THE COMPLEXITY OF EXPANSION PROBLEMS

A Thesis
Presented to
The Academic Faculty

by

Anand Louis

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
Algorithms, Combinatorics, and Optimization

Georgia Institute of Technology
August 2014

THE COMPLEXITY OF EXPANSION PROBLEMS

Approved by:

Santosh Vempala, Advisor
School of Computer Science
Georgia Institute of Technology

Antonius Dieker
School of Industrial and Systems Engineering
Georgia Institute of Technology

Prasad Raghavendra
Computer Science Division
University of California, Berkeley

Dana Randall
School of Computer Science
Georgia Institute of Technology

Prasad Tetali
School of Mathematics
Georgia Institute of Technology

Date Approved: 12 June 2014

To my Parents

TABLE OF CONTENTS

DEDICATION . . . . **iii**

LIST OF TABLES . . . . **vii**

LIST OF FIGURES . . . . **viii**

SUMMARY . . . . **ix**

I INTRODUCTION **1**

1.1 Contributions of this Thesis 5

II PRELIMINARIES **10**

2.1 Definitions of Problems 12

2.2 Related Work 14

2.3 Cheeger's Inequality 17

2.4 The Small Set Expansion Hypothesis 19

2.5 Probabilistic Inequalities 21

2.6 Miscellaneous Inequalities 23

PART I SPECTRAL BOUNDS

III HIGHER ORDER CHEEGER INEQUALITIES FOR GRAPHS **26**

3.1 Introduction 26

3.2 Min Sum k-partition 32

3.3 k sparse-cuts 36

3.4 Lower bound for k Sparse-Cuts 46

3.5 Gap examples 48

3.6 Conclusion 50

IV SPECTRAL PROPERTIES OF HYPERGRAPHS **52**

4.1 Introduction 52

4.2 The Hypergraph Markov Operator 54

4.3 Overview of Proofs 66

4.4 Spectral Gap of Hypergraphs 70
4.5 The Hypergraph Dispersion Process 77
4.6 Higher Eigenvalues and Hypergraph Expansion 93
4.7 Reduction from Vertex Expansion in Graphs to Hypergraph Expansion 100
4.8 Hypergraph Tensor Forms 100
4.9 An Exponential Time Algorithm for computing Eigenvalues 102
4.10 Conclusion 103

PART II APPROXIMATION ALGORITHMS

V APPROXIMATION ALGORITHM FOR SPARSEST K-PARTITION 105

5.1 Introduction 105
5.2 Main Algorithm 108
5.3 From Disjoint Sets to Partitioning 120
5.4 Proof of Theorem 5.1.3 121
5.5 Orthogonal Separators with ℓ_2 Distortion 124
5.6 Integrality Gap for the Assignment SDP 125
5.7 Conclusion 126

VI APPROXIMATION ALGORITHMS FOR VERTEX EXPANSION AND HYPERGRAPH EXPANSION 128

6.1 Introduction 128
6.2 An *Optimal* Algorithm for Vertex Expansion 133
6.3 Approximation Algorithms for Hypergraph Eigenvalues 138
6.4 Algorithm for Hypergraph Balanced Separator 142
6.5 Conclusion 144

VII APPROXIMATION ALGORITHMS FOR SMALL SET EXPANSION PROBLEMS 146

7.1 Introduction 146
7.2 Algorithm for Hypergraph Small Set Expansion 152
7.3 Generating Hypergraph Orthogonal Separators 156

7.4 ℓ_2–ℓ_2^2 Hypergraph Orthogonal Separators 163

7.5 Algorithm for Hypergraph Small Set Expansion via ℓ_2–ℓ_2^2 Hypergraph Orthogonal Separators 168

7.6 SDP Intgrality Gap 171

7.7 Reduction from Vertex Expansion to Hypergraph Expansion 172

7.8 Conclusion 173

PART III COMPUTATIONAL LOWER BOUNDS

VIII HARDNESS OF VERTEX EXPANSION PARAMETERS 175

8.1 Proof Overview 178

8.2 Preliminaries 184

8.3 Reduction between Vertex Expansion and Symmetric Vertex Expansion 187

8.4 Isoperimetry of the Gaussian Graph 190

8.5 Dictatorship Testing Gadget 194

8.6 Hardness Reduction from SSE 203

8.7 Reduction from Analytic d-Vertex Expansion to Vertex Expansion . 214

8.8 Hardness of Vertex Expansion 219

8.9 Conclusion 221

IX CONCLUSION 223

9.1 Future Directions 224

REFERENCES 229

LIST OF TABLES

1 Higher Order Cheeger Inequalities 223
2 Approximation Algorithms 224
3 Computing Eigenvalues 224
4 Mixing Time Bounds 225

LIST OF FIGURES

1	Recursive Algorithm for MIN SUM K-PARTITION	33
2	The K SPARSE-CUTS Algorithm	38
3	Recursive algorithm can give many sets with very small expansion	48
4	k-partition can have sparsity much larger than $\Omega(\sqrt{\lambda_k}\,\mathsf{polylog} k)$	50
5	The hypergraph Marko Operator	55
6	Continuous Time Hypergraph Dispersion Process	60
7	Projected Continuous Time Hypergraph Dispersion Process	77
8	Rounding Algorithm for Hypergraph Small set Expansion	94
9	Rounding Algorithm for Many Sparse Cuts	104
10	Reduction from Vertex Expansion in graphs to Hypergraph Expansion	104
11	SDP Relaxation for Sparsest k-Partition	110
12	Algorithm for generating $k' \geqslant (1-\varepsilon)k$ disjoint sets P_i.	113
13	Assignment SDP	126
14	SDP Relaxation for λ_∞.	133
15	Rounding Algorithm	134
16	SDP Relaxation for for γ_k.	140
17	Rounding Algorithm for γ_k.	140
18	SDP Relaxation for for Hypergraph Balanced Separator.	143
19	Rounding Algorithm for Hypergraph Balanced Separator	144
20	SDP relaxation for H-SSE	153
21	Hypergraph Orthogonal Separator	160
22	Reduction from SSE to Vertex Expansion	179
23	Gadget for the reduction	194
24	Hardness Reduction	204

SUMMARY

Graph-partitioning problems are a central topic of research in the study of algorithms and complexity theory. They are of interest to theoreticians with connections to error correcting codes, sampling algorithms, metric embeddings, among others, and to practitioners, as algorithms for graph partitioning can be used as fundamental building blocks in many applications. One of the central problems studied in this field is the sparsest cut problem, where we want to compute the cut which has the least ratio of number of edges cut to size of smaller side of the cut. This ratio is known as the *expansion* of the cut. In spite of over 3 decades of intensive research, the approximability of this parameter remains an open question. The study of this optimization problem has lead to powerful techniques for both upper bounds and lower bounds for various other problems [68, 12, 11, 31], and interesting conjectures such as the SSE conjecture [86].

Cheeger's Inequality, a central inequality in Spectral Graph Theory, establishes a bound on expansion via the spectrum of the graph. This inequality and its many (minor) variants have played a major role in the design of algorithms as well as in understanding the limits of computation.

In this thesis we study three notions of expansion, namely *edge expansion* in graphs, *vertex expansion* in graphs and *hypergraph expansion*. We define suitable notions of spectra w.r.t. these notions of expansion. We show how the notion Cheeger's Inequality goes across these three problems. We study higher order variants of these notions of expansion (i.e. notions of expansion corresponding to partitioning the graph/hypergraph into more than two pieces, etc.) and relate them to higher eigenvalues of graphs/hypergraphs. We also study approximation algorithms for these

problems.

Unlike the case of graph eigenvalues, the eigenvalues corresponding to vertex expansion and hypergraph expansion are intractable. We give optimal approximation algorithms and computational lower bounds for computing them.

CHAPTER I

INTRODUCTION

Graph partitioning refers broadly to the task of partitioning the vertex set of a graph into two or more pieces. There are numerous ways to quantify the quality of a partition; most of them are functions of the sizes of the various pieces, the fraction of the edges that are cut by the partitioning and/or the number of vertices in the *boundary* of the partition. Graph-partitioning problems are a central topic of research in the study of algorithms and complexity theory. They are of interest to theoreticians with connections to error correcting codes [97], sampling algorithms [95], metric embeddings [68], among others, and to practitioners, as algorithms for graph partitioning can be used as fundamental building blocks in many applications such as image segmentation [94], clustering [37] , parallel computation [60] and VLSI placement and routing [4].

Some of the standard measures for quantifying the quality of a partition are k-median [59], k-cut [91], minimum diameter [15], expansion etc. Kannan, Vempala and Vetta [51] show that several of these measures fail to capture the natural clustering in simple examples, and argue that expansion is one of the best objective functions for measuring the quality of a cluster. Given an edge-weighted graph $G = (V, E, w)$, the *expansion* or *edge-expansion* or *conductance* of a subset $S \subset V$ of vertices, denoted by $\phi_G(S)$, is defined as the ratio the total weight of edges leaving it to the size of the set,

$$\phi_G(S) \stackrel{\text{def}}{=} \frac{w(S, \bar{S})}{\min\left\{w(S), w(\bar{S})\right\}}$$

where by $w(S)$ we denote the total weight of edges incident to vertices in S and $w(S,T)$ is the total weight of edges between vertex subsets S and T. The expansion of the graph G is defined as

$$\phi_G \stackrel{\text{def}}{=} \min_{S \subset V} \phi_G(S)\,.$$

Finding the optimal subset that minimizes expansion $\phi_G(S)$ is known as the SPARSEST CUT problem.

The expansion of a graph and the problem of approximating it have been highly influential in the study of algorithms and complexity, and have exhibited deep connections to many other areas of mathematics. In particular, motivated by its applications and the NP-hardness of the problem, the study of approximation algorithms for sparsest cut has been a very fruitful area of research, leading, in particular, to the theory of metric embeddings and more recently the Unique Games conjecture and the Small-set Expansion hypothesis.

Building on the work of Cheeger [29], Alon and Milman [3, 1] proved the discrete Cheeger Inequality, a central inequality in Spectral Graph Theory. This inequality establishes a bound on expansion via the spectrum of the graph:

$$\frac{\lambda_2}{2} \leqslant \phi_G \leqslant \sqrt{2\lambda_2}$$

where λ_2 is the second smallest eigenvalue of the normalized Laplacian[1] matrix of the graph. 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 [78, 96, 97, 43, 38, 13, 8]. We refer the reader to [46] for a comprehensive survey. The proof of Cheeger's inequality is algorithmic, using the eigenvector corresponding to λ_2 to find a set S satisfying $\phi(S) \leqslant \sqrt{2\lambda_2}$.

Some applications of graph partitioning require finding clusters in graphs/networks which have a small number of nodes in the boundary of the parts. This is captured by the *Vertex Expansion* of the graph, a notion of expansion similar to edge-expansion. The vertex expansion of a set of vertices in a graph is defined as the ratio of the number of vertices in the boundary of the set to the size of the set. As in the case of

[1] The normalized Laplacian matrix is defined as $\mathcal{L}_G \stackrel{\text{def}}{=} D^{-1/2}(D - A)D^{-1/2}$ where A is the adjacency matrix of the graph and D is the diagonal matrix whose $(i,i)^{th}$ entry is equal to the degree of vertex i.

edge expansion, the vertex expansion of the graph is defined as the minimum value of vertex expansion over all sets of size at most half of the size of the graph. Vertex Expansion has applications in image segmentation [94], parallel computation [60] and VLSI placement and routing [5], among others and is a major primitive for many graph algorithms, specifically for those that are based on the divide and conquer paradigm [64].

There is an abundant spectral and approximation theory for edge expansion problems, but surprisingly little is known about their vertex expansion counterparts. An approximation algorithm for vertex expansion implies one with the same approximation guarantee for edge expansion. However the converse is not known to be true, which indicates that vertex expansion might be harder than edge expansion.

As in the case of edge expansion, the vertex expansion of a graph is also NP-hard to compute. Therefore, one can only hope to compute an approximation in polynomial time. The problem of approximating edge or vertex expansion can be studied at various regimes of parameters of interest. Perhaps the simplest possible version of the problem is to distinguish whether a given graph is an expander. For an absolute constant δ_0, a graph is a δ_0-vertex (resp. edge) expander if its vertex (resp. edge) expansion is at least δ_0. The problem of recognizing a vertex (resp. edge) expander can be stated as follows: Given a graph G, distinguish between the following two cases (a) (Non-Expander) the expansion is $< \epsilon$, and (b) (Expander) the expansion is $> \delta_0$ for some absolute constant δ_0. Notice that if there is some sufficiently small absolute constant ϵ (depending on δ_0), for which the above problem is easy, then we could argue that it is easy to "recognize" a vertex expander. For the edge case, Cheeger's inequality yields an algorithm to recognize an edge expander. In fact, it is possible to distinguish a δ_0 edge expander graph from a graph whose edge expansion is $< \delta_0^2/2$, by just computing the second eigenvalue of the graph Laplacian. It is natural to ask if there is an efficient algorithm with an analogous guarantee for vertex expansion. More precisely, is there some sufficiently small ϵ (an arbitrary function of δ_0), so that one can efficiently distinguish between a graph with vertex expansion $> \delta_0$ from one with vertex expansion $< \epsilon$. Bobkov, Houdré and Tetali [21] proved a Cheeger like inequality for Vertex Expansion in graphs, relating a Poincairé-type functional graph parameter called λ_∞ to vertex expansion. Unlike the case of edge expansion, this inequality does not yield an algorithm to recognize vertex expanders, as the computation of λ_∞ appears to be intractable.

While studying graphs has been a fruitful approach in modeling many practically relevant problems, some problems require more general mathematical models. A hypergraph is a generalization of a graph in which an edge can connect any number of vertices. Formally, a hypergraph H is a pair $H = (V, E)$ where V is a set of elements called nodes or vertices, and $E \subseteq 2^V \setminus \{\emptyset\}$ is a set of non-empty subsets of V called hyperedges or edges. Hypergraph expansion can be defined in manner similar to edge expansion in graphs, it is defined as the least among all cuts in the hypergraph of the ratio of the number of the hyperedges cut to the size of the smaller side of the cut. Hypergraph partitioning problems are of immense practical importance, having applications in parallel and distributed computing [25], VLSI circuit design and computer architecture [52, 42], scientific computing [36] and other areas. Inspite of this, there hasn't been much theoretical work on them. There is a rich spectral theory of graphs, based on studying the eigenvalues and eigenvectors of the adjacency matrix (and other related matrices) of graphs [3, 1, 2, 8] (we refer the reader to [33] for a comprehensive survey on Spectral Graph Theory). However, it has remained open to define a spectral model of hypergraphs, whose spectra can be used to estimate hypergraph parameters à la Spectral Graph Theory. Spectral graph partitioning algorithms are widely used in practice for their efficiency and the high quality of solutions that they often provide [18, 44]. Besides being of natural theoretical interest, a spectral theory of hypergraphs might also be relevant for practical applications.

In this thesis, we study these three notions of expansion, namely edge expansion in graphs, vertex expansion in graphs and hypergraph expansion. We show how the notion of Laplacian eigenvalues and Cheeger's Inequality go across these three problems. We study higher order notions of these notions of expansion (i.e. expansion corresponding to partitioning the graph/hypergraph into more than two pieces, etc.) and relate them to higher eigenvalues of graphs/hypergraphs.Unlike the case of graph eigenvalues, the eigenvalues corresponding to vertex expansion and hypergraph expansion are intractable. We give optimal approximation algorithms and computational lower bounds for computing them (under a complexity theoretic assumption).

1.1 Contributions of this Thesis

1.1.1 Graph Partitioning and Higher Eigenvalues

The normalized Laplacian matrix of a graph G, denoted by $\mathcal{L}_G$ is defined as $\mathcal{L}_G \stackrel{\text{def}}{=} D^{-1/2}(D - A)D^{-1/2}$ where A is the adjacency matrix of the graph and D is the diagonal matrix whose $(i,i)^{th}$ entry is equal to the degree of vertex i. Let us denote the eigenvalues of $\mathcal{L}_G$ by $0 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_n$. A basic fact in spectral graph theory is that a graph is disconnected if and only if λ_2, the second smallest eigenvalue of its normalized Laplacian matrix, is zero. Cheeger's Inequality can be viewed as robust version of this fact; qualitatively, it says that a graph has a "sparse" cut if and only if λ_2 is "small". Similarly, it can be shown that the graph has k components if and only if λ_k is zero. A natural question to ask is if a robust version of this fact can be proved. We address this question in Chapter 3 in two ways. First, we show that a graph can be partitioned into k pieces such that the total fraction of edges cut is $\mathcal{O}\left(\sqrt{\lambda_k}\log k\right)$. This shows that if λ_k is "small", then the graph can be partitioned in the k pieces while cutting a "small" fraction of edges Next, our main result, is that there exists an absolute constant $c \in (0,1)$ such that a graph can be partitioned into ck pieces such that each piece has expansion $\mathcal{O}\left(\sqrt{\lambda_k \log k}\right)$. Complementing this, we show that any

k-partition of the vertex set of a graph will have at least one piece whose expansion is at least $\lambda_k/2$. This shows a graph can be partitioned into roughly k pieces each having "small" expansion if and only if λ_k is "small". The latter result is the best possible in terms of the eigenvalues up to constant factors.

The underlying problem of partitioning a graph in k pieces, say $S_1, \ldots, S_k$, while minimizing $\phi_G^k(\{S_1, \ldots, S_k\}) \stackrel{\text{def}}{=} \max_i \phi(S_i)$ seems to be a natural clustering problem in its own right, which can be used to model the existence of several well-formed clusters in a graph. Our upper and lower bounds imply a bi-criteria $\mathcal{O}\left(\sqrt{\mathsf{OPT} \log k}\right)$ approximation bound for ϕ_G^k. However, many practical applications require multiplicative approximation algorithms for graph expansion parameters. We present a $\mathcal{O}\left(\sqrt{\log n \log k}\right)$-approximation algorithm for computing ϕ_G^k in Chapter 5.

The spectral bound on ϕ_G^k implies that for any k, there is a subset S whose size is at most a $\mathcal{O}(1/k)$ fraction of the graph and $\phi(S) = \mathcal{O}\left(\sqrt{\lambda_k \log k}\right)$. This gives a bound for the small-set expansion problem which, for a parameter k, asks to compute a set of vertices of size $1/k$ fraction of the graph and having the least expansion. This problem was posed by Raghavendra and Steurer [86], and was shown to be intimately connected to the Unique Games problem. We describe the significance of this problem in greater detail in Section 2.4.

1.1.2 Vertex Expansion in Graphs

Bobkov, Houdré and Tetali [21] proved a Cheeger like inequality for vertex expansion in graphs, relating a Poincairé-type functional graph parameter called λ_∞ to vertex expansion (we formally define λ_∞ in Chapter 6). λ_∞ appears to be hard to compute exactly. We study the computational aspects of λ_∞ and of vertex expansion in graphs. In Chapter 6 we give a natural SDP relaxation for λ_∞; using a simple random projection based rounding algorithm, we get a $\mathcal{O}(\log d)$ approximation to λ_∞, where d is the largest vertex degree of the graph. We use this to construct an algorithm to approximate vertex expansion to within $\mathcal{O}\left(\sqrt{\phi^{\mathsf{V}} \log d}\right)$. This improves the $\mathcal{O}\left(\sqrt{d \cdot \phi^{\mathsf{V}}}\right)$ approximation bound of Alon [1].

It is natural to ask if the approximation bound of $\mathcal{O}\left(\sqrt{\phi^{\mathsf{V}} \log d}\right)$ for vertex expansion is best that can be obtained in polynomial time, or in other words, is there a matching lower bound of $\Omega\left(\sqrt{\phi^{\mathsf{V}} \log d}\right)$ for the computation of vertex expansion? Most known computational lower bounds for vertex expansion are those that follow from the computational lower bounds for edge expansion. Since Cheeger's inequality yields a $\mathcal{O}\left(\sqrt{\mathsf{OPT}}\right)$ approximation bound for edge expansion, any computational lower bound for edge expansion can not be used to obtain an optimal computational lower bound for vertex expansion. In this thesis, we show a reduction from SSE to the problem of distinguishing between the case when vertex expansion of the graph is at most ε and the case when the vertex expansion is at least $\Omega(\sqrt{\varepsilon \log d})$. We give the formal definition of SSE in Chapter 2. This immediately implies that it is SSE-hard to find a subset of vertex expansion less than $C\sqrt{\phi^{\mathsf{V}} \log d}$ for some constant C, thereby implying that our approximation bound for vertex expansion is optimal (upto constant factors). Moreover this implies for all constant $\varepsilon > 0$, it is SSE-hard to distinguish whether the vertex expansion $< \varepsilon$ or at least an absolute constant. (The analogous threshold for edge expansion is $\sqrt{\phi}$ with no dependence on the degree). Thus our results suggest that vertex expansion is harder to approximate than edge expansion. In particular, while Cheegers Inequality can certify constant edge expansion, it is SSE-hard to certify constant vertex expansion in graphs.

In Chapter 4, we give a factor-preserving reduction from vertex expansion in graphs to hypergraph expansion. We show that λ_∞ as defined by Bobkov et. al. coincides with the second smallest eigenvalue of a certain Markov operator on the resulting hypergraph.

1.1.3 Hypergraph Expansion

Unlike graphs, hypergraphs do not have any canonical matrix structure that can be studied. The canonical tensor forms of hypergraphs have been studied, but without much success (see Section 4.1.1 for a brief survey); we show that the spectral properties of such tensors are unrelated to the expansion properties of hypergraphs. We also show that there can be no linear operator for hypergraphs whose spectra captures hypergraph expansion in a Cheeger-like manner. Our main contribution is the definition of a new hypergraph Markov operator $M : \mathbb{R}^n \to \mathbb{R}^n$ (generalizing the adjacency matrix of graphs). We describe this operator in Chapter 4. The corresponding Laplacian operator is defined as $L \stackrel{\text{def}}{=} I - M$ where I is the identity operator. As in the case of graphs, the smallest eigenvalue of this Laplacian operator is zero and the second smallest eigenvalue is zero if and only if the hypergraph is disconnected. We show that eigenvalues of this Laplacian operator can be used to bound many combinatorial properties of graphs. In particular, we prove a *Cheeger*-like inequality for hypergraphs, relating the second smallest eigenvalue of this operator to the expansion of the hypergraph. We bound other hypergraph expansion parameters, like small set expansion, ϕ^k, etc, via higher eigenvalues of this operator. We give bounds on the diameter of the hypergraph as a function of the second smallest eigenvalue of the Laplacian operator. We also prove a hypergraph *Expander Mixing Lemma* showing that hypergraph expanders behave like random hypergraphs.

Any Markov operator defines a canonical Markov process as follows. Starting with an initial distribution on the vertices $\mu^0 : V \to \mathbb{R}^+$, we can recursively define $\mu^{t+1} \stackrel{\text{def}}{=} M(\mu^t)$. In this case, the Markov process can be viewed as a dispersion process on the vertices of the hypergraph, and can be used to model rumour spreading in networks, Brownian motion, etc., and might be of independent interest. We bound the *Mixing-time* of this process as a function of the second smallest eigenvalue of the Laplacian operator. All these results are generalizations of the corresponding results for graphs.

Our Laplacian operator is non-linear and thus computing its eigenvalues exactly appears to be intractable. For any $k \in \mathbb{Z}_{\geqslant 0}$, we give a polynomial time approximation algorithm to compute an approximation to the k^{th} smallest eigenvalue of the operator. We show that this approximation factor is optimal under the SSE hypothesis for constant values of k.

We give approximation algorithms for hypergraph expansion and hypergraph small-set expansion problems in Chapter 7.

CHAPTER II

PRELIMINARIES

We will denote graphs by $G = (V, E, w)$ where V is the set vertices, $E \subseteq V^2$ is the set of edges and $w : E \to \mathbb{R}^+$ gives the weights on the edges. We will denote hypergraphs by $H = (V, E, w)$, where V is the set vertices, $E \subseteq 2^V \setminus \{\emptyset\}$ is the set of hyperedges (we will often refer to hyperedges as just edges) and $w : E \to \mathbb{R}^+$ gives the weights on the edges. For graphs and hypergraphs, we will use $n \stackrel{\text{def}}{=} |V|$ to denote the number of vertices, $m \stackrel{\text{def}}{=} |E|$ to denote the number of edges and $r \stackrel{\text{def}}{=} \max_{e \in E} |e|$ to denote the size of the largest hyperedge. The (weighted) degree of a vertex $v \in V$ is defined as $d_v \stackrel{\text{def}}{=} \sum_{e \in E : v \in e} w(e)$. The degrees of the vertices define a canonical probability distribution on the vertices. We use $\mu^* : V \to [0, 1]$ to denote this probability distribution, i.e.

$$\mu^*(i) \stackrel{\text{def}}{=} \frac{d_i}{\sum_{i \in V} d_i} .$$

We say that a graph/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. We use D to denote the $n \times n$ diagonal matrix whose $(i, i)^{th}$ entry is d_i.

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 graph/hypergraph is *connected* if for each pair of vertices $u, v \in V$, there exists a path connecting them. The *diameter* of a graph/hypergraph, denoted by $\mathsf{diam}(H)$, is the smallest value $l \in \mathbb{Z}_{\geqslant 0}$, such that each pair of vertices $u, v \in V$ have a path of length at most l connecting them.

Matrices related to Graphs. For a graph G, we denote its weighted adjacency matrix by A_G, i.e. A_G is the $n \times n$ matrix whose rows and columns are indexed by V such that

$$A(i,j) \stackrel{\text{def}}{=} \begin{cases} w\left(\{i,j\}\right) & \text{if } \{i,j\} \in E \\ 0 & \text{otherwise} \end{cases}.$$

The Laplacian matrix L of G is defined as

$$L \stackrel{\text{def}}{=} D - A\,.$$

The normalized Laplacian matrix $\mathcal{L}$ of a graph G is defined as

$$\mathcal{L} \stackrel{\text{def}}{=} D^{-\frac{1}{2}}\left(D - A\right)D^{-\frac{1}{2}}\,.$$

It is easy to see that both L and $\mathcal{L}$ as positive semidefinite.

Fact 2.0.1.

$$\mathcal{L} \succeq 0\,.$$

Proof. Fix any $X \in \mathbb{R}^n$. Let Y denote $Y = D^{-\frac{1}{2}}X$. Then we have

$$X^T \mathcal{L}\, X = Y^T L\, Y = \sum_{i \in V} d_i Y_i^2 - 2\sum_{i \sim j} w\left(\{i,j\}\right) Y_i Y_j = \sum_{i \sim j} w\left(\{i,j\}\right)\left(Y_i - Y_j\right)^2 \geqslant 0\,.$$

□

Eigenvalues of $\mathcal{L}$. Since $L \succeq 0$, all its eigenvalues are non-negative. An easy fact to show is that the smallest eigenvalue of L is 0. This can be seen as follows. Let $\mathbf{1} \in \mathbb{R}^n$ denote the vector which has 1 in every coordinate. Then

$$L\,\mathbf{1} = 0\,.$$

Similarly, the smallest eigenvalue of $\mathcal{L}$ is also 0 as evidenced by the vector $D^{\frac{1}{2}}\mathbf{1}$. We will denote the eigenvalues of $\mathcal{L}$ by $0 \leqslant \lambda_2 \leqslant \lambda_3 \leqslant \ldots \leqslant \lambda_n$.

Remark 2.0.2. A folklore result in linear algebra is that the matrices $D^{-1}A$ and $D^{-1/2}AD^{-1/2}$ have the same set of eigenvalues. This can be seen as follows; let v be an eigenvector of $D^{-1}A$ with eigenvalue λ, then for the vector $\left(D^{1/2}v\right)$

$$D^{-1/2}AD^{-1/2}\left(D^{1/2}v\right) = \left(D^{1/2}\right)\cdot\left(D^{-1}A\right)v = \left(D^{1/2}\right)\cdot(\lambda\, v) = \lambda\left(D^{1/2}v\right) .$$

Hence, $D^{1/2}v$ will be an eigenvector of $D^{-1/2}AD^{-1/2}$ having the same eigenvalue λ.

2.1 Definitions of Problems

Since we will be studying many notions of expansion, to avoid ambiguity, we will refer to the usual notion of expansion as Edge Expansion. We define it again formally.

Definition 2.1.1 (Edge Expansion in Graphs). Given a graph $G = (V, E, w)$, we define the expansion of a set $S \subset V$ as follows.

$$\phi_G(S) \stackrel{\text{def}}{=} \frac{w(S,\bar{S})}{\min\left\{w(S), w(\bar{S})\right\}}$$

where $w(S,T)$ is the total weight of edges between vertex subsets S and T and $w(S)$ denotes the total weight of edges incident to vertices in S. We will also denote the latter quantity of $\mathsf{vol}(S)$. We will drop the subscript G whenever the graph is clear from the context.

The expansion of the graph G is defined as

$$\phi_G \stackrel{\text{def}}{=} \min_{S\subset V} \phi_G(S) .$$

The problem of computing ϕ_G is also referred to as the Sparsest Cut problem.

Definition 2.1.2 (Vertex Expansion). Given a graph $G = (V, E)$, the vertex boundary of a set $S \subseteq V$ of vertices is defined as

$$N(S) \stackrel{\text{def}}{=} \left\{v \in \bar{S} \;\middle|\; \exists\, u \in S \text{ such that } \{u, v\} \in E\right\} .$$

The vertex expansion of S, denoted by $\phi^{\mathsf{V}}(S)$, is defined as the ratio of the size of the vertex boundary of S to the size of S

$$\phi^{\mathsf{V}}_G(S) \stackrel{\text{def}}{=} |V| \cdot \frac{|N(S)|}{|S|\,|\bar{S}|}\,.$$

We will drop the subscript G whenever the graph is clear from the context. The vertex expansion of the graph G is defined as the least value of $\phi^{\mathsf{V}}(S)$ over all sets S

$$\phi^{\mathsf{V}}_G \stackrel{\text{def}}{=} \min_{S \subset V} \phi^{\mathsf{V}}(S)\,.$$

For our proofs, the notion of Symmetric Vertex Expansion is useful.

Definition 2.1.3 (Symmetric Vertex Expansion). Given a graph $G = (V, E)$, we define the the symmetric vertex expansion of a set $S \subset V$ as follows.

$$\Phi^{\mathsf{V}}_G(S) \stackrel{\text{def}}{=} |V| \cdot \frac{\left|N_G(S) \cup N_G(\bar{S})\right|}{|S|\,|\bar{S}|}\,.$$

Definition 2.1.4 (Balanced Vertex Expansion). Given a graph G and balance parameter b, we define the *b-balanced vertex expansion* of G as follows.

$$\phi^{\mathsf{V,bal}}_b \stackrel{\text{def}}{=} \min_{S : |S|\left|\bar{S}\right| \geqslant bn^2} \phi^{\mathsf{V}}(S).$$

and

$$\Phi^{\mathsf{V,bal}}_b \stackrel{\text{def}}{=} \min_{S : |S|\left|\bar{S}\right| \geqslant bn^2} \Phi^{\mathsf{V}}(S).$$

We define $\phi^{\mathsf{V,bal}} \stackrel{\text{def}}{=} \phi^{\mathsf{V,bal}}_{1/100}$ and $\Phi^{\mathsf{V,bal}} \stackrel{\text{def}}{=} \Phi^{\mathsf{V,bal}}_{1/100}$.

Definition 2.1.5 (Hypergraph Expansion). Given a hypergraph $H = (V, E, w)$, and a set $S \subset V$, we denote by $E(S, T)$, the edges which have at least one end point in S, and at least one end point in T, i.e.

$$E(S,T) \stackrel{\text{def}}{=} \{e \in E : e \cap S \neq \emptyset \text{ and } e \cap T \neq \emptyset\}\,.$$

We define the expansion of S as

$$\phi_H(S) \stackrel{\text{def}}{=} \frac{\sum_{e \in E(S,\bar{S})} w(e)}{\min\left\{w(S), w(\bar{S})\right\}}$$

where $w(S) = \sum_{i \in S} d_i$ as before. We will drop the subscript H whenever the hypergraph is clear from the context. We define the expansion of the hypergraph H as

$$\phi_H \stackrel{\text{def}}{=} \min_{S \subset V} \phi(S)\,.$$

Problem 2.1.6 (Hypergraph Balanced Separator)**.** Given a hypergraph $H = (V, E, w)$, and a *balance* parameter $c \in (0, 1/2]$, a set $S \subset V$ is said to be c-balanced if $cn \leqslant |S| \leqslant (1-c)n$. The c-Hypergraph Balanced Separator problem asks to compute the c-balanced set $S \subset V$ which has the least *sparsity* $\mathsf{sp}(S)$ defined as follows.

$$\mathsf{sp}(S) \stackrel{\text{def}}{=} n \cdot \frac{w\left(E(S, \bar{S})\right)}{|S|\left|\bar{S}\right|}\,.$$

Small Set Expansion. We will be studying the "small set" versions of Edge Expansion, Vertex Expansion, and Hypergraph Expansion. We define this as follows.

Problem 2.1.7 (Small Set Expansion)**.** Given a graph/hypergraph (V, E, w), and a parameter $\delta \in (0, 1/2]$, its Small Set Expansion is defined as

$$\alpha_\delta \stackrel{\text{def}}{=} \min_{S : \mu^*(S) \leqslant \delta} \alpha(S)$$

where $\alpha(\cdot)$ denotes $\phi(\cdot)$ in the case of edge expansion in graphs, $\phi^{\mathsf{V}}(\cdot)$ in the case of vertex expansion in graphs and $\phi(\cdot)$ in the case of hypergraph expansion.

2.2 Related Work

Approximation Algorithms. The approximability of the Sparsest Cut problem has been studied extensively in the literature. The algorithmic proof of Cheeger's Inequality yields a $\mathcal{O}\left(\sqrt{\mathsf{OPT}}\right)$ algorithmic bound for the Sparsest Cut problem. The first multiplicative approximation algorithm for this problem was due to Leighton and Rao [64], who gave a $\mathcal{O}(\log n)$-approximation algorithm. In a seminal work, Arora, Rao and Vazirani [13] gave a $\mathcal{O}\left(\sqrt{\log n}\right)$-approximation algorithm for this problem. This is currently the best known approximation guarantee for the Sparsest Cut problem.

General Sparsest Cut. A more general version of the Sparsest Cut problem, referred to as the General Sparsest Cut problem, has also been extensively explored in the literature. The problem is defined as follows. Let $G = (V, E, w)$ be a graph. Assume that we are given k pairs of vertices $\{s_1, t_1\}, \ldots, \{s_k, t_k\}$ and corresponding *demands* $D_1, \ldots, D_k \geqslant 1$. The sparsity of a cut $(S, \bar{S})$ for a set $S \subset V$ is defined as

$$\Phi(S) \stackrel{\text{def}}{=} \frac{w(E(S,\bar{S}))}{\sum_{i=1}^{k} D_i \left|\mathbb{I}_S[s_i] - \mathbb{I}_S[t_i]\right|}$$

where $\mathbb{I}_S[\cdot]$ is the indicator function of S. The sparsest cut is the cut having least sparsity among all cuts

$$\Phi_G \stackrel{\text{def}}{=} \min_{S \subset V} \Phi(S) \,.$$

The special case when all pairs of vertices are demand pairs, is closely related to graph expansion. Auman and Rabani [14], and Linial, London and Rabinovich [68] showed that the General Sparsest Cut problem (and the Sparsest Cut problem) is closely related to metric embeddings, and used this connection to obtain a $\mathcal{O}(\log k)$-approximation algorithm for this problem. Subsequently, Chawla, Gupta and Racke [28] gave a $\mathcal{O}\left(\log^{\frac{3}{4}} k\right)$-approximation algorithm, and Arora, Lee and Naor [11] gave the currently best known approximation guarantee of $\mathcal{O}\left(\sqrt{\log k} \log\log k\right)$.

Computational Lowerbounds. The Sparsest Cut problem was shown to be NP-hard by Matula and Shahrokhi [79]. Ambhul, Mastrolilli and Svensson [6], building on the work of Khot [55], showed that this problem can not have a PTAS assuming the Exponential Time Hypothesis (ETH). Raghavendra, Steurer and Tulsiani [88] showed a lower bound of $\Omega\left(\sqrt{\mathsf{OPT}}\right)$ for this problem assuming the Small-set Expansion

Hypothesis (see Section 2.4 for a formal description of this hypothesis). Khot and Vishnoi [56] showed that the General Sparsest Cut problem is can not be approximated to within any constant factor assuming the Unique Games conjecture.

Integrality Gaps. The study of linear and semi-definite relaxations of problems has been a fruitful approach towards designing approximation algorithms. The approximation factor achieved is bounded by the ratio of the optimal solution of the problem to the optimal solution of the relaxation. The $\mathcal{O}\left(\sqrt{\log n}\right)$-approximation algorithm of Arora et. al. [13] is based on semi-definite programming. The standard SDP relaxation for the Sparsest Cut problem has an intergrality gap of $\Omega(1/\sqrt{\lambda_2})$. However, Arora et. al. [13] showed that adding *triangle inequality* constraints between every triplet of vertices breaks the integrality gap and the resulting SDP solution can be rounded to an integral solution that is at most $\mathcal{O}\left(\sqrt{\log n}\right)$ times the cost of the SDP solution. This leads one to speculate if a better approximation factor can be obtained by a better rounding algorithm for this stronger SDP relaxation or by strengthening the SDP for Sparsest Cut by adding more constraints.

In a break-through work, Khot and Vishnoi [56] showed that the SDP relaxation for General Sparsest Cut with triangle inequality constraints, has an integrality gap of at least $\Omega\left((\log\log n)^{1/6-o(1)}\right)$. Following a series of works [80, 58], the current best integrality gap known is $(\log n)^{\Omega(1)}$ due to Cheeger, Kleiner and Naor [30]. Building on the work of Khot and Vishnoi [56], Devanur et. al. [35] showed that the SDP relaxation for (uniform) Sparsest Cut with triangle inequality constraints, has an integrality gap of at least $\Omega(\log\log n)$. In a recent work, Kane and Meka [50] gave a family of instances having integrality gap at least $e^{\Omega\left(\sqrt{\log\log n}\right)}$.

Vertex Expansion. The work of Alon [1] implies a polynomial time algorithm to compute a set having vertex expansion at most $\mathcal{O}\left(\sqrt{d\cdot\phi^{\mathsf{V}}}\right)$, where d is the largest vertex degree in the graph. Leighton and Rao [64] gave a $\mathcal{O}(\log n)$-approximation algorithm for computing the vertex expansion. Subsequently Feige, Hajiaghayi and Lee [39], building on the work of [13], gave a $\mathcal{O}\left(\sqrt{\log n}\right)$-approximation algorithm for this problem.

2.3 Cheeger's Inequality

For the sake of completeness, we give a proof of the Cheeger's Inequality.

Theorem 2.3.1 ([1, 3])**.** *For any graph $G = (V, E, w)$,*

$$\frac{\lambda_2}{2} \leqslant \phi_G \leqslant \sqrt{2\lambda_2}$$

Towards proving this theorem, we first prove the following lemma. The proof of this lemma can be found in [33].

