, we have

$$\Pr[\sum_{e \in E} w_e \max_{u,v \in e} |x(u) - x(v)| \leq 96\sqrt{\log r} \cdot \Phi_H] \geq 1 - \frac{1}{24} . \tag{9.4}$$

For the denominator, observing that $x(s_i) - x(t_i)$ has a normal distribution with mean 0 and variance $\|f(s_i) - f(t_i)\|^2$ and a random variable Z having distribution $\mathcal{N}(0,1)$ satisfies $\mathbb{E}[|Z|] = \sqrt{\frac{2}{\pi}}$, we have

$$\mathbb{E}\left[\sum_{i \in [k]} D_i \cdot |x(s_i) - x(t_i)|\right] = \sqrt{\frac{2}{\pi}} \sum_{i \in [k]} D_i \|f(s_i) - f(t_i)\| \geq \sqrt{\frac{2}{\pi}} \cdot \frac{1}{\Lambda} \sum_{i \in [k]} D_i \cdot \|\bar{s}_i - \bar{t}_i\|^2 = \sqrt{\frac{2}{\pi}} \cdot \frac{1}{\Lambda},$$

where the inequality follows from the distortion of $f|_U$ as guaranteed by Fact 9.3, and the last equality follows from (9.1).

We next prove a variant of Fact 8.7.

Claim 9.4 *Let $Y_1, \ldots, Y_m$ be normal random variables (that are not necessarily independent) having mean 0. Denote $R := \sum_i |Y_i|$. Then,*

$$\Pr\left[R \geq \frac{1}{2}\mathbb{E}\left[R\right]\right] \geq \frac{1}{12}\,.$$

Proof: For each i, let $\sigma_i^2 = \mathbb{E}\left[Y_i\right]$. Then, $\mathbb{E}\left[R\right] = \sqrt{\frac{2}{\pi}}\sum_i \sigma_i$.

Moreover, we have

$\mathbb{E}\left[R^2\right] = \sum_{i,j} \mathbb{E}\left[|Y_i| \cdot |Y_j|\right] \leq \sum_{i,j} \sqrt{\mathbb{E}\left[Y_i^2\right] \cdot \mathbb{E}\left[Y_j^2\right]} = \sum_{i,j} \sigma_i \sigma_j = \frac{\pi}{2} \cdot \mathbb{E}\left[R\right]^2,$

where the inequality follows from Cauchy-Schwarz.

Finally, using the Paley-Zygmund Inequality, we have

$$\Pr\left[R \geq \frac{1}{2} \cdot \mathbb{E}\left[R\right]\right] \geq \left(\frac{1}{2}\right)^2 \cdot \frac{\mathbb{E}\left[R\right]^2}{\mathbb{E}\left[R^2\right]} \geq \frac{1}{12}\,.$$

■

Hence, using Fact 8.7, we get

$$\Pr[\sum_{i \in [k]} D_i \left|x(s_i) - x(t_i)\right| \geq \sqrt{\frac{1}{2\pi}} \cdot \frac{1}{\Lambda}] \geq \frac{1}{12}. \tag{9.5}$$

Using (9.4) and (9.5), we get that with probability at least $\frac{1}{24}$,

$$\varphi(x) = \frac{\sum_e w_e \max_{u,v \in e} |x(u) - x(v)|}{\sum_{i \in [k]} D_i \cdot |x(s_i) - x(t_i)|} \leq \mathcal{O}\left(\sqrt{\log r}\right) \cdot \Lambda\, \Phi_H\,.$$

We next apply an analysis that is similar to Proposition 6.1. For $r \in \mathbb{R}$, define $S_r := \{v \in V : x(v) \leq r\}$. Observe that if r is sampled uniformly at random from the interval $[\min_v x(v), \max_v x(v)]$, then two vertices u and v are separated by the cut $(S_r, \overline{S_r})$ with probability proportional to $|x(u) - x(v)|$.

Hence, an averaging argument implies that there exists $r \in \mathbb{R}$ such that $\Phi(S_r) \leq \varphi(x) \leq \mathcal{O}\left(\sqrt{\log k \log r}\log\log k\right)\Phi_H$, as required in the output of Step 4. ■

Acknowledgments The dispersion process associated with our Laplacian operator was suggested to us by Prasad Raghavendra in the context of understanding vertex expansion in graphs, and was the starting point of this project. We would like to thank Ravi Kannan, Konstantin Makarychev, Yury Makarychev, Yuval Peres, Prasad Raghavendra, Nikhil Srivastava, Piyush Srivastava, Prasad Tetali, Santosh Vempala, David Wilson and Yi Wu for helpful discussions. We would also like to thank Jingcheng Liu and Alistair Sinclair for giving us valuable comments on an earlier version of this paper.