Lemma 2.3.2. *Let $X \in (\mathbb{R}^+)^n$ be a vector such that $|\mathsf{supp}(X)| \leqslant n/2$ and*

$$\frac{\sum_{i \sim j} w\left(\{i,j\}\right)|X_i - X_j|}{\sum_i d_i X_i} \leqslant \varepsilon\,.$$

Then one of the level sets of X, say S, satisfies $\phi_G(S) \leqslant \varepsilon$.

Proof. W.l.o.g. we may assume that $X_1 \geqslant X_2 \geqslant \ldots \geqslant X_n \geqslant 0$. Let S_i denote the set consisting of the first i vertices in this ordering (breaking ties arbitrarily). Then,

$$\begin{aligned}
\frac{\sum_{i \sim j} w\left(\{i,j\}\right)|X_i - X_j|}{\sum_i d_i X_i} &= \frac{\sum_{i=1}^{n} \sum_{\substack{j \sim i \\ j > i}} w\left(\{i,j\}\right) \sum_{l=i}^{j-1} X_l - X_{l+1}}{\sum_i d_i X_i} \\
&= \frac{\sum_{i=1}^{n} (X_i - X_{i+1}) w(E(S_i, \bar{S}_i))}{\sum_{i=1}^{n} (X_i - X_{i+1}) w(S_i)} \\
&\geqslant \min_{\substack{i \in [n] \\ X_i - X_{i+1} > 0}} \frac{w(E(S_i, \bar{S}_i))}{w(S_i)} \\
&= \min_{\substack{i \in [n] \\ X_i - X_{i+1} > 0}} \phi(S_i)\,.
\end{aligned}$$

□

Next, we show the following lemma.

Lemma 2.3.3. *Let $X \in \mathbb{R}^n$ be any vector. Then for some level set $S \subseteq \mathsf{supp}(X)$ satisfies*

$$\phi(S) \leqslant \sqrt{2\frac{\sum_{i\sim j} w(\{i,j\})(X_i - X_j)^2}{\sum_i d_i X_i^2}}\,.$$

Proof.

$$\begin{aligned}
&\frac{\sum_{i\sim j} w(\{i,j\})\left|X_i^2 - X_j^2\right|}{\sum_i d_i X_i^2} \\
&= \frac{\sum_{i\sim j} w(\{i,j\})\left|X_i - X_j\right| \cdot \left|X_i + X_j\right|}{\sum_i d_i X_i^2} \\
&\leqslant \frac{\sqrt{\sum_{i\sim j} w(\{i,j\})(X_i - X_j)^2}\sqrt{\sum_{i\sim j} w(\{i,j\})(X_i + X_j)^2}}{\sum_i d_i X_i^2} \qquad \text{(Cauchy-Schwarz)} \\
&\leqslant \frac{\sqrt{\sum_{i\sim j} w(\{i,j\})(X_i - X_j)^2}\sqrt{2\sum_i d_i X_i^2}}{\sum_i d_i X_i^2} \\
&= \sqrt{2\frac{\sum_{i\sim j} w(\{i,j\})(X_i - X_j)^2}{\sum_i d_i X_i^2}}\,.
\end{aligned}$$

Invoking Lemma 2.3.2 with the vector X^2 finishes the proof of this lemma. □

We are now ready to finish the proof of Theorem 2.3.1.

Proof of Theorem 2.3.1. 1. Let $S \subset V$ be any set such that $\mathsf{vol}(S) \leqslant \mathsf{vol}(V)/2$, and let $X \in \mathbb{R}^n$ be the indicator vector of S. Let Y be the component of X orthogonal to μ^*. Then

$$\begin{aligned}
\lambda_2 &\leqslant \frac{\sum_{i\sim j} w(\{i,j\})(Y_i - Y_j)^2}{\sum_i d_i Y_i^2} = \frac{\sum_{i\sim j} w(\{i,j\})(X_i - X_j)^2}{\sum_i d_i X_i^2 - (\sum_i d_i X_i)^2/(\sum_i d_i)} \\
&= \frac{w(E(S,\bar{S}))}{\mathsf{vol}(S) - \mathsf{vol}(S)^2/\mathsf{vol}(V)} = \frac{\phi(S)}{1 - \mathsf{vol}(S)/\mathsf{vol}(V)} \\
&\leqslant 2\,\phi(S)\,.
\end{aligned}$$

Since the choice of the set S was arbitrary, we get

$$\frac{\lambda_2}{2} \leqslant \phi_G\,.$$

2. By the definition of v_2 we have

$$\left(D^{-\frac{1}{2}}AD^{-\frac{1}{2}}\right)\mathsf{v}_2 = (1-\lambda_2)\mathsf{v}_2\,.$$

W.l.o.g. we may assume that $\mu^*\left(\mathsf{supp}(\mathsf{v}_2^+)\right) \leqslant \mu^*\left(\mathsf{supp}(\mathsf{v}_2^-)\right)$. Since all the entries of A are non-negative, we have

$$\left(D^{-\frac{1}{2}}AD^{-\frac{1}{2}}\right)\mathsf{v}_2^+ \geqslant (1-\lambda_2)\mathsf{v}_2^+ \qquad \text{(coordinate wise)}$$

and hence

$$\left(I - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}\right)\mathsf{v}_2^+ \leqslant \lambda_2\mathsf{v}_2^+ \qquad \text{(coordinate wise)}\,.$$

Therefore,

$$\frac{(\mathsf{v}_2^+)^T\mathcal{L}\mathsf{v}_2^+}{(\mathsf{v}_2^+)^T\mathsf{v}_2^+} \leqslant \lambda_2 \quad \text{and} \quad \mu^*\left(\mathsf{supp}(\mathsf{v}_2^+)\right) \leqslant \frac{1}{2}\,.$$

Invoking Lemma 2.3.3 on v_2^+, we get a set $S \subseteq \mathsf{supp}(\mathsf{v}_2^+)$ having $\phi(S) \leqslant \sqrt{2\lambda_2}$. Therefore,

$$\phi_G \leqslant \sqrt{2\lambda_2}\,.$$

This finishes the proof of the theorem.

□

2.4 *The Small Set Expansion Hypothesis*

A more refined measure of the edge expansion of a graph is its expansion profile. Specifically, for a graph G the expansion profile is given by the curve

$$\phi_\delta = \min_{\mu^*(S)\leqslant\delta} \phi(S) \qquad\qquad \forall\delta \in [0, 1/2]\,.$$

The problem of approximating the expansion profile has received much less attention, and is seemingly far less tractable. In summary, the current state-of-the-art algorithms for approximating the expansion profile of a graph are still far from satisfactory. Specifically, the following hypothesis is consistent with the known algorithms for approximating expansion profile.

Hypothesis (Small-Set Expansion Hypothesis, [86])**.** *For every constant $\eta > 0$, there exists sufficiently small $\delta > 0$ such that given a graph G it is NP-hard to distinguish the cases,*

Yes: *there exists a vertex set S with volume $\mu(S) = \delta$ and expansion $\phi(S) \leqslant \eta$,*

No: *all vertex sets S with volume $\mu(S) = \delta$ have expansion $\phi(S) \geqslant 1 - \eta$.*

Apart from being a natural optimization problem, the Small Set Expansion problem is closely tied to the Unique Games Conjecture. Recent work by Raghavendra and Steurer [86] established a reduction from the Small Set Expansion problem in graphs to the well known Unique Games problem, thereby showing that Small-Set Expansion Hypothesis implies the Unique Games Conjecture. This result suggests that the problem of approximating expansion of small sets lies at the combinatorial heart of the Unique Games problem.

In a breakthrough work, Arora, Barak, and Steurer [8] showed that the problem Small Set Expansion admits a subexponential time algorithm, namely an algorithm that runs in time $\exp(n^{\eta}/\delta)$. However, such an algorithm does not refute the hypothesis that the problem Small Set Expansion(η, δ) might be hard for every constant $\eta > 0$ and sufficiently small $\delta > 0$.

The Unique Games Conjecture is not known to imply hardness results for problems closely tied to graph expansion such as Balanced Separator. The reason being that the hard instances of these problems are required to have certain global structure namely expansion. Gadget reductions from a unique games instance preserve the global properties of the unique games instance such as lack of expansion. Therefore, showing hardness for graph expansion problems often required a stronger version of the Expanding Unique Games, where the instance is guaranteed to have good expansion. To this end, several such variants of the conjecture for expanding graphs have been defined in literature, some of which turned out to be false [10]. The Small-Set Expansion Hypothesis could possibly serve as a natural unified assumption that yields all the implications of expanding unique games and, in addition, also hardness results for other fundamental problems such as Balanced Separator. In fact, Raghavendra, Steurer and Tulsiani [88] show that the the SSE hypothesis implies that the Cheeger's algorithm yields the best approximation for the Sparsest Cut problem.

2.5 Probabilistic Inequalities

We collect here some standard probabilistic inequalities that we will make use of.

Fact 2.5.1 (One-sided Chebychev Inequality). *For a random variable X with mean μ and variance σ^2 and any $t > 0$,*

$$\mathbb{P}[X < \mu - t\sigma] \leqslant \frac{1}{1+t^2}.$$

Fact 2.5.2 (Paley-Zygmund Inequality). *For a random variable $Z \geqslant 0$ with finite variance, and any $t \in (0,1)$,*

$$\mathbb{P}[Z \geqslant t\,\mathbb{E}[Z]] \geqslant (1-t)^2 \frac{\mathbb{E}[Z]^2}{\mathbb{E}[Z^2]}.$$

Fact 2.5.3 (Hoeffding's Inequality). *Let $X_1, \ldots, X_n$ be independent random variables, such that each X_i is bounded almost surely, i.e.*

$$\mathbb{P}[X_i \in [a_i, b_i]] = 1 \qquad \text{for some } a_i, b_i \in \mathbb{R}.$$

Then the mean $\bar{X} \stackrel{\text{def}}{=} \left(\sum_i X_i\right)/n$ satisfies

$$\mathbb{P}\left[\bar{X} - \mathbb{E}\left[\bar{X}\right] \geqslant t\right] \leqslant \exp\left(-\frac{2n^2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$

Properties of Gaussian Variables. The next few facts are folklore about Gaussians. Let $t_{1/k}$ denote the $(1/k)^{th}$ cap of a standard normal variable, i.e., $t_{1/k} \in \mathbb{R}$ is the number such that for a standard normal random variable X, $\mathbb{P}\left[X \geqslant t_{1/k}\right] = 1/k$.

Fact 2.5.4. *For a standard normal random variable X and for every $k > 100$,*

$$t_{1/k} \approx \sqrt{2\log k - \log\log k}$$

Fact 2.5.5. *Let $X_1, X_2, \ldots, X_k$ be k independent standard normal random variables. Let Y be the random variable defined as $Y \stackrel{\text{def}}{=} \max\{X_i | i \in [k]\}$. Then*

1. $t_{1/k} \leqslant \mathbb{E}[Y] \leqslant 2\sqrt{\log k}$

2. $\mathbb{E}[Y^2] \leqslant 4\log k$

3. $\mathbb{E}[Y^4] \leqslant 4e\log^2 k$

Proof. For any $Z_1, \ldots, Z_k \in \mathbb{R}^+$ and any $p \in \mathbb{Z}^+$, we have $\max_i Z_i \leqslant (\sum_i Z_i^p)^{\frac{1}{p}}$. Now $Y^4 = (\max_i X_i)^4 \leqslant \max_i X_i^4$.

$$\begin{aligned} \mathbb{E}\left[Y^4\right] &\leqslant \mathbb{E}\left[\left(\sum_i X_i^{4p}\right)^{\frac{1}{p}}\right] \leqslant \left(\mathbb{E}\left[\sum_i X_i^{4p}\right]\right)^{\frac{1}{p}} \qquad (\text{ Jensen's Inequality }) \\ &\leqslant \left(\sum_i (\mathbb{E}\left[X_i^2\right])\frac{(4p)!}{(2p)!2^{2p}}\right)^{\frac{1}{p}} \leqslant 4p^2 k^{\frac{1}{p}} \quad (\text{using } (4p)!/(2p)! \leqslant (4p)^{2p}\ ) \end{aligned}$$

Picking $p = \log k$ gives $\mathbb{E}[Y^4] \leqslant 4e\log^2 k$.

Therefore $\mathbb{E}[Y^2] \leqslant \sqrt{\mathbb{E}[Y^4]} \leqslant 4\log k$ and $\mathbb{E}[Y] \leqslant \sqrt{\mathbb{E}[Y^2]} \leqslant 2\sqrt{\log k}$.

□

The next lemma bounds the probability that a sum of standard normal random variables is not too small.

Lemma 2.5.6. *Suppose $z_1, \ldots, z_m$ are gaussian random variables (not necessarily independent) such $\mathbb{E}[\sum_i z_i^2] = 1$ then*

$$\mathbb{P}\left[\sum_i z_i^2 \geqslant \frac{1}{2}\right] \geqslant \frac{1}{12}$$

Proof. We will bound the variance of the random variable $R = \sum_i z_i^2$ as follows,

$$\begin{aligned}
\mathbb{E}[R^2] &= \sum_{i,j} E[z_i^2 z_j^2] \\
&\leqslant \sum_{i,j} \left(E[z_i^4]\right)^{\frac{1}{2}} \left(E[z_j^4]\right)^{\frac{1}{2}} \\
&= \sum_{i,j} 3E[z_i^2]E[z_j^2] \qquad \left(\text{Using } \mathbb{E}[g^4] = 3\left(\mathbb{E}[g^2]\right)^2 \text{ for gaussians }\right) \\
&= 3\left(\sum_i E[z_i^2]\right)^2 = 3
\end{aligned}$$

By the Paley-Zygmund inequality (Fact 2.5.2),

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

□

Lemma 2.5.7 ([27])**.** *Let $X_1, \ldots, X_k$ and $Y_1, \ldots, Y_k$ be i.i.d. standard normal random variables such that for all $i \in [k]$, the covariance of X_i and Y_i is at least $1 - \varepsilon^2$. Then*

$$\mathbb{P}\left[\mathsf{argmax}_i X_i \neq \mathsf{argmax}_i Y_i\right] \leqslant c_1 \left(\varepsilon\sqrt{\log k}\right)$$

for some absolute constant c_1.

Lemma 2.5.8 ([71])**.** *Given r standard normal random variables $g_1, \ldots, g_r$, with pairwise covariance at least $1 - \varepsilon^2$,*

$$\mathbb{P}\left[g_i \geqslant t_{1/k} \text{ and } g_j < t_{1/k} \text{ for some } i, j \in [r]\right] \leqslant c_1 \frac{\min\{r, k\}}{k} \varepsilon\sqrt{\log k \log r}\,.$$

2.6 Miscellaneous Inequalities

Next, we recall Weyl's Inequality.

Lemma 2.6.1 (Weyl's Inequality)**.** *Given a Hermitian matrix B with eigenvalues $\lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_n$, and a positive semidefinite matrix E, if $\lambda'_1 \leqslant \lambda'_2 \leqslant \ldots \leqslant \lambda'_n$ denote the eigenvalues of $B' \stackrel{\text{def}}{=} B - E$, then $\lambda'_i \leqslant \lambda_i$.*

Proof. The i^{th} eigenvalue of B' can be written as

$$\begin{aligned}
\lambda'_i &= \max_{S:rank(S)=i} \min_{x \in S} \frac{x^T B' x}{x^T x} \\
&= \max_{S:rank(S)=i} \min_{x \in S} \frac{x^T B x - x^T E x}{x^T x} \\
&\leqslant \max_{S:rank(S)=i} \min_{x \in S} \frac{x^T B x}{x^T x} \\
&= \lambda_i.
\end{aligned}$$

□

Proposition 2.6.2. *For any two non zero vectors u_i and u_j, if $\tilde{u_i} = u_i / \|u_i\|$ and $\tilde{u_j} = u_j / \|u_j\|$ then*

$$\|\tilde{u_i} - \tilde{u_j}\| \sqrt{\|u_i\|^2 + \|u_j\|^2} \leqslant 2 \|u_i - u_j\|$$

Proof. Note that $2 \|u_i\| \|u_j\| \leqslant \|u_i\|^2 + \|u_j\|^2$. Hence,

$$\begin{aligned}
\|\tilde{u_i} - \tilde{u_j}\|^2 (\|u_i\|^2 + \|u_j\|^2) &= (2 - 2 \langle \tilde{u_i}, \tilde{u_j} \rangle)(\|u_i\|^2 + \|u_j\|^2) \\
&\leqslant 2(\|u_i\|^2 + \|u_j\|^2 - (\|u_i\|^2 + \|u_j\|^2) \langle \tilde{u_i}, \tilde{u_j} \rangle)
\end{aligned}$$

If $\langle \tilde{u_i}, \tilde{u_j} \rangle \geqslant 0$, then

$$\|\tilde{u_i} - \tilde{u_j}\|^2 (\|u_i\|^2 + \|u_j\|^2) \leqslant 2(\|u_i\|^2 + \|u_j\|^2 - 2 \|u_i\| \|u_j\| \langle \tilde{u_i}, \tilde{u_j} \rangle) \leqslant 2 \|u_i - u_j\|^2$$

Else if $\langle \tilde{u_i}, \tilde{u_j} \rangle < 0$, then

$$\|\tilde{u_i} - \tilde{u_j}\|^2 (\|u_i\|^2 + \|u_j\|^2) \leqslant 4(\|u_i\|^2 + \|u_j\|^2 - 2 \|u_i\| \|u_j\| \langle \tilde{u_i}, \tilde{u_j} \rangle) \leqslant 4 \|u_i - u_j\|^2$$

□

2.6.1 Notation

We use μ to denote a probability distribution on vertices of the graph/hypergraph. For a set of vertices S, we define $\mu(S) = \int_{x \in S} \mu(x)$. We use $\mu_{|S}$ to denote the distribution

μ restricted to the set $S \subset V(G)$. For the sake of simplicity, we sometimes say that vertex $v \in V(G)$ has weight $w(v)$, in which case we define $\mu(v) = w(v)/\sum_{u \in V} w(u)$. We denote the weight of a set $S \subseteq V$ by $w(S)$.

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

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

We denote the 2-norm of a vector by $\|\cdot\|$, and its 1 norm by $\|\cdot\|_1$.

We use $\Pi(\cdot)$ to denote projection operators. For a subspace S, we denote by $\Pi_S : \mathbb{R}^n \to \mathbb{R}^n$ the projection operator that maps a vector to its projection on S. We denote by $\Pi_S^\perp : \mathbb{R}^n \to \mathbb{R}^n$ the projection operator that maps a vector to its projection orthogonal to S.

THE COMPLEXITY OF EXPANSION PROBLEMS

PART I

Spectral Bounds

CHAPTER III

HIGHER ORDER CHEEGER INEQUALITIES FOR GRAPHS

3.1 Introduction

In this chapter we study extensions of Edge Expansion in graphs to more than one subset. We study multiple natural generalizations of Sparsest Cut problem. All these generalizations are parametrized by a positive integer k, and reduce to the Edge Expansion problem when restricted to the case $k = 2$. A natural question is whether these problems are connected to higher eigenvalues of the graph. We obtain upper and lower bounds for these generalizations of Sparsest Cut using higher eigenvalues. In the rest of the section, we briefly describe each generalization and present our results.

Problem 3.1.1 (Min Sum k-partition)**.** Given a weighted undirected graph $G = (V, E, w)$ and an integer $k > 1$, find the k-partition of V with the least sum-sparsity, where the sum-sparsity of a k-partition $\mathcal{P} = \{S_1, \ldots, S_k\}$ is defined as the ratio of the weight of edges between different parts to the sum of the weights of smallest $k-1$ parts in $\mathcal{P}$, i.e.,

$$\phi^{\mathsf{sum}}(\mathcal{P}) \stackrel{\text{def}}{=} \frac{\sum_{i \neq j} w(V_i, V_j)}{\min_{j \in [k]} w(V \backslash V_j)} .$$

Variants of the Min Sum k-partition have been considered in the literature. Closer to this is the k-cut problem which asks to partition a graph into k pieces so as to minimize the fraction of edges cut. Saran and Vazirani [91] gave a 2-approximation algorithm for this problem.

It is easy to see that the lower bound in Cheeger's inequality implies a lower bound on $\phi^{\mathsf{sum}}(\cdot)$,

$$\phi^{\mathsf{sum}}(\mathcal{P}) \geqslant \lambda_2/2 \qquad \forall \text{ partitions } \mathcal{P} .$$

As it turns out, this lower bound cannot be strengthened for $k > 2$. To see this, consider the following simple construction: construct a graph G by taking $k-1$ cliques $C_1, C_2, \ldots, C_{k-1}$ each on $(n-1)/(k-1)$ vertices along with an additional vertex v. Let the cliques $C_1, \ldots, C_{k-1}$ be connected to v by a single edge. Now, the graph G will have $k-1$ eigenvalues close to 1 because of the $k-1$ cuts $(\{v\}, C_i)$ for $i \in [k-1]$. However, the k^{th} eigenvalue will be close to 0, since any other cut which is not a linear combination of these $k-1$ cuts will have to cut through one of the cliques. Therefore, λ_k is a constant smaller than $1/2$. But $\min_{\mathcal{P}} \phi^{\mathsf{sum}}(\mathcal{P}) = (k-1)/((k-2)(n/k)^2) \approx k^2/n^2$. Thus, $\lambda_k \gg \min_{\mathcal{P}} \phi^{\mathsf{sum}}(\mathcal{P})$ for small enough values of k.

Our main result is an upper bound on the Sparsest k-Partition via the higher eigenvalues. Specifically, we show the following.

Theorem 3.1.2. *For any edge-weighted graph $G = (V, E, w)$, and any integer $1 \leqslant k \leqslant n$, there exists a k-partition $S_1, \ldots, S_k$ of the vertices such that*

$$\phi^{\mathsf{sum}}(\{S_1, \ldots, S_k\}) \leqslant 8\sqrt{\lambda_k} \log k\,.$$

Moreover, such a partition can be identified in polynomial time.

The proof of Theorem 3.1.5 is based on a simple recursive partitioning algorithm that might be of independent interest.

The second problem we study is the following.

Problem 3.1.3 (k sparse-cuts)**.** Given an edge weighted graph $G = (V, E, w)$ and an integer $k > 1$, find k disjoint non-empty subsets $S_1, \ldots, S_k$ of V such that $\max_i \phi_G(S_i)$ is minimized.

Note that the sets $S_1, \ldots, S_k$ need not form a partition of the set of vertices, i.e., there could be vertices that do not belong to any of the sets. Therefore problem models the existence of several well-formed *clusters* in a graph without the clusters being required to form a partition.

Along the lines of lower bound in Cheeger's inequality, it is not hard to show that the k^{th} smallest eigenvalue of the normalized Laplacian of the graph gives a lower bound to the K SPARSE-CUTS problem. Formally, we prove the following lower bound.

Proposition 3.1.4. *For any edge-weighted graph $G = (V, E, w)$, for any integer $1 \leqslant k \leqslant n$, and for any k disjoint subsets $S_1, \ldots, S_k \subset V$*

$$\max_i \phi_G(S_i) \geqslant \frac{\lambda_k}{2}$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of the normalized Laplacian of G.

Complementing the lower bound, we show the following upper bound on K SPARSE-CUTS problem in terms of λ_k.

Theorem 3.1.5. *For absolute constants c, C, the following holds: For every edge-weighted graph $G = (V, E, w)$, and any integer $1 \leqslant k \leqslant n$, there exist ck disjoint subsets $S_1, \ldots, S_{ck}$ of vertices such that*

$$\max_i \phi_G(S_i) \leqslant C\sqrt{\lambda_k \log k}$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of the normalized Laplacian of G. Moreover, the sets $S_1, \ldots, S_{ck}$ satisfying the inequality can be identified in polynomial time.

The proof of Theorem 3.1.5 is algorithmic and is based on spectral projection. Starting with the embedding given by the smallest k eigenvectors of the (normalized) Laplacian of the graph, a simple randomized rounding procedure is used to produce k vectors having disjoint support, and then a Cheeger cut is obtained from each of these vectors. The running time is dominated by the time taken to compute the smallest k eigenvectors of the normalized Laplacian.

In general, one can not prove an upper bound better than $\mathcal{O}(\sqrt{\lambda_k \log k})$ for K SPARSE-CUTS. This bound is matched by the family of *Gaussian graphs*. For a constant $\varepsilon \in (-1, 1)$, let $N_{k,\varepsilon}$ denote the infinite graph over $\mathbb{R}^k$ where the weight of an edge (x, y) is the probability density that two standard Gaussian random vectors X, Y with correlation [1] $1 - \varepsilon$ equal x and y respectively. The first k eigenvalues of the Laplacian of $N_{k,\varepsilon}$ are at most ε ([88]). The following lemma bounds the expansion of small sets in $N_{k,\varepsilon}$.

Lemma 3.1.6 ([23])**.** *For any set $S \subset \mathbb{R}^k$ with Gaussian probability measure at most $1/k$,*

$$\phi_{N_{k,\varepsilon}}(S) = \Omega\left(\sqrt{\varepsilon \log k}\right) .$$

For any k disjoint subsets $S_1, \ldots, S_k$ of the Gaussian graph $N_{k,\varepsilon}$, at least one of the sets has measure smaller than $\frac{1}{k}$, thus implying

$$\max_i \phi_{N_{k,\varepsilon}}(S_i) = \Omega\left(\sqrt{\epsilon \log k}\right) = \Omega\left(\sqrt{\lambda_k \log k}\right) .$$

It is natural to wonder if the above bounds extend to the case when the k-sets are required to form a partition. First, it is easy to see that Theorem 3.1.5 also implies an upper bound of $\mathcal{O}(\sqrt{\lambda_k \log k})$ on $\max_i \phi(S_i)$ for the case when the sets are required to form a partition of the vertex set.

Corollary 3.1.7. *For any edge-weighted graph $G = (V, E, w)$ and any integer $1 \leqslant k \leqslant n$, there exists a partition of the vertex set V into ck parts $S_1, \ldots, S_{ck}$ such that*

$$\max_i \phi(S_i) \leqslant C\sqrt{\lambda_k \log k}$$

for absolute constants c, C.

Complementing the above bound, we show that for a k-partition $S_1, S_2, \ldots, S_k$, the quantity $\max_i \phi_G(S_i)$ cannot be bounded by $\mathcal{O}(\sqrt{\lambda_k}\,\mathsf{polylog} k)$ in general. We view this as further evidence suggesting that the K SPARSE-CUTS problem is the right generalization of SPARSEST CUT to multiple subsets.

[1] ε correlated Gaussians can be constructed as follows : $X \sim \mathcal{N}(0,1)^k$ and $Y \sim (1-\varepsilon)X + \sqrt{2\varepsilon - \varepsilon^2} Z$ where $Z \sim \mathcal{N}(0,1)^k$.

Theorem 3.1.8. *There exists a family of graphs such that for any k-partition $\{S_1, \ldots, S_k\}$ of the vertex set*

$$\max_i \phi_G(S_i) \geqslant C \min\left\{\frac{k^2}{\sqrt{n}}, n^{\frac{1}{12}}\right\} \sqrt{\lambda_k}.$$

We also recall the Small Set Expansion problem (Problem 2.1.7).

Problem 3.1.9 (Small Set Expansion)**.** Given an edge weighted graph $G = (V, E, w)$ and $k > 1$, find a subset of vertices S such that $w(S) \leqslant w(V)/k$ and $\phi_G(S)$ is minimized.

As an immediate consequence of Theorem 3.1.5, we get the following optimal bound on the small-set expansion problem.

Corollary 3.1.10. *For any edge-weighted graph $G = (V, E, w)$ and any integer $1 \leqslant k \leqslant n$, there is a subset S with $w(S) = \mathcal{O}(1/k)w(V)$ and $\phi_G(S) \leqslant C\sqrt{\lambda_k \log k}$ for an absolute constant C.*

3.1.1 Related work

The classic Sparsest Cut problem has been extensively studied, and is closely connected to metric geometry [68, 14]. Leighton and Rao [64] gave an $\mathcal{O}(\log n)$ factor approximation algorithm via an LP relaxation. The same approximation factor can also be achieved using using properties of embeddings of metrics into Euclidean space [68, 14]. This was improved to $\mathcal{O}(\sqrt{\log n})$ via a semi-definite relaxation and embeddings of special metrics by Arora, Rao and Vazirani [12]. In many contexts, and in practice, the eigenvector approach is often preferred in spite of a higher worst-case approximation factor.

Arora, Barak and Steurer [8] showed that the expansion of sets of size at most n/k can be bounded by $\mathcal{O}(\sqrt{\lambda_{k^{100}} \log_k n})$. Using a semidefinite programming relaxation, Raghavendra, Steurer and Tetali [87] gave an algorithm that outputs a small set with expansion at most $\sqrt{\mathsf{OPT} \log k}$ where OPT is the sparsity of the optimal set of size at most n/k. Bansal et.al. [16] obtained an $\mathcal{O}(\sqrt{\log n \log k})$ approximation algorithm also using a semidefinite programming relaxation.

A problem closely related to the Min Sum k-partition problem is the k-cut problem that asks for a k-partition which minimizes the fraction of edges cut. Saran and Vazirani [91] gave a 2-approximation algorithm for this problem.

In an independent work, [63] have obtained results similar to Theorem 3.1.5 with different techniques. They also studied a close variant of the problem we consider, and show that every graph G has a k partition such that each part has expansion at most $\mathcal{O}(k^3\sqrt{\lambda_k})$. Other generalizations of the sparsest cut problem have been considered for special classes of graphs ([20, 53, 98]).

A randomized rounding step similar to the one in our algorithm was used previously in the context of rounding semidefinite programs for unique games ([27]).

3.1.2 Notation

Recall that we use $0 = \lambda_1 \leqslant \lambda_2 \leqslant \ldots \lambda_n$ to denote the eigenvalues of $\mathcal{L}_G$ and use $\mathsf{v}_1, \mathsf{v}_2, \ldots, \mathsf{v}_n$ to denote the corresponding eigenvectors. Let $v_i \stackrel{\text{def}}{=} D^{-\frac{1}{2}}\mathsf{v}_i$ for each $i \in [n]$. Then,

$$\mathsf{v}_i^T \mathcal{L}_G \mathsf{v}_i = \sum_{x \sim y} w\left(\{x,y\}\right)\left(v_i(x) - v_j(y)\right)^2.$$

Since $\forall i \neq j$ $\langle \mathsf{v}_i, \mathsf{v}_j \rangle = 0$, $\sum_l d_l v_i(l) v_j(l) = 0$.

Given a k-partition $\mathcal{P} = \{S_1, \ldots, S_k\}$ we denote the sum of the weights of the edges with endpoints in different pieces by $E(\mathcal{P})$. More formally,

$$E(\mathcal{P}) \stackrel{\text{def}}{=} \frac{1}{2} \sum_{e \in E(S_i, \bar{S}_i)} w(e)$$

Organization. We study the Min Sum k-partition problem and prove Theorem 3.1.2 in Section 3.2. We prove our upper bound for k sparse-cuts (Theorem 3.1.5) in Section 3.3 and we prove the lower bound for k sparse-cuts (Proposition 3.1.4) in Section 3.4.

3.2 Min Sum k-partition

3.2.1 Recursive partitioning algorithm

We propose the following recursive algorithm for finding a k-partitioning of G. Use the second eigenvector of $\mathcal{L}$ to find a sparse cut $(C, \bar{C})$. Let $G' = (V, E')$ be the graph obtained by removing the edges in the cut $(C, \bar{C})$ from G and adding self loops at the endpoints of the edges removed. Let $\mathcal{L}'$ be the normalized Laplacian of the graph obtained. The matrix $\mathcal{L}'$ is block-diagonal with two blocks for the two components of G'. The spectrum of $\mathcal{L}'$ (eigenvalues, eigenvectors) is the union of the spectra of the two blocks. The first two eigenvalues of $\mathcal{L}'$ are now 0 and we use the third largest eigenvector of $\mathcal{L}'$ to find a sparse cut in G'. This is the second eigenvector in one of the two blocks and partitions that block. We repeat the above process till we have at least k connected components. This can be viewed as a recursive algorithm, where at each step one of the current components is partitioned into two; the component partitioned is the one that has the lowest second eigenvalue among all the current components. The precise algorithm appears in Algorithm 3.2.1.

3.2.2 Analysis

In this section, we analyze the recursive partitioning algorithm given in Algorithm 3.2.1. Our analysis will also be a proof of Theorem 3.1.2. We begin with some monotonicity properties of eigenvalues.

Monotonicity of Eigenvalues. We first prove a lemma about the monotonicity of eigenvalues on removing edges from the graph.

Lemma 3.2.2. *Let $\mathcal{L}$ be the normalized Laplacian matrix of the graph G. Let F be any subset of edges of G. For every pair $\{i, j\} \in F$, remove the edge $\{i, j\}$ from G and add self loops at i and j to get the graph G'. Let $\mathcal{L}'$ be the normalized Laplacian matrix of G'. Let the eigenvalues of $\mathcal{L}$ be $0 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_n$ and let the eigenvalues of*

Algorithm 3.2.1.
Input: Graph $G = (V, E, w)$, $k \in \mathbb{Z}_{\geqslant 0}$ such that $1 < k < n$.
Initialize $i := 2$, and $G_i = G$, $\mathcal{L}_i =$ normalized Laplacian matrix of G_i.

1. Find a sparse cut $(C_i, \bar{C}_i)$ in G_i using the i^{th} eigenvector of $\mathcal{L}_i$ (the first $i-1$ are all equal to 0).
2. Set $V(G_{i+1}) = V$ and
$$E(G_{i+1}) := \big(E(G_i) \setminus E_{G_i}(C_i, \bar{C}_i)\big) \cup \big\{\{v, v\} \mid \exists\, u \text{ such that } \{u, v\} \in E_{G_i}(C, \bar{C})\big\}$$
with $w(\{v, v\}) = \sum_{\{u,v\} \in E_{G_i}(C,\bar{C})} w(\{u, v\})$.
3. If $i = k$ then output the connected components of G_{i+1} and End else
4. Let $\mathcal{L}_{i+1}$ be the normalized Laplacian matrix of G_{i+1}.

Figure 1: Recursive Algorithm for Min Sum k-partition

$\mathcal{L}'$ be $0 \leqslant \lambda'_2 \leqslant \lambda'_3 \leqslant \ldots \leqslant \lambda'_n$. Then $\lambda'_i \leqslant \lambda_i$ $\forall i \in [n]$.

Proof. Let $C \stackrel{\text{def}}{=} \mathcal{L} - \mathcal{L}'$ is the matrix corresponding to the edge subset F. It has non-negative entries along its diagonal and non-positive entries elsewhere such that $\forall i$ $c_{ii} = -\sum_{j \neq i} c_{ij}$. C is symmetric and positive semi-definite as for any vector x of appropriate dimension, we have

$$x^T C x = \sum_{ij} c_{ij} x_i x_j = -\frac{1}{2} \sum_{i \neq j} c_{ij} (x_i - x_j)^2 \geqslant 0.$$

Using Lemma 2.6.1, we get that $\lambda'_i \leqslant \lambda_i$ $\forall i \in [n]$. □

Lemma 3.2.2 shows that the eigenvalues of $\mathcal{L}_i$ are monotonically non-increasing with i. This will show that $\phi_{G_i}(C_i) \leqslant \sqrt{2\lambda_k}$. We are now ready to prove Theorem 3.1.2.

Proof of Theorem 3.1.2 . Let $\mathcal{P}$ be the partition output by the algorithm and let $S(\mathcal{P})$ denote the sum of weights of the smallest $k-1$ pieces in $\mathcal{P}$. Note that we need only the smaller side of a cut to bound the size of the cut:

$$w(E_G(S, \bar{S})) \leqslant \phi_G \; w(S)\,.$$

We define the notion of a $\mathsf{cut-tree}$ $T = (V(T), E(T))$ as follows: $V(T) = \{V\} \cup \{C_i | i \in [k]\}$ (For any cut $(C_i, \bar{C}_i)$ we denote the part with the smaller weight by C_i and the part with the larger weight by $\bar{C}_i$. We break ties arbitrarily). We put an edge between $S_1, S_2 \in V(T)$ if $\nexists S \in V(T)$ such that $S_1 \subsetneq S \subsetneq S_2$ or $S_2 \subsetneq S \subsetneq S_1$, (one can view S_1 as a 'top level' cut of S_2 in the former case).

Clearly, T is connected and is a tree. We call V the root of T. We define the *level* of a node in T to be its depth from the root. We denote the level of node $S \in V(T)$ by $L(S)$. The root is defined to be at level 0. Observe that $S_1 \in V(T)$ is a descendant of $S_2 \in V(T)$ if and only if $S_1 \subsetneq S_2$. Now

$$E(\mathcal{P}) = \cup_i E_{G_i}(C_i, \bar{C}_i) = \cup_i \cup_{j:L(C_j)=i} E_{G_j}(C_j, \bar{C}_j)\,.$$

We make the following claim.

Claim 3.2.3.

$$w(\cup_{j:L(C_j)=i} E(C_j, \bar{C}_j)) \leqslant 2\sqrt{\lambda_k}\; S(\mathcal{P})$$

Proof. By definition of level, if $L(C_i) = L(C_j)$, $i \neq j$, then the node corresponding to C_i in the T can not be an ancestor or a descendant of the node corresponding to C_j. Hence, $C_i \cap C_j = \phi$. Therefore, all the sets of vertices in level i are pairwise disjoint. Using Cheeger's inequality we get that

$$E(C_j, \bar{C}_j) \leqslant 2\sqrt{\lambda_k} w(C_j)\,.$$

Therefore

$$w(\cup_{j:L(C_j)=i} E(C_j, \bar{C}_j)) \leqslant 2\sqrt{\lambda_k} \sum_{j:L(C_j)=i} w(C_j) \leqslant 2\sqrt{\lambda_k} S(\mathcal{P})$$

□

This claim implies that $\phi(\mathcal{P}) \leqslant 2\sqrt{\lambda_k}\ \mathsf{height}(T)$.

The height of T might be as much as k. But we will show that we can assume height(T) to be $\log k$. For any path in the tree $v_1, v_2, \ldots, v_{p-1}, v_p$ such that $\mathsf{deg}(v_1) > 2$, $\mathsf{deg}(v_i) = 2$ (i.e. v_i has only 1 child in T) for $1 < i < k$, we have $w(C_{v_{i+1}}) \leqslant w(C_{v_i})/2$, as v_{i+1} being a child of v_i in the T implies that $C_{v_{i+1}}$ was obtained by cutting C_{v_i} using it's second eigenvector. Thus

$$\sum_{i=2}^{p} w(C_{v_i}) \leqslant w(C_{v_1})\,.$$

Hence we can modify the T as follows : make the nodes $v_3, \ldots, v_p$ children of v_2. The nodes $v_3, \ldots, v_{p-1}$ now become leaves whereas the subtree rooted at v_p remains unchanged. We also assign the level of each node as its new distance from the root. In this process we might have destroyed the property that a node is obtained from by cutting its parent, but we make the following claim.

Claim 3.2.4.

$$w(\cup_{j:L(C_j)=i} E(C_j, \bar{C}_j)) \leqslant 4\sqrt{\lambda_k}\ S(\mathcal{P})$$

Proof. If the nodes in level i are unchanged by this process, then the claim clearly holds. If any node v_j in level i moved to a higher level, then the nodes replacing v_j in level i would be descendants of v_j in the original T and hence would have weight at most $w(C_{v_j})$. If the descendants of some node v_j got added to level i, then, as seen above, their combined weight would be at most $w(C_{v_j})$. Hence,

$$w(\cup_{j:L(C_j)=i} E(C_j, \bar{C}_j)) \leqslant 2\left(2\sqrt{\lambda_k} \sum_{j:L(C_j)=i} w(C_j)\right) \leqslant 4\sqrt{\lambda_k}\ S(\mathcal{P})$$

.

□

Repeating this process we can ensure that no two adjacent nodes in the T have degree 2. Hence, there are at most $\log k$ vertices along any path starting from the root which have exactly one child. Thus the height of the new $\mathsf{cut}-\mathsf{tree}$ is at most $2\log k$ and hence

$$E(\mathcal{P}) \leqslant 8\sqrt{\lambda_k}\log k\ S(\mathcal{P}) \qquad \text{and} \qquad \phi^{\mathsf{sum}} \leqslant \frac{E(\mathcal{P})}{S(\mathcal{P})} \leqslant 8\sqrt{\lambda_k}\log k\,.$$

□

3.3 k sparse-cuts

3.3.1 Geometric Embeddings

Rayleigh Quotient. Recall that for a graph $G = (V, E, w)$, and an embedding $f : V \to \mathbb{R}$ of G on to $\mathbb{R}$, the Rayleigh quotient of f is given by

$$\mathcal{R}(f) = \frac{f^T \mathcal{L}_G f}{f^T f}$$

and if we let $g = D^{-1/2} f$, then

$$\mathcal{R}(f) = \frac{g^T L g}{g^T D g} = \frac{\sum_{i\sim j} w(\{i,j\})(g_i - g_j)^2}{\sum_i d_i g_i^2}\,.$$

We can generalize this definition to higher dimensional embeddings. Let $f : V \to \mathbb{R}^d$ be an embedding of the graph G in to d-dimensional Euclidean space $\mathbb{R}^d$ for some positive integer d. We will often use f_i to denote $f(i)$. Again, we let $g_i = f_i/\sqrt{d_i}$. Then

$$\mathcal{R}(f) \stackrel{\text{def}}{=} \frac{\sum_{i\sim j} w(\{i,j\}) \|g(i) - g(j)\|_2^2}{\sum_i d_i \|g(i)\|^2}\,.$$

It is clear that the Rayleigh quotient of an embedding f measures the ratio between the averaged squared length of the edges to the average squared length of vectors in the embedding.

Spectral Embedding. The eigenvectors of $\mathcal{L}$ $\mathsf{v}_1, \ldots \mathsf{v}_n$ form an orthonormal set of vectors, i.e.,

$$\langle \mathsf{v}_a, \mathsf{v}_b \rangle = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{otherwise} \end{cases}\,.$$

By orthonormality of the vectors $\{\mathsf{v}_a\}$ we will have,

$$\sum_{i \in V} d_i v_a(i) v_b(i) = \langle \mathsf{v}_a, \mathsf{v}_b \rangle = \delta_{ab} \,.$$

Hence for each $\ell \in [n]$, the set of vectors $v_1, \ldots v_\ell$ yield an isotropic embedding of the graph in to $\mathbb{R}^\ell$. For the sake of concreteness, we state this observation formally below.

Lemma 3.3.1. *For $k \in [n]$, the embedding $u : V \to \mathbb{R}^k$ given by the top k-eigenvectors* $\mathsf{v}_1, \ldots, \mathsf{v}_k$ *,i.e.,*

$$u(i) = \frac{1}{\sqrt{d_i}} (\mathsf{v}_1(i), \ldots, \mathsf{v}_k(i))$$

is an isotropic embedding satisfying

$$\sum_{i,j \in V} d_i d_j \langle u(i), u(j) \rangle^2 = k \qquad\qquad \sum_{i \in V} d_i \|u(i)\|^2 = k$$

and

$$\mathcal{R}_G(u) \leqslant \lambda_k \,.$$

Proof. It follows from the definition of u that

$$\sum_i d_i \|u(i)\|^2 = \sum_{l=1}^{k} \sum_{i \in V} \mathsf{v}_l(i)^2 = k \,.$$

Next,

$$\begin{aligned}
\sum_{i,j \in V} d_i d_j \langle u(i), u(j) \rangle^2 &= \sum_{i,j \in V} \left(\sum_{l=1}^{k} \mathsf{v}_l(i) \mathsf{v}_l(j) \right)^2 \\
&= \sum_{i,j \in V} \sum_{l_1, l_2 \in [k]} \mathsf{v}_{l_1}(i) \mathsf{v}_{l_1}(j) \mathsf{v}_{l_2}(i) \mathsf{v}_{l_2}(j) \\
&= \sum_{l_1, l_2 \in [k]} \langle \mathsf{v}_{l_1}, \mathsf{v}_{l_2} \rangle^2 \\
&= k \,.
\end{aligned}$$

Here the last equality follows from the fact that $\{\mathsf{v}_i : i \in [n]\}$ is set of orthonormal vectors.

□

3.3.2 Gaussian Projection Algorithm

We now present the rounding algorithm that we will use to prove Theorem 3.1.5. At a high level our algorithm can be viewed as follows. Given a graph $G = (V, E, w)$, we start with the spectral embedding $u : V \to \mathbb{R}^k$ (Lemma 3.3.1). Next we pick k random directions $g_1, \ldots, g_k$ (For technical reasons, we pick $g_1, \ldots, g_k$ to be independent Gaussian vectors). We perform a preliminary partitioning of the vertices by assigning a vertex i to the set represented by the direction g_l along which $u(i)$ has the largest projection. We refine these sets using a standard local-search approach based on the lengths of the vectors $u(i)$ (Lemma 2.3.2). We show that with constant probability, $\Omega(k)$ of the resulting sets have expansion at most $\mathcal{O}\left(\sqrt{\mathcal{R}(u) \log k}\right)$.

Algorithm 3.3.2.
Input: Graph $G = (V, E, w)$, parameter k and an isotropic embedding $u : V \to \mathbb{R}^d$.

1. Pick k independent Gaussian vectors $g_1, g_2, \ldots, g_k \sim \mathcal{N}(0, 1)^d$. Construct vectors $h_1, h_2, \ldots, h_k \in \mathbb{R}^n$ as follows:
$$h_i(a) = \begin{cases} \|u_a\|^2 & \text{if } i = \mathsf{argmax}_{i \in [k]} \{\langle u_a, g_i \rangle\} \\ 0 & \text{otherwise.} \end{cases}$$
2. For $j = 1, \ldots, k$, sort the coordinates of h_j according to their magnitude, and pick the level set having the least expansion (Lemma 2.3.2).
3. Output all subsets with expansion smaller than $C\sqrt{\mathcal{R}(u) \log k}$ for an appropriately chosen constant C.

Figure 2: The K SPARSE-CUTS Algorithm

3.3.3 Analysis

In this section, we will present the analysis of the Random Projection algorithm (Algorithm 3.3.2). We begin with an outline of the argument. We summarize the analysis as follows.

Theorem 3.3.3. *For a graph $G = (V, E, w)$, parameter $k \in \mathbb{Z}_{\geqslant 0}$ and an embedding $u : V \to \mathbb{R}^d$ such that*

$$\sum_{i \in V} d_i \left\|u_i\right\|^2 = k \qquad \text{and} \qquad \sum_{i,j \in V} d_i d_j \left\langle u_i, u_j \right\rangle^2 = k\,.$$

Then, with constant probability, Algorithm 3.3.2 outputs ck non-empty disjoint sets each having expansion at most $C\sqrt{\mathcal{R}(u)\log k}$ for some universal constants $c, C > 0$.

Theorem 3.1.5 follows directly from Theorem 3.3.3 and Lemma 3.3.1.

Theorem 3.1.5. Invoking Theorem 3.3.3 with the spectral embedding given by the top k eigenvectors (Lemma 3.3.1) yields that G has ck non-empty disjoint subsets each having expansion at most $C\sqrt{\lambda_k \log k}$. □

3.3.3.1 Proof Outline of Theorem 3.3.3

Notice that the vectors $h_1, h_2, \ldots, h_k$ have disjoint support since for each coordinate j, exactly one of the $\langle u_j, g_i \rangle$ is maximum. Therefore, the cuts obtained by the vectors h_i yield k disjoint sets. It is sufficient to show that a constant fraction of the sets so produced have small expansion. We will show that for each $i \in \{1, \ldots, k\}$, the vector h_i has a constant probability of yielding a cut with small expansion. The outline of the proof is as follows. Let f denote the vector h_1. The choice of the index 1 is arbitrary and the same analysis is applicable to all other indices $i \in [k]$. We recall Lemma 2.3.2.

Lemma 3.3.4 (Restatement of Lemma 2.3.2)**.** *Let $X \in \mathbb{R}^n$ be a vector such that $|\mathsf{supp}(X)| \leqslant n/2$ and*

$$\frac{\sum_{i \sim j} w\left(\{i, j\}\right) |X_i - X_j|}{\sum_i d_i X_i} \leqslant \varepsilon\,.$$

Then one of the level sets of X, say S, satisfies $\phi_G(S) \leqslant \varepsilon$.

Applying Lemma 2.3.2, the expansion of the set retrieved from $f = h_1$ is upper bounded by,

$$\frac{\sum_{i \sim j} w\left(\{i, j\}\right) |f_i - f_j|}{\sum_i d_i f_i}\,.$$

Both the numerator and denominator are random variables depending on the choice of random Gaussians $g_1, \ldots, g_k$. It is a fairly straightforward calculation to bound the expected value of the denominator.

Lemma 3.3.5.

$$\mathbb{E}\left[\sum_i d_i f_i\right] = 1\,.$$

Bounding the expected value of the numerator is more subtle. We show the following bound on the expected value of the numerator.

Lemma 3.3.6.

$$\mathbb{E}\left[\sum_{i\sim j} w\left(\{i,j\}\right)|f_i - f_j|\right] \leqslant \mathcal{O}\left(\sqrt{\mathcal{R}\left(u\right)\log k}\right)\,.$$

Notice that the ratio of their expected values is $\mathcal{O}(\sqrt{\mathcal{R}\left(u\right)\log k})$, as intended. To control the ratio of the two quantities, the numerator is to be bounded from above, and the denominator is to be bounded from below. A simple Markov inequality can be used to upper bound the probability that the numerator is much larger than its expectation. To control the denominator, we bound its variance. Specifically, we will show the following bound on the variance of the denominator.

Lemma 3.3.7.

$$\mathsf{Var}\left[\sum_i d_i f_i\right] \leqslant 1\,.$$

The above moment bounds are sufficient to conclude that with constant probability, the ratio

$$\frac{\sum_{i\sim j}|f_i - f_j|}{\sum_i d_i f_i} = \mathcal{O}\left(\sqrt{\mathcal{R}\left(u\right)\log k}\right)\,.$$

Therefore, with constant probability over the choice of the Gaussians $g_1, \ldots, g_k$, $\Omega(k)$ of the vectors $h_1, \ldots h_k$ yield sets of expansion $\mathcal{O}(\sqrt{\mathcal{R}\left(u\right)\log k})$.

3.3.3.2 *Main Proofs*

Let f denote the vector h_1. The choice of the index 1 is arbitrary and the same analysis is applicable to all other indices $i \in [k]$. We first separately bound the expectations of the numerator and denominator of the sparsity of each cut, and then the variance of the denominator. The proofs of these bounds will follow their application in the proof of our main theorem.

Expectation of the Denominator. Bounding the expectation of the denominator is a straightforward calculation as shown below.

Lemma 3.3.8 (Restatement of Lemma 3.3.5)**.**

$$\mathbb{E}\left[\sum_i d_i f_i\right] = 1\,.$$

Proof of Lemma 3.3.5. For any $i \in [n]$, recall that

$$f_i = \begin{cases} \|u_i\|^2 & \text{if } \langle \tilde{u}_i, g_1 \rangle \geqslant \langle \tilde{u}_i, g_j \rangle \ \ \forall j \in [k] \\ 0 & \text{otherwise} \end{cases}.$$

The first case happens with probability $1/k$ and so $f_i = 0$ with the remaining probability. Therefore

$$\mathbb{E}\left[\sum_i d_i f_i\right] = \sum_i d_i \frac{1}{k} \|u_i\|^2 = 1\,.$$

□

Expectation of the Numerator. For bounding the expectation of the numerator we will need a lemma that is a direct consequence of Lemma 2.5.7 about the maximum of k i.i.d normal random variables.

Corollary 3.3.9. *For any $i, j \in [n]$,*

$$\mathbb{P}\left[f_i > 0 \text{ and } f_j = 0\right] \leqslant c_1 \left(\|\tilde{u}_i - \tilde{u}_j\| \frac{\sqrt{\log k}}{k} \right).$$

The following lemma is the main technical lemma in bounding the expected value of the numerator.

Lemma 3.3.10. *For any indices $i, j \in [n]$,*

$$\mathbb{E}\left[|f_i - f_j|\right] \leqslant \frac{(2c_1+1)\sqrt{\log k}}{k} \|u_i - u_j\| \left(\|u_i\| + \|u_j\|\right)$$

Proof.

$$\begin{aligned}
\mathbb{E}\left[|f_i - f_j|\right] &= \|u_i\|^2 \, \mathbb{P}\left[f_i > 0 \text{ and } f_j = 0\right] + \|u_j\|^2 \, \mathbb{P}\left[f_j > 0 \text{ and } f_i = 0\right] \\
&\quad + \left| \|u_i\|^2 - \|u_j\|^2 \right| \mathbb{P}\left[f_i, f_j > 0\right] \\
&\leqslant c_1 \left(\|\tilde{u_i} - \tilde{u_j}\| \frac{\sqrt{\log k}}{k} \right) \left(\|u_i\|^2 + \|u_j\|^2\right) + \left| \|u_i\|^2 - \|u_j\|^2 \right| \frac{1}{k} \\
&\qquad \text{(Using Corollary 3.3.9)} \\
&\leqslant \frac{2c_1\sqrt{\log k}}{k} \|u_i - u_j\| \sqrt{\|u_i\|^2 + \|u_j\|^2} + \frac{1}{k} \left|\langle u_i - u_j, u_i + u_j \rangle\right| \\
&\qquad \text{(Using Proposition 2.6.2)} \\
&\leqslant \frac{2c_1\sqrt{\log k}}{k} \|u_i - u_j\| \left(\|u_i\| + \|u_j\|\right) + \frac{1}{k} \|u_i - u_j\| \left(\|u_i\| + \|u_j\|\right) \\
&\qquad \text{(Using the Cauchy-Schwarz inequality)}
\end{aligned}$$

□

We are now ready to bound the expectation of the numerator, we restate the lemma for the convenience of the reader.

Lemma 3.3.11. *(Restatement of Lemma 3.3.6)*

$$\mathbb{E}\left[\sum_{i \sim j} w\left(\{i, j\}\right) |f_i - f_j|\right] \leqslant 2(2c_1 + 1)\sqrt{\mathcal{R}\left(u\right) \log k}\,.$$

Proof of Lemma 3.3.6.

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{i\sim j} w\left(\{i,j\}\right)|f_i - f_j|\right] \\
&\leqslant \frac{(2c_1+1)\sqrt{\log k}}{k}\sum_{i\sim j} w\left(\{i,j\}\right)\|u_i - u_j\|\left(\|u_i\| + \|u_j\|\right) \qquad \text{(Lemma 3.3.10)} \\
&\leqslant \frac{(2c_1+1)\sqrt{\log k}}{k}\sqrt{\sum_{i\sim j} w\left(\{i,j\}\right)\|u_i - u_j\|^2}\sqrt{\sum_{i\sim j} w\left(\{i,j\}\right)\left(\|u_i\| + \|u_j\|\right)^2} \\
&\qquad\qquad \text{(Using the Cauchy-Schwarz inequality)} \\
&\leqslant \frac{(2c_1+1)\sqrt{\log k}}{k}\sqrt{\mathcal{R}\left(u\right)\cdot\left(\sum_i d_i\|u_i\|^2\right)}\sqrt{\sum_{i\sim j} w\left(\{i,j\}\right)2\left(\|u_i\|^2 + \|u_j\|^2\right)} \\
&\leqslant \frac{2(2c_1+1)\sqrt{\log k}}{k}\sqrt{\mathcal{R}\left(u\right)}\left(\sum_i d_i\|u_i\|^2\right) \\
&= 2(2c_1+1)\sqrt{\mathcal{R}\left(u\right)\log k} \qquad \left(\text{Using } \sum_i d_i\|u_i\|^2 = k\ \right)
\end{aligned}
$$

□

Variance of the Denominator. Here too we will need some groundwork. Let $\mathcal{G}$ denote the Gaussian space. The Hermite polynomials $\{H_i\}_{i\in\mathbb{Z}_{\geqslant 0}}$ form an orthonormal basis for real valued functions over the Gaussian space $\mathcal{G}$, i.e., $\mathbb{E}_{g\in\mathcal{G}}[H_i(g)H_j(g)] = 1$ if $i = j$ and 0 otherwise. The k-wise tensor product of the Hermite basis forms an orthonormal basis for functions over $\mathcal{G}^k$. Specifically, for each $\alpha \in \mathbb{Z}^k_{\geqslant 0}$ define the polynomial H_α as

$$H_\alpha(x_1,\ldots,x_k) = \prod_{i=1}^{k} H_{\alpha_i}(x_i).$$

The functions $\{H_\alpha\}_{\alpha\in\mathbb{Z}^k_{\geqslant 0}}$ form an orthonormal basis for functions over $\mathcal{G}^k$. The degree of the polynomial $H_\alpha(x)$ denote by $|\alpha|$ is $|\alpha| = \sum_i \alpha_i$.

The Hermite polynomials form a complete eigenbasis for the noise operator on the Gaussian space (Ornstein-Uhlenbeck operator). In particular, they are known to satisfy the following property (see e.g. the book of Ledoux and Talagrand [62], Section 3.2).

Fact 3.3.12. *Let $(g_i, h_i)_{i=1}^k$ be k independent samples from two ρ-correlated Gaussians, i.e., $\mathbb{E}[g_i^2] = \mathbb{E}[h_i^2] = 1$, and $\mathbb{E}[g_i h_i] = \rho$. Then for all $\alpha \in \mathbb{Z}_{\geqslant 0}^k$,*

$$\mathbb{E}[H_\alpha(g_1, \ldots, g_k) H_{\alpha'}(h_1, \ldots, h_k)] = \rho^{|\alpha|} \text{ if } \alpha = \alpha' \text{ and } 0 \text{ otherwise}$$

Let $B : \mathcal{G}^k \longrightarrow \mathbb{R}$ be the function defined as follows,

$$B(x) = \begin{cases} 1 & \text{if } (x_1 \geqslant x_i \ \forall i \in [k]) \text{ or } (x_1 \leqslant x_i \ \forall i \in [k]) \\ 0 & \text{otherwise} \end{cases} .$$

Then

$$\mathbb{E}[B] = \mathbb{E}[B^2] = \frac{1}{k} .$$

Lemma 3.3.13. *Let u, v be unit vectors and $g_1, \ldots, g_k$ be i.i.d Gaussian vectors. Then,*

$$\mathbb{E}[B(\langle u, g_1 \rangle, \ldots, \langle u, g_k \rangle) B(\langle v, g_1 \rangle, \ldots, \langle v, g_k \rangle)] \leqslant \frac{1}{k^2} + \langle u, v \rangle^2 \frac{1}{k} .$$

Proof. The function B on the Gaussian space can be written in the Hermite expansion $B(x) = \sum_\alpha B_\alpha H_\alpha(x)$ such that

$$\sum_\alpha B_\alpha^2 = \mathbb{E}[B^2] = \frac{1}{k} .$$

Using Fact 3.3.12, we can write

$$\mathbb{E}[B(\langle u, g_1 \rangle, \ldots, \langle u, g_k \rangle) B(\langle v, g_1 \rangle, \ldots, \langle v, g_k \rangle)] = (\mathbb{E}[B])^2 + \sum_{\alpha \in \mathbb{Z}_{\geqslant 0}^k, |\alpha| > 0} B_\alpha^2 \rho^{|\alpha|}$$

where $\rho = \langle u, v \rangle$. Since B is an even function, only the even degree coefficients are non-zero, i.e., $B_\alpha = 0$ for all $|\alpha|$ odd. Along with $\rho \leqslant 1$, this implies that

$$\begin{aligned}\mathbb{E}[B(\langle u, g_1 \rangle, \ldots, \langle u, g_k \rangle) B(\langle v, g_1 \rangle, \ldots, \langle v, g_k \rangle)] &\leqslant (\mathbb{E}[B])^2 + \rho^2 \left(\sum_{\alpha, |\alpha| \geqslant 2} B_\alpha^2 \right) \\ &= \frac{1}{k^2} + \langle u, v \rangle^2 \frac{1}{k} .\end{aligned}$$

□

Next we bound the variance of the denominator.

Proof of Lemma 3.3.7.

$$\mathbb{E}\left[\sum_{i,j} d_i d_j f_i f_j\right]$$
$$= \sum_{i,j} d_i d_j \|u_i\|^2 \|u_j\|^2 \, \mathbb{E}\left[\frac{f_i}{\|u_i\|^2}\frac{f_j}{\|u_j\|^2}\right]$$
$$\leqslant \sum_{i,j} d_i d_j \|u_i\|^2 \|u_j\|^2 \, \mathbb{E}\left[B(\langle \tilde{u}_i, g_1\rangle, \ldots, \langle \tilde{u}_i, g_k\rangle) B(\langle \tilde{u}_j, g_1\rangle, \ldots, \langle \tilde{u}_j, g_k\rangle)\right]$$
$$\leqslant \sum_{i,j} d_i d_j \|u_i\|^2 \|u_j\|^2 \cdot \left(\frac{1}{k^2} + \frac{1}{k}\langle \tilde{u}_i, \tilde{u}_j\rangle^2\right) \qquad (\textit{Lemma } 3.3.13)$$
$$= \left(\frac{1}{k}\sum_{i,j} d_i d_j \langle u_i, u_j\rangle^2 + \frac{1}{k^2}\left(\sum_i d_i \|u_i\|^2\right)^2\right)$$
$$= \left(\frac{1}{k}\cdot k + \frac{1}{k^2}\cdot k^2\right) \qquad (\text{Using Lemma 3.3.1 })$$
$$\leqslant 2\,.$$

Therefore

$$\mathsf{Var}\left[\sum_i d_i f_i\right] = \mathbb{E}\left[\sum_{i,j} d_i d_j f_i f_j\right] - \left(\mathbb{E}\left[\sum_i d_i f_i\right]\right)^2 \leqslant 1\,.$$

□

Putting It Together

Proof of Theorem 3.3.3. For each $l \in [k]$, from Lemma 3.3.5 and Lemma 3.3.7 we get that

$$\mathbb{E}\left[\sum_i d_i h_l(i)\right] = 1 \qquad \text{and} \qquad \mathsf{Var}\left[\sum_i d_i h_l(i)\right] \leqslant 1\,.$$

Therefore, from the One-sided Chebyshev inequality (Fact 2.5.1), we get

$$\mathbb{P}\left[\sum_i d_i h_l(i) \geqslant \frac{1}{2}\right] \geqslant \frac{\left(\frac{\mathbb{E}\left[\sum_i d_i h_l(i)\right]}{2}\right)^2}{\left(\frac{\mathbb{E}\left[\sum_i d_i h_l(i)\right]}{2}\right)^2 + \mathsf{Var}\left[\sum_i d_i h_l(i)\right]} \geqslant c' \qquad (1)$$

where c' is some absolute constant. Therefore, with constant probability, for $\Omega(k)$ indices $l \in [k]$, $\sum_i d_i h_l(i) \geqslant 1/2$. Next, for each l, using Markov's inequality

$$\mathbb{P}\left[\sum_{i\sim j} w\left(\{i,j\}\right)\left|h_l(i) - h_l(j)\right| \leqslant 2/c'\, \mathbb{E}\left[\sum_{i\sim j} w\left(\{i,j\}\right)\left|h_l(i) - h_l(j)\right|\right]\right] \geqslant 1 - c'/2\,. \tag{2}$$

Therefore, with constant probability, for a constant fraction, say c, of the indices $l \in [k]$, we have

$$\frac{\sum_{i\sim j} w\left(\{i,j\}\right)\left|h_l(i) - h_l(j)\right|}{\sum_i d_i h_l(i)} \leqslant \frac{4}{c'} \frac{\mathbb{E}\left[\sum_{i\sim j} w\left(\{i,j\}\right)\left|h_l(i) - h_l(j)\right|\right]}{\mathbb{E}\left[\sum_i d_i h_l(i)\right]} = C\sqrt{\mathcal{R}\left(u\right)\log k}$$

for some constant C. Applying Lemma 2.3.2 on the vectors with those indices will give ck disjoint sets $S_1, \ldots, S_{ck}$ such that $\phi_G(S_i) = C\sqrt{\lambda_k \log k}\ \forall i \in [ck]$. This completes the proof of Theorem 3.3.3. □

3.4 *Lower bound for* k *Sparse-Cuts*

In this section, we prove a lower bound for the k sparse-cuts in terms of higher eigenvalues (Proposition 3.1.4) thereby generalizing the lower bound side of the Cheeger's inequality.

Proposition 3.4.1 (Restatement of Proposition 3.1.4)**.** *For any edge-weighted graph $G = (V, E)$, for any integer $1 \leqslant k \leqslant n$, and for any k disjoint subsets $S_1, \ldots, S_k \subset V$*

$$\max_i \phi_G(S_i) \geqslant \frac{\lambda_k}{2}\,.$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of the normalized Laplacian of G.

Proof. Let α denote $\max_i \phi_G(S_i)$. Let $T \stackrel{\text{def}}{=} V \backslash (\cup_i S_i)$. Let G' be the graph obtained by shrinking each piece in the partition $\{T, S_i : i \in [k]\}$ of V to a single vertex. We denote the vertex corresponding to S_i by $s_i\ \forall i$ and the vertex corresponding to T by

t. Let $\mathcal{L}'$ be the normalized Laplacian matrix corresponding to G'. Note that, by construction, the expansion of every set in G' not containing t is at most α.

Let $U \stackrel{\text{def}}{=} \left\{D^{\frac{1}{2}} X_{S_i} : i \in [k]\right\}$. Here X_S is the incidence vector of the set $S \subset V$. Since all the vectors in U are orthogonal to each other, we have

$$\lambda_k = \min_{S:\mathsf{rank}(S)=k} \max_{X \in S} \frac{X^T \mathcal{L} X}{X^T X} \leqslant \max_{X \in \mathsf{span}(U)} \frac{X^T \mathcal{L} X}{X^T X} = \max_{Y \in \mathbb{R}^k * \{0\}} \frac{\sum_{i,j} w'(\{i,j\})(Y_i - Y_j)^2}{\sum_i d'_i Y_i^2} .$$

For any $x \in \mathbb{R}$, let $x^+ \stackrel{\text{def}}{=} \max\{x, 0\}$ and $x^- \stackrel{\text{def}}{=} \max\{-x, 0\}$. Then it is easily verified that for any $Y_i, Y_j \in \mathbb{R}$,

$$(Y_i - Y_j)^2 \leqslant 2((Y_i^+ - Y_j^+)^2 + (Y_i^- - Y_j^-)^2) .$$

Therefore,

$$\begin{aligned}
&\sum_i \sum_{j>i} w'(\{i,j\})(Y_i - Y_j)^2 \\
&\leqslant 2\left(\sum_i \sum_{j>i} w'(\{i,j\})(Y_i^+ - Y_j^+)^2 + \sum_i \sum_{j>i} w'(\{i,j\})(Y_j^- - Y_i^-)^2\right) \\
&\leqslant 2\left(\sum_i \sum_{j>i} w'(\{i,j\})\left|(Y_i^+)^2 - (Y_j^+)^2\right| + \sum_i \sum_{j>i} w'(\{i,j\})\left|(Y_j^-)^2 - (Y_i^-)^2\right|\right) .
\end{aligned}$$

Without loss of generality, we may assume that $Y_1^+ \geqslant Y_2^+ \geqslant \ldots \geqslant Y_k^+ \geqslant Y_t = 0$. Let $T_i = \{s_1, \ldots, s_i\}$ for each $i \in [k]$. Therefore, we have

$$\begin{aligned}
\sum_i \sum_{j>i} w'(\{i,j\})\left|(Y_i^+)^2 - (Y_j^+)^2\right| &\leqslant \sum_{i=1}^{k}((Y_i^+)^2 - (Y_{i+i}^+)^2) w'(E(T_i, \bar{T}_i)) \\
&\leqslant \alpha \sum_{i=1}^{k}((Y_i^+)^2 - (Y_{i+i}^+)^2) w'(T_i) \\
&= \alpha \sum_i d'_i (Y_i^+)^2 .
\end{aligned}$$

Here we are using the fact that $w'(E(T_i, \bar{T}_i)) \leqslant \alpha w'(T_i)$ which follows from the definition of α and that $w'(T_{i+1}) - w'(T_i) = d'_{i+1}$. Similiarly, we get that

$$\sum_i \sum_{j>i} w'(\{i,j\})\left|(Y_j^-)^2 - (Y_i^-)^2)\right| \leqslant \alpha \sum_i d'_i (Y_i^-)^2 .$$

Putting these two inequalities together we get that

$$\sum_{j>i} w'(\{i,j\})(Y_i - Y_j)^2 \leqslant 2\alpha \sum_i d'_i Y_i^2 \ .$$

Therefore, $\lambda_k(\mathcal{L}) \leqslant 2 \max_i \phi_G(S_i)$. □

3.5 Gap examples

In this section, we present constructions of graphs that serve as lower-bounds against natural classes of algorithms. We begin with a family of graphs on which the performance of recursive partitioning algorithms is poor for the k-Sparse cuts problem.

3.5.1 Recursive Algorithms

Recursive algorithms are one of most commonly used techniques in practice for graph multi-partitioning. However, we show that partitioning a graph into k pieces using a simple recursive algorithm can yield as many $k(1-o(1))$ sets with expansion much larger than $\mathcal{O}\left(\sqrt{\lambda_k}\,\mathsf{polylog}\,k\right)$. Thus this is not an effective method for finding many sparse cuts.

The following construction (Figure 3) shows that partition of V obtained using the recursive algorithm in Algorithm 3.2.1 can give as many as $k(1-o(1))$ sets have expansion $\Omega(1)$ while $\lambda_k \leqslant \mathcal{O}(k^2/n^2)$.

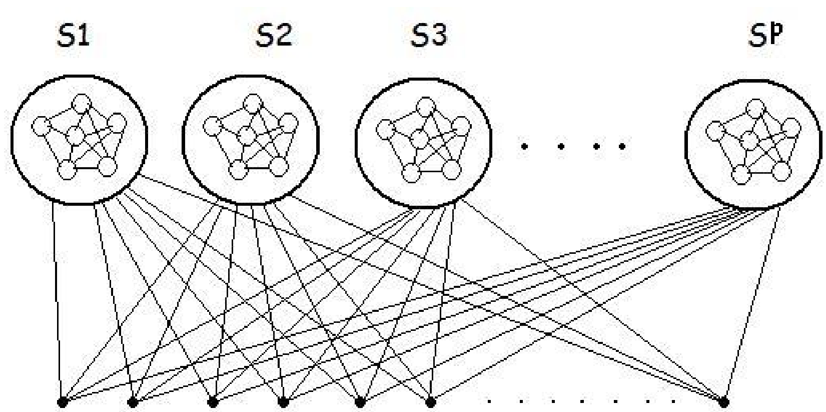

Figure 3: Recursive algorithm can give many sets with very small expansion

In this graph, there are $p \stackrel{\text{def}}{=} k^{\varepsilon}$ sets S_i for $1 \leqslant i \leqslant k^{\varepsilon}$. We will fix the value of ε later. Each of the S_i has $k^{1-\varepsilon}$ cliques $\{S_{ij} : 1 \leqslant j \leqslant k^{1-\varepsilon}\}$ of size n/k which are sparsely connected to each other. The total weight of the edges from S_{ij} to $S_i \backslash S_{ij}$ is equal to a constant c. In addition to this, there are also $k - k^{\varepsilon}$ vertices $v_i : 1 \leqslant i \leqslant k - k^{\varepsilon}$. The weight of edges from S_i to v_j is equal to $k^{-\varepsilon}$.

Claim 3.5.1. *1.* $\phi(S_{ij}) \leqslant (c+1)k^2/n^2 \; \forall i, j$

2. $\phi(S_i) \leqslant 1/(c+1)\phi(S_{ij}) \; \forall i, j$

3. $\lambda_k = \mathcal{O}(k^2/n^2)$

Proof. 1.

$$\phi(S_{ij}) = \frac{c + \frac{(k-k^{\varepsilon})k^{-\varepsilon}}{k^{1-\varepsilon}}}{(\frac{n}{k})^2 + c + \frac{(k-k^{\varepsilon})k^{-\varepsilon}}{k^{1-\varepsilon}}} \leqslant \frac{(c+1)k^2}{n^2}$$

2. $w(S_i) = \sum_j w(S_{ij})$, but for each S_{ij} only $1/(c+1)$ fraction of edges incident at S_{ij} are also incident at S_i. Therefore, $\phi(S_i) \leqslant 1/(c+1)\phi(S_{ij})$.

3. Follows from (1) and Proposition 3.1.4.

□

For appropriate values of ε and k, the partition output by the recursive algorithm will be $\{S_i : i \in [k^{\varepsilon}]\} \cup \{v_i : i \in [k - k^{\varepsilon}]\}$. Hence, $k(1 - o(1))$ sets have expansion equal to 1.

3.5.2 k-partition

In this section, we give a constructive proof of Theorem 3.1.8, i.e., we construct a family of graphs such that for any k-partition $\{S_1, \ldots, S_k\}$ of the graph, $\max_i \phi(S_i) > \Theta(k^2 \sqrt{\frac{p}{n}})$. We view this as further evidence suggesting that the k sparse-cuts problem is the right generalization of sparsest cut.

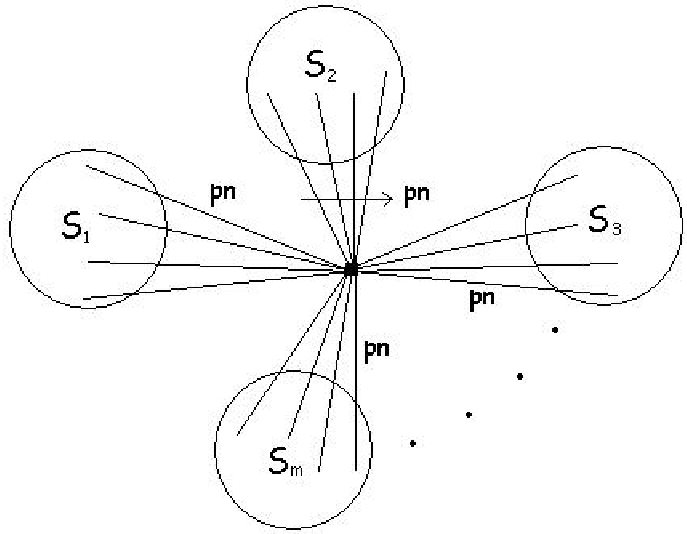

Figure 4: k-partition can have sparsity much larger than $\Omega(\sqrt{\lambda_k}\mathsf{polylog}k)$

Lemma 3.5.2. *For the graph G in Figure 4, and for any k-partition $S_1, \ldots, S_k$ of its vertex set,*

$$\frac{\max_i \phi_G(S_i)}{\sqrt{\lambda_k}} = \Theta\left(k^2\sqrt{\frac{p}{n}}\right) .$$

Proof. In Figure 4, $\forall i \in [k]$, S_i is a clique of size $(n-1)/k$ (pick n so that $k|(n-1)$). There is an edge from central vertex v to every other vertex of weight pn. Here p is some absolute constant. Let $\mathcal{P}' \stackrel{\text{def}}{=} \{S_1 \cup \{v\}, S_2, S_3, \ldots, S_k\}$. For $n > k^3$, it is easily verified that the optimum k-partition is isomorphic to $\mathcal{P}'$. For $k < o(n^{\frac{1}{3}})$, we have

$$\max_{S_i \in \mathcal{P}'} \phi_G(S_i) = \phi_G(S_1 \cup \{v\}) = \frac{pnk}{\left(\frac{n-1}{k}\right)^2 + pnk} = \Theta\left(\frac{pk^3}{n}\right)$$

Applying Proposition 3.1.4 to $S_1, \ldots, S_k$, we get that $\lambda_k = \mathcal{O}(pk^2/n)$. Thus we have the lemma. □

3.6 Conclusion

We exhibited new connections between higher eigenvalues of the graph Laplacian and higher order graph partitions à la Cheeger's Inequality. Crucial to our proofs, is a new bound on the covariance of two truncated Gaussian random variables (Lemma 3.3.13).

A natural question to ask is if our bounds are optimal? We show that our bounds for k sparse-cuts is tight upto constant factors in the number of sets, and our bound for Small Set Expansion is tight upto constant factors in the size of the set. We prove an upper bound of $\mathcal{O}\left(\sqrt{\lambda_k}\log k\right)$ for Min Sum k-partition, however we do not know if this is tight (the corresponding factor for the Gaussian graph is $\Theta\left(\sqrt{\lambda_k \log k}\right)$). We leave these questions as open problems.

Problem 3.6.1. Does every graph $G = (V, E)$, for every parameter $k \in [n]$ have k disjoint non-empty subsets, say $S_1, \ldots, S_k$, such that

$$\max_{i \in [k]} \phi(S_i) \leqslant \mathcal{O}\left(\sqrt{\lambda_k \log k}\right)?$$

Problem 3.6.2. Does every graph $G = (V, E)$, for every parameter $k \in [n]$ have a k-partition, say $S_1, \ldots, S_k$, such that

$$\phi^{\mathsf{sum}}\left(\{S_1, \ldots, S_k\}\right) \leqslant \mathcal{O}\left(\sqrt{\lambda_k \log k}\right)?$$

Acknowledgements. The results in this chapter were obtained in joint work with Prasad Raghavendra, Prasad Tetali and Santosh Vempala [72, 73].

CHAPTER IV

SPECTRAL PROPERTIES OF HYPERGRAPHS

4.1 Introduction

There is a rich spectral theory of graphs, based on studying the eigenvalues and eigenvectors of the adjacency matrix (and other related matrices) of graphs. Cheeger's Inequality and its many (minor) variants have played a major role in the design of algorithms as well as in understanding the limits of computation [96, 97, 38, 13, 8]. We refer the reader to [46] for a comprehensive survey.