References

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

[AC88] Noga Alon and Fan RK Chung. Explicit construction of linear sized tolerant networks. *Annals of Discrete Mathematics*, 38:15–19, 1988.

[AK95] Charles J Alpert and Andrew B Kahng. Recent directions in netlist partitioning: a survey. *Integration, the VLSI journal*, 19(1):1–81, 1995.

[ALN08] Sanjeev Arora, James Lee, and Assaf Naor. Euclidean distortion and the sparsest cut. *Journal of the American Mathematical Society*, 21(1):1–21, 2008.

[Alo86] Noga Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986.

[AM85] Noga Alon and V. D. Milman. λ_1, isoperimetric inequalities for graphs, and superconcentrators. *J. Comb. Theory, Ser. B*, 38(1):73–88, 1985.

[ARV09] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings and graph partitioning. *Journal of the ACM (JACM)*, 56(2):5, 2009.

[BBC$^+$15] Oana Denisa Balalau, Francesco Bonchi, T.-H. Hubert Chan, Francesco Gullo, and Mauro Sozio. Finding subgraphs with maximum total density and limited overlap. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, WSDM 2015, Shanghai, China, February 2-6, 2015*, pages 379–388, 2015.

[BHT00] Sergey Bobkov, Christian Houdré, and Prasad Tetali. λ_∞ vertex isoperimetry and concentration. *Combinatorica*, 20(2):153–172, 2000.

[BS94] Stephen T Barnard and Horst D Simon. Fast multilevel implementation of recursive spectral bisection for partitioning unstructured problems. *Concurrency: Practice and Experience*, 6(2):101–117, 1994.

[BTL84] Sandeep N Bhatt and Frank Thomson Leighton. A framework for solving vlsi graph layout problems. *Journal of Computer and System Sciences*, 28(2):300–343, 1984.

[BV09] S Charles Brubaker and Santosh S Vempala. Random tensors and planted cliques. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 406–419. Springer, 2009.

[CA99] Umit V Catalyurek and Cevdet Aykanat. Hypergraph-partitioning-based decomposition for parallel sparse-matrix vector multiplication. *Parallel and Distributed Systems, IEEE Transactions on*, 10(7):673–693, 1999.

[CD12] Joshua Cooper and Aaron Dutle. Spectra of uniform hypergraphs. *Linear Algebra and its Applications*, 436(9):3268–3292, 2012.

[CDP10] L Elisa Celis, Nikhil R Devanur, and Yuval Peres. Local dynamics in bargaining networks via random-turn games. In *Internet and Network Economics*, pages 133–144. Springer, 2010.

[Cha00] Moses Charikar. Greedy approximation algorithms for finding dense components in a graph. In *Approximation Algorithms for Combinatorial Optimization, Third International Workshop, APPROX 2000, Saarbrücken, Germany, September 5-8, 2000, Proceedings*, pages 84–95, 2000.

[Chu93] F Chung. The laplacian of a hypergraph. *Expanding graphs (DIMACS series)*, pages 21–36, 1993.

[Chu97] Fan Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.

[CTZ15] T.-H. Hubert Chan, Zhihao Gavin Tang, and Chenzi Zhang. Spectral properties of laplacian and stochastic diffusion process for edge expansion in hypergraphs. *CoRR*, abs/1510.01520, 2015.

[DBH+06] Karen D Devine, Erik G Boman, Robert T Heaphy, Rob H Bisseling, and Umit V Catalyurek. Parallel hypergraph partitioning for scientific computing. In *Parallel and Distributed Processing Symposium, 2006. IPDPS 2006. 20th International*, pages 10–pp. IEEE, 2006.

[Din07] Irit Dinur. The pcp theorem by gap amplification. *Journal of the ACM (JACM)*, 54(3):12, 2007.

[EN14] Alina Ene and Huy L Nguyen. From graph to hypergraph multiway partition: Is the single threshold the only route? In *Algorithms-ESA 2014*, pages 382–393. Springer Berlin Heidelberg, 2014.

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

[FW95] Joel Friedman and Avi Wigderson. On the second eigenvalue of hypergraphs. *Combinatorica*, 15(1):43–65, 1995.

[GGLP00] Patrick Girard, L Guiller, C Landrault, and Serge Pravossoudovitch. Low power bist design by hypergraph partitioning: methodology and architectures. In *Test Conference, 2000. Proceedings. International*, pages 652–661. IEEE, 2000.

[HL95] Bruce Hendrickson and Robert Leland. An improved spectral graph partitioning algorithm for mapping parallel computations. *SIAM Journal on Scientific Computing*, 16(2):452–469, 1995.