It has remained open to define a spectral model of hypergraphs, whose spectra can be used to estimate hypergraph parameters à la Spectral Graph Theory. Hypergraph expansion and related hypergraph partitioning problems are of immense practical importance, having applications in parallel and distributed computing [25], VLSI circuit design and computer architecture [52, 42], scientific computing [36] and other areas. Inspite of this, there hasn't been much theoretical work on them (see Section 4.1.1). Spectral graph partitioning algorithms are widely used in practice for their efficiency and the high quality of solutions that they often provide [18, 44]. 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 $H = (V, E, w)$ on the vertex set V and having edge set $E \subseteq V^r$, one can define the canonical r-tensor form A as follows.

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

This tensor form and its minor variants have been explored in the literature (see Section 4.1.1 for a brief survey), but have not been understood very well. Optimizing over tensors is NP-hard [45]; even getting good approximations might be intractable [24]. The spectral properties of tensors seem to be unrelated to combinatorial properties of hypergraphs (See Section 4.8).

Another way to study a hypergraph, say $H = (V, E, w)$, is to replace each hyperedge $e \in E$ by complete graph or a low degree expander on the vertices of e to obtain a 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 can not 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 8.0.4).

In this chapter, we define a new Markov operator for hypergraphs, obtained by generalizing the random-walk operator on graphs. Our operator is simple and 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 4.2 (See Definition 4.2.1). We present our main results about this hypergraph operator in Section 4.2.1 and Section 4.2.3. 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 graphs.

4.1.1 Related Work

Freidman and Wigderson [40] study the canonical tensors of hypergraphs. They bound the second eigenvalue of such tensors for hypergraphs drawn randomly from various distributions and show their connections to randomness dispersers. Rodriguez [90] studies the eigenvalues of 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 [34] study the roots of the characteristic polynomial of hypergraphs and relate it to its chromatic number. [47, 48] also study the canonical tensor form of the hypergraph and relate its eigenvectors to some configured components of that hypergraph. Lenz and Mubayi [65, 66, 67] relate the eigenvector corresponding to the second largest eigenvalue of the canonical tensor to hypergraph quasi-randomness. Chung [32] defines a notion of Laplacians for hypergraphs and studies the relationship between its eigenvalues and a very different notion of hypergraph cuts and homologies. [84, 99, 83, 82] study the relation of hypergraphs to rather different notion of Laplacian forms and prove isoperimetric inequalities, study homologies and mixing times.

Peres et. al.[85] study a "tug of war" Laplacian operator on graphs that is similar to our hypergraph Markov 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 to for analyzing the rate of convergence of local dynamics in bargaining networks [26].

4.2 The Hypergraph Markov Operator

We now formally define the hypergraph Markov operator $M : \mathbb{R}^n \to \mathbb{R}^n$ (Definition 4.2.1). For a hypergraph H, we denote its Markov operator by M_H. We drop the subscript whenever the hypergraph is clear from the context. We note that unlike

Definition 4.2.1 (The Hypergraph Markov Operator)**.**
Given a vector $X \in \mathbb{R}^n$, $M(X)$ is computed as follows.

1. For each hyperedge $e \in E$, let $(i_e, j_e) := \mathsf{argmax}_{i,j \in e} |X_i - X_j|$, breaking ties arbitrarily (See Remark 4.5.2).

2. We now construct the weighted graph G_X on the vertex set V as follows. We add edges $\{\{i_e, j_e\} : e \in E\}$ having weight $w(\{i_e, j_e\}) := w(e)$ to G_X. Next, to each vertex v we add self-loops of sufficient weight such that its degree in G_X is equal to d_v; more formally we add self-loops of weight
$$w(\{v, v\}) := d_v - \sum_{e \in E : v \in \{i_e, j_e\}} w(e) \,.$$

3. We define A_X to be the random walk matrix of G_X, i.e., A_X is obtained from the adjacency matrix of G_X by dividing the entries of the i^{th} row by the degree of vertex i in G_X.

Then,
$$M(X) \stackrel{\text{def}}{=} A_X X \,.$$

Figure 5: The hypergraph Marko Operator

most of spectral models for hypergraphs considered in the literature, our Markov operator M does not require the hypergraph to be uniform (i.e. it does not require all hyperedges to have the same number of vertices in them).

Remark 4.2.2. Let G_X denote the adjacency matrix of the graph in Definition 4.2.1. Then, by construction, $A_X = D^{-1} G_X$, where D is the diagonal matrix whose $(i,i)^{th}$ entry is d_i. We will often study $D^{-1/2} G_X D^{-1/2}$ in the place of studying $D^{-1} G_X$ (see Remark 2.0.2).

Definition 4.2.3 (Hypergraph Laplacian)**.** Given a hypergraph H, we define its Laplacian operator L as
$$L \stackrel{\text{def}}{=} I - M \,.$$

Here, I is the identity operator and M is the hypergraph Markov operator. The action of L on a vector X is $L(X) \stackrel{\text{def}}{=} X - M(X)$. We define the matrix $L_X \stackrel{\text{def}}{=} I - A_X$ (See Remark 4.2.2). We define the *Rayleigh quotient* $\mathcal{R}(\cdot)$ of a vector X as

$$\mathcal{R}(X) \stackrel{\text{def}}{=} \frac{X^T L(X)}{X^T X} .$$

Our definition of M is inspired by the ∞-Harmonic functions studied by [85]. We note that M is a generalization of the random-walk matrix for graphs to hypergraphs; if all hyperedges had exactly two vertices, then $\{i_e, j_e\} = e$ for each hyperedge e and M would be the random-walk matrix (i.e. the normalized adjacency matrix).

Let us consider the special case when the hypergraph $H = (V, E, w)$ is d-regular. We can also view the operator M as a collection of maps $\{f_r : \mathbb{R}^r \to \mathbb{R}^r\}_{r \in \mathbb{Z}_{\geqslant 0}}$ as follows. We define the action of f_r on a tuple $(x_1, \ldots, x_r)$ as follows. It picks the coordinates $i, j \in [r]$ which have the highest and the lowest values respectively. Then it decreases the value at the i^{th} coordinate by $(x_i - x_j)/d$ and increases the value at the j^{th} coordinate by $(x_i - x_j)/d$, whereas all other coordinates remain unchanged. For a vector $X \in \mathbb{R}^n$, the computation of $M(X)$ in Definition 4.2.1 can be viewed as applying these maps to X, where for each hyperedge $e \in E$, $f_{|e|}$ is applied to the tuple corresponding to the coordinates of X represented by the vertices in e.

Comparison to other operators. A natural question to ask is if any other set of maps, say $\{g_r : \mathbb{R}^r \to \mathbb{R}^r\}_{r \in \mathbb{Z}_{\geqslant 0}}$, used in this manner gives a 'better' Markov operator? A natural set of maps that one would be tempted to try are the *averaging* maps which map an r-tuple $(x_1, \ldots, x_r)$ to $(\sum_i x_i/r, \ldots, \sum_i x_i/r)$.

If we consider the embedding of the vertices of a hypergraph $H = (V, E, w)$ on $\mathbb{R}$, given by the vector $X \in \mathbb{R}^V$, then the length $l(\cdot)$ of a hyperedge $e \in E$ is $l(e) \stackrel{\text{def}}{=} \max_{i,j \in e} |X_i - X_j|$. We believe that $l(e)$ is the most essential piece of information about the hyperedge e. As a motivating example, consider the special case when all the entries of X are in $\{0,1\}$. In this case, the vector X defines a cut $(S,\bar{S})$, where $S = \mathsf{supp}(X)$, and the $l(e)$ indicates whether e is cut by S or not. Building on this idea, we can use the average length of edges to bound expansion of sets. We will be studying the length of the hyperedges in the proofs of all the results in this chapter. A well known fact from Statistical Information Theory is that moving in the direction of ∇l will yield the most *information* about the function in question. We refer the reader to [81, 19, 92] for the formal statement and proof of this fact, and for a comprehensive discussion on this topic. Our set of maps move a tuple precisely in the direction of ∇l, thereby achieving this goal.

For an hyperedge $e \in E$ the averaging maps will yield information about the function $\mathbb{E}_{i,j\in e}|X_i - X_j|$ and not about $l(e)$. In particular, the averaging maps will have a gap of factor $\Omega(r)$ between the hypergraph expansion[1] and the square root spectral gap[2] of the operator. In general, if a set of maps changes r' out of r coordinates, it will have a gap of $\Omega(r')$ between hypergraph expansion and the square root of the spectral gap.

Our set of maps $\{f_r\}_{r\in\mathbb{Z}_{\geqslant 0}}$ are the very natural *greedy* maps which bring the pair of coordinates which are farthest apart slightly closer to each other. Let us consider the continuous dispersion process where we repeatedly apply the markov operator $((1-\mathsf{dt})I + \mathsf{dt}\, M)$ (for an infinitesimally small value of dt) to an arbitrary starting probability distribution on the vertices (see Definition 4.2.9). In the case when the maximum value (resp. minimum value) in the r-tuple is much higher (resp. much lower) than the second maximum value (resp. second minimum value), then these set of greedy maps are essentially the best we can hope for, as they will lead to the greatest decrease in variance of the values in the tuple. In the case when the maximum value (resp. minimum value) in the tuple, located at some coordinate $i_1 \in [r]$ is close to the

[1] See Definition 2.1.5.

[2] The spectral gap of a Laplacian operator is defined as its second smallest eigenvalue. See Definition 4.2.7 for the definition of eigenvalues of the Markov operator M.

second maximum value (resp. second minimum value), located at some coordinate $i_2 \in [r]$, the dispersion process is likely to decrease the value at coordinate i_1 till it equals the value at coordinate i_2 after which these two coordinates will decrease at the same rate (see Section 4.5 and Remark 4.5.2). Therefore, our set of greedy maps addresses all cases satisfactorily.

4.2.1 Hypergraph Eigenvalues

As in the case of graphs, it is easy to see that the hypergraph Laplacian operator is positive semidefinite.

Proposition 4.2.4. *Given a hypergraph H and its Laplacian operator L,*

$$X^T L(X) \geqslant 0 \qquad \forall X \in \mathbb{R}^n .$$

Proof. $X^T L(X) = X^T (I - A_X) X$. Since A_X is a random-walk matrix, $I - A_X \succeq 0$. Hence, the proposition follows. □

Stationary Distribution. A probability distribution μ on V is said to be *stationary* if $M(\mu) = \mu$. We define the probability distribution μ^* as follows.

$$\mu^*(i) = \frac{d_i}{\sum_{j \in V} d_j} \qquad \text{for } i \in V .$$

μ^* is a stationary distribution of M, as it is an eigenvector with eigenvalue 1 of A_X $\forall X \in \mathbb{R}^n$.

Laplacian Eigenvalues. An operator L is said to have an eigenvalue $\lambda \in \mathbb{R}$ if for some vector $X \in \mathbb{R}^n$, $L(X) = \lambda X$. It follows from the definition of L that λ is an eigenvalue of L if and only if $1 - \lambda$ is an eigenvalue of M. In the case of graphs, the Laplacian Matrix and the adjacency matrix have n orthogonal eigenvectors. However for hypergraphs, the Laplacian operator L (respectively M) is a highly non-linear operator. In general non-linear operators can have a lot more more than n eigenvalues or a lot fewer than n eigenvalues.

From the definition of stationary distribution we get that μ^* is an eigenvector of M with eigenvalue 1. Therefore, μ^* is an eigenvector of L with eigenvalue 0.

We start by showing that L has at least one non-trivial eigenvalue.

Theorem 4.2.5. *Given a hypergraph H, there exists a vector $v \in \mathbb{R}^n$ and a $\lambda \in \mathbb{R}$ such that $\langle v, \mu^* \rangle = 0$ and $L(v) = \lambda\, v$.*

Given that a non-trivial eigenvector exists, we can define the second smallest eigenvalue γ_2 as the smallest eigenvalue from Theorem 4.2.5. We define v_2 to be the corresponding eigenvector.

It is not clear if L has any other eigenvalues. We again remind the reader that in general, non-linear operators can have very few eigenvalues or sometimes even have no eigenvalues at all. We leave as an open problem the task of investigating if other eigenvalues exist. We study the eigenvalues of L when restricted to certain subspaces. We prove the following theorem (see Theorem 4.5.6 for formal statement).

Theorem 4.2.6 (Informal Statement)**.** *Given a hypergraph H, for every subspace S of $\mathbb{R}^n$, the operator $\Pi_S L$ has an eigenvector, i.e. there exists a vector $\mathsf{v} \in S$ and a $\gamma \in \mathbb{R}$ such that*

$$\Pi_S L(\mathsf{v}) = \gamma\, \mathsf{v}\,.$$

Given that L restricted to any subspace has an eigenvalue, we can now define higher eigenvalues of L à la Principal Component Analysis (PCA).

Definition 4.2.7. Given a hypergraph H, we define its k^{th} smallest eigenvalue γ_k and the corresponding eigenvector v_k recursively as follows. The basis of the recursion is $\mathsf{v}_1 = \mu^*$ and $\gamma_1 = 0$. Now, let $S_k := \mathsf{span}\left(\{\mathsf{v}_i : i \in [k]\}\right)$. We define γ_k to be the smallest non-trivial[3] eigenvalue of $\Pi^{\perp}_{S_{k-1}} L$ and v_k to be the corresponding eigenvector.

We will often use the following formulation of these eigenvalues.

[3]By non-trivial eigenvalue of $\Pi^{\perp}_{S_{k-1}} L$, we mean vectors in $\mathbb{R}^n \backslash S_{k-1}$ as guaranteed by Theorem 4.2.6.

Proposition 4.2.8. *The eigenvalues defined in Definition 4.2.7 satisfy*

$$\gamma_k = \min_X \frac{X^T \Pi^{\perp}_{S_{k-1}} L(X)}{X^T \Pi^{\perp}_{S_{k-1}} X} = \min_{X \perp \mathsf{v}_1, \ldots, \mathsf{v}_{k-1}} \mathcal{R}(X) \ .$$

$$\mathsf{v}_k = \mathsf{argmin}_X \frac{X^T \Pi^{\perp}_{S_{k-1}} L(X)}{X^T \Pi^{\perp}_{S_{k-1}} X} = \mathsf{argmin}_{X \perp \mathsf{v}_1, \ldots, \mathsf{v}_{k-1}} \mathcal{R}(X) \ .$$

4.2.2 Hypergraph Dispersion Processes

A Dispersion Process on a vertex set V starts with some distribution of mass on the vertices, and moves mass around according to some predefined rule. Usually mass moves from vertex having a higher concentration of mass to a vertex having a lower concentration of mass. A random walk on a graph is a dispersion process, as it can be viewed as a process moving *probability-mass* along the edges of the graph. We define the canonical dispersion process based on the hypergraph Markov operator (Definition 4.2.9). This dispersion process can be viewed as the hypergraph analogue

Definition 4.2.9 (Continuous Time Hypergraph Dispersion Process)**.** Given a hypergraph $H = (V, E, w)$, a starting probability distribution μ^0 on V, we (recursively) define the probability distribution on the vertices at time $t + \mathsf{dt}$, for an infinitesimal time duration dt, as a function of the distribution at time t as follows.

$$\mu^{t+\mathsf{dt}} = ((1 - \mathsf{dt})I + \mathsf{dt}\, M) \circ \mu^t \ .$$

Figure 6: Continuous Time Hypergraph Dispersion Process

of the random walk on graphs; indeed, when all hyperedges have cardinality 2 (i.e. the hypergraph is a graph), the action of the hypergraph Markov operator M on a vector X is equivalent to the action of the (normalized) adjacency matrix of the graph on X. This process can be used as an algorithm to estimate size of a hypergraph and for sampling vertices from it, in the same way as random walks are used to accomplish these tasks in graphs. We further believe that this dispersion process will have numerous applications in counting/sampling problems on hypergraphs, in the same way that random walks on graphs have applications in counting/sampling problems on graphs.

A fundamental parameter associated with the dispersion processes is its *Mixing Time*.

Definition 4.2.10 (Mixing Time)**.** Given a hypergraph $H = (V, E, w)$, a probability distribution μ is said to be δ-mixed if

$$\|\mu - \mu^*\|_1 \leqslant \delta \,.$$

Given a starting probability distribution μ^0, we define its *Mixing time* $\mathsf{t}_\delta^{\mathsf{mix}}(\mu^0)$ as the smallest time t such that

$$\left\|\mu^t - \mu^*\right\|_1 \leqslant \delta$$

where the μ^t are as given by the hypergraph Dispersion Process (Definition 4.2.9).

We will show that in some hypergraphs on 2^k vertices, the mixing time can be $\mathcal{O}(\mathsf{poly}(k))$ (Theorem 4.2.17). We believe that this fact will have applications in counting/sampling problems on hypergraphs à la MCMC (Markov chain monte carlo) algorithms on graphs.

4.2.3 Summary of Results

We first show that the Laplacian operator L has eigenvalues (see Theorem 4.2.6 and Proposition 4.2.8). We relate these eigenvalues to other properties of hypergraphs as follows.

4.2.3.1 Spectral Gap of Hypergraphs

A basic fact in spectral graph theory is that a graph is disconnected if and only if λ_2, the second smallest eigenvalue of its normalized Laplacian matrix, is zero. Cheeger's Inequality is a fundamental inequality which can be viewed as robust version of this fact.

We prove a generalization of Cheeger's Inequality to hypergraphs.

Theorem 4.2.11 (Hypergraph Cheeger Inequality)**.** *Given a hypergraph H,*

$$\frac{\gamma_2}{2} \leqslant \phi_H \leqslant \sqrt{2\gamma_2}\,.$$

Expander Mixing Lemma. The Expander Mixing Lemma [2] for graphs says that expanders behave like random graphs, in respect to the number of edges that cross any cut. More formally, given a graph $G = (V, E, w)$, for any two non-empty sets $S, T \subset V$

$$\left| |E(S,T)| - \frac{d\,|S|\;|T|}{n} \right| \leqslant (1 - \lambda_2)\sqrt{|S|\;|T|}$$

where λ_2 is the second smallest eigenvalue of the graph Laplacian. We prove the hypergraph version of this Lemma.

Theorem 4.2.12. *Given a d-regular hypergraph $H = (V, E, w)$, for any two non-empty sets $S, T \subset V$*

$$\left| |E(S,T)| - \frac{d\,|S|\;|T|}{n} \right| \leqslant (1 - \gamma_2) d\sqrt{|S|\;|T|}\,.$$

Hypergraph Diameter. A well known fact about graphs is that the diameter of a graph G is at most $\mathcal{O}\left(\log n / \left(\log(1/(1 - \lambda_2))\right)\right)$ where λ_2 is the second smallest eigenvalue of the graph Laplacian. Here we prove a generalization of this fact to hypergraphs.

Theorem 4.2.13. *Given a hypergraph $H = (V, E, w)$ with all its edges having weight 1, its diameter is at most*

$$\mathsf{diam}(H) \leqslant \mathcal{O}\left(\frac{\log |V|}{\log \frac{1}{1-\gamma_2}} \right).$$

4.2.3.2 Higher Order Cheeger Inequalities.

A well known fact in spectral graph theory is that a graph has at least k components if and only if λ_k, the k^{th} smallest eigenvalue of its normalized Laplacian matrix, is zero.

It is easy to see that the analogous fact for hypergraphs is also true. The following is a robust version of this fact for graphs.

Theorem 4.2.14. *[63, 70] For any graph $G = (V, E, w)$ and any integer $k < |V|$, there exists a k-partition of V into $\{S_1, \ldots, S_k\}$ such that*

$$\max_{i \in [k]} \phi(S_i) \leqslant \mathcal{O}\left(k^3 \sqrt{\lambda_k}\right) .$$

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

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

We prove a slightly weaker generalization to hypergraphs.

Theorem 4.2.15. *For any hypergraph $H = (V, E, w)$ and any integer $k < |V|$, there exists a k-partition of V into $\{S_1, \ldots, S_k\}$ such that*

$$\max_{i \in [k]} \phi(S_i) \leqslant \mathcal{O}\left(k^4 \sqrt{\gamma_k \log r}\right) .$$

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

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

Small-set Expansion. Recall that the SMALL SET EXPANSION problem (Problem 2.1.7) asks to compute the set of size at most $|V|/k$ vertices having the least expansion. Corollary 3.1.10 bounds small-set expansion in graphs via higher eigenvalues of the graph Laplacians as follows. It says that for a graph G and a parameter $k \in \mathbb{Z}_{\geqslant 0}$, there exists a set $S \subset V$ of size $\mathcal{O}(n/k)$ such that

$$\phi(S) \leqslant \mathcal{O}\left(\sqrt{\lambda_k \log k}\right) .$$

We prove a generalization of this bound to hypergraphs (see Theorem 4.6.1 for formal statement).

Theorem 4.2.16 (Informal Statement)**.** *Given hypergraph $H = (V, E, w)$ and parameter $k < |V|$, there exists a set $S \subset V$ such that $|S| \leqslant \mathcal{O}(|V|/k)$ satisfying*

$$\phi(S) \leqslant \tilde{\mathcal{O}}\left(\min\{r, k\}\sqrt{\gamma_k}\right)$$

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

4.2.3.3 Mixing Time Bounds

A well known fact in spectral graph theory is that a random walk on a graph mixes in time at most $\mathcal{O}(\log n/\lambda_2)$ where λ_2 is the second smallest eigenvalue of graph Laplacian. Moreover, every graph has some vertex such that a random walk starting from that vertex takes at least $\Omega(1/\lambda_2)$ time to mix, thereby proving that the dependence of the mixing time on λ_2 is optimal. We prove a generalization of the first fact to hypergraphs and a slightly weaker generalization of the second fact to hypergraphs. Both of them together show that dependence of the mixing time on γ_2 is optimal. Further, we believe that Theorem 4.2.17 will have applications in counting/sampling problems on hypergraphs à la MCMC (Markov chain monte carlo) algorithms on graphs.

Theorem 4.2.17 (Upper bound on Mixing Time)**.** *Given a hypergraph $H = (V, E, w)$, for all starting probability distributions $\mu^0 : V \to [0, 1]$, the Hypergraph Dispersion Process satisfies*

$$\mathsf{t}_{\delta}^{\mathsf{mix}}\left(\mu^0\right) \leqslant \frac{\log(n/\delta)}{\gamma_2}\,.$$

Theorem 4.2.18 (Lower bound on Mixing Time)**.** *Given a hypergraph $H = (V, E, w)$, there exists a probability distribution μ^0 on V such that $\left\|\mu^0 - \mathbf{1}/n\right\|_1 \geqslant 1/2$ and*

$$\mathsf{t}_{\delta}^{\mathsf{mix}}\left(\mu^0\right) \geqslant \frac{\log(1/\delta)}{16\,\gamma_2}\,.$$

We view the condition in Theorem 4.2.18 that the starting distribution μ^0 satisfy $\left\|\mu^0 - \mathbf{1}/n\right\|_1 \geqslant 1/2$ as the analogue of a random walk in a graph starting from some vertex (i.e. it is far from being mixed).

4.2.3.4 Vertex Expansion in Graphs and Hypergraph Expansion

We present a factor preserving reduction from vertex expansion in graphs to hypergraph expansion. Recall that the notion of Vertex Expansion and Symmetric Vertex Expansion are computationally equivalent up to constant factors (Theorem 8.3.1 and Theorem 8.3.2).

Theorem 4.2.19. *Given a graph $G = (V, E)$ of maximum degree d and minimum degree $c_1 d$ (for some constant c_1), there exists a polynomial time computable hypergraph $H = (V, E')$ on the same vertex set having the hyperedges of cardinality at most $d + 1$ such that for all sets $S \subset V$,*

$$c_1 \phi_H(S) \leqslant \frac{1}{d} \cdot \Phi^{\mathsf{V}}(S) \leqslant \phi_H(S)\,.$$

Remark 4.2.20. The dependence on the degree in Theorem 4.2.19 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.

Theorem 4.2.19 implies that all our results for hypergraphs directly extend to vertex expansion in graphs. More formally, we have a Markov operator M and a Laplacian operator L, whose eigenvalues satisfy the vertex expansion (in graphs) analogs of Theorem 4.2.11[4], Theorem 4.2.12, Theorem 4.2.13, Theorem 4.2.15, Theorem 4.2.16, Theorem 4.2.17, Theorem 4.2.18, and Theorem 6.1.5.

4.2.3.5 Discussion

We stress that none of our bounds have a polynomial dependence on r, the size of the largest hyperedge (Theorem 4.2.16 has a dependence on $\min\{r, k\}$) . In many of the

[4] A Cheeger-type Inequality for vertex expansion in graphs was also proven by [21].

practical applications, the typical instances have $r = \Theta(n^\alpha)$ for some $\alpha = \Omega(1)$; in such cases, bounds of $\mathsf{poly}(r)$ would not be of any practical utility.

We also stress that all our results generalize the corresponding results for graphs.

4.2.4 Organization

We begin with an overview of the proofs in Section 4.3. We prove Theorem 4.2.6 (formally Theorem 4.5.6) and Proposition 4.2.8 in Section 4.5. We prove Theorem 4.2.11, Theorem 4.2.12 and Theorem 4.2.13 in Section 4.4.1. We prove Theorem 4.2.15 and Theorem 4.2.16 in Section 4.6. We prove Theorem 4.2.17 and Theorem 4.2.18 in Section 4.5. We prove Theorem 4.9.1 in Section 4.9. We prove Theorem 4.2.19 in Section 4.7.

4.3 Overview of Proofs

Hypergraph Eigenvalues. To prove that hypergraph eigenvalues exist (Theorem 4.2.6 and Proposition 4.2.8), we study the hypergraph dispersion process in a more general setting (Definition 4.5.1). We start the dispersion process with an arbitrary vector $\mu^0 \in \mathbb{R}^n$. Our main tool here is to show that the Rayleigh quotient (as a function of the time) monotonically decreases with time. More formally, we show that the Rayleigh quotient of $\mu^{t+\mathsf{dt}}$, the vector at time $t + \mathsf{dt}$ (for some infinitesimally small dt), is not larger than the Rayleigh quotient of μ^t, the vector at time t. If the under lying matrix A_{μ^t} did not change between times t and $t + \mathsf{dt}$, then this fact can be shown using simple linear algebra. If the under lying matrix A_{μ^t} changes between t and $t + \mathsf{dt}$, then proof requires a lot more work. Our proof involves studying the limits of the Rayleigh quotient in the neighborhoods of the time instants at which the support matrix changes, and exploiting the continuity properties of the process.

To show that eigenvectors exist, we start with a candidate eigenvector, say X, that satisfies the conditions of Proposition 4.2.8. We study a slight variant of hypergraph dispersion process starting with this vector X. We use the monotonicity of the

Rayleigh quotient to conclude that $\forall t \geqslant 0$, the vector at time t of this process, say X^t, also satisfies the conditions of Proposition 4.2.8. Then we use the fact that the number of possible support matrices $|\{A_Y : Y \in \mathbb{R}^n\}| < \infty$ to argue that there exists a time interval of positive Lebesgue measure during which the support matrix does not change. We use this to conclude that the vectors X^t during that time interval must also not change (the proof of this uses the previous conclusion that all X^t the conditions of Proposition 4.2.8) and hence must be an eigenvector.

Mixing Time Bounds. To prove a lower bound on the mixing time of the Hypergraph Dispersion process (Theorem 4.2.18), we need to exhibit a probability distribution that is far from being mixed and takes a long time to mix. To show that a distribution μ takes a long time to mix, it would suffice to show that $\mu - \mathbf{1}/n$ is "close" to v_2, as we can then use our previous assertion about the monotonicity of the Rayleigh quotient to prove a lower bound on the mixing time. As a first attempt at constructing such a distribution, one might be tempted to consider the vector $\mathbf{1}/n + \mathsf{v}_2$. But this vector might not even be a probability distribution if $\mathsf{v}_2(i) < -1/n$ for some coordinate i. A simple fix for this would to consider the vector $\mu \stackrel{\text{def}}{=} \mathbf{1}/n + \mathsf{v}_2/(n \left\lVert \mathsf{v}_2 \right\rVert_\infty)$. But then $\left\lVert \mu - \mathbf{1}/n \right\rVert_1 = \left\lVert \mathsf{v}_2/(n \left\lVert \mathsf{v}_2 \right\rVert_\infty) \right\rVert_1$ which could be very small depending on $\left\lVert \mathsf{v}_2 \right\rVert_\infty$. Our proof involves starting with v_2 and carefully "chopping" of the vector at some points to control its infinity-norm while maintaining that its Rayleigh quotient is still $\mathcal{O}(\gamma_2)$.

The main idea used in proving the upper bound on the mixing time of (Theorem 4.2.17) is that the support matrix at any time t has a spectral gap of at least γ_2. Therefore, after every unit of the time, the component of the vector μ^t that is orthogonal to $\mathbf{1}$, decreases in ℓ_2-norm by a factor of at least $1 - \gamma_2$ (irrespective of the fact that the support matrix might be changing infinitely many times during that time interval).

Hypergraph Diameter. Our proof strategy for Theorem 4.2.13 is as follows. Let $M' \stackrel{\text{def}}{=} I/2 + M/2$ be a lazy version of M. Fix some vertex $u \in V$. Consider the vector $M'(\chi_u)$. This vector will have non-zero values at exactly those coordinates which correspond to vertices that are at a distance of at most 1 from u. Building on this idea, it follows that the vector $M'^t(\chi_u)$ will have non-zero values at exactly those coordinates which correspond to vertices that are at a distance of at most t from u. Therefore, the diameter of H is the smallest value $t \in \mathbb{Z}_{\geqslant 0}$ such that the vectors $\{M'^t(\chi_u) : u \in V\}$ have non-zero entries in all coordinates. We will upper bound the value of such a t. The key insight in this step is that the support matrix A_X of any vector $X \in \mathbb{R}^n$ has a spectral gap of at least γ_2, irrespective of what the vector X is.

Hypergraph Cheeger Inequality. We appeal to the formulation of eigenvalues in Proposition 4.2.8 to prove Theorem 4.2.11.

$$\gamma_2 = \min_{X \perp \mathbf{1}} \frac{X^T L(X)}{X^T X} = \frac{\sum_{e \in E} w(e) \max_{i,j \in E} (X_i - X_j)^2}{d \sum_i X_i^2} .$$

First, observe that if all the entries of the vector X were in $\{0, 1\}$, then the support of this vector X, say S, will have expansion equal to $\mathcal{R}(X)$. Building on this idea, we start with the vector v_2, and use it to construct a line-embedding of the vertices of the hypergraph, such that the average "distortion" of the hyperedges is at most $\mathcal{O}\left(\sqrt{\gamma_2}\right)$. Next, we represent this average distortion as an average over cuts in the hypergraph and conclude that at least one of these cuts must have expansion at most this average value. Overall, we follow the strategy of proving Cheeger's Inequality for graphs. However, we need some new ideas to handle hyperedges.

Higher Order Cheeger Inequalities. Proving our bound for hypergraph small-set expansion (Theorem 4.2.16) requires a lot more work. We start with the spectral embeddings, the canonical embedding of the vertex set into $\mathbb{R}^k$ given by the top k eigenvectors. As a first attempt, one might try to "round" this embedding using the rounding algorithms for small set expansion on graphs, namely the algorithms of [16] or [87]. However, the rounding algorithm of [16] uses the fact that the vectors should satisfy ℓ_2^2-triangle inequality and more crucially uses the fact that the inner product between any two vectors is non-negative. Neither of these properties are satisfied by the spectral embedding[5]. The rounding algorithm of [87] crucially uses the fact that the Rayleigh quotient of the vector X_l obtained by picking the l^{th} coordinate from each vector of the spectral embedding be "small" for at least one coordinate l. It is easy to show that this fact holds for graphs, but this is not true for hypergraphs because of the "max" in the definition of the eigenvalues.

Our proof starts with the spectral embedding and uses a simple random projection step to produce a vector X. This step is similar to the rounding algorithm of [70], who studied a variant of small-set expansion in graphs. We then bound the length[6] of the hyperedges. Here we deviate from [70], as hyperedges have more than two vertices and can not be analyzed in the same way as edges in graphs. We handle the hyperedges whose vertices have roughly equal lengths by bounding the variance of their projections in the random projection step. We handle the hyperedges whose vertices have very large disparity in lengths by showing that they must be having a large contribution to the Rayleigh quotient. This suffices to bound the expansion of the set obtained by our rounding algorithm (Algorithm 4.6.2). To show that the set is small, we use a combination of the techniques studied in [73] and [70]. This gives uses the desired bound for small-set expansion. To get a bound on hypergraph multi-partitioning (Theorem 4.2.15), at a high level, we use a stronger form of our hypergraph small-set expansion bound together with the framework of [70].

[5]If the v_i's are the spectral embedding vectors, then one could also try to round the vectors $v_i \otimes v_i$. This will have the property $\langle v_i \otimes v_i, v_j \otimes v_j \rangle \geqslant 0$. However, by rounding these vectors one can only hope to prove a $\mathcal{O}\left(\sqrt{\gamma_{k^2} \,\mathsf{polylog}\, k}\right)$ (see [72]).

[6]Length of an edge e under X is defined as $\max_{i,j \in e} |X_i - X_j|$.

4.4 Spectral Gap of Hypergraphs

We define the *Spectral Gap* of a hypergraph to be γ_2, the second smallest eigenvalue of its Laplacian operator.

4.4.1 Hypergraph Cheeger Inequality

In this section we prove the hypergraph Cheeger Inequality Theorem 4.2.11.

Theorem 4.4.1 (Restatement of Theorem 4.2.11)**.** *Given a hypergraph H,*

$$\frac{\gamma_2}{2} \leqslant \phi_H \leqslant \sqrt{2\gamma_2}\,.$$

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

Proposition 4.4.2. *Let $H = (V, E, w)$ be a hypergraph with edge weights $w : E \to \mathbb{R}^+$ and let $Y \in [0,1]^{|V|}$. Then there exists a set $S \subseteq \mathsf{supp}(Y)$ such that*

$$\phi(S) \leqslant \frac{\sum_{e \in E} w(e) \max_{i,j \in e} |Y_i - Y_j|}{\sum_i d_i Y_i}$$

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

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

Let S_r denote the support of the vector $F_r(Y)$. For any $a \in [0,1]$ it is easy to see that

$$\int_0^1 F_r(a)\,\mathsf{dr} = a\,. \tag{3}$$

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

$$\int_0^1 |F_r(a) - F_r(b)|\,\mathsf{dr} = \left| \int_0^1 F_r(a)\mathsf{dr} - \int_0^1 F_r(b)\mathsf{dr} \right| = |a - b|\ . \tag{4}$$

Also, for a hyperedge $e = \{a_i : i \in [r]\}$ if $|a_1 - a_2| \geqslant |a_i - a_j| \, \forall a_i, a_j \in e$, then

$$|F_r(a_1) - F_r(a_2)| \geqslant |F_r(a_i) - F_r(a_j)| \quad \forall r \in [0,1] \text{ and } \forall a_i, a_j \in e\,. \tag{5}$$

Therefore,

$$\begin{aligned}
&\frac{\int_0^1 \sum_e w(e) \max_{i,j \in e} |F_r(Y_i) - F_r(Y_j)| \, \mathsf{dr}}{\int_0^1 \sum_i d_i F_r(Y_i) \mathsf{dr}} \\
&\geqslant \frac{\sum_e w(e) \max_{i,j \in e} \int_0^1 |F_r(Y_i) - F_r(Y_j)| \, \mathsf{dr}}{\int_0^1 \sum_i d_i F_r(Y_i) \mathsf{dr}} && \text{(Using (5))} \\
&= \frac{\sum_e w(e) \max_{i,j \in e} \left| \int_0^1 F_r(Y_i) - \int_0^1 F_r(Y_j) \right| \mathsf{dr}}{\sum_i d_i \int_0^1 F_r(Y_i) \mathsf{dr}} && \text{(Using (4))} \\
&= \frac{\sum_e w(e) \max_{i,j \in e} |Y_i - Y_j|}{\sum_i d_i Y_i} && \text{(Using (3))}\,.
\end{aligned}$$

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

$$\frac{\sum_e w(e) \max_{i,j \in e} |F_{r'}(Y_i) - F_{r'}(Y_j)|}{\sum_i d_i F_{r'}(Y_i)} \leqslant \frac{\sum_e w(e) \max_{i,j \in e} |Y_i - Y_j|}{\sum_i d_i Y_i}\,.$$

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

$$\frac{\sum_e w(e) \max_{i,j \in e} |F_{r'}(Y_i) - F_{r'}(Y_j)|}{\sum_{i \in V} d_i F_{r'}(Y_i)} = \frac{\sum_e w(e) \mathbb{I}\left[e \text{ is cut by } S_{r'}\right]}{\sum_{i \in S_{r'}} d_i} = \phi(S_{r'})\,.$$

Therefore,

$$\phi(S_{r'}) \leqslant \frac{\sum_e w(e) \max_{i,j \in e} |Y_i - Y_j|}{\sum_i d_i Y_i} \qquad \text{and} \qquad S_{r'} \subset \mathsf{supp}(Y)\,.$$

□

Proposition 4.4.3. *Given a hypergraph $H = (V, E, w)$ and a vector $Y \in \mathbb{R}^{|V|}$ such that $\langle Y, \mu^* \rangle = 0$, there exists a set $S \subset V$ such that*

$$\phi(S) \leqslant \sqrt{2 \mathcal{R}(Y)}\,.$$

Proof. Since $\langle Y, \mu^* \rangle = 0$, we have

$$\mathcal{R}(Y) = \frac{\sum_{e \in E} w(e) \max_{i,j \in e} (Y_i - Y_j)^2}{\sum_i d_i Y_i^2 - (\sum_i d_i Y_i)^2 / (\sum_i d_i)} = \frac{\sum_{e \in E} w(e) \max_{i,j \in e} (Y_i - Y_j)^2}{\sum_{i,j} d_i d_j \left(Y_i - Y_j\right)^2 / (\sum_i d_i)}\,.$$

Let $X = Y + c\mathbf{1}$ for an appropriate $c \in \mathbb{R}$ such that $|\mathsf{supp}(X^+)| = |\mathsf{supp}(X^-)| = n/2$. Then we get

$$\mathcal{R}(Y) = \frac{\sum_{e \in E} w(e) \max_{i,j \in e}(X_i - X_j)^2}{\sum_{i,j} d_i d_j \left(X_i - X_j\right)^2 / (\sum_i d_i)} = \frac{\sum_{e \in E} w(e) \max_{i,j \in e}(X_i - X_j)^2}{\sum_i d_i X_i^2 - (\sum_i d_i X_i)^2 / (\sum_i d_i)} \geqslant \mathcal{R}(X) \ .$$

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

$$(a^+ - b^+)^2 + (a^- - b^-)^2 \leqslant (a - b)^2$$

Therefore we have

$$\begin{aligned}
\mathcal{R}(Y) \geqslant \mathcal{R}(X) &= \frac{\sum_{e \in E} w(e) \max_{i,j \in e}(X_i - X_j)^2}{\sum_i d_i X_i^2} \\
&\geqslant \frac{\left(\sum_{e \in E} w(e) \max_{i,j \in e}(X_i^+ - X_j^+)^2\right) + \left(\sum_{e \in E} w(e) \max_{i,j \in e}(X_i^- - X_j^-)^2\right)}{\sum_i d_i (X_i^+)^2 + \sum_i d_i (X_i^-)^2} \\
&\geqslant \min \left\{ \frac{\sum_{e \in E} w(e) \max_{i,j \in e}(X_i^+ - X_j^+)^2}{\sum_i d_i (X_i^+)^2}, \frac{\sum_{e \in E} w(e) \max_{i,j \in e}(X_i^- - X_j^-)^2}{\sum_i d_i (X_i^-)^2} \right\}
\end{aligned}$$

Let $Z \in \{X^+, X^-\}$ be the vector corresponding the minimum in the previous inequality.

$$\begin{aligned}
\sum_{e \in E} w(e) \max_{i,j \in e} \left| Z_i^2 - Z_j^2 \right| &= \sum_{e \in E} w(e) \max_{i,j \in e} |Z_i - Z_j| \left(Z_i + Z_j\right) \\
&\leqslant \sqrt{\sum_{e \in E} w(e) \max_{i,j \in e} (Z_i - Z_j)^2} \sqrt{2 \sum_i d_i Z_i^2}
\end{aligned}$$

Therefore,

$$\frac{\sum_{e \in E} w(e) \max_{i,j \in e} \left| Z_i^2 - Z_j^2 \right|}{\sum_i d_i Z_i^2} \leqslant \sqrt{2 \mathcal{R}(Z)} \leqslant \sqrt{2 \mathcal{R}(Y)} \, .$$

Invoking Proposition 4.4.2 with vector Z, we get that there exists a set $S \subset V$ such that

$$\phi(S) \leqslant \sqrt{2 \mathcal{R}(Y)} \, .$$

□

We are now ready to prove Theorem 4.2.11.

Proof of Theorem 4.2.11.

1. Let $S \subset V$ be any set such that $\mathsf{vol}(S) \leqslant \mathsf{vol}(V)/2$, and let $X \in \mathbb{R}^n$ be the indicator vector of S. Let Y be the component of X orthogonal to μ^*. Then

$$\begin{aligned}
\gamma_2 &\leqslant \frac{\sum_e w(e) \max_{i,j\in e}(Y_i - Y_j)^2}{\sum_i d_i Y_i^2} = \frac{\sum_e w(e) \max_{i,j\in e}(X_i - X_j)^2}{\sum_i d_i X_i^2 - (\sum_i d_i X_i)^2/(\sum_i d_i)} \\
&= \frac{w(E(S,\bar{S}))}{\mathsf{vol}(S) - \mathsf{vol}(S)^2/\mathsf{vol}(V)} = \frac{\phi(S)}{1 - \mathsf{vol}(S)/\mathsf{vol}(V)} \\
&\leqslant 2\phi(S)\,.
\end{aligned}$$

   Since the choice of the set S was arbitrary, we get

$$\frac{\gamma_2}{2} \leqslant \phi_H\,.$$

2. Invoking Proposition 4.4.3 with v_2 we get that

$$\phi_H \leqslant \sqrt{2\,\mathcal{R}\left(\mathsf{v}_2\right)} = \sqrt{2\,\gamma_2}\,.$$

□

4.4.2 Hypergraph Expander Mixing Lemma

Theorem 4.4.4 (Restatement of Theorem 4.2.12)**.** *Given a d-regular hypergraph $H = (V, E, w)$, for any two non-empty sets $S, T \subset V$*

$$\left| |E(S,T)| - \frac{d\,|S|\,\,|T|}{n} \right| \leqslant (1-\gamma_2)d\sqrt{|S|\,\,|T|}\,.$$

Proof. Fix non-empty sets $S, T \subset V$. We construct a graph $G = (V, E', w)$ as follows. For each hyperedge $e \in E$, we add an edge to E' as follows. If $e \in E$ is cut by both S and T, then we pick any one vertex from $e \cap S$ and any one vertex from $e \cap T$. If e is cut only by S (resp. T), then we pick any one vertex from $e \cap S$ (resp. $e \cap T$) and any one vertex from $e \cap \bar{S}$ (resp. $e \cap \bar{T}$). If e is cut neither by S nor by T, then we pick any pair of vertices. Next, we add sufficient self-loops at each vertex to make G d-regular. We let A be the normalized adjacency matrix of G. By construction, it is easily verified that

$$\chi_S^T A \chi_T = \tfrac{1}{d} \cdot |E(S,T)| \tag{6}$$

$$\chi_S^T (I-A)\chi_S = \phi(S)\, \chi_S^T \chi_S \geqslant \tfrac{\gamma_2}{2}\, \chi_S^T \chi_S \qquad \text{(Using Theorem 4.2.11)} \tag{7}$$

$$\chi_T^T (I-A)\chi_T = \phi(T)\, \chi_T^T \chi_T \geqslant \tfrac{\gamma_2}{2}\, \chi_T^T \chi_T \qquad \text{(Using Theorem 4.2.11)} \tag{8}$$

Let Y_S be the component of χ_S orthogonal to $\mathbf{1}$, i.e.

$$Y_S \stackrel{\text{def}}{=} \chi_S - \left\langle \chi_S, \mathbf{1}/\sqrt{n} \right\rangle \mathbf{1}/\sqrt{n} = \chi_S - \frac{|S|}{n}\mathbf{1}\,.$$

Then,

$$\begin{aligned} Y_S^T A Y_S &= Y_S^T Y_S - Y_S^T (I-A) Y_S = Y_S^T Y_S - \chi_S^T (I-A)\chi_S \\ &\leqslant (1-\gamma_2/2)\chi_S^T \chi_S \qquad \text{(Using } \|Y_S\| \leqslant \|\chi_S\| \text{ and (7))} \end{aligned} \tag{9}$$

Similarly,

$$Y_T^T A Y_T \leqslant (1-\gamma_2/2)\chi_T^T \chi_T \tag{10}$$

Now,

$$\begin{aligned} \frac{1}{d} \cdot |E(S,T)| = \chi_S^T A \chi_T &= \left(\frac{|S|}{n}\mathbf{1} + Y_S\right)^T A \left(\frac{|T|}{n}\mathbf{1} + Y_T\right) \\ &= |S|\;|T| \cdot \frac{1}{n^2} \cdot \mathbf{1}^T A \mathbf{1} + \frac{|T|}{n} Y_S^T A \mathbf{1} + \frac{|S|}{n} Y_T^T A \mathbf{1} + Y_S^T A Y_T \\ &= |S|\;|T| \cdot \frac{1}{n^2} \cdot n + 0 + 0 + Y_S^T A Y_T \qquad (A\mathbf{1} = \mathbf{1} \text{ and } \langle Y_S, \mathbf{1}\rangle = 0) \end{aligned}$$

Since $A \succeq 0$,

$$\begin{aligned} \left| \frac{1}{d} \cdot |E(S,T)| - \frac{|S|\;|T|}{n} \right| &\leqslant \sqrt{Y_S^T A Y_S}\sqrt{Y_T^T A Y_T} \qquad \text{(Cauchy-Schwarz Inequality)} \\ &\leqslant (1-\gamma_2/2)\,\|\chi_S\|\;\|\chi_T\| \qquad \text{(Using (9) and (10))} \\ &\leqslant (1-\gamma_2/2)\sqrt{|S|\;|T|} \qquad (\|\chi_S\| \leqslant \sqrt{|S|})\,. \end{aligned}$$

This finishes the proof of the theorem.

□

4.4.3 Hypergraph Diameter

In this section we prove Theorem 4.2.13.

Theorem 4.4.5 (Restatement of Theorem 4.2.13)**.** *Given a hypergraph $H = (V, E, w)$ with all its edges having weight* 1*, its diameter is at most*

$$\mathsf{diam}(H) \leqslant \mathcal{O}\left(\frac{\log n}{\log \frac{1}{1-\gamma_2}}\right).$$

Remark 4.4.6. A weaker bound on the diameter follows from Theorem 4.2.17

$$\mathsf{diam}(H) \leqslant \mathcal{O}\left(\frac{\log n}{\gamma_2}\right).$$

We start by defining the notion of operator powering.

Definition 4.4.7 (Operator Powering)**.** For a $t \in \mathbb{Z}_{\geqslant 0}$, and an operator $M : \mathbb{R}^n \to \mathbb{R}^n$, for a vector $X \in \mathbb{R}^n$ we define $M^t(X)$ as follows

$$M^t(X) \stackrel{\text{def}}{=} M(M^{t-1}(X)) \qquad \text{and} \qquad M^1(X) \stackrel{\text{def}}{=} M(X).$$

Next, we state bound the norms of powered operators.

Lemma 4.4.8. *For vector $\omega \in \mathbb{R}^n$, such that $\langle \omega, \mathbf{1} \rangle = 0$,*

$$\left\| M^t(\omega) \right\| \leqslant (1 - \gamma_2)^{t/2} \left\| \omega \right\|.$$

Proof. We prove this by induction on t. Let $v_1, \ldots, v_n$ be the eigenvectors of A_ω and let $\lambda_1, \ldots, \lambda_n$ be the the corresponding eigenvalues. Let $\omega = \sum_{i=1}^n c_i v_i$ for appropriate constants $c_i \in \mathbb{R}$. Then, for $t = 1$,

$$\frac{\|M(\omega)\|}{\|\omega\|} = \frac{\|A_\omega\, \omega\|}{\|\omega\|} = \sqrt{\frac{\sum_i c_i^2 \lambda_i^2}{\sum_i c_i^2}} \leqslant \sqrt{\frac{\sum_i c_i^2 \lambda_i}{\sum_i c_i^2}} \qquad \text{(Since each } \lambda_i \in [0,1],\ \lambda_i^2 \leqslant \lambda_i)$$

$$= \sqrt{\frac{\omega^T M(\omega)}{\omega^T \omega}} \leqslant \sqrt{1 - \gamma_2}\,. \tag{11}$$

Similarly, for $t > 1$.

$$\left\| M^t(\omega) \right\| = \left\| M(M^{t-1}(\omega)) \right\| \leqslant (1 - \gamma_2)^{1/2} \left\| M^{t-1}(\omega) \right\| \leqslant (1 - \gamma_2)^{t/2} \left\| \omega \right\|$$

where the last inequality follows from the induction hypothesis. ☐

Proof of Theorem 4.2.13. For the sake of simplicity, we will assume that the hypergraph is regular. Our proof easily extends to the general case. We define the operator $M' \stackrel{\text{def}}{=} I/2 + M/2$. Then the eigenvalues of M' are $1/2 + \gamma_i/2$, and the corresponding eigenvectors are v_i, for $i \in [n]$.

Our proof strategy is as follows. Fix some vertex $u \in V$. Consider the vector $M'(\chi_u)$. This vector will have non-zero values at exactly those coordinates which correspond to vertices that are at a distance of at most 1 from u (see also Remark 4.5.2). Building on this idea, it follows that the vector $M'^t(\chi_u)$ will have non-zero values at exactly those coordinates which correspond to vertices that are at a distance of at most t from u. Therefore, the diameter of H is the smallest value $t \in \mathbb{Z}_{\geqslant 0}$ such that the vectors $\{M'^t(\chi_u) : u \in V\}$ have non-zero entries in all coordinates. We will upper bound the value of such a t.

Fix two vertices $u, v \in V$. Let χ_u, χ_v be their respective characteristic vectors and let ω_u, ω_v be the components of χ_u, χ_v orthogonal to $\mathbf{1}$ respectively

$$\omega_u \stackrel{\text{def}}{=} \chi_u - \frac{\mathbf{1}}{n} \qquad \text{and} \qquad \omega_v \stackrel{\text{def}}{=} \chi_v - \frac{\mathbf{1}}{n}.$$

Then

$$\|\omega_u\| = \sqrt{\left(\chi_u - \frac{\mathbf{1}}{n}\right)^T \left(\chi_u - \frac{\mathbf{1}}{n}\right)} = \sqrt{1 - \frac{1}{n} - \frac{1}{n} + \frac{n}{n^2}} = \sqrt{1 - \frac{1}{n}}. \tag{12}$$

Since $\mathbf{1}$ is invariant under M' we get

$$\begin{aligned}\chi_u^T M'^t(\chi_v) &= \left(\frac{\mathbf{1}}{n} + \omega_u\right)^T M'^t \left(\frac{\mathbf{1}}{n} + \omega_v\right) = \left(\frac{\mathbf{1}}{n} + \omega_u\right)^T \left(\frac{\mathbf{1}}{n} + M'^t(\omega_v)\right) \\ &= \frac{1}{n} + 0 + \frac{1}{n}\mathbf{1}^T M'^t(\omega_v) + \omega_u^T M'^t(\omega_v).\end{aligned}$$

Now since M' is a dispersion process, if $\langle \omega_u, \mathbf{1} \rangle = 0$, then $\langle M'(\omega_u), \mathbf{1} \rangle = 0$ and hence $\langle M'^t(\omega_u), \mathbf{1} \rangle = 0$. Therefore,

$$\chi_u^T M'^t \chi_v = \frac{1}{n} + \omega_u^T M'^t(\omega_v). \tag{13}$$

Now,

$$\left|\omega_u^T M'^t(\omega_v)\right| \leqslant \|\omega_u\| \; \left\|M'^t(\omega_v)\right\| \leqslant \left(\frac{1-\gamma_2}{2}\right)^{t/2} \|\omega_u\| \; \|\omega_v\| \qquad \text{(Using Lemma 4.4.8).}$$

Therefore, from (13) and (12),

$$\chi_u^T M'^t \chi_v \geqslant \frac{1}{n} - \left(\frac{1-\gamma_2}{2}\right)^{t/2} \|\omega_u\| \; \|\omega_v\| \geqslant \frac{1}{n} - \left(\frac{1-\gamma_2}{2}\right)^{t/2} \left(1 - \frac{1}{n}\right). \qquad (14)$$

Therefore, for

$$t \geqslant \frac{2\log(n/2)}{\log\left(\frac{2}{1-\gamma_2}\right)},$$

we have $\chi_u^T M'^t \chi_v > 0$. Therefore,

$$\mathsf{diam}(H) \leqslant \frac{\log n}{\log\left(\frac{1}{1-\gamma_2}\right)}.$$

□

4.5 The Hypergraph Dispersion Process

In this section we will prove Theorem 4.2.6, Proposition 4.2.8, Theorem 4.2.17 and Theorem 4.2.18. For the sake of simplicity, we assume that the hypergraph is regular. All our proofs easily extend to the general case.

Definition 4.5.1 (Projected Continuous Time Hypergraph Dispersion Process)**.** Given a hypergraph $H = (V, E, w)$, a projection operator $\Pi_S : \mathbb{R}^n \to \mathbb{R}^n$ for some subspace S of $\mathbb{R}^n$ and a function $\omega^0 : V \to \mathbb{R}$ such that $\omega^0 \in S$, we (recursively) define the functions on the vertices at time $t + \mathsf{dt}$, for an infinitesimal time duration dt, as a function of ω^t as follows

$$\omega^{t+\mathsf{dt}} \stackrel{\text{def}}{=} \Pi_S \left((1-\mathsf{dt})I + \mathsf{dt}\, M\right) \circ \omega^t .$$

Figure 7: Projected Continuous Time Hypergraph Dispersion Process

Remark 4.5.2. We make a remark about the matrices A_X for vectors $X \in \mathbb{R}^n$ in Definition 4.2.1 when being used in the continuous time processes of Definition 4.2.9

and Definition 4.5.1. For a hyperedge $e \in E$, we compute the pair of vertices

$$(i_e, j_e) = \mathsf{argmax}_{i,j \in e} \left(X_i - X_j \right)$$

and add an edge between them in the graph G_X. If the pair is not unique, then we define

$$S_e^t \stackrel{\text{def}}{=} \left\{ i \in e : \omega^t(i) = \max_{j \in e} \omega^t(j) \right\} \qquad \text{and} \qquad R_e^t \stackrel{\text{def}}{=} \left\{ i \in e : \omega^t(i) = \min_{j \in e} \omega^t(j) \right\}$$

and add to G_X a complete weighted bipartite graph on $S_e^t \times R_e^t$ with each edge having weight $w(e)/\left(|S^t|\,|R^t|\right)$.

A natural thing one would try first is to pick a vertex, say i_1, from S_e^t and a vertex, say j_1, from R_e^t and add an edge between $\{i_1, j_1\}$. However, in such a case, after 1 infinitesimal time unit, the pair (i_1, j_1) will no longer have the largest difference in values of X among the pairs in $e \times e$, and we will need to pick some other suitable pair from $S_e^t \times R_e^t \setminus \{(i_1, j_1)\}$. We will have to repeat this process of picking a different pair of vertices after each infinitesimal time unit. Moreover, each of these infinitesimal time units will have Lebesgue measure 0. Therefore, we avoid this difficulty by adding a suitably weighted complete graph on $S_e^t \times R_e^t$ without loss of generality.

Note that when $\Pi_S = I$, then Definition 4.5.1 is the same as Definition 4.2.9. We need to study the Dispersion Process in this generality to prove Theorem 4.2.6 and Proposition 4.2.8.

Lemma 4.5.3 (Main Technical Lemma)**.** *Given a hypergraph $H = (V, E, w)$, and a function $\omega^0 : V \to \mathbb{R}$, the Dispersion process in Definition 4.5.1 satisfies the following properties.*

1.

$$\frac{\mathrm{d}\left\| \omega^t \right\|^2}{\mathsf{dt}} = -2\,\mathcal{R}\left(\omega^t\right) \left\| \omega^t \right\|^2 \qquad \forall t \geqslant 0\,. \tag{15}$$

2.

$$\mathcal{R}\left(\omega^{t+\mathsf{dt}}\right) \leqslant \mathcal{R}\left(\omega^t\right) \qquad \forall t, \mathsf{dt} \geqslant 0\,. \tag{16}$$

Proof. Fix a time $t \geqslant 0$.

1. Let
$$A \stackrel{\text{def}}{=} A_{\omega^{\mathsf{t}}} \qquad \text{and} \qquad A' = (1-\mathsf{dt})I + \mathsf{dt}\, A\,.$$
Then
$$\left\|\omega^t\right\|^2 - \left\|\omega^{t+\mathsf{dt}}\right\|^2 = \left\langle \omega^t - \omega^{t+\mathsf{dt}}, \omega^t + \omega^{t+\mathsf{dt}} \right\rangle = (\omega^t)^T (I - \Pi_S A')(I + \Pi_S A')\omega^t$$
Now, $\lim_{\mathsf{dt}\to 0}(I + \Pi_S A') = I + \Pi_S$. By construction, we have $\omega^t \in S$. Therefore,
$$\left\|\omega^t\right\|^2 - \left\|\omega^{t+\mathsf{dt}}\right\|^2 = 2\mathsf{dt}\,(\omega^t)^T (I - A)\omega^t\,.$$
Therefore
$$\frac{\mathrm{d}\left\|\omega^t\right\|^2}{\mathsf{dt}} = -2\,\mathcal{R}\left(\omega^t\right)\left\|\omega^t\right\|^2\,.$$

2. Let
$$A_1 \stackrel{\text{def}}{=} A_{\omega^{\mathsf{t}}}, \qquad A_1' \stackrel{\text{def}}{=} (1-\mathsf{dt})I + \mathsf{dt}\, A_1, \qquad A_2 \stackrel{\text{def}}{=} A_{\omega^{\mathsf{t+dt}}}\,.$$
Then
$$\mathcal{R}\left(\omega^t\right) = \frac{(\omega^t)^T(I - A_1)\omega^t}{(\omega^t)^T(\omega^t)} \qquad \text{and} \qquad \mathcal{R}\left(\omega^{t+\mathsf{dt}}\right) = \frac{(\omega^{t+\mathsf{dt}})^T(I - A_2)\omega^{t+\mathsf{dt}}}{(\omega^{t+\mathsf{dt}})^T(\omega^{t+\mathsf{dt}})}\,.$$
From the definition of the process, we have $\omega^{t+\mathsf{dt}} = \Pi_S A_1'\omega^t$ and therefore
$$\mathcal{R}\left(\omega^{t+\mathsf{dt}}\right) = \frac{(\omega^t)^T A_1' \Pi_S (I - A_2)\Pi_S A_1' \omega^t}{(\omega^t)^T A_1' \Pi_S A_1' (\omega^t)}\,.$$
If $A_1 = A_2$, then we can finish the proof of this lemma by using Proposition 4.5.5. Thefore, we will assume that $A_1 \neq A_2$.

   We make the following claim.

   Claim 4.5.4. *For S_e^t, R_e^t as in Remark 4.5.2, $f_e(t)$ defined as follows.*
$$f_e(t) \stackrel{\text{def}}{=} \frac{w(e)}{|S_e^t|\,|R_e^t|} \sum_{i \in S_e^t, j \in R_e^t} \left(\omega^t(i) - \omega^t(j)\right)^2$$
is a continuous function of t $\forall t \geqslant 0$.

Proof. This follows from the definition of process. The projection operator Π_S, being a linear operator, is continuous. Being a projection operator, it has operator norm at most 1. For a fixed edge e, and vertex $v \in e$, the rate of change of mass at v due to edge e is at most $\omega^t(v)/d$ (from Definition 4.5.1). Since, v belongs to at most d edges, the total rate of change of mass at v is at most $\omega^t(v)$.

Therefore, for any fix any time t_0 and for every $\varepsilon > 0$,

$$|f_e(t) - f_e(t_0)| \leqslant \varepsilon \qquad \forall\, |t - t_0| < \frac{\varepsilon}{2d}\,.$$

$\square$

We will construct a matrix A such that

$$\mathcal{R}\left(\omega^t\right) \geqslant \frac{(\omega^t)^T (I - A)\omega^t}{(\omega^t)^T (\omega^t)}$$

and

$$\mathcal{R}\left(\omega^{t+\mathsf{dt}}\right) \leqslant \frac{(\omega^t)^T A' \Pi_S (I - A) \Pi_S A' \omega^{t+\mathsf{dt}}}{(\omega^t)^T A' \Pi_S A' (\omega^t)}$$

where $A' = (1 - \mathsf{dt})I + \mathsf{dt}\, A$. This will suffice to prove this lemma by using Proposition 4.5.5.

We will start with an empty graph (i.e. no edges) G on the vertex set V and add weighted edges to it. At the end we will let A be the normalized adjacency matrix of G.

Recall from Remark 4.5.2 that

$$S_e^t \stackrel{\text{def}}{=} \left\{ i \in e : \omega^t(i) = \max_{j \in e} \omega^t(j) \right\} \quad \text{and} \quad R_e^t \stackrel{\text{def}}{=} \left\{ i \in e : \omega^t(i) = \min_{j \in e} \omega^t(j) \right\}.$$

The contribution of e to the numerator of $\mathcal{R}\left(\omega^t\right)$ is

$$\frac{w(e)}{|S_e^t|\,|R_e^t|} \sum_{i \in S_e^t, j \in R_e^t} \left(\omega^t(i) - \omega^t(j)\right)^2 .$$

If $S_e^t \subseteq S_e^{t+\mathsf{dt}}$ and $R_e^t \subseteq R_e^{t+\mathsf{dt}}$, then

$$\begin{aligned}\frac{w(e)}{|S_e^{t+\mathsf{dt}}|\,|R_e^{t+\mathsf{dt}}|} &\sum_{i\in S_e^{t+\mathsf{dt}},j\in R_e^{t+\mathsf{dt}}} \left(\omega^{t+\mathsf{dt}}(i) - \omega^{t+\mathsf{dt}}(j)\right)^2 \\ &= \frac{w(e)}{|S_e^t|\,|R_e^t|} \sum_{i\in S_e^t,j\in R_e^t} \left(\omega^{t+\mathsf{dt}}(i) - \omega^{t+\mathsf{dt}}(j)\right)^2 . \qquad (17)\end{aligned}$$

In this case we add to G the complete weighted bipartite graph on $S_e^t \times R_e^t$ with each edge having weight $w(e)/\left(|S_e^t|\,|R_e^t|\right)$.

Next, we consider the case when $S_e^t \not\subset S_e^{t+\mathsf{dt}}$ for some $e \in E$ (the case $R^t \not\subset R^{t+\mathsf{dt}}$ can be handled in the same way). Let $B \subset E$ be the set of all such edges. By taking dt to be small enough and breaking ties arbitrarily we can assume that $S_e^{t+\mathsf{dt}} \subsetneq S_e^t\ \forall e \in B$. By making dt sufficiently small, we may assume that for each $e \in B$, $\exists v \in S_e^t \setminus S_e^{t+\mathsf{dt}}$ such that $v \notin S^{t+\varepsilon}\ \forall \varepsilon \in (0, \mathsf{dt}]$. We define the following limiting quantities.

$$f_e^{\lim} \stackrel{\text{def}}{=} \lim_{\varepsilon\to 0^+} f_e(t+\varepsilon), \qquad S_e^{\lim} \stackrel{\text{def}}{=} \cap_{\varepsilon>0} S_e^{t+\varepsilon} \qquad R_e^{\lim} \stackrel{\text{def}}{=} \cap_{\varepsilon>0} R_e^{t+\varepsilon} .$$

Then, by construction,

$$S_e^{\lim} = S_e^{t+\mathsf{dt}} \qquad \text{and} \qquad R_e^{\lim} = R_e^{t+\mathsf{dt}} . \qquad (18)$$

Then, from Claim 4.5.4, we get

$$\begin{aligned}f_e(t) = f_e^{\lim} = \frac{w(e)}{|S_e^{\lim}|\,|R_e^{\lim}|} &\sum_{i\in S_e^{\lim},j\in R_e^{\lim}} \left(\omega^t(i) - \omega^t(j)\right)^2 \\ &= \frac{w(e)}{|S_e^{t+\mathsf{dt}}|\,|R_e^{t+\mathsf{dt}}|} \sum_{i\in S_e^{t+\mathsf{dt}},j\in R_e^{t+\mathsf{dt}}} \left(\omega^t(i) - \omega^t(j)\right)^2 . \qquad (19)\end{aligned}$$

In this case we add to G a complete weighted bipartite graph on $S_e^{\lim} \times R_e^{\lim}$ with each edge having weight $w(e)/\left(\left|S_e^{\lim}\right|\left|R_e^{\lim}\right|\right)$.

We add self loops at each vertex of G to make this graph d-regular. And we let A be the normalized adjacency matrix of G. Note that A is also the limit point of $A_{\omega^{t+\varepsilon}}$ (the limit is well defined as, $S_e^{\lim}, R_e^{\lim}$ are well defined as shown above):

$$A = \lim_{\varepsilon \to 0^+} A_{\omega^{t+\varepsilon}} \tag{20}$$

and using (19)

$$(\omega^t)^T (I - A)\omega^t = \sum_{e \in B} f_e^{\lim} + \sum_{e \in E \setminus B} f_e(t) = (\omega^t)^T (I - A_1)\omega^t \,. \tag{21}$$

Therefore,

$$\begin{aligned}
\mathcal{R}\left(\omega^{t+\mathsf{dt}}\right) &= \frac{(\omega^t)^T A_1' \Pi_S (I - A_2) \Pi_S A_1' (\omega^t)}{(\omega^t)^T A_1' \Pi_S A_1' (\omega^t)} \\
&= \frac{(\omega^t)^T A_1' \Pi_S (I - A) \Pi_S A_1' (\omega^t)}{(\omega^t)^T A_1' \Pi_S A_1' (\omega^t)} \qquad \text{(By construction; using (17), (18))} \\
&= \frac{(\omega^t)^T A' \Pi_S (I - A) \Pi_S A' (\omega^t)}{(\omega^t)^T A' \Pi_S A' (\omega^t)} \qquad \text{(From (20))}
\end{aligned}$$

By definition, we have $\omega^t \in S$ and $\Pi_S A' \omega^t \in S$. Using this and $I - A' = \mathsf{dt}(I - A)$ we get

$$\frac{(\omega^t)^T A' \Pi_S (I - A) \Pi_S A' (\omega^t)}{(\omega^t)^T A' \Pi_S A' (\omega^t)} = \frac{1}{\mathsf{dt}} \cdot \frac{(\omega^t)^T (\Pi_S A' \Pi_S)(I - (\Pi_S A' \Pi_S))(\Pi_S A' \Pi_S)(\omega^t)}{(\omega^t)^T (\Pi_S A' \Pi_S)(\Pi_S A' \Pi_S)(\omega^t)} \,.$$

Therefore,

$$\begin{aligned}
\mathcal{R}\left(\omega^{t+\mathsf{dt}}\right) &= \frac{1}{\mathsf{dt}} \frac{(\omega^t)^T (\Pi_S A' \Pi_S)(I - (\Pi_S A' \Pi_S))(\Pi_S A' \Pi_S)(\omega^t)}{(\omega^t)^T (\Pi_S A' \Pi_S)(\Pi_S A' \Pi_S)(\omega^t)} \\
&\leqslant \frac{1}{\mathsf{dt}} \frac{(\omega^t)^T (I - \Pi_S A' \Pi_S)(\omega^t)}{(\omega^t)^T (\omega^t)} \\
&\qquad \text{(Using Proposition 4.5.5 with } \Pi_S A' \Pi_S \text{)} \\
&= \frac{(\omega^t)^T (I - A)(\omega^t)}{(\omega^t)^T (\omega^t)} \qquad \text{(Using } 1 - A' = \mathsf{dt}(I - A) \text{ and } \omega^t \in S \text{)} \\
&= \frac{(\omega^t)^T (I - A_1)(\omega^t)}{(\omega^t)^T (\omega^t)} \qquad \text{(Using (21))} \\
&= \mathcal{R}\left(\omega^t\right)
\end{aligned}$$

□

Proposition 4.5.5. *Let A be a symmetric $n \times n$ matrix with eigenvalues $\alpha_1, \ldots, \alpha_n$ and corresponding eigenvectors $v_1, \ldots, v_n$ such that $A \succeq 0$. Then, for any $X \in \mathbb{R}^n$*

$$\frac{X^T(I-A)X}{X^TX} - \frac{X^TA^T(I-A)AX}{X^TA^TAX} = 2\frac{\sum_{i,j} c_i^2 c_j^2 (\alpha_i - \alpha_j)^2 (\alpha_i + \alpha_j)}{\sum_i c_i^2 \sum_i c_i^2 \alpha_i^2} \geqslant 0$$

where $X = \sum_i c_i v_i$.

Proof. We first note that the eigenvectors of $I - A$ are also $v_1, \ldots, v_n$ with $1 - \alpha_1, \ldots, 1 - \alpha_n$ being the corresponding eigenvalues.

$$\begin{aligned}
&\frac{X^T(I-A)X}{X^TX} - \frac{X^TA^T(I-A)AX}{X^TA^TAX} \\
&= \frac{\sum_i c_i^2(1-\alpha_i)}{\sum_i c_i^2} - \frac{\sum_i c_i^2 \alpha_i^2 (1-\alpha_i)}{\sum_i c_i^2 \alpha_i^2} \\
&= 2\frac{\sum_{i \neq j} c_i^2 c_j^2 \left((1-\alpha_i)\alpha_j^2 + (1-\alpha_j)\alpha_i^2 - (1-\alpha_i)\alpha_i^2 - (1-\alpha_j)\alpha_j^2\right)}{\sum_i c_i^2 \sum_i c_i^2 \alpha_i^2} \\
&= 2\frac{\sum_{i,j} c_i^2 c_j^2 (\alpha_i - \alpha_j)^2 (\alpha_i + \alpha_j)}{\sum_i c_i^2 \sum_i c_i^2 \alpha_i^2}
\end{aligned}$$

□

4.5.1 Eigenvalues in Subspaces

Theorem 4.5.6 (Formal statement of of Theorem 4.2.6)**.** *Given a hypergraph H, for every subspace S of $\mathbb{R}^n$, the operator $\Pi_S L$ has a eigenvector, i.e. there exists a vector $\mathsf{v} \in S$ and a $\gamma \in \mathbb{R}$ such that*

$$\Pi_S L(\mathsf{v}) = \gamma\, \mathsf{v} \qquad \textit{and} \qquad \gamma = \min_{X \in S} \frac{X^T \Pi_S L(X)}{X^T X} .$$

Proof. Fix a subspace S of $\mathbb{R}^n$. Then γ is also fixed as above. We define the set of vectors U_γ as follows.

$$U_\gamma \stackrel{\text{def}}{=} \left\{ X \in S : X^T X = 1 \text{ and } X^T \Pi_S L(X) = \gamma \right\} . \tag{22}$$

From the definition of γ, we get that U_γ is non-empty. Now, the set U_γ could potentially have many vectors. We will show that at least one of them will be an eigenvector. As a warm up, let us first consider the case when $|U_\gamma| = 1$. Let v denote the unique vector in U_γ. We will show that v is an eigenvector of $\Pi_S L$. To see this, we define the unit vector v' as follows.

$$\mathsf{v}' \stackrel{\text{def}}{=} \frac{\Pi_S M(\mathsf{v})}{\|\Pi_S M(\mathsf{v})\|}\,.$$

Since v is the vector in S having the smallest value of $\mathcal{R}(\cdot)$, we get

$$\mathcal{R}(\mathsf{v}) \leqslant \mathcal{R}(\mathsf{v}')\ .$$

But from Lemma 4.5.3(2), we get the $\mathcal{R}(\cdot)$ is a monotonic function, i.e. $\mathcal{R}(\mathsf{v}') \leqslant \mathcal{R}(\mathsf{v})$. Therefore

$$\mathcal{R}(\mathsf{v}) = \mathcal{R}(\mathsf{v}')\ .$$

Therefore, v' also belongs to U_γ. But we assumed that $|U_\gamma| = 1$. Therefore, $\mathsf{v}' = \mathsf{v}$, or in other words v is an eigenvector of $\Pi_S L$.

$$\Pi_S L(\mathsf{v}) = (1 - \|\Pi_S M(\mathsf{v})\|)\,\mathsf{v} = \gamma\,\mathsf{v}\,.$$

The general case when $|U_\gamma| > 1$ requires more work, as the operator L is non-linear. We follow the general idea of the case when $|U_\gamma| = 1$. We let $\omega^0 \stackrel{\text{def}}{=} \mathsf{v}$ for any $\mathsf{v} \in U_\gamma$. We define the set of unit vectors $\{\omega^t\}_{t\in[0,1]}$ recursively as follows (for an infinitesimally small dt).

$$\omega^{t+\mathsf{dt}} \stackrel{\text{def}}{=} \frac{((1-\mathsf{dt})I + \mathsf{dt}\,\Pi_S M) \circ \omega^t}{\|((1-\mathsf{dt})I + \mathsf{dt}\,\Pi_S M) \circ \omega^t\|}\,. \tag{23}$$

As before, we get that

$$\omega^t \in U_\gamma \qquad \forall t \geqslant 0\,. \tag{24}$$

If for any t, $\omega^t = \omega^{t'}$ $\forall t' \in [t, t+\mathsf{dt}]$, then $\omega^t = \omega^{t'}$ $\forall t' \geqslant t$, and we have that ω^t is an eigenvector of $\Pi_S M$, and hence also of $\Pi_S L$ (of eigenvalue γ). Therefore, let us assume that $\omega^t \neq \omega^{t+\mathsf{dt}}$ $\forall t \geqslant 0$.

Let A_ω be the set of support matrices of $\{\omega^t\}_{t\geqslant 0}$, i.e.

$$A_\omega \stackrel{\text{def}}{=} \{A_{\omega^t} : t \geqslant 0\}\ .$$

Note that unlike the set $\{\omega^t\}_{t\geqslant 0}$ which could potentially be of uncountably infinite cardinality, the A_ω is of finite size. A matrix A_X is only determined by the subsets of maximal and minimal vertices (under X) in each hyperedge. Therefore,

$$|A_\omega| \leqslant \left(2^{2r}\right)^m < \infty\,.$$

Now, since $|A_\omega|$ is finite, (using Lemma 4.5.7) there exists $p, q \in [0,1]$, $p < q$ such that

$$A_{\omega^{\mathsf{t}}} = A_{\omega^{\mathsf{p}}} \qquad \forall t \in [p,q]\,.$$

For the sake of brevity let $A \stackrel{\text{def}}{=} A_{\omega^{\mathsf{p}}}$ denote this matrix.

We now show that ω^p is an eigenvector of $\Pi_S L$. From (24), we get that for infinitesimally small dt (in fact anything smaller than $q-p$ will suffice),

$$\mathcal{R}\left(\omega^p\right) - \mathcal{R}\left(\omega^{p+\mathsf{dt}}\right) = 0\,.$$

Let $\alpha_1, \ldots, \alpha_n$ be the eigenvalues of $A' \stackrel{\text{def}}{=} ((1-\mathsf{dt})I + \mathsf{dt}\,A)$ and let $v_1, \ldots, v_n$ be the corresponding eigenvectors. Since A is a stochastic matrix,

$$A \succeq (1-2\mathsf{dt})I \succeq \frac{1}{2}I \qquad \text{or} \qquad \alpha_i \geqslant \frac{1}{2}\ \forall i\,. \tag{25}$$

Let $c_1, \ldots, c_n \in \mathbb{R}$ be appropriate constants such that

$$\omega^p = \sum_i c_i v_i\,.$$

Then using Proposition 4.5.5, we get that

$$\begin{aligned} 0 &= \mathcal{R}\left(\omega^p\right) - \mathcal{R}\left(\omega^{p+\mathsf{dt}}\right) \\ &= \frac{1}{\mathsf{dt}} \cdot \left(\frac{(\omega^p)^T (I - \Pi_S A')\omega^p}{(\omega^p)^T \omega^p} - \frac{(\omega^p)^T A' \Pi_S (I - \Pi_S A') \Pi_S A' \omega^p}{(\omega^p)^T A' \Pi_S A' \omega^p} \right) \\ &= \frac{1}{\mathsf{dt}} 2 \frac{\sum_{i,j} c_i^2 c_j^2 (\alpha_i - \alpha_j)^2 (\alpha_i + \alpha_j)}{\sum_i c_i^2 \sum_i c_i^2 \alpha_i^2}\,. \end{aligned}$$

Since, all $\alpha_i \geqslant 1/2$ (from (25)), the last term can be zero if and only if for some eigenvalue $\alpha \in \{\alpha_i : i \in [n]\}$,

$$c_i \neq 0 \text{ if and only if } \alpha_i = \alpha\,.$$

Or equivalently, ω^p is an eigenvector of A, and $\omega^t = \omega^p\ \forall t \in [p, q]$. Hence, by recursion

$$\omega^t = \omega^p \qquad \forall t \geqslant p\,.$$

Therefore,

$$\Pi_S L(\omega^p) = \left(\frac{1-\alpha}{\mathsf{dt}}\right)\omega^p$$

Since we have already established that $\mathcal{R}\left(\omega^p\right) = \gamma$, this finishes the proof of the theorem.

□

Proposition 4.2.8 follows from Theorem 4.5.6 as a corollary.