[HL13] Christopher J Hillar and Lek-Heng Lim. Most tensor problems are np-hard. *Journal of the ACM (JACM)*, 60(6):45, 2013.

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006.

[HQ13] Shenglong Hu and Liqun Qi. The laplacian of a uniform hypergraph. *Journal of Combinatorial Optimization*, pages 1–36, 2013.

[HQ14] Shenglong Hu and Liqun Qi. The eigenvectors associated with the zero eigenvalues of the laplacian and signless laplacian tensors of a uniform hypergraph. *Discrete Applied Mathematics*, 169:140–151, 2014.

[KAKS99] George Karypis, Rajat Aggarwal, Vipin Kumar, and Shashi Shekhar. Multilevel hypergraph partitioning: applications in vlsi domain. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 7(1):69–79, 1999.

[KKL14] Tali Kaufman, David Kazhdan, and Alexander Lubotzky. Ramanujan complexes and bounded degree topological expanders. In *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on*, pages 484–493. IEEE, 2014.

[KLL+13] Tsz Chiu Kwok, Lap Chi Lau, Yin Tat Lee, Shayan Oveis Gharan, and Luca Trevisan. Improved cheeger's inequality: analysis of spectral partitioning algorithms through higher order spectral gap. In *STOC*, pages 11–20, 2013.

[Lei80] Charles E Leiserson. Area-efficient graph layouts. In *Foundations of Computer Science, 1980., 21st Annual Symposium on*, pages 270–281. IEEE, 1980.

[LM12] John Lenz and Dhruv Mubayi. Eigenvalues and quasirandom hypergraphs. *arXiv preprint arXiv:1208.4863*, 2012.

[LM13] John Lenz and Dhruv Mubayi. Eigenvalues of non-regular linear quasirandom hypergraphs. *arXiv preprint arXiv:1309.3584*, 2013.

[LM14a] Anand Louis and Konstantin Makarychev. Approximation algorithm for sparsest k-partitioning. In *SODA*, pages 1244–1255. SIAM, 2014.

[LM14b] Anand Louis and Yury Makarychev. Approximation Algorithms for Hypergraph Small Set Expansion and Small Set Vertex Expansion. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2014)*, volume 28, pages 339–355, 2014.

[LM15] John Lenz and Dhruv Mubayi. Eigenvalues and linear quasirandom hypergraphs. In *Forum of Mathematics, Sigma*, volume 3, page e2. Cambridge Univ Press, 2015.

[LOT12] James R Lee, Shayan OveisGharan, and Luca Trevisan. Multi-way spectral partitioning and higher-order cheeger inequalities. In *Proceedings of the 44th symposium on Theory of Computing*, pages 1117–1130. ACM, 2012.

[Lou15] Anand Louis. Hypergraph markov operators, eigenvalues and approximation algorithms. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 713–722, 2015.

[LRTV11] Anand Louis, Prasad Raghavendra, Prasad Tetali, and Santosh Vempala. Algorithmic extensions of cheegers inequality to higher eigenvalues and partitions. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 315–326. Springer, 2011.

[LRTV12] Anand Louis, Prasad Raghavendra, Prasad Tetali, and Santosh Vempala. Many sparse cuts via higher eigenvalues. In *Proceedings of the 44th symposium on Theory of Computing*, pages 1131–1140. ACM, 2012.

[LRV13] Anand Louis, Prasad Raghavendra, and Santosh Vempala. The complexity of approximating vertex expansion. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pages 360–369. IEEE, 2013.

[LT80] Richard J Lipton and Robert Endre Tarjan. Applications of a planar separator theorem. *SIAM journal on computing*, 9(3):615–627, 1980.

[Mer69] Robert C Merton. Lifetime portfolio selection under uncertainty: The continuous-time case. *The review of Economics and Statistics*, pages 247–257, 1969.

[Mer71] Robert C Merton. Optimum consumption and portfolio rules in a continuous-time model. *Journal of economic theory*, 3(4):373–413, 1971.

[MT06] Ravi R Montenegro and Prasad Tetali. *Mathematical aspects of mixing times in Markov chains*. Now Publishers Inc, 2006.

[Øks14] Bernt Øksendal. *Stochastic Differential Equations: An Introduction with Applications (Universitext)*. Springer, 6th edition, 2014.

[Par13] Ori Parzanchevski. Mixing in high-dimensional expanders. *arXiv preprint arXiv:1310.6477*, 2013.

[PR12] Ori Parzanchevski and Ron Rosenthal. Simplicial complexes: spectrum, homology and random walks. *arXiv preprint arXiv:1211.6775*, 2012.