Proof of Proposition 4.2.8 . We will prove this by induction on k. The proposition is trivially true of $k = 1$. Let us assume that the proposition holds for $k - 1$. We will show that it holds for k. Recall that v_k is defined as

$$\mathsf{v}_k = \mathsf{argmin}_X \frac{X^T \Pi^{\perp}_{S_{k-1}} L(X)}{X^T \Pi^{\perp}_{S_{k-1}} X}\,.$$

Then from Theorem 4.5.6, we get that v_k is indeed an eigenvector of $\Pi^{\perp}_{S_{k-1}} L$ with eigenvalue

$$\gamma_k = \min_X \frac{X^T \Pi^{\perp}_{S_{k-1}} L(X)}{X^T \Pi^{\perp}_{S_{k-1}} X}\,.$$

□

Lemma 4.5.7. *Let $f : [0, 1] \to \{1, 2, \ldots, k\}$ be any discrete function. Then there exists an interval $(a, b) \subset [0, 1]$, $a \neq b$, such that for some $\alpha \in \{1, 2, \ldots, k\}$*

$$f(x) = \alpha \qquad \forall x \in (a, b)\,.$$

Proof. Let $\upsilon(\cdot)$ denote the standard Lebesgue measure on the real line. Then since f is a discrete function on $[0, 1]$ we have

$$\sum_{i=1}^{k} \upsilon\left(f^{-1}(i)\right) = 1\,.$$

Then, for some $\alpha \in \{1, 2, \ldots, k\}$

$$\upsilon\left(f^{-1}(\alpha)\right) \geqslant \frac{1}{k}\,.$$

Therefore, there is some interval $(a, b) \subset f^{-1}(\alpha)$ such that

$$\upsilon\left((a, b)\right) > 0\,.$$

This finishes the proof of the lemma. □

4.5.2 Upper bounds on the Mixing Time

Theorem 4.5.8 (Restatement of Theorem 4.2.17)**.** *Given a hypergraph $H = (V, E, w)$, for all starting probability distributions $\mu^0 : V \to [0, 1]$, the Hypergraph Dispersion Process (Definition 4.2.9) satisfies*

$$\mathsf{t}_\delta^{\mathsf{mix}}\left(\mu^0\right) \leqslant \frac{\log(n/\delta)}{\gamma_2}\,.$$

Proof. Fix a distribution μ^0 on V. For the sake of brevity, let A_t denote $A_{\mu^{\mathsf{t}}}$ and let A'_t denote $((1 - \mathsf{dt})I + \mathsf{dt}\, A_{\mu^{\mathsf{t}}})$. We first note that

$$A'_t \succeq (1 - 2\mathsf{dt})I + \succeq 0 \qquad \forall t\,. \tag{26}$$

This follows from the fact that A_t being a stochastic matrix, satisfies $I \succeq A_t \succeq -I$. Let $1 \geqslant \alpha_2 \geqslant \ldots \geqslant \alpha_n$ be the eigenvalues of A_t and let $\mathbf{1}/\sqrt{n}, v_2, \ldots, v_n$ be the corresponding eigenvectors. Let $\alpha'_i \stackrel{\text{def}}{=} (1 - \mathsf{dt}) + \mathsf{dt}\, \alpha_i$ for $i \in [n]$ be the eigenvalues of A'_t. Writing μ^t in this eigen-basis, let $c_1, \ldots, c_n \in \mathbb{R}$ be appropriate constants such that $\mu^t = \sum_i c_i v_i$. Since μ^t is a probability distribution on V, its component along the first eigenvector $v_1 = \mathbf{1}/\sqrt{n}$ is

$$c_1 v_1 = \left\langle \mu^t, \frac{\mathbf{1}}{\sqrt{n}} \right\rangle \frac{\mathbf{1}}{\sqrt{n}} = \frac{\mathbf{1}}{n}\,.$$

Then, using the fact that $\alpha'_1 = (1 - \mathsf{dt}) + \mathsf{dt} \cdot 1 = 1$.

$$\mu^{t+\mathsf{dt}} = A'_t\, \mu^t = \sum_{i=1}^{n} \alpha'_i c_i v_i = \frac{\mathbf{1}}{n} + \sum_{i=2}^{n} \alpha'_i c_i v_i\,. \tag{27}$$

Note that at all times $t \geqslant 0$, the component of μ^t along $\mathbf{1}$ (i.e. $c_1 v_1$) remains unchanged. Since for regular hypergraphs $\mu^* = \mathbf{1}/n$,

$$\left\| \mu^{t+\mathsf{dt}} - \mu^* \right\| = \left\| \mu^{t+\mathsf{dt}} - \mathbf{1}/n \right\| = \left\| \sum_{i=2}^{n} \alpha'_i c_i v_i \right\| = \sqrt{\sum_{i=2}^{n} \alpha'^2_i c_i^2}\,. \tag{28}$$

Since all the $\alpha'_i \geqslant 0$ (using (26)) and $\alpha_2 \geqslant \alpha_i \ \forall i \geqslant 2$, $\alpha'^2_2 \geqslant \alpha'^2_i \ \forall i \geqslant 2$. Therefore, from (28)

$$\left\| \mu^{t+\mathsf{dt}} - \mathbf{1}/n \right\| \leqslant \alpha'_2 \sqrt{\sum_{i=2}^{n} c_i^2} = \alpha'_2 \left\| \mu^t - \mathbf{1}/n \right\| \,. \tag{29}$$

We defined γ_2 to the second smallest eigenvalue of L. Therefore, from the definition of L, it follows that $(1 - \gamma_2)$ is the second largest eigenvalue of M. In this context, this implies that

$$\alpha_2 \leqslant 1 - \gamma_2 \,.$$

Therefore, from the definition of α'_2

$$\alpha'_2 = (1 - \mathsf{dt}) + \mathsf{dt}\, \alpha_2 \leqslant (1 - \mathsf{dt}) + \mathsf{dt}\, (1 - \gamma_2) = 1 - \mathsf{dt}\, \gamma_2 \,.$$

Therefore, from (29),

$$\left\| \mu^{t+\mathsf{dt}} - \mathbf{1}/n \right\| \leqslant (1 - \mathsf{dt}\, \gamma_2) \left\| \mu^t - \mathbf{1}/n \right\| \leqslant e^{-\mathsf{dt}\, \gamma_2} \left\| \mu^t - \mathbf{1}/n \right\| \,.$$

Integrating with respect to time, from time 0 to t,

$$\left\| \mu^t - \mathbf{1}/n \right\| \leqslant e^{-\gamma_2 t} \left\| \mu^0 - \mathbf{1}/n \right\| \leqslant 2 e^{-\gamma_2 t} \,.$$

Therefore, for $t \geqslant \log(n/\delta)/\gamma_2$,

$$\left\| \mu^t - \mathbf{1}/n \right\| \leqslant \frac{\delta}{\sqrt{n}} \qquad \text{and} \qquad \left\| \mu^t - \mathbf{1}/n \right\|_1 \leqslant \sqrt{n} \cdot \left\| \mu^t - \mathbf{1}/n \right\| \leqslant \delta \,.$$

Therefore,

$$\mathsf{t}^{\mathsf{mix}}_{\delta} \left(\mu^0 \right) \leqslant \frac{\log(n/\delta)}{\gamma_2} \,.$$

□

Remark 4.5.9. Theorem 4.2.17 can also be proved directly by using Lemma 4.5.3, but we believe that this proof is more intuitive.

4.5.3 Lower bounds on Mixing Time

Next we prove Theorem 4.2.18

Theorem 4.5.10 (Restatement of Theorem 4.2.18)**.** *Given hypergraph $H = (V, E, w)$, there exists a probability distribution μ^0 on V such that $\|\mu^0 - \mathbf{1}/n\|_1 \geqslant 1/2$ and*

$$\mathsf{t}_\delta^{\mathsf{mix}}\left(\mu^0\right) \geqslant \frac{\log(1/\delta)}{16\,\gamma_2}\,.$$

In an attempt to motivate why Theorem 4.2.18 is true, we first prove the following (weaker) lower bound.

Theorem 4.5.11. *Given a hypergraph $H = (V, E, w)$, there exists a probability distribution μ^0 on V such that $\|\mu^0 - \mathbf{1}/n\|_1 \geqslant 1/2$ and*

$$\mathsf{t}_\delta^{\mathsf{mix}}\left(\mu^0\right) \geqslant \frac{\log(1/\delta)}{\phi_H}\,.$$

Proof Sketch. Let $S \subset V$ be the set which has the least value of $\phi_H(S)$. Let $\mu^0 : V \to [0, 1]$ be the probability distribution supported on S that is stationary on S, i.e.

$$\mu^0(i) = \begin{cases} \frac{1}{|S|} & i \in S \\ 0 & i \notin S \end{cases}$$

Then, for an infinitesimal time duration dt, only the edges in $E(S, \bar{S})$ will be active in the dispersion process, and for each edge $e \in E(S, \bar{S})$, the vertices in $e \cap S$ will be sending $1/d$ fraction of their mass to the vertices in $e \cap \bar{S}$. Therefore,

$$\mu^0(S) - \mu^{\mathsf{dt}}(S) = \sum_{e \in E(S,\bar{S})} \frac{1}{d} \cdot \frac{1}{|S|}\,\mathsf{dt} = \frac{\left|E(S,\bar{S})\right|}{d\,|S|}\,\mathsf{dt} = \phi_H\,\mathsf{dt}\,.$$

In other words, mass escapes from S at the rate of ϕ_H initially. It is easy to show that the rate at which mass escapes from S is a non-increasing function of time. Therefore, it will take at least $\Omega(1/\phi_H)$ units of time to remove $1/2$ of the mass from the S. Thus the lower bound follows.

□

Now, we will work towards proving Theorem 4.2.18.

Lemma 4.5.12. *For any hypergraph $H = (V, E, w)$ and any probability distribution μ^0 on V, let $\alpha = \|\mu^0 - \mathbf{1}/n\|^2$. Then*

$$\mathsf{t}_\delta^{\mathsf{mix}}\left(\mu^0\right) \geqslant \frac{\log(\alpha/\delta)}{4\mathcal{R}\left(\mu^0 - \mathbf{1}/n\right)}\,.$$

Proof. For a probability distribution μ^t on V, let ω^t be its component orthogonal to $\mu^* = \mathbf{1}/\sqrt{n}$

$$\omega^t \stackrel{\text{def}}{=} \mu^t - \left\langle \mu^t, \frac{\mathbf{1}}{\sqrt{n}} \right\rangle \frac{\mathbf{1}}{\sqrt{n}} = \mu^t - \frac{\mathbf{1}}{n}\,.$$

As we saw before (in (27)), only ω^t, the component of μ^t orthogonal to $\mathbf{1}$, changes with time; the component of μ^t along $\mathbf{1}$ does not change with time. For the sake of brevity, let $\lambda = \mathcal{R}\left(\mu^0 - \mathbf{1}/n\right)$. Then, using Lemma 4.5.3(2) and the definition of ω, we get that

$$\mathcal{R}\left(\omega^t\right) \leqslant \mathcal{R}\left(\omega^0\right) = \lambda \qquad \forall t \geqslant 0\,.$$

Now, using this and Lemma 4.5.3(1) we get

$$\frac{\mathrm{d}\left\|\omega^t\right\|^2}{\left\|\omega^t\right\|^2} = -2\,\mathcal{R}\left(\omega^t\right)\mathsf{dt} \geqslant -2\lambda\,\mathsf{dt}\,.$$

Integrating with respect to time from 0 to t, we get

$$\log\left\|\omega^t\right\|^2 - \log\left\|\omega^0\right\|^2 \geqslant -2\lambda\, t\,.$$

Therefore

$$e^{-2\lambda t} \leqslant \frac{\left\|\omega^t\right\|^2}{\left\|\omega^0\right\|^2} = \frac{\left\|\mu^t - \mathbf{1}/n\right\|^2}{\left\|\mu^0 - \mathbf{1}/n\right\|^2} = \frac{\left\|\mu^t - \mathbf{1}/n\right\|^2}{\alpha} \qquad \forall t \geqslant 0\,.$$

Hence

$$\left\|\mu^t - \mathbf{1}/n\right\|_1 \geqslant \left\|\mu^t - \mathbf{1}/n\right\| \geqslant 2\delta \qquad \text{for } t \leqslant \frac{\log(\alpha/\delta)}{4\lambda}\,.$$

Thus

$$\mathsf{t}_\delta^{\mathsf{mix}}\left(\mu^0\right) \geqslant \frac{\log(\alpha/\delta)}{4\mathcal{R}\left(\mu^0 - \mathbf{1}/n\right)}\,.$$

□

Lemma 4.5.13. *Given a hypergraph $H = (X, E)$ and a vector $X \in \mathbb{R}^V$, there exists a polynomial time algorithm to compute a probability distribution μ on V satisfying*

$$\|\mu - \boldsymbol{1}/n\|_1 \geqslant \frac{1}{2} \qquad \text{and} \qquad \mathcal{R}(\mu - \boldsymbol{1}/n) \leqslant 4\mathcal{R}(X - \langle X, \boldsymbol{1}\rangle \boldsymbol{1}/n) .$$

Proof. For the sake of building intuition, let us consider the case when $\langle X, \mathbf{1}\rangle = 0$. As a first attempt, one might be tempted to consider the vector $\mathbf{1}/n + X$. This vector might not be a probability distribution if $X(i) < -1/n$ for some coordinate i. A simple fix for this would to consider the vector $\mu' \stackrel{\text{def}}{=} \mathbf{1}/n + X/(n\|X\|_\infty)$. This is clearly a probability distribution on the vertices, but

$$\left\|\mu' - \frac{\mathbf{1}}{n}\right\|_1 = \left\|\frac{X}{n\|X\|_\infty}\right\|_1 = \frac{\|X\|_1}{n\|X\|_\infty}$$

and $\|X\|_1/(n\|X\|_\infty) \ll 1/2$ depending on X, for e.g. when X is very sparse. Therefore, we must proceed differently.

Since we only care about $\mathcal{R}(X - \langle X, \mathbf{1}\rangle \mathbf{1}/n)$, w.l.o.g. we may assume that $|\mathsf{supp}(X^+)| = |\mathsf{supp}(X^-)|$ by simply setting $X := X + c\mathbf{1}$ for some appropriate constant c. W.l.o.g. we may also assume that $\|X^+\| \geqslant \|X^-\|$. Let ω be the component of X^+ orthogonal to $\mathbf{1}$

$$\omega \stackrel{\text{def}}{=} X^+ - \frac{\langle X^+, \mathbf{1}\rangle}{n}\mathbf{1} = X^+ - \frac{\|X^+\|_1}{n}\mathbf{1} .$$

By definition, we get that $\langle \omega, \mathbf{1}\rangle = 0$. Now,

$$\|\omega\|_1 \geqslant \sum_{i \in \mathsf{supp}(\omega^-)} |\omega(i)| \geqslant \sum_{i \in \mathsf{supp}(X^-)} |\omega(i)| \geqslant \frac{n}{2}\frac{\|X^+\|_1}{n} \geqslant \frac{\|X^+\|_1}{2} . \tag{30}$$

We now define the probability distribution μ on V as follows.

$$\mu \stackrel{\text{def}}{=} \frac{\mathbf{1}}{n} + \frac{\omega}{2\|\omega\|_1} .$$

We now verify that μ is indeed a probability distribution, i.e. $\mu(i) \geqslant 0 \ \forall i \in V$. If vertex $i \in \mathsf{supp}(X^+)$, then clearly $\mu(i) \geqslant 0$. Lets consider an $i \in \mathsf{supp}(X^-)$.

$$\frac{\omega(i)}{2\|\omega\|_1} = \frac{-|X^+|/n}{2\|\omega\|_1} \geqslant -\frac{1}{n} \qquad \text{(Using (30))} .$$

Therefore, $\mu(i) = 1/n + \omega(i)/(2\|\omega\|_1) \geqslant 0$ in this case as well. Thus, μ is a probability distribution on V. Next, we work towards bounding $\mathcal{R}(\mu - \mathbf{1}/n)$.

$$\sum_e w(e) \max_{i,j\in e} (\mu(i) - \mu(j))^2 = \frac{1}{4\|\omega\|_1^2} \cdot \sum_e w(e) \max_{i,j\in e} (\omega(i) - \omega(j))^2$$
$$\leqslant \frac{1}{4\|\omega\|_1^2} \cdot \sum_e w(e) \max_{i,j\in e} (X(i) - X(j))^2 . \quad (31)$$

We now bound $\|\omega\|_2$.

$$\|\omega\|_2^2 = \left\|X^+ - \left\langle X^+, \mathbf{1}\right\rangle \mathbf{1}/n\right\|^2 = \left\|X^+\right\|^2 - \frac{\langle X^+, \mathbf{1}\rangle^2}{n} = \left\|X^+\right\|^2 - \frac{\|X^+\|_1^2}{n} . \quad (32)$$

Since $|\mathsf{supp}(X^+)| \leqslant n/2$,

$$\left\|X^+\right\|_1^2 \leqslant \frac{n}{2} \left\|X^+\right\|^2 .$$

Combining this with (32), and using our assumption that $\|X^+\| \geqslant \|X^-\|$, we get

$$\|\omega\|_2^2 = \left\|X^+\right\|^2 - \frac{\|X^+\|_1^2}{n} \geqslant \frac{\|X^+\|^2}{2} \geqslant \frac{\|X\|^2}{4} .$$

Therefore,

$$\|\mu - \mathbf{1}/n\|^2 = \frac{\|\omega\|^2}{4\|\omega\|_1^2} \geqslant \frac{1}{4\|\omega\|_1^2} \cdot \frac{\|X\|^2}{4} \geqslant \frac{1}{4\|\omega\|_1^2} \cdot \frac{\|X - \langle X, \mathbf{1}\rangle \mathbf{1}/n\|^2}{4} . \quad (33)$$

Therefore, using (31) and (33), we get

$$\mathcal{R}(\mu - \mathbf{1}/n) \leqslant 4\mathcal{R}(X - \langle X, \mathbf{1}\rangle \mathbf{1}/n)$$

and by construction

$$\|\mu - \mathbf{1}/n\|_1 = \left\|\frac{\omega}{2\|\omega\|_1}\right\|_1 = \frac{1}{2}$$

□

We are now ready to prove Theorem 4.2.18.

Proof of Theorem 4.2.18.

Let $X = \mathsf{v}_2$. Using Lemma 4.5.13, there exists a probability distribution μ on V such that

$$\|\mu - \mathbf{1}/n\|_1 \geqslant \frac{1}{2} \qquad \text{and} \qquad \mathcal{R}(\mu - \mathbf{1}/n) \leqslant 4\gamma_2$$

and for this distribution μ, using Lemma 4.5.12, we get

$$\mathsf{t}_{\delta}^{\mathsf{mix}}(\mu) \geqslant \frac{\log(1/\delta)}{16\,\gamma_2}\,.$$

□

Remark 4.5.14. The distribution in Theorem 4.2.18 is not known to be computable in polynomial time. We can compute a probability distribution μ in polynomial time such

$$\|\mu - \mathbf{1}/n\|_1 \geqslant \frac{1}{2} \qquad \text{and} \qquad \mathsf{t}_{\delta}^{\mathsf{mix}}(\mu) \geqslant \frac{\log(1/\delta)}{c\gamma_2 \log r}$$

for some absolute constant c. Using Theorem 6.1.5, we get a vector $X \in \mathbb{R}^n$ such that $\mathcal{R}(X) \leqslant c_1\gamma_2 \log r$ for some absolute constant c_1. Using Lemma 4.5.13, we compute a probability distribution ν on V such that

$$\|\nu - \mathbf{1}/n\|_1 \geqslant \frac{1}{2} \qquad \text{and} \qquad \mathcal{R}(\nu - \mathbf{1}/n) \leqslant 4c_1\gamma_2 \log r\,.$$

and for this distribution ν, using Lemma 4.5.12, we get

$$\mathsf{t}_{\delta}^{\mathsf{mix}}(\nu) \geqslant \frac{\log(1/\delta)}{4c_1\gamma_2 \log r}\,.$$

4.6 Higher Eigenvalues and Hypergraph Expansion

In this section we will prove Theorem 4.2.16 and Theorem 4.2.15.

4.6.1 Small Set Expansion

Theorem 4.6.1 (Formal Statement of Theorem 4.2.16)**.** *There exists an absolute constant C such that every hypergraph $H = (V, E, w)$ and parameter $k < |V|$, there exists a set $S \subset V$ such that $|S| \leqslant 16\,|V|/k$ satisfying*

$$\phi(S) \leqslant C \min\left\{\sqrt{r \log k},\, k \log k \log\log k \sqrt{\log r}\right\} \sqrt{\gamma_k}$$

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

Our proof will be via a simple randomized polynomial time algorithm (Algorithm 4.6.2) to compute a set S satisfying the conditions of the theorem. Let $t_{1/k}$ denote the $(1/k)^{th}$ cap of the standard normal random variables, i.e., $t_{1/k} \in R$ is the number such that for a standard normal random variable X, $\mathbb{P}\left[X \geqslant t_{1/k}\right] = 1/k$.

Algorithm 4.6.2.

1. **Spectral Embedding**. We first construct a mapping of the vertices in $\mathbb{R}^k$ using the first k eigenvectors. We map a vertex $i \in V$ to the vector u_i defined as follows.
$$u_i(l) = \frac{1}{\sqrt{d_i}} \mathsf{v}_l(i) .$$
In other words, we map the vertex i to the vector formed by taking the i^{th} coordinate from the first k eigenvectors.

2. **Random Projection**. We sample a random Gaussian vector $g \sim \mathcal{N}(0,1)^k$ and define the vector $X \in \mathbb{R}^n$ as follows.
$$X(i) \stackrel{\text{def}}{=} \begin{cases} \|u_i\|^2 & \text{if } \langle \tilde{u}_i, g \rangle \geqslant t_{1/k} \\ 0 & \text{otherwise} \end{cases} .$$

3. **Sweep Cut**. Sort the entries of the vector X in decreasing order and output the level set having the least expansion (See Proposition 4.4.2).

Figure 8: Rounding Algorithm for Hypergraph Small set Expansion

We prove some basic facts about the Spectral Embedding (Lemma 4.6.3). The analogous facts for graphs are well known (folklore).

Lemma 4.6.3 (Spectral embedding)**.**

1.
$$\frac{\sum_{e \in E} \max_{i,j \in e} w(e) \|u_i - u_j\|^2}{\sum_i d_i \|u_i\|^2} \leqslant \gamma_k .$$

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

3.

$$\sum_{i,j \in V} d_i d_j \langle u_i, u_j \rangle^2 = k\,.$$

Proof. The proof of this is identical to the proof of Lemma 3.3.1.

□

We will use the following variant of Lemma 3.3.13.

Lemma 4.6.4. *Given two unit vectors $\tilde{u}_i, \tilde{u}_j \in \mathbb{R}^n$,*

$$\mathop{\mathbb{P}}_{g \sim \mathcal{N}(0,1)^n} \left[\langle \tilde{u}_i, g \rangle \geqslant t_{1/k} \text{ and } \langle \tilde{u}_j, g \rangle \geqslant t_{1/k} \right] \leqslant \frac{1}{k} \langle \tilde{u}_i, \tilde{u}_j \rangle^2 + \frac{1}{k^2}\,.$$

Main Analysis. To prove that Algorithm 4.6.2 outputs a set which meets the requirements of Theorem 4.6.1, we will show that the vector X meets the requirements of Proposition 4.4.3. We will need an upper bound on the numerator of *cut-value* of the vector X (Lemma 4.6.5), and a lower bound on the denominator of the *cut-value* of the vector X (Lemma 4.6.6).

Lemma 4.6.5.

$$\mathbb{E}\left[\sum_{e \in E} w(e) \max_{i,j \in e} |X_i - X_j| \right] \leqslant \tilde{\mathcal{O}}\left(k \sqrt{\gamma_k \log r} \right)\,.$$

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

$$\mathbb{E}\left[\max_{i,j \in e} |X_i - X_j| \right] \leqslant \max_{i,j \in e} \left| \|u_i\|^2 - \|u_j\|^2 \right| \mathop{\mathbb{P}}_{g \sim \mathcal{N}(0,1)^k} \left[\langle \tilde{u}_i, g \rangle \geqslant t_{1/k} \ \forall i \in e \right]$$
$$+ \max_{i \in e} \|u_i\|^2 \mathop{\mathbb{P}}_{g \sim \mathcal{N}(0,1)^n} \left[\langle \tilde{u}_i, g \rangle \geqslant t_{1/k} \text{ and } \langle \tilde{u}_j, g \rangle < t_{1/k} \text{ for some } i, j \in [r] \right] \quad (34)$$

The first term can be bounded by

$$\frac{1}{k} \max_{i,j \in e} \left| \|u_i\|^2 - \|u_j\|^2 \right| \leqslant \frac{1}{k} \max_{i,j \in e} \|u_i - u_j\| \cdot \|u_i + u_j\| \leqslant$$
$$2\frac{1}{k} \max_{i,j \in e} \|u_i - u_j\| \max_{i \in e} \|u_i\|\,. \quad (35)$$

Now for a hyperedge $e \in E$, using Lemma 2.5.8,

$$\mathop{\mathbb{P}}_{g \sim \mathcal{N}(0,1)^n} \left[\langle \tilde{u}_i, g \rangle \geqslant t_{1/k} \text{ and } \langle \tilde{u}_j, g \rangle < t_{1/k} \text{ for some } i,j \in e \right] \leqslant c_1 \frac{k \log k \log \log k}{k} \max_{i,j \in e} \|\tilde{u}_i - \tilde{u}_j\| \sqrt{\log r} \,. \quad (36)$$

To bound the second term in (34), we will divide the edge set E into two parts E_1 and E_2 as follows.

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

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. For a hyperedge $e \in E_1$, using Proposition 2.6.2 and (36), the second term in (34) can be bounded by

$$\frac{2c_1 k \log k \log \log k}{k} \max_{l \in e} \|u_l\|^2 \max_{i,j \in e} \frac{\|u_i - u_j\|}{\sqrt{\|u_i\|^2 + \|u_j\|^2}} \sqrt{\log r} \leqslant \frac{2c_1 k \log k \log \log k}{k} \max_{l \in e} \|u_l\| \max_{i,j \in e} \|u_i - u_j\| \sqrt{\log r} \,. \quad (37)$$

Let us analyze the edges in E_2. Fix any $e \in E_2$. Let $e = \{u_1, \ldots, u_r\}$ such that $\|u_1\| \geqslant \|u_2\| \geqslant \ldots \geqslant \|u_r\|$. Then from the definition of E_2 we have that

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

Rearranging, we get

$$\begin{aligned} \|u_1\|^2 &\leqslant 2 \left(\|u_1\|^2 - \|u_r\|^2 \right) = 2 \langle u_1 - u_r, u_1 + u_r \rangle \leqslant 2 \|u_1 + u_r\| \, \|u_1 - u_r\| \\ &\leqslant 2\sqrt{2} \max_{i \in e} \|u_i\| \max_{i,j \in e} \|u_i - u_j\| \,. \end{aligned}$$

Therefore for an edge $e \in E_2$, using this and (36), the second term in (34) can be bounded by

$$\frac{4}{k} \max_{i \in e} \|u_i\| \max_{i,j \in e} \|u_i - u_j\| \,. \quad (38)$$

Using (34), (35), (37) and(38) we get

$$\mathbb{E} \left[\max_{i,j \in e} |X_i - X_j| \right] \leqslant \frac{8c_1 k \log k \log \log k}{k} \max_{l \in e} \|u_l\| \max_{i,j \in e} \|u_i - u_j\| \sqrt{\log r} \,. \quad (39)$$

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{e\in E} w(e)\max_{i,j\in e}|X_i - X_j|\right] \\
&\leqslant \frac{8c_1 k\log k\log\log k\sqrt{\log r}}{k}\sum_{e\in E} w(e)\max_{i\in e}\|u_i\|\max_{i,j\in e}\|u_i - u_j\| \\
&\leqslant \frac{8c_1 k\log k\log\log k\sqrt{\log r}}{k}\sqrt{\sum_{e\in E} w(e)\max_{i\in e}\|u_i\|^2}\sqrt{\sum_{e\in E} w(e)\max_{i,j\in e}\|u_i - u_j\|^2} \\
&\leqslant \frac{8c_1 k\log k\log\log k\sqrt{\log r}}{k}\sqrt{\sum_{i\in V} d_i\|u_i\|^2}\sqrt{\sum_{e\in E} w(e)\max_{i,j\in e}\|u_i - u_j\|^2} \\
&\leqslant 8c_1 k\log k\log\log k\sqrt{\gamma_k\log r} \qquad \text{(Using Lemma 4.6.3)}
\end{aligned}
$$

□

Lemma 4.6.6.

$$
\mathbb{P}\left[\sum_{i\in V} d_i X_i > \frac{1}{2}\right] \geqslant \frac{1}{8}.
$$

Proof. For the sake of brevity, we define $D \stackrel{\text{def}}{=} \sum_{i\in V} d_i X_i$. We first bound $\mathbb{E}[D]$ as follows.

$$
\begin{aligned}
\mathbb{E}[D] &= \sum_{i\in V} d_i\|u_i\|^2 \mathop{\mathbb{P}}_{g\sim\mathcal{N}(0,1)^k}\left[\langle\tilde{u}_i, g\rangle \geqslant t_{1/k}\right] \\
&= \sum_{i\in V} d_i\|u_i\|^2\cdot\frac{1}{k} && \text{(From the definition of } t_{1/k}) \\
&= k\cdot\frac{1}{k} = 1 && \text{(Using Lemma 4.6.3)}\,.
\end{aligned}
$$

Next we bound the variance of D.

$$
\begin{aligned}
\mathbb{E}\left[D^2\right] &= \sum_{i,j} d_i d_j\|u_i\|^2\|u_j\|^2\,\mathbb{P}\left[\langle\tilde{u}_i, g\rangle > t_{1/k} \text{ and } \langle\tilde{u}_i, g\rangle > t_{1/k}\right] \\
&\leqslant \sum_{i,j} d_i d_j\|u_i\|^2\|u_j\|^2\left(\frac{1}{k}\langle\tilde{u}_i,\tilde{u}_j\rangle^2 + \frac{1}{k^2}\right) && \text{(Using Lemma 4.6.4)} \\
&= \frac{1}{k}\sum_{i,j} d_i d_j\langle u_i, u_j\rangle^2 + \frac{1}{k^2}\left(\sum_i d_i\|u_i\|^2\right)^2 \\
&= \frac{1}{k}\cdot k + \frac{1}{k^2}\cdot k^2 = 2 \qquad \text{(Using Lemma 4.6.3)}\,.
\end{aligned}
$$

Since D is a non-negative random variable, we get using the Paley-Zygmund inequality (Fact 2.5.2) that

$$\mathbb{P}\left[D \geqslant \frac{1}{2}\,\mathbb{E}\left[D\right]\right] \geqslant \left(\frac{1}{2}\right)^2 \frac{\mathbb{E}\left[D\right]^2}{\mathbb{E}\left[D^2\right]} = \frac{1}{4}\cdot\frac{1}{2} = \frac{1}{8}\,.$$

This finishes the proof of the lemma.

□

We are now ready finish the proof of Theorem 4.6.1.

Proof of Theorem 4.6.1. By definition of Algorithm 4.6.2,

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

Therefore, by Markov's inequality,

$$\mathbb{P}\left[|\mathsf{supp}(X)| \leqslant 16\frac{n}{k}\right] \geqslant 1 - \frac{1}{16}\,. \tag{40}$$

Using Markov's inequality and Lemma 4.6.5,

$$\mathbb{P}\left[\sum_{e\in E}\max_{i,j\in e}|X_i - X_j| \leqslant 256 c_1 k \log k \log\log k \sqrt{\gamma_k \log r}\right] \geqslant 1 - \frac{1}{32}\,. \tag{41}$$

Therefore, using a union bound over (40), (41) and Lemma 4.6.5, we get that

$$\mathbb{P}\left[\frac{\sum_{e\in E} w(e)\max_{i,j\in e}|X_i - X_j|}{\sum_i d_i X_i} \leqslant \tilde{\mathcal{O}}\left(k\sqrt{\gamma_k \log r}\right) \text{ and } |\mathsf{supp}(X)| \leqslant 16\frac{n}{k}\right] \geqslant \frac{1}{32}\,.$$

Invoking Proposition 4.4.3 on this vector X, we get that with probability at least $1/32$, Algorithm 4.6.2 outputs a set S such that

$$\phi(S) \leqslant \tilde{\mathcal{O}}\left(k\sqrt{\gamma_k \log r}\right) \qquad \text{and} \qquad |S| \leqslant 16\frac{n}{k}\,. \tag{42}$$

This finishes the proof of the theorem. □

4.6.2 Hypergraph Multi-partition

In this section we only give a sketch of the proof of Theorem 4.2.15, as this theorem can be proven by essentially using Theorem 4.6.1 and some ideas studied in [70, 73].

Theorem 4.6.7 (Restatement of Theorem 4.2.15). *For any hypergraph $H = (V, E, w)$ and any integer $k < |V|$, there exists a k-partition of V into $\{S_1, \ldots, S_k\}$ such that*

$$\max_{i \in [k]} \phi(S_i) \leqslant \mathcal{O}\left(k^4 \sqrt{\gamma_k \log r}\right) .$$

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

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

Proof Sketch. The first part of the theorem can be proved in a manner similar to Theorem 4.6.1, additionally using techniques from [70]. As before, we will start with the spectral embedding and then round it to get k-partition where each piece has small expansion (Algorithm 4.6.8). Note that Algorithm 4.6.8 can be viewed as a recursive application of Algorithm 4.6.2; the algorithm computes a "small" set having small expansion, removes it and recurses on the remaining graph.

Note that step 3a of Algorithm 4.6.8 is somewhat different from step 2 of Algorithm 4.6.2. Nevertheless, with some more work, we can bound the expansion of the set obtained at the end of step 3b by $\mathcal{O}\left(k^3 \sqrt{\gamma_k \log r}\right)$. The proof of this bound on expansion follows from stronger forms of Lemma 4.6.5 and Lemma 4.6.6.

Once we have this, we can finish the proof of this theorem in a manner similar to [70]. [70] studied k-partitions in graphs and gave an alternate proof of the graph version of this theorem (Theorem 4.2.14). They implicitly show how to use an algorithm for computing small-set expansion to compute a k-partition in graphs where each piece has small expansion. A similar analysis can be used for hypergraphs as well, but incurs an additional factor of $\mathcal{O}(\min\{r, k\})$ in the bound on the expansion of the sets.

□

4.7 Reduction from Vertex Expansion in Graphs to Hypergraph Expansion

Theorem 4.7.1 (Restatement of Theorem 4.2.19)**.** *Given a graph $G = (V, E)$ of maximum degree d and minimum degree $c_1 d$ (for some constant c_1), there exists a polynomial time computable hypergraph $H = (V, E')$ on the same vertex set having the hyperedges of cardinality at most $d+1$ such that for all sets $S \subset V$,*

$$c_1 \phi_H(S) \leqslant \frac{1}{d} \cdot \Phi^{\mathsf{V}}(S) \leqslant \phi_H(S)\,.$$

Proof. We present the reduction as follows (Figure 10).

By construction, all hyperedges in E' have cardinality at most $d+1$. Fix an arbitrary set $S \subset V$.

We first show that $\Phi^{\mathsf{V}}(S) \leqslant d\phi_H(S)$. Consider the vertices $N^{\mathsf{in}}(S)$. Each vertex in $v \in N^{\mathsf{in}}(S)$ has a neighbor, say u, in $\bar{S}$. Therefore the hyperedge $\{v\} \cup N^{\mathsf{out}}(\{v\})$ is cut by S in H. Similarly, for each vertex $v \in N^{\mathsf{out}}(S)$, the hyperedge $\{v\} \cup N^{\mathsf{out}}(\{v\})$ is cut by S in H. By construction it follows that all these hyperedges are disjoint. Therefore,

$$\Phi^{\mathsf{V}}(S) = \frac{\left|N^{\mathsf{in}}(S)\right| + \left|N^{\mathsf{out}}(S)\right|}{|S|} \leqslant d \cdot \frac{\left|E_H(S, \bar{S})\right|}{d\,|S|} \leqslant d\phi_H(S)\,.$$

Now we verify that $\phi_H(S) \leqslant \Phi^{\mathsf{V}}(S)/(c_1 d)$. For any hyperedge $(\{v\} \cup N^{\mathsf{out}}(\{v\})) \in E_H(S, \bar{S})$, the vertex v has to belong to either $N^{\mathsf{in}}(S)$ or $N^{\mathsf{out}}(S)$. Therefore,

$$\phi_H(S) \leqslant \frac{\left|E_H(S, \bar{S})\right|}{c_1 d\,|S|} \leqslant \frac{\left|N^{\mathsf{in}}(S)\right| + \left|N^{\mathsf{out}}(S)\right|}{c_1 d\,|S|} = \frac{1}{c_1 d}\Phi^{\mathsf{V}}(S)\,.$$

□

4.8 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 4.8.1. We define λ_1, the largest eigenvalue of a tensor A as follows.

$$\lambda_1 \stackrel{\text{def}}{=} \max_{X \in \mathbb{R}^n} \frac{\sum_{i_1,i_2,\ldots,i_r} A_{i_1 i_2 \ldots i_r} X_{i_1} X_{i_2} \ldots X_{i_r}}{\|X\|_\square}$$

$$v_1 \stackrel{\text{def}}{=} \mathsf{argmax}_{X \in \mathbb{R}^n} \frac{\sum_{i_1,i_2,\ldots,i_r} A_{i_1 i_2 \ldots i_r} X_{i_1} X_{i_2} \ldots X_{i_r}}{\|X\|_\square}$$

We inductively define successive eigenvalues $\lambda_2 \geqslant \lambda_3 \geqslant \ldots$ as follows.

$$\lambda_k \stackrel{\text{def}}{=} \max_{X \perp \{v_1,\ldots,v_{k-1}\}} \frac{\sum_{i_1,i_2,\ldots,i_r} A_{i_1 i_2 \ldots i_r} X_{i_1} X_{i_2} \ldots X_{i_r}}{\|X\|_\square}$$

$$v_k \stackrel{\text{def}}{=} \mathsf{argmax}_{x \perp \{v_1,\ldots,v_{k-1}\}} \frac{\sum_{i_1,i_2,\ldots,i_r} A_{i_1 i_2 \ldots i_r} X_{i_1} X_{i_2} \ldots X_{i_r}}{\|X\|_\square}$$

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 any 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 related to expansion in a Cheeger-like manner.

Proposition 4.8.2. *For any $k \in \mathbb{Z}_{\geqslant 0}$, 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.

$$H_2 \stackrel{\text{def}}{=} (V(A) \cup V(B), (E(A) \setminus \{a\}) \cup (E(B) \setminus \{b\}) \cup \{(a_1 \cup b_1), (a_2 \cup b_2)\})$$

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 the -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. □

4.9 An Exponential Time Algorithm for computing Eigenvalues

Theorem 4.9.1. *Given a hypergraph $H = (V, E, w)$, there exists an algorithm running in time $\tilde{\mathcal{O}}\left(2^{rm}\right)$ which outputs all eigenvalues and eigenvectors of M.*

Proof. Let X be an eigenvector M with eigenvalue γ. Then

$$\gamma X = M(X) = A_X X \,.$$

Therefore, X is also an eigenvector of A_X. Therefore, the set of eigenvalues of M is a subset of the set of eigenvalues of all the support matrices $\{A_X : X \in \mathbb{R}^n\}$. Note that a support matrix A_X is only determined by the subsets of maximal and minimal vertices (under X) in each hyperedge. Therefore,

$$|\{A_X : X \in \mathbb{R}^n\}| \leqslant (2^r)^m \,.$$

Therefore, we can compute all the eigenvalues and eigenvectors of M by enumerating over all 2^{rm} matrices.