[PRT12] Ori Parzanchevski, Ron Rosenthal, and Ran J Tessler. Isoperimetric inequalities in simplicial complexes. *arXiv preprint arXiv:1207.0638*, 2012.

[PSSW09] Yuval Peres, Oded Schramm, Scott Sheffield, and David Wilson. Tug-of-war and the infinity laplacian. *Journal of the American Mathematical Society*, 22(1):167–210, 2009.

[Rod09] J.A. Rodriguez. Laplacian eigenvalues and partition problems in hypergraphs. *Applied Mathematics Letters*, 22(6):916 – 921, 2009.

[RS10] Prasad Raghavendra and David Steurer. Graph expansion and the unique games conjecture. In *Proceedings of the 42nd ACM symposium on Theory of computing*, pages 755–764. ACM, 2010.

[RST12] Prasad Raghavendra, David Steurer, and Madhur Tulsiani. Reductions between expansion problems. In *Computational Complexity (CCC), 2012 IEEE 27th Annual Conference on*, pages 64–73. IEEE, 2012.

[SJ89] Alistair Sinclair and Mark Jerrum. Approximate counting, uniform generation and rapidly mixing markov chains. *Information and Computation*, 82(1):93–133, 1989.

[SKM14] John Steenbergen, Caroline Klivans, and Sayan Mukherjee. A cheeger-type inequality on simplicial complexes. *Advances in Applied Mathematics*, 56:56–77, 2014.

[SM00] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):888–905, 2000.

[SS96] Michael Sipser and Daniel A. Spielman. Expander codes. *IEEE Transactions on Information Theory*, 42:1710–1722, 1996.

A Hypergraph Tensor Forms

Let A be an r-tensor. For any suitable norm $\|\cdot\|_{\square}$, e.g. $\|.\|_2^2$, $\|.\|_r^r$, we define tensor eigenvalues as follows.

Definition A.1 *We define λ_1, the largest eigenvalue of a tensor A as follows.*

$$\lambda_1 := \max_{X \in \mathbb{R}^n} \frac{\sum_{i_1,i_2,\ldots,i_r} A_{i_1 i_2 \ldots i_r} X_{i_1} X_{i_2} \ldots X_{i_r}}{\|X\|_{\square}},$$

$$v_1 := \mathsf{argmax}_{X \in \mathbb{R}^n} \frac{\sum_{i_1,i_2,\ldots,i_r} A_{i_1 i_2 \ldots i_r} X_{i_1} X_{i_2} \ldots X_{i_r}}{\|X\|_{\square}}.$$

We inductively define successive eigenvalues $\lambda_2 \geq \lambda_3 \geq \ldots$ as follows.

$$\lambda_k := \max_{X \perp \{v_1,\ldots,v_{k-1}\}} \frac{\sum_{i_1,i_2,\ldots,i_r} A_{i_1 i_2 \ldots i_r} X_{i_1} X_{i_2} \ldots X_{i_r}}{\|X\|_{\square}},$$

$$v_k := \mathsf{argmax}_{x \perp \{v_1,\ldots,v_{k-1}\}} \frac{\sum_{i_1,i_2,\ldots,i_r} A_{i_1 i_2 \ldots i_r} X_{i_1} X_{i_2} \ldots X_{i_r}}{\|X\|_{\square}}.$$

Informally, the Cheeger's Inequality states that a graph has a sparse cut if and only if the gap between the two largest eigenvalues of the adjacency matrix is small; in particular, a graph is disconnected if and only if its top two eigenvalues are equal. In the case of the hypergraph tensors, we show that there exist hypergraphs having no gap between many top eigenvalues while still being connected. This shows that the tensor eigenvalues are not relatable to expansion in a Cheeger-like manner.

Proposition A.1 *For any $k \in \mathbb{N}$, there exist connected hypergraphs such that $\lambda_1 = \ldots = \lambda_k$.*

Proof: Let $r = 2^w$ for some $w \in \mathbb{Z}^+$. Let H_1 be a large enough complete r-uniform hypergraph. We construct H_2 from two copies of H_1, say A and B, as follows. Let $a \in E(A)$ and $b \in E(B)$ be any two hyperedges. Let $a_1 \subset a$ (resp. $b_1 \subset b$) be a set of any $r/2$ vertices. We are now ready to define H_2.

$$H_2 := (V(H_1) \cup V(H_2), (E(H_1) \setminus \{a\}) \cup (E(H_2) \setminus \{b\}) \cup \{(a_1 \cup b_1), (a_2 \cup b_2)\})$$

Similarly, one can recursively define H_i by joining two copies of H_{i-1} (this can be done as long as $r > 2^{2i}$). The construction of H_k can be viewed as a *hypercube of hypergraphs*.

Let A_H be the tensor form of hypergraph H. For H_2, it is easily verified that $v_1 = \mathbf{1}$. Let X be the vector which has $+1$ on the vertices corresponding to A and -1 on the vertices corresponding to B. By construction, for any hyperedge $\{i_1, \ldots, i_r\} \in E$

$$X_{i_1} \ldots X_{i_r} = 1$$

and therefore,

$$\frac{\sum_{i_1,i_2,\ldots,i_r} A_{i_1 i_2 \ldots i_r} X_{i_1} X_{i_2} \ldots X_{i_r}}{\|X\|_\Box} = \lambda_1 .$$

Since $\langle X, \mathbf{1} \rangle = 0$, we get $\lambda_2 = \lambda_1$ and $v_2 = X$. Similarly, one can show that $\lambda_1 = \ldots = \lambda_k$ for H_k. This is in sharp contrast to the fact that H_k is, by construction, a connected hypergraph. ■

B Examples

We give examples of hypergraphs to show that some properties are not satisfied. For convenience, we consider the properties in terms of the weighted space. We remark that the examples could also be formulated equivalently in the normalized space. In our examples, the procedural minimizers are discovered by trial-and-error using programs. However, we only describe how to use Mathematica to verify them. Our source code can be downloaded at the following link:

`http://i.cs.hku.hk/~algth/project/hyper_lap/main.html`

Verifying Procedural Minimizers. In our examples, we need to verify that we have the correct value for $\gamma_k := \min_{\vec{\mathbf{0}} \neq f \perp_w \{f_1, f_2, \ldots, f_{k-1}\}} \mathsf{D}_w(f)$, and a certain non-zero vector f_k attains the minimum.

We first check that f_k is perpendicular to $\{f_1, \ldots, f_{k-1}\}$ in the weighted space, and $\mathsf{D}_w(f_k)$ equals γ_k.

Then, it suffices to check that for all $\vec{\mathbf{0}} \neq f \perp_w \{f_1, f_2, \ldots, f_{k-1}\}$, $\mathsf{D}_w(f) \geq \gamma_k$. As the numerator in the definition of $\mathsf{D}_w(f)$ involves the maximum operator, we use a program to consider all cases of the relative order of the vertices with respect to f.

For each permutation $\sigma : [n] \to V$, for $e \in E$, we define $S_\sigma(e) := \sigma(\max\{i : \sigma(i) \in e\})$ and $I_\sigma(e) := \sigma(\min\{i : \sigma(i) \in e\})$.

We consider the mathematical program $P(\sigma) := \min \sum_{e \in E} w_e \cdot (f(S_\sigma(e)) - f(I_\sigma(e)))^2 - \gamma_k \cdot \sum_{u \in V} w_u f(u)^2$ subject to $f(\sigma(n)) \geq f(\sigma(n-1)) \geq \cdots f(\sigma(1))$ and $\forall i \in [k-1], \langle f_i, f \rangle = 0$. Since the objective function is a polynomial, and all constraints are linear, the Mathematica function Minimize can solve the program.

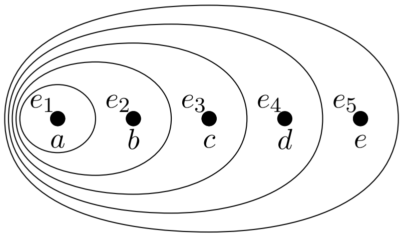
Moreover, the following two statements are equivalent.

1. For all $\vec{\mathbf{0}} \neq f \perp_w \{f_1, f_2, \ldots, f_{k-1}\}$, $\mathsf{D}_w(f) \geq \gamma_k$.
2. For all permutations σ, $P(\sigma) \geq 0$.

Hence, to verify the first statement, it suffices to use Mathematica to solve $P(\sigma)$ for all permutations σ.

Example B.1 *The sequence $\{\gamma_k\}$ generated by the procedural minimizers is not unique.*

Proof: Consider the following hypergraph with 5 vertices and 5 hyperedges each with unit weight.