□

4.10 Conclusion

We introduced a new hypergraph Markov operator generalizing the random-walk operator on graphs. We studied the eigenvalues of this operator, and showed that we can prove numerous relations between them and the combinatorial properties of graphs. All such relations generalize the corresponding relations for graphs. However, many open problems remain. In short, we ask what properties of graphs and random walks generalize to hypergraphs and this Markov operator? We pose here two concrete open problems.

Problem 4.10.1 (K SPARSE-CUTS)**.** Does every hypergraph $H = (V, E)$, for every parameter $k \in [n]$ have k disjoint non-empty subsets, say $S_1, \ldots, S_k$, such that

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

Problem 4.10.2 (SMALL SET EXPANSION)**.** Does every hypergraph $H = (V, E)$, for every parameter $k \in [n]$ have a set, say S, of size at most $n/k^{\Omega(1)}$ and

$$\phi(S) \leqslant \mathcal{O}\left(\sqrt{\gamma_k \log_k n}\right)?$$

The results in this chapter appear in [69].

Algorithm 4.6.8. Define $k' \stackrel{\text{def}}{=} k^2$.

1. Initialize $t := 1$ and $V_t := V$ and $C := \phi$.
2. **Spectral Embedding**. We first construct a mapping of the vertices in $\mathbb{R}^k$ using the first k eigenvectors. We map a vertex $i \in V$ to the vector u_i defined as follows.
$$u_i(l) = \frac{1}{\sqrt{d_i}} \mathsf{v}_l(i) .$$
3. While $l \leqslant 100k^3$
   (a) **Random Projection**. We sample a random Gaussian vector $g \sim \mathcal{N}(0,1)^k$ and define the vector $X \in \mathbb{R}^n$ as follows.
$$X(i) \stackrel{\text{def}}{=} \begin{cases} \|u_i\|^2 & \text{if } \langle \tilde{u}_i, g \rangle \geqslant t_{1/k'} \text{ and } i \in V_l \\ 0 & \text{otherwise} \end{cases} .$$
   (b) **Sweep Cut**. Sort the entries of the vector X in decreasing order and compute the set S having the least expansion (See Proposition 4.4.2). If
$$\sum_{i \in S} \|u_i\|^2 > 1 + \frac{1}{2k} \qquad \text{or} \qquad \phi(S) > 10^5 k^3 \sqrt{\gamma_k \log r}$$
   then discard S, else $C \leftarrow C \cup \{S\}$ and $V_{l+1} \leftarrow V_l \setminus S$.
   (c) $l \leftarrow l + 1$ and repeat.
4. Output C.

Figure 9: Rounding Algorithm for Many Sparse Cuts

Input: Graph $G = (V, E)$ having maximum degree d.
We construct hypergraph $H = (V, E')$ as follows. For every vertex $v \in V$, we add the hyperedge $\{v\} \cup N^{\mathsf{out}}(\{v\})$ to E'.

Figure 10: Reduction from Vertex Expansion in graphs to Hypergraph Expansion

THE COMPLEXITY OF EXPANSION PROBLEMS

PART II

Approximation Algorithms

CHAPTER V

APPROXIMATION ALGORITHM FOR SPARSEST K-PARTITION

5.1 Introduction

In this chapter, we present approximation algorithms for the SPARSEST k-PARTITION problem in graphs. We define the problem formally as follows.

Problem 5.1.1 (SPARSEST k-PARTITION). Given a graph $G = (V, E, w)$ and a parameter k, compute a partition $\{P_1, \ldots, P_k\}$ of V into k non-empty pieces so as to minimize

$$\phi^k_G(\{P_1, \ldots, P_k\}) \stackrel{\text{def}}{=} \max_i \phi_G(P_i).$$

The optimal value is called the k-sparsity of G and is denoted by ϕ^k_G.

This problem is very similar to the K SPARSE-CUTS problem studied in Chapter 3. It differs from K SPARSE-CUTS only in requiring that the sets form a partition of the vertex set. Recall that Theorem 3.1.8 shows that ϕ^k_G can not be bounded by $\mathcal{O}\left(\sqrt{\lambda_k}\,\mathsf{polylog}\,k\right)$. Therefore, we study this problem with the view of obtaining approximation algorithms for it.

Since we are not trying to relate $\phi^k_G(G)$ to the graph spectra, we can afford to work with a more general notion of expansion. Given a graph $G = (V, E, w)$, where $w : V \cup E \to \mathbb{R}^+$, we define the expansion of a set $S \subset V$ as

$$\phi(S) \stackrel{\text{def}}{=} \frac{w(E(S, \bar{S}))}{\sum_{u \in V} w(u)}\,.$$

Note that this definition of expansion coincides with our previous definition of expansion when $w(u) = d_u$ for each $u \in V$. The main results of this chapter are as follows.

Theorem 5.1.2. *There exists a randomized polynomial-time algorithm that given an undirected graph $G = (V, E, w)$ and parameters $k \in \mathbb{Z}^+$ $(k \geqslant 2)$, $\varepsilon > 0$, w.h.p. outputs a $k' \geqslant (1-\varepsilon)k$ partition such that each set has expansion at most $\mathcal{O}_\varepsilon\left(\sqrt{\log n \log k}\ \phi_G^k\right)$.*

Theorem 5.1.3. *There exists a randomized polynomial-time algorithm that given an undirected graph $G = (V, E, w)$ with vertex weights $w_u = d_u$ (d_u is the degree of the vertex u) and parameters $k \in \mathbb{N}$ $(k \geqslant 2)$, $\varepsilon > 0$, w.h.p. outputs a $k' \geqslant (1-\varepsilon)k$ partition such that each set has expansion at most $\mathcal{O}_\varepsilon\left(\sqrt{\phi_G^k \log k}\right)$.*

Note that for $k = 2$, Theorem 5.1.2 gives the same guarantee as that of Arora, Rao and Vazirani [12] for Edge Expansion and Theorem 5.1.3 gives the same guarantee as that of Cheeger's inequality for Edge Expansion. A direct corollary of the work of Raghavendra, Steurer and Tulsiani [88] is that Theorem 5.1.3 is optimal under the SSE hypothesis.

SDP Relaxation. The proofs of our main theorems go via an SDP relaxation of ϕ_G^k and a rounding algorithm for it. As a first attempt, one would try an assignment SDP à la Unique Games (as used in [54, 102, 27, 31]), but such relaxations have a large integrality gap (see Section 5.6). The main difficulty in constructing an integer programming formulation of Sparsest k-partition is that we do not know the sizes of the sets in the optimal partition. We use a novel SDP relaxation which gets around this obstacle. In this SDP , we manage to encode a partitioning of the graph as well as a special measure on the vertices. This measure tells us how large every set must be. Roughly speaking, we expect that in the solution obtained by the algorithm, the measure of every set is approximately 1, irrespective of its size. We give a formal description of the SDP in Section 5.2.1.

A natural assignment SDP relaxation has a large integrality gap (see Section 5.6). To round our new SDP (see Section 5.2.1), one can try to adopt the rounding algorithms of Lee et al. [63] and Algorithm 3.3.2. (Both [63] and the proof of Algorithm 3.3.2

construct an embedding of the graph into $\mathbb{R}^k$ as a first step. The proofs of their main theorems can be viewed as an algorithm to round these vectors into sets) . However, these algorithms could only possibly give an approximation guarantee of the form $\mathcal{O}(\sqrt{\mathsf{OPT}\log k})$. To get rid of the square root, we need to embed the SDP solution from ℓ_2^2 to ℓ_2. This step distorts the vectors, so that they no longer satisfy SDP constraints and no longer have properties required by these algorithms.

5.1.1 Extensions

Our SDP formulation and rounding algorithm can be used to solve other problems as well. Consider the balanced version of Sparsest k-Partition.

Problem 5.1.4 (Balanced Sparsest k-Partitioning Problem)**.** Given a graph $G = (V, E, w)$ and a parameter k, compute a partition $\{P_1, \dots, P_k\}$ of V into k non-empty pieces each of weight $w(G)/k$ so as to minimize $\max_i \phi_G(P_i)$.

Using our techniques, we can prove the following theorems.

Theorem 5.1.5. *There exists a randomized polynomial-time algorithm that given an undirected graph $G = (V, E, w)$ and parameters $k \in \mathbb{N}$ ($k \geqslant 2$), $\varepsilon > 0$, w.h.p. outputs $k' \geqslant (1-\varepsilon)k$ disjoint sets (not necessarily a partition) such that the weight of each set is in the range $[w(G)/(2k), (1+\varepsilon)w(G)/k]$, and the expansion of each set is at most $\mathcal{O}_\varepsilon\left(\sqrt{\log n \log k}\ \mathsf{OPT}\right)$.*

Theorem 5.1.6. *There exists a randomized polynomial-time algorithm that given an undirected graph $G = (V, E, w)$ with vertex weights $w_u = d_u$ (d_u is the degree of the vertex u) and parameters $k \in \mathbb{N}$ ($k \geqslant 2$), $\varepsilon > 0$, w.h.p. outputs $k' \geqslant (1-\varepsilon)k$ disjoint sets (not necessarily a partition) such that the weight of each set is in the range $[w(G)/(2k), (1+\varepsilon)w(G)/k]$, and the expansion of each set is at most $\mathcal{O}_\varepsilon\left(\sqrt{\mathsf{OPT}\log k}\right)$.*

Note that the algorithms above return k' disjoint sets that do not have to cover all vertices. The proofs of these theorems are similar to the proofs of our main results

– Theorem 5.1.2 and Theorem 5.1.3. We refer the reader to Section 5.2.6 for more details. In fact, the assumption that all sets in the optimal solution have the same size makes the balanced problem much simpler. Theorem 5.1.5 also follows (possibly with slightly worse guarantees) from the result of Krauthgamer, Naor, and Schwartz [57], who gave a bi-criteria $O(\sqrt{\log n \log k})$ approximation algorithm for the k-Balanced Partitioning Problem (with the "min-sum" objective).

Organization. We prove Theorem 5.1.2 in Section 5.2.4. We present the SDP relaxation of Sparsest k-partition in Section 5.2.1 and the main rounding algorithm in Section 5.2.4. We prove Theorem 5.1.3 in Appendix 5.4.

5.2 Main Algorithm

We first prove a slightly weaker result. We give an algorithm that finds at least $(1-\varepsilon)k$ disjoint sets each with expansion at most $\mathcal{O}_\varepsilon\left(\sqrt{\log n \log k}\ \phi_G^k\right)$. Note that we do not require that these sets cover all vertices in V.

Theorem 5.2.1. *There exists a randomized polynomial-time algorithm that given an undirected graph G and parameters $k \in \mathbb{N}$ ($k \geqslant 2$), $\varepsilon > 0$, outputs $k' \geqslant (1-\varepsilon)k$ disjoint sets $P_1, \ldots, P_{k'}$ such that*

$$\mathbb{E}\left[\max_i \phi(S_i)\right] \leqslant \mathcal{O}_\varepsilon\left(\sqrt{\log n \log k}\ \phi_G^k\right) .$$

Then, in Section 5.3, we show how using $k' \geqslant (1-\varepsilon)k$ such sets, we can find a partitioning of V into $k'' \geqslant (1-2\varepsilon)k$ sets with each set having expansion at most $\mathcal{O}_\varepsilon\left(\sqrt{\log n \log k}\ \phi_G^k\right)$.

Our algorithm works in several phases. First, it solves the SDP relaxation, which we present in Section 5.2.1. Then it transforms all vectors to unit vectors and defines a measure $\mu(\cdot)$ on vertices of the graph. We give the details of this transformation in Section 5.2.2. Succeeding this, in the main phase, the algorithm samples many independent *orthogonal separators* $S_1, \ldots, S_T$ and then extracts $k' > (1-\varepsilon)k$ disjoint subsets from them. We describe this phase in Section 5.2.4. Finally, the algorithm merges some of these sets with the left over vertices to obtain a $k'' \geqslant (1-\varepsilon)k'$ partition. We describe this phase in Section 5.2.6.

5.2.1 SDP Relaxation

We employ a novel SDP relaxation for the SPARSEST k-PARTITION problem. The main challenge in writing an SDP relaxation is that we do not know the sizes of the sets in advance, so we cannot write standard spreading constraints or spreading constraints used in the paper of Bansal et. al.[16]. For each vertex u, we introduce a vector $\bar{u}$. In the integral solution corresponding to the optimal partitioning $P_1, \dots, P_k$, each vector $\bar{u}$ has k coordinates, one for every set P_i:

$$\bar{u}(i) = \begin{cases} \frac{1}{\sqrt{w(P_i)}} & \text{if } u \in P_i; \\ 0 & \text{otherwise.} \end{cases}$$

Observe, that the integral solution satisfies two crucial properties: for each set P_i,

$$\sum_{u \in P_i} w_u \|\bar{u}\|^2 = \sum_{u \in P_i} \frac{w_u}{w(P_i)} = 1, \tag{43}$$

and for every vertex $u \in P_i$,

$$\sum_{v \in V} w_v \langle \bar{u}, \bar{v} \rangle = \sum_{v \in P_i} \frac{w_v}{w(P_i)} + \sum_{v \notin P_i} 0 = 1. \tag{44}$$

(43) gives us a way to measure sets. Given a set of vectors $\{\bar{u}\}$, we define a measure $\mu(\cdot)$ on vertices as follows

$$\mu(S) = \sum_{u \in S} w_u \|\bar{u}\|^2. \tag{45}$$

For the intended solution, we have $\mu(P_i) = 1$, and hence $\mu(V) = k$. This is the first constraint we add to the SDP :

$$\mu(V) \equiv \sum_{u \in V} w_u \|\bar{u}\|^2 = k.$$

From (44), we get a spreading constraint:

$$\sum_{v \in V} w_v \langle \bar{u}, \bar{v} \rangle = 1.$$

We also add ℓ_2^2 triangle inequalities to the SDP . It is easy to check that they are satisfied in the intended solution (since they are satisfied for each coordinate).

Finally, we need to write the objective function that measures the expansion of the sets. In the intended solution, if $u, v \in P_i$ (for some i), then $\bar{u} = \bar{v}$, and $\|\bar{u} - \bar{v}\|^2 = 0$. If $u \in P_i$ and $v \in P_j$ (for $i \neq j$), then

$$\|\bar{u} - \bar{v}\|^2 = \|\bar{u}\|^2 + \|\bar{v}\|^2 = 1/w(P_i) + 1/w(P_j)\,.$$

Hence,

$$\begin{aligned}\frac{1}{k} \sum_{\{u,v\} \in E} w(\{u,v\}) \|\bar{u} - \bar{v}\|^2 &= \frac{1}{k} \sum_{i<j} \sum_{\substack{\{u,v\} \in E \\ u \in P_i \\ v \in P_j}} \left(\frac{1}{w(P_i)} + \frac{1}{w(P_j)} \right) w(\{u,v\}) \\ &= \frac{1}{k} \sum_i \frac{w(E(P_i, V \setminus P_i))}{w(P_i)} = \frac{1}{k} \sum_i \phi_G(P_i) \leqslant \phi_G^k\,. \quad (46)\end{aligned}$$

We get the following SDP relaxation for the problem.

SDP 5.2.2.

$$\min \frac{1}{k} \sum_{\{u,v\} \in E} w(\{u,v\}) \|\bar{u} - \bar{v}\|^2\ .$$

$$\begin{array}{rcll} \sum_{u \in V} w_u \|\bar{u}\|^2 & = & k & \\ \sum_{v \in V} w_v \langle \bar{u}, \bar{v} \rangle & = & 1 & \forall u \in V \\ \|\bar{u} - \bar{x}\|^2 + \|\bar{x} - \bar{v}\|^2 & \geqslant & \|\bar{u} - \bar{v}\|^2 & \forall u, v, x \in V \\ 0 \leqslant \langle \bar{u}, \bar{v} \rangle & \leqslant & \|\bar{u}\|^2 & \forall u, v \in V \end{array}$$

Figure 11: SDP Relaxation for Sparsest k-Partition

5.2.2 Normalization

After the algorithm solves the SDP 5.2.2, we define the measure μ using (45), and "normalize" all vectors using a transformation ψ from the paper of Chlamtac, Makarychev and Makarychev [31]. The transformation ψ defines the inner products between $\psi(\bar{u})$ and $\psi(\bar{v})$ as follows (all vectors $\bar{u}$ are nonzero in our SDP relaxation):

$$\langle \psi(\bar{u}), \psi(\bar{v}) \rangle = \frac{\langle \bar{u}, \bar{v} \rangle}{\max\{\|\bar{u}\|^2, \|\bar{v}\|^2\}}.$$

This uniquely defines vectors $\psi(\bar{u})$ (up to an isometry of ℓ_2). Chlamtac, Makarychev and Makarychev showed that the image $\psi(X)$ of any ℓ_2^2 space X is an ℓ_2^2 space, and the following condtions hold.

- For all non-zero vectors $\bar{u} \in X$, $\|\psi(\bar{u})\|^2 = 1$.

- For all non-zero vectors $u, v \in X$,

$$\|\psi(\bar{u}) - \psi(\bar{v})\|^2 \leqslant \frac{2\,\|\bar{u} - \bar{v}\|^2}{\max\left\{\|\bar{u}\|^2, \|\bar{v}\|^2\right\}}.$$

5.2.3 Orthogonal Separators

Our algorithm uses the notion of orthogonal separators introduced by Chlamtac, Makarychev, and Makarychev [31]. Let X be an ℓ_2^2 space. We say that a distribution over subsets of X is a k-orthogonal separator of X with distortion D, probability scale $\alpha > 0$ and separation threshold $\beta < 1$, if the following conditions hold for $S \subset X$ chosen according to this distribution:

1. For all $\bar{u} \in X$, $\mathbb{P}[\bar{u} \in S] = \alpha\,\|\bar{u}\|^2$.

2. For all $\bar{u}, \bar{v} \in X$ with $\langle \bar{u}, \bar{v} \rangle \leqslant \beta \max\left\{\|\bar{u}\|^2, \|\bar{v}\|^2\right\}$,

$$\mathbb{P}[\bar{u} \in S \text{ and } \bar{v} \in S] \leqslant \frac{\alpha \min\left\{\|\bar{u}\|^2, \|\bar{v}\|^2\right\}}{k}.$$

3. For all $u, v \in X$

$$\mathbb{P}\left[I_S(\bar{u}) \neq I_S(\bar{v})\right] \leqslant \alpha D \left\|\bar{u} - \bar{v}\right\|^2.$$

Here I_S is the indicator function of the set S.

Theorem 5.2.3 ([31, 16]). *There exists a polynomial-time randomized algorithm that given a set of vectors X, a parameter k, and $\beta < 1$ generates a k-orthogonal separator with distortion $D = \mathcal{O}_\beta\left(\sqrt{\log |X| \log k}\right)$ and scale $\alpha \geqslant 1/p(|X|)$ for some polynomial p.*

In the algorithm, we sample orthogonal separators from the set of normalized vectors $\{\psi(\bar{u}) : u \in V\}$. For simplicity of exposition we assume that an orthogonal separator S contains not vectors $\bar{u}$, but the corresponding vertices. That is, for an orthogonal separator $\tilde{S}$, we consider the set of vertices $S = \{u \in V : \psi(\bar{u}) \in \tilde{S}\}$.

5.2.4 Algorithm

We give an algorithm for generating $k' \geqslant (1-\varepsilon)k$ disjoint sets P_i in Figure 12.

5.2.5 Properties of Sets S_i''

We prove that (a) the edge boundaries of the sets S_i'' are *small*; and (b) the sets S_i'' form a partition of V w.h.p. The following lemma makes these statements precise.

Lemma 5.2.5. *For a set $S \subset V$, define*

$$\nu(S) \stackrel{\text{def}}{=} \sum_{\substack{\{u,v\} \in E(S, V\setminus S) \\ u \in S, v \notin S}} w\left(\{u,v\}\right) \|\bar{u}\|^2 + \sum_{\substack{\{u,v\} \in E \\ u,v \in S}} w\left(\{u,v\}\right) \left| \|\bar{u}\|^2 - \|\bar{v}\|^2 \right|. \qquad (47)$$

Then, sets S_i'' satisfy the following conditions:

1.

$$\mathbb{E}\left[\sum_i \nu(S_i'')\right] \leqslant (8D+1)k \cdot \mathsf{SDPval},$$

where $D = \mathcal{O}_\varepsilon(\sqrt{\log n \log k})$ is the distortion of $(12k/\varepsilon)$-orthogonal separator, and SDPval *is the value of the* SDP *solution.*

Algorithm 5.2.4.

1. Solve SDP 5.2.2 and obtain vectors $\{\bar{u}\}$.
2. Compute normalized vectors $\psi(\bar{u})$, and define the measure $\mu(\cdot)$ (see Section 5.2.2 and Eq. (45)).
3. Sample $T = 2n/\alpha$ independent $(12k/\varepsilon)$-orthogonal separators $S_1, \dots, S_T$ for vectors $\psi(\bar{u})$ $(u \in V)$
with separation threshold $\beta = 1 - \varepsilon/4$.
4. For each i, define S_i' as follows:
$$S_i' \stackrel{\text{def}}{=} \begin{cases} S_i & \text{if } \mu(S_i) \leqslant 1 + \varepsilon/2; \\ \varnothing & \text{otherwise.} \end{cases}$$
5. For each i, let $S_i'' = S_i' \setminus \left(\cup_t^{i-1} S_t'\right)$ be the set of yet uncovered vertices in S_i'.
6. For each i, set $P_i = \{u \in S_i'' : \|\bar{u}\|^2 \geqslant r_i\}$, where the parameter r_i is chosen to minimize the expansion $\phi_G(P_i)$ of the set P_i.
7. Output $(1-\varepsilon)k$ non-empty sets P_i with the smallest expansion $\phi_G(P_i)$.

Figure 12: Algorithm for generating $k' \geqslant (1-\varepsilon)k$ disjoint sets P_i.

2. All sets S_i'' are disjoint; and

$$\mathbb{P}\left[\mu(\cup S_i'') = k\right] \geqslant 1 - ne^{-n}.$$

Proof. (a) Let E_{cut} be the set of edges cut by the partitioning $S_1'', \dots, S_T'', V \setminus (\cup S_i'')$. Observe, that each cut edge $\{u,v\}$ contributes $\|\bar{u}\|^2 + \|\bar{v}\|^2$ to the sum $\sum \nu(S_i'')$, and each uncut edge contributes either $|\|\bar{u}\|^2 - \|\bar{v}\|^2|$, or 0. Hence,

$$\mathbb{E}\left[\sum_i \nu(S_i'')\right] \leqslant \mathbb{E}\left[\sum_{\{u,v\}\in E_{cut}} w(\{u,v\})\left(\|\bar{u}\|^2 + \|\bar{v}\|^2\right)\right] + \sum_{\{u,v\}\in E} w(\{u,v\})\left|\|\bar{u}\|^2 - \|\bar{v}\|^2\right|.$$

The second term is bounded by

$$\sum_{\{u,v\}\in E} w(\{u,v\})\,\|\bar{u} - \bar{v}\|^2 = k \cdot \mathsf{SDPval},$$

since

$$\|\bar{u}\|^2 - \|\bar{v}\|^2 = \|\bar{u}-\bar{v}\|^2 - 2(\|\bar{v}\|^2 - \langle \bar{u}, \bar{v}\rangle) \leqslant \|\bar{u}-\bar{v}\|^2.$$

The inequality follows from the SDP constraint $\|\bar{v}\|^2 \geqslant \langle \bar{u}, \bar{v}\rangle$. We now bound the first term. To do so we need the following lemma.

Lemma 5.2.6. *For every vertex $u \in V$ and $i \in \{1, \dots, T\}$, we have $\mathbb{P}[u \in S'_i] \geqslant \alpha/2$.*

We give the proof of Lemma 5.2.6 after we finish the proof of Lemma 5.2.5. Let us estimate the probability that an edge $\{u, v\}$ is cut. Let $U_t = \cup_{i \leqslant t} S'_i$ be the set of vertices covered by the first t sets S'_i. Note, that $S''_i = S'_i \setminus U_{i-1}$. We say that the edge $\{u, v\}$ is cut by the set S'_t, if S'_t is the first set containing u or v, and it contains only one of these vertices. Then,

$$\begin{aligned}
\mathbb{P}[\{u,v\} \in E_{cut}] &= \sum_{i=1}^{T} \mathbb{P}[\{u,v\} \text{ is cut by } S'_i] \\
&= \sum_{i=1}^{T} \mathbb{P}\left[u, v \notin U_{i-1} \text{ and } I_{S'_i}(u) \neq I_{S'_i}(v)\right] \\
&\leqslant \sum_{i=1}^{T} \mathbb{P}[u \notin U_{i-1} \text{ and } I_{S_i}(u) \neq I_{S_i}(v)] \\
&= \sum_{i=1}^{T} \mathbb{P}[u \notin U_{i-1}]\, \mathbb{P}[I_{S_i}(u) \neq I_{S_i}(v)].
\end{aligned}$$

Now, by Lemma 5.2.6, $\mathbb{P}[u \notin U_{i-1}] \leqslant (1-\alpha/2)^{i-1}$, and, by Property 3 of orthogonal separators,

$$\begin{aligned}
\mathbb{P}[I_{S_i}(u) \neq I_{S_i}(v)] &\leqslant \alpha D \|\psi(u) - \psi(v)\|^2 \\
&\leqslant \frac{2\alpha D \|\bar{u}-\bar{v}\|^2}{\max\{\|\bar{u}\|^2, \|\bar{v}\|^2\}}.
\end{aligned}$$

Thus (using $\sum_i (1-\alpha/2)^i \leqslant 2/\alpha$),

$$\mathbb{P}[\{u,v\} \in E_{cut}] \leqslant \frac{4D\,\|\bar{u}-\bar{v}\|^2}{\max\{\|\bar{u}\|^2, \|\bar{v}\|^2\}}.$$

We are almost done,

$$\begin{aligned}
&\mathbb{E}\left[\sum_{\{u,v\}\in E_{cut}} w(\{u,v\})\,(\|\bar{u}\|^2+\|\bar{v}\|^2)\right] \\
&= \sum_{\{u,v\}\in E} w(\{u,v\})\,\mathbb{P}\left[\{u,v\}\in E_{cut}\right](\|\bar{u}\|^2+\|\bar{v}\|^2) \\
&\leqslant \sum_{\{u,v\}\in E} w(\{u,v\})\,\frac{4D\,\|\bar{u}-\bar{v}\|^2}{\max\{\|\bar{u}\|^2,\|\bar{v}\|^2\}}\cdot(\|\bar{u}\|^2+\|\bar{v}\|^2) \\
&\leqslant \sum_{\{u,v\}\in E} 8D\,w(\{u,v\})\;\|\bar{u}-\bar{v}\|^2 \\
&= 8kD\cdot\mathsf{SDPval}\,.
\end{aligned}$$

Thus we get that

$$\mathbb{E}\left[\sum_i \nu(S_i'')\right] \leqslant (8D+1)k\cdot\mathsf{SDPval}\,.$$

(b) The sets S_i'' are disjoint by definition. By Lemma 5.2.6, the probability that a vertex is not covered by any set S_i is $(1-\alpha/2)^T = (1-\alpha/2)^{2n/\alpha} < e^{-n}$. So with probability at least $1-ne^{-n}$ all vertices are covered. □

It remains to prove Lemma 5.2.6.

Proof of Lemma 5.2.6. We adopt a slightly modified argument from the paper of Bansal et al. [16] (Theorem 2.1, arXiv). If $u\in S_i$, then $u\in S_i'$ unless $\mu(S_i)>1+\varepsilon/2$, hence

$$\begin{aligned}
\mathbb{P}\left[u\in S_i'\right] &= \mathbb{P}\left[u\in S_i\right](1-\mathbb{P}\left[\mu(S_i)>1+\varepsilon/2 \mid u\in S_i\right]) \\
&= \alpha(1-\mathbb{P}\left[\mu(S_i)>1+\varepsilon/2 \mid u\in S_i\right]).
\end{aligned}$$

Here, we used that $\mathbb{P}\left[u\in S_i\right]=\alpha\|\psi(\bar{u})\|^2=\alpha$ (see Property 1 of orthogonal separators). We need to show that $\mathbb{P}\left[\mu(S_i)>1+\varepsilon/2 \mid u\in S_i\right]\leqslant 1/2$. Let us define the sets A_u and B_u as follows.

$$A_u=\{v\in V:\langle\psi(\bar{u}),\psi(\bar{v})\rangle\geqslant\beta\}$$

and

$$B_u = \{v \in V : \langle \psi(\bar{u}), \psi(\bar{v}) \rangle < \beta\}.$$

Now,

$$\begin{aligned}
\mu(A_u) &= \sum_{v \in A_u} w_v \|\bar{v}\|^2 \leqslant \frac{1}{\beta} \sum_{v \in V} w_v \|\bar{v}\|^2 \langle \psi(\bar{u}), \psi(\bar{v}) \rangle \\
&= \frac{1}{\beta} \sum_{v \in V} w_v \|\bar{v}\|^2 \frac{\langle \bar{u}, \bar{v} \rangle}{\max\{\|\bar{v}\|^2, \|\bar{v}\|^2\}} \\
&\leqslant \frac{1}{\beta} \sum_{v \in V} w_v \langle \bar{u}, \bar{v} \rangle \overset{\diamond}{=} \frac{1}{\beta} \leqslant 1 + \frac{\varepsilon}{3}.
\end{aligned}$$

Equality "$\diamond$" follows from the SDP constraint $\sum_{v \in V} w_v \langle \bar{u}, \bar{v} \rangle = 1$. For any $v \in B_u$, we have $\langle \psi(\bar{u}), \psi(\bar{v}) \rangle < \beta$. Hence, by Property 2 of orthogonal separators,

$$\mathbb{P}[v \in S_i \mid u \in S_i] \leqslant \frac{\varepsilon}{12k}$$

Therefore,

$$\mathbb{E}[\mu(S_i \cap B_u) \mid u \in S_i] \leqslant \frac{\varepsilon \mu(B_u)}{12k} \leqslant \frac{\varepsilon \mu(V)}{12k} = \frac{\varepsilon}{12}.$$

By Markov's inequality, $\mathbb{P}[\mu(S_i \cap B_u) \geqslant \varepsilon/6 \mid u \in S_i] \leqslant 1/2$. Since $\mu(S_i) = \mu(S_i \cap A_u) + \mu(S_i \cap B_u)$, we get $\mathbb{P}[\mu(S_i) \geqslant 1 + \varepsilon/2 \mid u \in S_i] \leqslant 1/2$. □

5.2.6 End of Proof

We are ready to finish the analysis of Algorithm 5.2.4 and prove Theorem 5.2.1 and Theorem 5.1.5.

Proofs of Theorem 5.2.1 and Theorem 5.1.5. We first prove Theorem 5.2.1, then we slightly modify Algorithm 5.2.4 and prove Theorem 5.1.5.

I. We show that Algorithm 5.2.4 outputs sets satisfying conditions of Theorem 5.2.1. The sets S_i'' are disjoint (see Lemma 5.2.5), thus sets P_i are also disjoint. We now need to prove that among sets P_i obtained at Step 6 of the algorithm, there are at least $(1-\varepsilon)k$ sets with expansion less than $O_\varepsilon(\sqrt{\log n \log k}\, OPT)$ (in expectation).

Let $Z = \frac{1}{k}\sum_i \nu(S_i'')$. By Lemma 5.2.5 we have,

$$\mathbb{E}\left[Z\right] \leqslant (8D+1)\, OPT$$

and S_i'' form a partition[1] of V. We throw away all empty sets S_i'', and set $\sigma_i = \mu(S_i'')/k$. Then $\sum_i \sigma_i = 1$, and

$$Z = \frac{1}{k}\sum_i \nu(S_i'') = \sum_i \sigma_i \cdot \frac{\nu(S_i'')}{\mu(S_i'')}.$$

Define $\mathcal{I} \stackrel{\text{def}}{=} \{i : \nu(S_i'')/\mu(S_i'') \leqslant 3Z/\varepsilon\}$. By Markov's inequality (we can think of σ_i as the weight of i),

$$\sum_{i\in\mathcal{I}} \sigma_i \geqslant 1 - \varepsilon/2\,. \tag{48}$$

Since each σ_i satisfies

$$\sigma_i = \mu(S_i'')/k \leqslant (1+\varepsilon/2)/k$$

the set $\mathcal{I}$ has at least $(1-\varepsilon/2)k/(1+\varepsilon/2) \geqslant (1-\varepsilon)k$ elements.

Now for any $S \subset V$, let us define a vector $X_S \in \mathbb{R}^n$ as follows.

$$X_S(u) \stackrel{\text{def}}{=} \begin{cases} \|\bar{u}\|^2 & \text{if } u \in S \\ 0 & \text{otherwise} \end{cases}.$$

Then for each $i \in \mathcal{I}$, using Lemma 2.3.2, we get that a set $P_i \subset S_i''$ such that

$$\phi(P_i) \leqslant \frac{\sum_{u\sim v} w\{(u,v)\}\,|X_u - X_v|}{\sum_w w_u X_u} = \frac{\nu(S_i'')}{\mu(S_i'')} \leqslant \frac{3Z}{\varepsilon}\,.$$

Therefore, we showed that there are at least $|\mathcal{I}| \geqslant (1-\varepsilon)k$ sets P_i with expansion at most $3Z/\varepsilon$. Therefore, the expansion of the sets returned by the algorithm is at most $3Z/\varepsilon$. This finishes the proof, since $\mathbb{E}\left[3Z/\varepsilon\right] = \mathcal{O}_\varepsilon(\sqrt{\log n \log k})\ \phi_G^k$.

II. To prove Theorem 5.1.5, we need to modify the algorithm. For simplicity, we rescale all weights w_u and assume that $w(G) = k$. Then our goal is to find k' disjoint sets P_i of weight in the range $[1/2, 1+\varepsilon]$ each. Since all sets in the optimal solution

[1] With an exponentially small probability the sets S_i'' do not cover all the vertices. In this unlikely event, the algorithm may output an arbitrary partition.

to the k-Balanced Sparsest Partitioning Problem have weight 1, we add the SDP constraint that all vectors $\bar{u}$ have length 1 (see Section 5.2.1): for all $u \in V$:

$$\|\bar{u}\|^2 = 1.$$

The intended solution satisfies this constraint.

For a random $r \in (0, R)$ and $L_r \stackrel{\text{def}}{=} \{u \in S_i'' : \|\bar{u}\|^2 \geqslant r\}$, we have

$$\mathop{\mathbb{E}}_r[w(L_r)] = \mu(S_i'') \tag{49}$$

as each u belongs to L_r with probability $\|\bar{u}\|^2$ and

$$\mathop{\mathbb{E}}_r[w\left(E(L_r, V \setminus L_r)\right)] = \nu(S_i'')$$

(since an edge in $S_i'' \times S_i''$ is cut with probability $|\|\bar{u}\|^2 - \|\bar{v}\|^2|$; and an edge $\{u, v\}$ with $u \in S_i''$ and $v \notin S_i''$ is cut with probability $\|\bar{u}\|^2$ — if and only if $u \in L_r$; compare with Definition 47). Therefore,

$$\mathop{\mathbb{E}}_r[w\left(E(L_r, V \setminus L_r)\right)] = \frac{\nu(S_i'')}{R} \leqslant \frac{3Z}{\varepsilon} \cdot \frac{\mu(S_i'')}{R} = \frac{3Z}{\varepsilon} \cdot \mathop{\mathbb{E}}_r[w(L_r)]\,.$$

We also change the way the algorithm picks the parameters r_i. The algorithm chooses r_i so as to minimize the expansion $\phi_G(P_i)$ subject to an additional constraint $\mu(P_i) \geqslant (1 - \varepsilon/2)\mu(S_i'')$. Finally, once the algorithm obtains sets P_i, it greedily merges sets of weight at most 1/2. The rest of the algorithm is the same as Algorithm 5.2.4.

From (49) and (50), we get

$$\begin{aligned} \mathop{\mathbb{E}}_r[w(L_r)] &\geqslant \frac{\varepsilon^2}{6Z} \mathop{\mathbb{E}}_r[w\left(E(L_r, V \setminus L_r)\right)] + \frac{(1 - \varepsilon/2)\mu(S_i'')}{R} \\ &\geqslant \max\left\{\frac{\varepsilon^2}{6Z} \mathop{\mathbb{E}}_r[w\left(E(L_r, V \setminus L_r)\right)], \frac{(1 - \varepsilon/2)\mu(S_i'')}{R}\right\}. \end{aligned}$$

Since $\|\bar{u}\|^2 = 1$ for all $u \in V$, we have $R = 1$ and $\mu(L_r) = w(L_r)$. Therefore,

$$\mathop{\mathbb{E}}_r[w(L_r)] \geqslant \max\left\{\frac{\varepsilon^2}{6Z} \mathop{\mathbb{E}}_r[w\left(E(L_r, V \setminus L_r)\right)], \left(1 - \frac{\varepsilon}{2}\right)\mu(S_i'')\right\},$$

and for some r^*,

$$w(L_{r^*}) \geqslant \frac{\varepsilon^2}{6Z} \mathop{\mathbb{E}}_r [w\left(E(L_{r^*}, V \setminus L_{r^*})\right)];$$
$$\mu(L_{r^*}) \geqslant (1 - \frac{\varepsilon}{2})\mu(S_i'').$$

Consequently, we get

$$\phi_G(P_i) \leqslant \phi_G(L_{r^*}) \leqslant \frac{6Z}{\varepsilon^2}.$$

Now, recall, that by (48), $\sum_{i\in\mathcal{I}} \sigma_i \geqslant 1 - \varepsilon/2$. Hence,

$$\begin{aligned} \sum_{i\in\mathcal{I}} w(P_i) &= \sum_{i\in\mathcal{I}} \mu(P_i) \geqslant (1-\varepsilon/2) \sum_{i\in\mathcal{I}} \mu(S_i'') = (1-\varepsilon/2) \sum_{i\in\mathcal{I}} k\sigma_i \\ &\geqslant (1-\varepsilon)k. \end{aligned}$$

We showed that the algorithm gets sets P_i satisfying the following properties:

1. the expansion
$$\phi_G(P_i) \leqslant \frac{6Z}{\varepsilon^2}.$$
2. $w(P_i) \leqslant (1 + \varepsilon/2)$ and
3. $\sum_i w(P_i) \geqslant (1 - \varepsilon)k$.

To get sets of weight in the range $[1/2, 1 + \varepsilon]$ the algorithm greedily merges sets P_i of weight at most $1/2$ and obtains a collection of new sets, which we denote by Q_i. The algorithm outputs all sets Q_i with weight at least $1/2$.