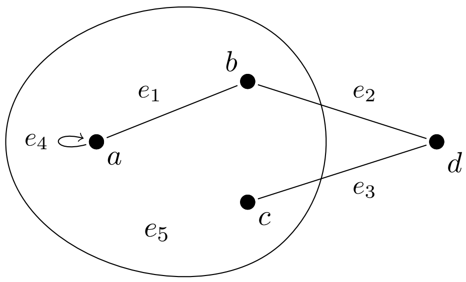
We have verified that different minimizers for γ_2 can lead to different values for γ_3.

i	γ_i	f_i^T	γ_i'	$f_i'^\mathsf{T}$
1	0	$(1,1,1,1,1)$	0	$(1,1,1,1,1)$
2	5/6	$(1,1,1,-4,-4)$	5/6	$(2,2,-3,-3,-3)$
3	113/99	$(2,2,-6,3,-6)$	181/165	$(4,-5,-5,5,5)$

■

Example B.2 *There exists a hypergraph such that $\xi_2 < \gamma_2$.*

Proof: Consider the following hypergraph $H = (V, E)$ with $V = \{a, b, c, d\}$ and $E = \{e_i : i \in [5]\}$. For $i \neq 3$, edge e_i has weight 1, and edge e_3 has weight 2. Observe that every vertex has weight 3.



We can verify that $\gamma_2 = \frac{2}{3}$ with the corresponding vector $f_2 := (1, 1, -1, -1)^\mathsf{T}$.

Recall that $\xi_2 = \min_{g_1,g_2} \max_{i\in[2]} \mathsf{D}_w(g_i)$, where the minimum is over all non-zero g_1 and g_2 such that $g_1 \perp_w g_2$. We can verify that $\xi_2 \leq \frac{1}{3}$ by considering the the two orthogonal vectors $g_1 = (0, 0, 1, 1)^\mathsf{T}$ and $g_2 = (1, 1, 0, 0)^\mathsf{T}$ in the weighted space. ■

Example B.3 (Issues with Distributing Hyperedge Weight Evenly) *Suppose $\overline{\mathsf{L}}_w$ is the operator on the weighted space that is derived from the Figure 3.1 by distributing the weight w_e evenly among $S_e(f)\times I_e(f)$. Then, there exists a hypergraph such that any minimizer f_2 attaining $\gamma_2 := \min_{\vec{\mathbf{0}}\neq f\perp_w \vec{\mathbf{1}}} \mathsf{D}_w(f)$ is not an eigenvector of $\overline{\mathsf{L}}_w$ or even $\Pi^w_{\vec{\mathbf{1}}^{\perp_w}} \overline{\mathsf{L}}_w$.*

Proof: We use the same hypergraph as in Example B.2. Recall that $\gamma_2 = \frac{2}{3}$ with the corresponding vector $f_2 := (1, 1, -1, -1)^\mathsf{T}$.

We next show that f_2 is the only minimizer, up to scalar multiplication, attaining γ_2.

According to the definition,

$$\gamma_2 = \min_{(a,b,c,d)\perp_w 1} \frac{(a-b)^2 + (b-d)^2 + 2(c-d)^2 + \max_{x,y\in e_5}(x-y)^2}{3(a^2+b^2+c^2+d^2)}.$$

Without loss of generality, we only need to consider the following three cases:

1. $a \geq b \geq c$: Then, by substituting $a = -b - c - d$,

$$\frac{(a-b)^2 + (b-d)^2 + 2(c-d)^2 + (a-c)^2}{3(a^2+b^2+c^2+d^2)} \geq \frac{2}{3}$$
$$\iff (c-d)^2 + 2(b+c)^2 \geq 0,$$

   and the equality is attained only when $a = b = -c = -d$.

2. $a \geq c \geq b$: Then, by substituting $d = -a - b - c$,

$$\frac{(a-b)^2 + (b-d)^2 + 2(c-d)^2 + (a-b)^2}{3(a^2+b^2+c^2+d^2)} \geq \frac{2}{3}$$
$$\iff (a+2b+c)^2 + 8c^2 + 4(a-c)(c-b) \geq 0,$$

   and the equality cannot be attained.

3. $b \geq a \geq c$: Then, by substituting $d = -a - b - c$,

$$\frac{(a-b)^2 + (b-d)^2 + 2(c-d)^2 + (b-c)^2}{3(a^2+b^2+c^2+d^2)} \geq \frac{2}{3}$$
$$\iff 4(b+c)^2 + 2(a+c)^2 + 2(b-a)(a-c) \geq 0,$$

   and the equality is attained only when $a = b = -c = -d$.

Therefore, all minimizers attaining γ_2 must be in $\mathrm{span}(f_2)$.

We next showt that f_2 is not an eigenvector of $\Pi^w_{\vec{1}^{\perp_w}} \overline{\mathsf{L}}_w$. Observe that only the hyperedge $e_5 = \{a, b, c\}$ involves more than 2 vertices. In this case, the weight of e_5 is distributed evenly between $\{a, c\}$ and $\{b, c\}$. All other edges keep their weights. Hence, the resulting weighted adjacency matrix $\overline{A}$ and $\mathsf{I} - \mathsf{W}^{-1}\overline{A}$ are as follows:

$$\overline{A} = \begin{pmatrix} \frac{3}{2} & 1 & \frac{1}{2} & 0 \\ 1 & \frac{1}{2} & \frac{1}{2} & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 & 2 \\ 0 & 1 & 2 & 0 \end{pmatrix} \text{ and } \mathsf{I} - \mathsf{W}^{-1}\overline{A} = \begin{pmatrix} \frac{1}{2} & -\frac{1}{3} & -\frac{1}{6} & 0 \\ -\frac{1}{3} & \frac{5}{6} & -\frac{1}{6} & -\frac{1}{3} \\ -\frac{1}{6} & -\frac{1}{6} & 1 & -\frac{2}{3} \\ 0 & -\frac{1}{3} & -\frac{2}{3} & 1 \end{pmatrix}.$$

Hence, $\overline{\mathsf{L}}_w f_2 = (\mathsf{I} - \mathsf{W}^{-1}\overline{A}) f_2 = (\frac{1}{3}, 1, -\frac{2}{3}, -\frac{2}{3})^\mathsf{T} \notin \mathrm{span}(f_2)$. Moreover, $\Pi^w_{\vec{1}^{\perp_w}} \overline{\mathsf{L}}_w f_2 = (\frac{1}{3}, 1, -\frac{2}{3}, -\frac{2}{3})^\mathsf{T} \notin \mathrm{span}(f_2)$.

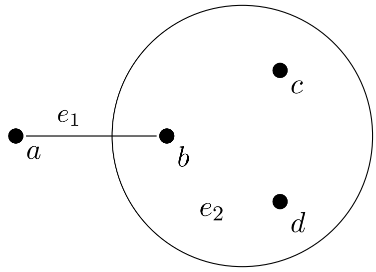
In comparison, in our approach, since b is already connected to d with edge e_2 of weight 1, it follows that the weight of e_5 should all go to the pair $\{a, c\}$. Hence, the resulting adjacency matrix is:

$$A = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 2 \\ 0 & 1 & 2 & 0 \end{pmatrix}.$$

One can verify that $\mathsf{L}_w f_2 = (\mathsf{I} - \mathsf{W}^{-1} A) f_2 = \frac{2}{3} f_2$, as claimed in Theorem 4.1. ■

Example B.4 (Third minimizer not eigenvector of Laplacian) *There exists a hypergraph such that for all procedural minimizers* $\{(f_i, \gamma_i)\}_{i \in [3]}$ *of* D_w*, the vector* f_3 *is not an eigenvector of* L_w *or even* $\Pi^w_{F_2^{\perp_w}} \mathsf{L}_w$*, where* L_w *is the operator on the weighted space defined in Lemma 4.8, and* $F_2 := \{f_1, f_2\}$.

Proof: Consider the following hypergraph with 4 vertices and 2 hyperedges each with unit weight.



We can verify the first 3 procedural minimizers.

i	γ_i	f_i^T
1	0	$(1, 1, 1, 1)$
2	$\frac{5-\sqrt{5}}{4}$	$(\sqrt{5}-1, \frac{3-\sqrt{5}}{2}, -1, -1)$
3	$\frac{11+\sqrt{5}}{8}$	$(\sqrt{5}-1, -1, 4-\sqrt{5}, -1)$
3′	$\frac{11+\sqrt{5}}{8}$	$(\sqrt{5}-1, -1, -1, 4-\sqrt{5})$

We next show that f_3 and $f_{3'}$ are the only minimizers, up to scalar multiplication, attaining γ_3. According to the definition,

$$\gamma_2 = \min_{(a,b,c,d)\perp_w 1} \frac{(a-b)^2 + \max_{x,y\in e_2}(x-y)^2}{a^2+2b^2+c^2+d^2}.$$

Observe that c and d are symmetric, we only need to consider the following two cases,

1. $c \geq b \geq d$: Then, by substituting $a = -2b - c - d$,

$$\frac{(a-b)^2+(c-d)^2}{a^2+2b^2+c^2+d^2} \geq 1$$
$$\iff 5b^2 + 2(c-b)(b-d) \geq 0.$$

2. $b \geq c \geq d$: Then, by substituting $a = -2b - c - d$,

$$\frac{(a-b)^2+(b-d)^2}{a^2+2b^2+c^2+d^2} \geq \frac{5-\sqrt{5}}{4}$$
$$\iff (5+3\sqrt{5})b^2 + (\sqrt{5}-3)c^2 + (\sqrt{5}-1)d^2 + (2\sqrt{5}+2)bc + (2\sqrt{5}-2)bd + (\sqrt{5}-1)cd \geq 0.$$

Let $f(b,c,d)$ denotes the function above. Since f is a quadratic function of c and the coefficient of c^2 is negative, the minimum must be achieved when $c = b$ or d. In other words,$f(b,c,d) \geq \min\{f(b,b,d), f(b,d,d)\}$. Note that

$$f(b,b,d) = (6\sqrt{5}+4)b^2 + (3\sqrt{5}-3)bd + (\sqrt{5}-1)d^2 \geq 0$$
$$\text{and } f(b,d,d) = (5+3\sqrt{5})b^2 + 4\sqrt{5}bd + (3\sqrt{5}-5)d^2 \geq 0.$$

and the equality holds only when $c = d = -\frac{3+\sqrt{5}}{2}b$.

Therefore, $\gamma_2 = \frac{5-\sqrt{5}}{4}, f_2^T = (\sqrt{5}-1, \frac{3-\sqrt{5}}{2}, -1, -1)$.

Now we are ready to calculate γ_3.

$$\gamma_3 = \min_{(a,b,c,d)\perp_w 1, f_2} \frac{(a-b)^2 + \max_{x,y\in e_2}(x-y)^2}{a^2+2b^2+c^2+d^2}.$$

Note that,

$$(a,b,c,d) \perp_w \vec{\mathbf{1}}, f_2 \iff \begin{cases} a+2d+c+d=0 \\ (\sqrt{5}-1)a+(3-\sqrt{5})b-c-d=0 \end{cases} \iff \begin{cases} a=(1-\sqrt{5})b \\ c+d=(\sqrt{5}-3)b \end{cases}$$

1. $c \geq b \geq d$: which is equivalent to $c \geq -\frac{\sqrt{5}+3}{4}(c+d) \geq d$, then

$$\frac{(a-b)^2+(c-d)^2}{a^2+2b^2+c^2+d^2} \geq \frac{11+\sqrt{5}}{8}$$
$$\iff (c - \frac{\sqrt{5}+3}{4}(c+d))(d - \frac{\sqrt{5}+3}{4}(c+d)) \leq 0.$$

2. $b \geq c \geq d$: which is equivalent to $(4-\sqrt{5})b+d \geq 0 \geq (3-\sqrt{5})b+2d$, then

$$\frac{(a-b)^2+(b-d)^2}{a^2+2b^2+c^2+d^2} \geq \frac{11+\sqrt{5}}{8}$$
$$\iff ((4-\sqrt{5})b+d)((3+\sqrt{5})((3-\sqrt{5})b+2d)-(\sqrt{5}-1)((4-\sqrt{5})b+d)) \leq 0.$$

Therefore, $\gamma_3 = \frac{11+\sqrt{5}}{8}$, and the corresponding $f_3^T = (\sqrt{5}-1,-1,4-\sqrt{5},-1)$ or $(\sqrt{5}-1,-1,-1,4-\sqrt{5})$.

We let $f = f_3 = (\sqrt{5}-1,-1,4-\sqrt{5},-1)^\mathsf{T}$, and we apply the procedure described in Lemma 4.8 to compute $\mathsf{L}_w f$.

Observe that $w_a = w_c = w_d = 1$ and $w_b = 2$, and $f(b) = f(d) < f(a) < f(c)$.

For edge e_1, $\Delta_1 = f(a)-f(b) = \sqrt{5}$ and $c_1 = w_1 \cdot \Delta_1 = \sqrt{5}$. For edge e_2, $\Delta_2 = f(c)-f(b) = 5-\sqrt{5}$, and $c_2 = w_2 \cdot \Delta_2 = 5-\sqrt{5}$. Hence, $r_a = -\frac{c_1}{w_a}$, $r_c = -\frac{c_2}{w_c}$, and $r_b = r_d = \frac{c_1+c_2}{w_b+w_d}$.

Therefore, $\mathsf{L}_w f = -r = (\sqrt{5}, -\frac{5}{3}, 5-\sqrt{5}, -\frac{5}{3})^\mathsf{T}$.

Moreover, $\Pi^w_{F_2^{\perp_w}} \mathsf{L}_w f = (-\frac{1}{2}+\frac{7}{6}\cdot\sqrt{5}, -\frac{4}{3}-\frac{1}{6}\cdot\sqrt{5}, \frac{59}{12}-\frac{11}{12}\cdot\sqrt{5}, -\frac{7}{4}+\frac{1}{12}\cdot\sqrt{5})^\mathsf{T} \notin \mathrm{span}(f)$.

The case when $f_3 = (\sqrt{5}-1,-1,-1,4-\sqrt{5})^\mathsf{T}$ is similar, with the roles of c and d reversed. ■