Note that for any two disjoint sets A and B,

$$\phi_G(A \cup B) \leqslant \max\left\{\phi_G(A), \phi_G(B)\right\}.$$

So

$$\phi_G(Q_i) \leqslant \max_j \phi_G(P_j) \leqslant \frac{6Z}{\varepsilon^2}.$$

All sets Q_i but possibly one have weight at least $1/2$. So the weight of sets Q_i output by the algorithm is at least $(1-\varepsilon)k - 1/2$. The maximum weight of sets Q_i is $1+\varepsilon/2$, so the number of sets Q_i is at least

$$\left\lceil \frac{(1-\varepsilon)k - 1/2}{1-\varepsilon/2} \right\rceil \geqslant \lceil (1-2\varepsilon)k - 1/2 \rceil \geqslant \lceil (1-4\varepsilon)k \rceil .$$

To verify the last inequality check two cases: if $2\varepsilon k \geqslant 1/2$, then

$$(1-2\varepsilon)k - 1/2 \geqslant (1-4\varepsilon)k;$$

if $2\varepsilon k < 1/2$, then

$$\lceil (1-2\varepsilon)k - 1/2 \rceil = k .$$

This finishes the proof. □

5.3 From Disjoint Sets to Partitioning

We now show how given $k' \geqslant (1-\varepsilon)$ sets $P_1, \ldots, P_{k'}$, we can obtain a true partitioning $P'_1, \ldots, P'_{k''}$ of V.

Proof of Theorem 5.1.2. To get the desired partitioning, we first run Algorithm 5.2.4 several times (say, n) to obtain disjoint non-empty sets $P_1, \ldots, P_{k'}$ that satisfy $\max_i \phi_G(P_i) \leqslant \mathcal{O}_\varepsilon(\sqrt{\log n \log k})\ \phi_G^k$ w.h.p. Let $Z = \max_i \phi_G(P_i)$. We sort sets P_i by weight $w(P_i)$. We output the smallest $k'' = \lfloor (1-\varepsilon)k' \rfloor$ sets P_i, and the compliment set $P' = V \setminus (\cup_{1 \leqslant i \leqslant k''} P_i)$.

Since sets P_i are disjoint and non-empty, the first k'' sets P_i and the set P' are also disjoint and non-empty. Moreover, $\phi_G(P_i) \leqslant Z$, so we only need to show that $\phi_G(P') \leqslant \mathcal{O}_\varepsilon(Z)$. Note, that $w(P') \geqslant \varepsilon w(V)$, since P' contains vertices in the $\lceil \varepsilon k \rceil$ largest sets P_i and all vertices not covered by sets P_i. Then,

$$E(P', V \setminus P') = \cup_{i \leqslant k''} E(P', P_i) \subset \cup_{i \leqslant k''} E(P_i, V \setminus P_i) .$$

So

$$\begin{aligned}\phi_G(P') &= \frac{w\left(E(P',V\setminus P')\right)}{w(P')} \leqslant \frac{\sum_{i=1}^{k''} w\left(E(P_i,V\setminus P_i)\right)}{w(P')} \\ &= \frac{\sum_{i=1}^{k''} w(P_i)\phi_G(P_i)}{\varepsilon w(V)} \leqslant \frac{\sum_{i=1}^{k''} w(P_i) Z}{\varepsilon w(V)} \\ &\leqslant \frac{Z w(V)}{\varepsilon w(V)} = \frac{Z}{\varepsilon}.\end{aligned}$$

This concludes the proof. □

5.4 Proof of Theorem 5.1.3

The proof of Theorem 5.1.3 is almost the same as the proof of Theorem 5.1.2. The only difference is that we need to replace orthogonal separators with a slightly different variant of orthogonal separators (implicitly defined in [31]).

Orthogonal Separators with ℓ_2 distortion. Let X be a set of unit vectors in ℓ_2. We say that a distribution over subsets of X is a k-orthogonal separator of X with ℓ_2 distortion D, probability scale $\alpha > 0$ and separation threshold $\beta < 1$, if the following conditions hold for $S \subset X$ chosen according to this distribution:

1. For all $\bar{u} \in X$, $\mathbb{P}\left[\bar{u} \in S\right] = \alpha$.
2. For all $\bar{u}, \bar{v} \in X$ with $\langle \bar{u}, \bar{v} \rangle \leqslant \beta \max\left\{\|\bar{u}\|^2, \|\bar{v}\|^2\right\}$,

$$\mathbb{P}\left[\bar{u} \in S \text{ and } \bar{v} \in S\right] \leqslant \frac{\alpha}{k}.$$

3. For all $u, v \in X$,

$$\mathbb{P}\left[I_S(\bar{u}) \neq I_S(\bar{v})\right] \leqslant \alpha D \|\bar{u} - \bar{v}\|.$$

Theorem 5.4.1 ([31])**.** *There exists a polynomial-time randomized algorithm that given a set of unit vectors X, a parameter k, and $\beta < 1$ generates a k-orthogonal separator with ℓ_2 distortion $D = \mathcal{O}_\beta\left(\sqrt{\log k}\right)$ and scale $\alpha \geqslant 1/n$.*

For completeness we sketch the proof of this lemma in Section 5.5. Algorithm 5.2.4′ is the same as Algorithm 5.2.4 except that at Step 3, it samples orthogonal separators with ℓ_2 distortion $\mathcal{O}_\varepsilon(\sqrt{\log k})$ using Theorem 5.4.1. The proof of Theorem 5.1.2 goes through for the new algorithm essentially as is. The only statement we need to take care of is Lemma 5.2.5 (a). We prove the following bound on $\mathbb{E}\left[\sum_i \nu(S_i'')\right]$.

Lemma 5.4.2. *The sets S_i'' satisfy the following condition: $\mathbb{E}\left[\sum_i \nu(S_i'')\right] \leqslant (8D+1)k \cdot \sqrt{\mathsf{SDPval}}$, where $D = \mathcal{O}_\varepsilon(\sqrt{\log k})$ is the ℓ_2 distortion of $(12k/\varepsilon)$-orthogonal separator, and* SDPval *is the value of the* SDP *solution.*

Proof. Let E_{cut} be the set of edges cut by the partitioning $S_1'', \ldots, S_T'', V \setminus (\cup S_i'')$. As before (in Lemma 5.2.5), we have

$$\begin{aligned}
\mathbb{E}\left[\sum_i \nu(S_i'')\right] &\leqslant \mathbb{E}\left[\sum_{\{u,v\}\in E_{cut}} w(\{u,v\})\left(\|\bar{u}\|^2 + \|\bar{v}\|^2\right)\right] \\
&\qquad + \sum_{\{u,v\}\in E} w(\{u,v\})\left|\|\bar{u}\|^2 - \|\bar{v}\|^2\right| \\
&\leqslant \mathbb{E}\left[\sum_{\{u,v\}\in E_{cut}} w(\{u,v\})\left(\|\bar{u}\|^2 + \|\bar{v}\|^2\right)\right] + k\,\mathsf{SDPval}\,.
\end{aligned}$$

We now bound the first term. Estimate the probability that an edge $\{u,v\}$ is cut. Let $U_t = \cup_{i\leqslant t} S_i'$ be the set of vertices covered by the first t sets S_i'. Note, that $S_i'' = S_i' \setminus U_{i-1}$. We say that the edge $\{u,v\}$ is cut by the set S_t', if S_t' is the first set containing u or v, and it contains only one of these vertices. Then,

$$\begin{aligned}
\mathbb{P}\left[\{u,v\} \in E_{cut}\right] &= \sum_i \mathbb{P}\left[\{u,v\} \text{ is cut by } S_i'\right] \\
&= \sum_i \mathbb{P}\left[u, v \notin U_{i-1} \text{ and } I_{S_i'}(u) \neq I_{S_i'}(v)\right] \\
&\leqslant \sum_i \mathbb{P}\left[u \notin U_{i-1} \text{ and } I_{S_i}(u) \neq I_{S_i}(v)\right] \\
&= \sum_i \mathbb{P}\left[u \notin U_{i-1}\right] \mathbb{P}\left[I_{S_i}(u) \neq I_{S_i}(v)\right].
\end{aligned}$$

Now, by Lemma 5.2.6, $\mathbb{P}\left[u \notin U_{i-1}\right] \leqslant (1-\alpha/2)^{i-1}$, and, by Property 3 of ℓ_2 orthogonal separators,

$$\mathbb{P}\left[I_{S_i}(u) \neq I_{S_i}(v)\right] \leqslant \alpha D \left\|\psi(u) - \psi(v)\right\| \leqslant \alpha D \frac{\sqrt{2}\,\left\|\bar{u} - \bar{v}\right\|}{\max\left\{\left\|\bar{u}\right\|, \left\|\bar{v}\right\|\right\}}.$$

Thus,

$$\mathbb{P}\left[\{u,v\} \in E_{cut}\right] \leqslant \frac{2\sqrt{2}\,D\,\|\bar{u} - \bar{v}\|}{\max\{\|\bar{u}\|, \|\bar{v}\|\}}.$$

Now, the proof deviates from the proof of Lemma 5.2.5:

$$\begin{aligned}
&\mathbb{E}\left[\sum_{\{u,v\}\in E_{cut}} w\left(\{u,v\}\right)\left(\|\bar{u}\|^2 + \|\bar{v}\|^2\right)\right] \\
&= \sum_{\{u,v\}\in E} w\left(\{u,v\}\right)\mathbb{P}\left[\{u,v\} \in E_{cut}\right]\left(\|\bar{u}\|^2 + \|\bar{v}\|^2\right) \\
&\leqslant \sum_{\{u,v\}\in E} w\left(\{u,v\}\right)\frac{2\sqrt{2}\,D\,\|\bar{u} - \bar{v}\|}{\max\{\|\bar{u}\|, \|\bar{v}\|\}} \cdot \left(\|\bar{u}\|^2 + \|\bar{v}\|^2\right) \\
&\leqslant 2\sqrt{2}\,D \sum_{\{u,v\}\in E} w\left(\{u,v\}\right)\|\bar{u} - \bar{v}\| \cdot \left(\|\bar{u}\| + \|\bar{v}\|\right).
\end{aligned}$$

By Cauchy–Schwarz,

$$\begin{aligned}
&2\sqrt{2}\,D \sum_{\{u,v\}\in E} w\left(\{u,v\}\right)\|\bar{u} - \bar{v}\| \cdot \left(\|\bar{u}\| + \|\bar{v}\|\right) \\
&\leqslant 2\sqrt{2}\,D \left(\sum_{\{u,v\}\in E} w\left(\{u,v\}\right)\|\bar{u} - \bar{v}\|^2\right)^{1/2}\left(\sum_{\{u,v\}\in E} w\left(\{u,v\}\right)\left(\|\bar{u}\| + \|\bar{v}\|\right)^2\right)^{1/2} \\
&\leqslant 4\,D \left(\sum_{\{u,v\}\in E} w\left(\{u,v\}\right)\|\bar{u} - \bar{v}\|^2\right)^{1/2}\left(\sum_{\{u,v\}\in E} w\left(\{u,v\}\right)\|\bar{u}\|^2 + \|\bar{v}\|^2\right)^{1/2} \\
&= 4\,D\ \left(k\,\mathsf{SDPval}\right)^{1/2}\left(\sum_{\{u,v\}\in E} d_u\|\bar{u}\|^2\right)^{1/2}.
\end{aligned}$$

Recall, that in Theorem 5.1.3, we assume that the weight of every vertex w_u equals its degree d_u. Hence, $\sum_{\{u,v\}\in E} d_u\|\bar{u}\|^2 = \mu(V) = k$. We get,

$$\mathbb{E}\left[\sum_{\{u,v\}\in E_{cut}} w\left(\{u,v\}\right)\left(\|\bar{u}\|^2 + \|\bar{v}\|^2\right)\right] \quad \leqslant \quad 4\,D\,k\sqrt{\mathsf{SDPval}}.$$

Since $\mathsf{SDPval} \leqslant \phi_G^k \leqslant 1$ (here we use that $d_u = w_u$), $\mathsf{SDPval} \leqslant \sqrt{\mathsf{SDPval}}$, and

$$\mathbb{E}\left[\sum_i \nu(S_i'')\right] \leqslant 8Dk\sqrt{\mathsf{SDPval}} + k\,\mathsf{SDPval} \leqslant (8D+1)k\sqrt{\mathsf{SDPval}}.$$

This concludes the proof. □

5.5 Orthogonal Separators with ℓ_2 Distortion

In this section, we sketch the proof of Theorem 5.4.1 which is proven in [31] as part of Lemma 4.9. Let us fix some notation. Let $\bar{\Phi}(t)$ be the probability that the standard $\mathcal{N}(0,1)$ Gaussian variable is greater than t. We will use the following easy lemma from [77].

Lemma 5.5.1 (Lemma 2.1. in [77]). *For every $t > 0$ and $\beta \in (0,1]$, we have*

$$\bar{\Phi}(\beta t) \leqslant \bar{\Phi}(t)^{\beta^2}.$$

We now describe an algorithm for m-orthogonal separators with ℓ_2 distortion (see Appendix 5.4). Let $\beta < 1$ be the separation threshold. Assume w.l.o.g. that all vectors $\bar{u}$ lie in $\mathbb{R}^n$. Fix $m' = m^{\frac{1+\beta}{1-\beta}}$ and $t = \bar{\Phi}^{-1}(1/m')$ (i.e., t such that $\bar{\Phi}(t) = 1/m'$). Sample a random Gaussian n dimensional vector γ in $\mathbb{R}^n$. Return the set

$$S = \{\bar{u} : \langle \bar{u}, \gamma \rangle \geqslant t\}.$$

We claim that S is an m-orthogonal separator with ℓ_2 distortion $\mathcal{O}(\sqrt{\log m})$ and scale $\alpha = 1/m'$. We now verify the conditions of orthogonal separators with ℓ_2 distortion.

1. For every $\bar{u}$,

$$\mathbb{P}\left[\bar{u} \in S\right] = \mathbb{P}\left[\langle \bar{u}, \gamma \rangle \geqslant t\right] = 1/m' \equiv \alpha.$$

Here we used that $\langle \bar{u}, \gamma \rangle$ is distributed as $\mathcal{N}(0,1)$, since $\bar{u}$ is a unit vector.

2. For every $\bar{u}$ and $\bar{v}$ with $\langle \bar{u}, \bar{v} \rangle \leqslant \beta$,

$$\begin{aligned}\mathbb{P}\left[\bar{u}, \bar{v} \in S\right] &= \mathbb{P}\left[\langle \bar{u}, \gamma \rangle \geqslant t \text{ and } \langle \bar{v}, \gamma \rangle \geqslant t\right] \\ &\leqslant \mathbb{P}\left[\langle \bar{u} + \bar{v}, \gamma \rangle \geqslant 2t\right].\end{aligned}$$

Note that $\|\bar{u}+\bar{v}\| = \sqrt{2+2\langle \bar{u},\bar{v}\rangle}$, hence $(\bar{u}+\bar{v})/\sqrt{2+2\langle \bar{u},\bar{v}\rangle}$ is a unit vector. We have

$$\begin{aligned}
\mathbb{P}[\bar{u},\bar{v}\in S] &\leqslant \mathbb{P}\Big[\Big\langle \frac{\bar{u}+\bar{v}}{\sqrt{2+2\langle \bar{u},\bar{v}\rangle}},\gamma\Big\rangle \geqslant \frac{2t}{\sqrt{2+2\langle \bar{u},\bar{v}\rangle}}\Big] \\
&= \bar{\Phi}\Big(\frac{\sqrt{2}t}{\sqrt{1+\langle \bar{u},\bar{v}\rangle}}\Big) \leqslant \bar{\Phi}\Big(\frac{\sqrt{2}t}{\sqrt{1+\beta}}\Big) \leqslant \bar{\Phi}(t)^{\frac{2}{1+\beta}} \\
&= \Big(\frac{1}{m'}\Big)^{\frac{2}{1+\beta}} = \frac{1}{m'}\cdot\Big(\frac{1}{m'}\Big)^{\frac{1-\beta}{1+\beta}} = \frac{\alpha}{m}.
\end{aligned}$$

3. The third property directly follows from Lemma A.2. in [31].

We note that this proof gives probability scale $\alpha = m^{-\frac{1+\beta}{1-\beta}}$. So, for some β, we may get $\alpha \ll 1/n$. However, it is easy to sample γ in such a way that $\mathbb{P}[\langle \bar{u},\gamma\rangle \geqslant 1/n]$ for every vector $\bar{u}$ in our set. To do so, we order vectors $\{\bar{u}\}$ in an arbitrary way: $\bar{u}_1,\ldots,\bar{u}_n$. Then, we pick a random index $\iota \in \{1,\ldots,n\}$, and sample a random Gaussian vector γ' conditional on $\langle \bar{u}_\iota,\gamma'\rangle \geqslant t$. We set $S' = \{\bar{u} : \langle \bar{u},\gamma'\rangle \geqslant t\}$ as in the algorithm above. Note that $\bar{u}_\iota$ always belongs to S'. We output $S'' = S'$ if S' does not contain vectors $\bar{u}_1,\ldots,\bar{u}_{\iota-1}$; and we output $S'' = \varnothing$ otherwise. It is easy to verify that $\mathbb{P}[\bar{u}\in S''] = 1/n$ for every $\bar{u}$, and, furthermore, for every non-empty set $S^* \neq \varnothing$,

$$\mathbb{P}[S'' = S^*] = \frac{1}{\alpha n}\,\mathbb{P}[S = S^*],$$

where S is the orthogonal separator from the proof above. So all properties of orthogonal separators hold for S'' with $\alpha' = \alpha/(\alpha n) = 1/n$.

5.6 *Integrality Gap for the Assignment* SDP

In this Section, we show that the standard *Assignment* SDP has high integrality gap.

Proposition 5.6.1. *SDP 13 has an unbounded integrality gap.*

Proof. Consider the following infinite family of graphs $\mathcal{G} = \{G_n : n \geqslant 0\}$. G_n consists of the two disjoint cliques of size $C_1 = K_{\lfloor n/2\rfloor}$ and $C_2 = K_{\lceil n/2\rceil}$. It is easy to see that for $\phi^k(G_n) = \Omega(1)$ for $k > 2$.

$$\min \alpha$$

$$\begin{aligned}
\sum_{\{u,v\} \in E} \|\bar{u}_i - \bar{v}_i\|^2 &\leqslant \alpha \sum_{u \in V} w_u \|\bar{u}_i\|^2 \qquad \forall i \in [k] \\
\sum_{i \in [k]} \|\bar{u}_i\|^2 &= 1 \\
\langle \bar{u}_i, \bar{u}_j \rangle &= 0 \qquad \forall i \neq j \text{ and } \forall u \in V \\
\left\langle \sum_i \bar{u}_i, I \right\rangle &= 1 \\
\|I\|^2 &= 1
\end{aligned}$$

Figure 13: Assignment SDP

For the sake of simplicity, let us assume that k is a multiple of 2. Let $e_1, \ldots, e_{k/2}$ be the standard basis vectors. Consider the following vector solution to SDP 13.

$$\bar{u}_i = \begin{cases} \sqrt{\frac{2}{k}} e_i & \text{if } u \in C_1 \text{ and } i \leqslant k/2 \\ \sqrt{\frac{2}{k}} e_{(i-k/2)} & \text{if } u \in C_2 \text{ and } i > k/2 \\ 0 & \text{otherwise} \end{cases}$$

and

$$I = \sqrt{\frac{2}{k}} \sum_{i=1}^{k/2} e_i \,.$$

It is easy to verify that this is a feasible solution with $\alpha = 0$. Therefore, SDP 13 has an unbounded integrality gap.

□

5.7 Conclusion

In this chapter we studied the Sparsest k-partition problem. Note that this differs from the k sparse-cuts problem (studied in Chapter 3) only in requiring that the k sets form a partition of the vertex set. Theorem 3.1.8 shows that ϕ_G^k can not be bounded by $\mathcal{O}\left(\sqrt{\lambda_k}\,\mathsf{polylog} k\right)$. In this chapter, we give an approximation algorithm for ϕ_G^k via a rounding algorithm for a novel SDP relaxation of ϕ_G^k. Our approximation algorithm is a bicriteria approximation algorithm. We leave it as an open problem to get a true $\mathcal{O}\left(\sqrt{\log n \log k}\right)$-approximation for ϕ_G^k.

Problem 5.7.1. Is there a randomized polynomial time algorithm that for every graph $G = (V, E, w)$, and for every parameter $k \in [n]$, outputs a k-partition, say $S_1, \ldots, S_k$, such that

$$\max_i \phi(S_i) \leqslant \mathcal{O}\left(\sqrt{\log n \log k}\, \phi_G^k\right)?$$

Acknowledgements. The results in this chapter were obtained in joint work with Konstantin Makarychev [70].

CHAPTER VI

APPROXIMATION ALGORITHMS FOR VERTEX EXPANSION AND HYPERGRAPH EXPANSION

6.1 Introduction

The problem of approximating Edge Expansion or Vertex Expansion, or Hypergraph Expansion can be studied at various regimes of parameters of interest. Perhaps the simplest possible version of the problem is to distinguish whether a given graph is an expander. Fix an absolute constant δ_0. A graph is a δ_0-vertex (edge) expander if its vertex (edge) expansion is at least δ_0. The problem of recognizing a vertex expander can be stated as follows:

Problem 6.1.1. Given a graph G, distinguish between the following two cases

(Non-Expander) the vertex expansion is $< \epsilon$

(Expander) the vertex expansion is $> \delta_0$ for some absolute constant δ_0.

Similarly, one can define the problem of recognizing an edge expander graph.

For the edge case, the Cheeger's inequality yields an algorithm to recognize an edge expander. In fact, it is possible to distinguish a δ_0 edge expander graph, from a graph whose edge expansion is $< \delta_0^2/2$, by just computing the second eigenvalue of the graph Laplacian.

It is natural to ask if there is an efficient algorithm with an analogous guarantee for vertex expansion. More precisely, is there some sufficiently small ϵ (an arbitrary function of δ_0), so that one can efficiently distinguish between a graph with vertex expansion $> \delta_0$ from one with vertex expansion $< \epsilon$. Bobkov et. al.[21] define a functional graph constant λ_∞ as follows.

$$\lambda_\infty \stackrel{\text{def}}{=} \min_{X \in \mathbb{R}^n} \frac{\sum_i \max_{j \sim i} (X_i - X_j)^2}{\sum_i X_i^2 - \frac{1}{n}(\sum_i X_i)^2} .$$

They also prove the following theorem relating λ_∞ to Φ^{V} in a Cheeger-like manner.

Theorem 6.1.2 ([21]). *For any unweighted, undirected graph G, we have*

$$\frac{\lambda_\infty}{2} \leqslant \Phi^{\mathsf{V}} \leqslant \sqrt{2\lambda_\infty}$$

We note that our definition of γ_2 for the hypergraph Laplacian Operator is very similar to the definition of λ_∞, and Theorem 6.1.2 is very similar to Theorem 4.2.11.

While Theorem 6.1.2 and Theorem 4.2.11 seem to suggest that vertex expanders and hypergraph expanders can be identified by computation of λ_∞ and γ_2 respectively, in the same way as edge expanders can be identified by computing λ_2, the computation of λ_∞ and γ_2 seems intractable. In Chapter 8, we show a hardness result suggesting that there is no efficient algorithm to recognize vertex expanders. More precisely, we prove a hardness result for the problem of approximating λ_∞ in graphs of bounded degree d. The hardness result shows that the approximability of vertex expansion degrades with the degree, and therefore the problem of recognizing expanders is hard for sufficiently large degree. We get similar hardness results for γ_2 and hypergraph expansion via the reduction from Vertex Expansion to Hypergraph Expansion (Theorem 4.2.19).

In this chapter we present approximation algorithms for λ_∞ and the Hypergraph Eigenvalues whose guarantee matches the hardness result up to constant factors. We use this to obtain approximation algorithms for Vertex Expansion and Hypergraph Expansion. We state our results formally in Section 6.1.1.

6.1.1 Formal Statement of Results.

Vertex Expansion. Our first result is a simple polynomial-time algorithm to obtain a $\mathcal{O}(\log d)$ approximation to λ_∞ in graphs having largest degree d. Via our algorithmic proof of Theorem 6.1.2, this directly implies an algorithm to obtain a subset of vertices S whose vertex expansion is at most $\mathcal{O}\left(\sqrt{\Phi^{\mathsf{V}} \log d}\right)$.

Theorem 6.1.3. *There exists a polynomial time algorithm which given a graph $G = (V, E)$ having vertex degrees at most d, outputs a vector $X \in \mathbb{R}^n$ such that*

$$\frac{\sum_{i \in V} \max_{j \sim i} (X_i - X_j)^2}{\sum_i X_i^2 - \frac{1}{n}(\sum_i X_i)^2} \leqslant \mathcal{O}\left(\lambda_\infty \log d\right)$$

and outputs a set $S \subset V$, such that $\Phi^{\mathsf{V}}(S) = \mathcal{O}\left(\sqrt{\Phi_G^{\mathsf{V}} \log d}\right)$.

In Chapter 8 we will show that Vertex Expansion and Symmetric Vertex Expansion are computationally equivalent upto constant factors (Theorem 8.3.1 and Theorem 8.3.2). Using this we get an approximation algorithm for Vertex Expansion as well.

Corollary 6.1.4 (Corollary to Theorem 6.1.3 and Theorem 8.3.2)**.** *There exists a polynomial time algorithm which given a graph $G = (V, E)$ having vertex degrees at most d, outputs a set $S \subset V$, such that $\phi^{\mathsf{V}}(S) = \mathcal{O}\left(\sqrt{\phi_G^{\mathsf{V}} \log d}\right)$.*

Hypergraph Expansion. Computing the eigenvalues of the hypergraph Markov operator (Definition 4.2.1) is intractable, as the operator is non-linear. We gave an exponential time algorithm to compute all the eigenvalues and eigenvectors of M and L (Theorem 4.9.1). We give a polynomial time $\mathcal{O}(k \log r)$-approximation algorithm to compute the k^{th} smallest eigenvalue, where r is the size of the largest hyperedge.

Theorem 6.1.5. *There exists a randomized polynomial time algorithm that given a hypergraph $H = (V, E, w)$ and a parameter $k < |V|$, outputs k orthonormal vectors $u_1, \ldots, u_k$ such that for each $i \in [k]$,*

$$\mathcal{R}(u_i) \leqslant \mathcal{O}(i \log r \ \gamma_i)$$

where r is the size of the largest hyperedge.

Theorem 4.2.11 gives a bound on ϕ_H in terms of γ_2. Obtaining a $\mathcal{O}(\log r)$-approximation to γ_2 from Theorem 6.1.5 gives us the following result directly.

Corollary 6.1.6 (Corollary to Theorem 4.2.11 and Theorem 6.1.5)**.** *There exists a randomized polynomial time algorithm that given a hypergraph $H = (V, E, w)$, outputs a set $S \subset V$ such that*

$$\phi(S) \leqslant \mathcal{O}\left(\sqrt{\phi_H \log r}\right)$$

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

Hypergraph Balanced Separator. We recall Hypergraph Balanced Separator problem (Problem 2.1.6).

Problem 6.1.7 (Hypergraph Balanced Separator)**.** Given a hypergraph $H = (V, E, w)$, and a *balance* parameter $c \in (0, 1/2]$, a set $S \subset V$ is said to be c-balanced if $cn \leqslant |S| \leqslant (1-c)n$. The c-Hypergraph Balanced Separator problem asks to compute the c-balanced set $S \subset V$ which has the least *sparsity* $\mathsf{sp}(S)$ defined as follows.

$$\mathsf{sp}(S) \stackrel{\text{def}}{=} n \cdot \frac{w\left(E(S, \bar{S})\right)}{|S|\left|\bar{S}\right|}.$$

In a seminal work, Arora, Rao and Vazirani [13] gave a $\mathcal{O}\left(\sqrt{\log n}\right)$ approximation algorithm for the Balanced Separator problem in graphs. We present an analog of this result for hypergraphs.

Theorem 6.1.8. *There exists a randomized polynomial time algorithm that given $H = (V, E, w)$, an instance of the c-Hypergraph Balanced Separator problem, outputs a c'-balanced set $S \subset V$ such that $\mathsf{sp}(S) = \mathcal{O}\left(\sqrt{\log n}\right) \mathsf{OPT}$, where OPT is the least sparsity of a c-balanced set and $c' \geqslant c/100$.*

Our algorithm for Hypergraph Balanced Separator is a bi-criteria algorithm in that it outputs a set of size at least $c'n$ instead of a set of size at least cn ($c > c'$).

We note that this is similar to algorithm for Arora, Rao and Vazirani [13] which also finds a set of size at least $c'n$ instead of a set of size at least cn .

Balanced Vertex Separator. Our techniques can also be used to obtain an approximation algorithm for Balanced Vertex Expansion in graphs (Definition 2.1.4).

Theorem 6.1.9 (Corollary to Theorem 6.1.8 and Theorem 7.1.5)**.** *There is a randomized polynomial-time algorithm that given a graph $G = (V, E)$, an instance of the c-*Balanced Vertex Expansion *problem, outputs a c'-balanced set $S \subset V$ such that $\mathsf{sp}(S) = \mathcal{O}\left(\sqrt{\log n}\right) \mathsf{OPT}$. Here $c' \geqslant c/100$.*

6.1.2 Proof Overview

We give a $\mathcal{O}(k \log r)$-approximation algorithm for γ_k (Theorem 6.1.5). Our algorithm proceeds inductively. We assume that we have computed $k-1$ orthonormal vectors $u_1, \ldots, u_{k-1}$ such that $\mathcal{R}(u_i) \leqslant \mathcal{O}(i \log r \, \gamma_i)$, and show how to compute γ_k. Our main idea is to show that there exists a unit vector $X \in \mathsf{span}\{\mathsf{v}_1, \ldots, \mathsf{v}_k\}$ which is orthogonal to $\mathsf{span}\{u_1, \ldots, u_{k-1}\}$ and has small Rayleigh quotient. Note that unlike the case of matrices, for an $X \in \mathsf{span}\{\mathsf{v}_1, \ldots, \mathsf{v}_k\}$, we can not bound $X^T L(X)$ by $\max_{i \in [k]} \mathsf{v}_i^T L(\mathsf{v}_i)$. The operator L is non-linear, and there is no reason to believe that something like the celebrated *Courant-Fischer Theorem* for matrices holds for this operator. In general, for an $X \in \mathsf{span}\{\mathsf{v}_1, \ldots, \mathsf{v}_k\}$, the Rayleigh quotient can be much larger than γ_k. We will show that for such an X, $\mathcal{R}(X) \leqslant k \, \gamma_k$. However, we still do not have a way to compute such a vector X. We given an SDP relaxation and a rounding algorithm to compute an "approximate" X.

Hypergraph Balanced Separator To prove Theorem 6.1.8, we start with an SDP relaxation of the Rayleigh quotient together with ℓ_2^2-triangle inequality constraints. We use the framework of Arora et. al.[13], to find two well separated sets in the SDP solution. We use these sets as "guides" and find a set with small sparsity in the same way as we do in the proof of the Hypergraph Cheeger's Inequality.

6.1.3 Organization

We give our approximation algorithm for λ_∞ and VERTEX EXPANSION in Section 6.2. We present our approximation algorithms for hypergraph eigenvalues in Section 6.3. We prove Theorem 6.1.8 in Section 6.4.

6.2 *An* Optimal *Algorithm for Vertex Expansion*

In this section we give a simple polynomial time algorithm which outputs a set S whose vertex expansion is at most $\mathcal{O}\left(\sqrt{\Phi^{\mathsf{V}} \log d}\right)$. We restate Theorem 6.1.3.

Theorem 6.2.1 (Restatement of Theorem 6.1.3). *There exists a polynomial time algorithm which given a graph $G = (V, E)$ having vertex degrees at most d, outputs a vector $X \in \mathbb{R}^n$ such that*

$$\frac{\sum_{i \in V} \max_{j \sim i} (X_i - X_j)^2}{\sum_i X_i^2 - \frac{1}{n}(\sum_i X_i)^2} \leqslant \mathcal{O}\left(\lambda_\infty \log d\right)$$

and outputs a set $S \subset V$, such that $\Phi^{\mathsf{V}}(S) = \mathcal{O}\left(\sqrt{\Phi_G^{\mathsf{V}} \log d}\right)$.

Consider the following SDP relaxation of λ_∞ (SDP 6.2.2).

SDP 6.2.2.

$$\mathsf{SDPval} \stackrel{\text{def}}{=} \min \sum_{i \in} \alpha_i$$

subject to:

$$\begin{aligned} \|v_j - v_i\|^2 &\leqslant \alpha_i \qquad \forall i \in V \text{ and } \forall j \sim i \\ \sum_i \|v_i\|^2 - \frac{1}{n}\left\|\sum_i v_i\right\|^2 &= 1 \end{aligned}$$

Figure 14: SDP Relaxation for λ_∞.

It's easy to see that this is a relaxation for λ_∞. We present a simple randomized rounding of this SDP which, with constant probability, outputs a set with vertex expansion at most $C\sqrt{\phi^{\mathsf{V}} \log d}$ for some absolute constant C.

Algorithm 6.2.3.

- *Input :* A graph $G = (V, E)$
- *Output :* A set S with vertex expansion at most $576\sqrt{\mathsf{SDPval} \log d}$ (with constant probability).

1. Solve SDP 6.2.2 for graph G.
2. Pick a random Gaussian vector $g \sim N(0,1)^n$. For each $i \in [n]$, define
$$x_i \stackrel{\text{def}}{=} \langle v_i, g \rangle \ .$$
3. Sort the x_i's in decreasing order $x_{i_1} \geqslant x_{i_2} \geqslant \ldots x_{i_n}$. Let S_j denote the set of the first j vertices appearing in the sorted order. Let l be the index such that
$$l = \mathsf{argmin}_{1 \leqslant j \leqslant n/2} \Phi^{\mathsf{V}}(S_j) \, .$$

Figure 15: Rounding Algorithm

We first prove a technical lemma which shows that we can a recover a set with small vertex expansion from a *good* line-embedding (Step 5 in Algorithm 6.2.3).

Lemma 6.2.4. *Let $Y \in (\mathbb{R}^+)^n$ be any vector. Then $\exists S \subseteq \mathsf{supp}(Y)$ such that*
$$\Phi^{\mathsf{V}}(S) \leqslant \frac{\sum_i \max_{j \sim i} |Y_j - Y_i|}{\sum_i Y_i} \, .$$
Moreover, such a set can be computed in polynomial time.

The proof of this lemma follows from Proposition 4.4.2. We prove it here again for completeness.

Proof of Lemma 6.2.4. W.l.o.g we may assume that $Y_1 \geqslant Y_2 \geqslant \ldots \geqslant Y_n \geqslant 0$. Let α denote
$$\alpha \stackrel{\text{def}}{=} \frac{\sum_i \max_{j \sim i} |Y_j - Y_i|}{\sum_i Y_i} \, .$$

Let $i_{\max} \stackrel{\text{def}}{=} \mathsf{argmax}_i\, Y_i > 0$, i.e. $i_{\max}$ be the largest index such that $Y_{i_{\max}} > 0$. Let $S_i \stackrel{\text{def}}{=} \{Y_1, \ldots, Y_i\}$. Let us consider the following case

$$\left|N^V(S_i) \cup N^V(\bar{S}_i)\right| > \alpha \left|S_i\right| \qquad \forall i < i_{\max} .$$

Then,

$$\begin{aligned}
\alpha = \frac{\sum_i \max_{j\sim i}(Y_j - Y_i)}{\sum_i Y_i} &\geqslant \frac{\sum_i \max_{j\sim i} \sum_{l=j}^{l=i-1}(Y_l - Y_{l+1})}{\sum_i Y_i} \\
&= \frac{\sum_i (Y_i - Y_{i+1}) \left|N^V(S_i) \cup N^V(\bar{S}_i)\right|}{\sum_i Y_i} \\
&> \alpha \frac{\sum_i (Y_i - Y_{i+1}) \left|S_i\right|}{\sum_i Y_i} \\
&= \alpha
\end{aligned}$$

Thus we get $\alpha > \alpha$ which is a contradition. Therefore, $\exists i \leqslant i_{\max}$ such that $\Phi^{\mathsf{V}}(S_i) \leqslant \alpha$. □

Next we show a λ_∞-like bound for the x_i's.

Lemma 6.2.5. *Let $x_1, \ldots, x_n$ be as defined in Algorithm 6.2.3. Then, with constant probability, we have*

$$\frac{\sum_i \max_{j\sim i}(x_i - x_j)^2}{\sum_i x_i^2 - \frac{1}{n}\left(\sum_i x_i\right)^2} \leqslant 96 \ \mathsf{SDPval} \log d .$$

Proof. Using Fact 2.5.5 we get,

$$\mathbb{E}\left[\max_{j\sim i}(x_j - x_j)^2\right] = \mathbb{E}\left[\max_{j\sim i} \langle v_i - v_j, g\rangle^2\right] \leqslant 2 \max_{j\sim i} \left\|v_j - v_i\right\|^2 \log d .$$

Therefore, $\mathbb{E}\left[\sum_i \max_{j\sim i}(x_j - x_j)^2\right] \leqslant 2\ \mathsf{SDPval} \log d$. Using Markov's Inequality we get

$$\mathbb{P}\left[\sum_i \max_{j\sim i}(x_j - x_j)^2 > 48\ \mathsf{SDPval} \log d\right] \leqslant \frac{1}{24} \tag{50}$$

For the denominator, using linearity of expectation, we get

$$\mathbb{E}\left[\sum_i x_i^2 - \frac{1}{n}\left(\sum_i x_i\right)^2\right] = \sum_i \|v_i\|^2 - \frac{1}{n}\left\|\sum_i v_i\right\|^2.$$

Also recall that the denominator can be re-written as

$$\sum_i x_i^2 - \frac{1}{n}\left(\sum_i x_i\right)^2 = \frac{1}{n}\sum_{i,j}(x_i - x_j)^2\,,$$

which is a sum of squares of Gaussian random variables. Now applying Lemma 2.5.6 to the denominator we conclude

$$\mathbb{P}\left[\sum_i x_i^2 - \frac{1}{n}\left(\sum_i x_i\right)^2 \geqslant \frac{1}{2}\right] \geqslant \frac{1}{12}. \tag{51}$$

Using (50) and (51) we get that

$$\mathbb{P}\left[\frac{\sum_i \max_{j\sim i}(x_i - x_j)^2}{\sum_i x_i^2 - \frac{1}{n}\left(\sum_i x_i\right)^2} \leqslant 96\ \mathsf{SDPval}\log d\right] > \frac{1}{24}.$$

□

We will use the following fact from [21]. For the sake of completeness, we prove it here again.

Lemma 6.2.6 ([21])**.** *Let $z_1, \ldots, z_n \in R$. Then there exists $x \in \mathbb{R}^n$ such that*

$$\frac{\sum_i \max_{j\sim i}\left|x_i^2 - x_j^2\right|}{\sum_i x_i^2 - \frac{1}{n}\left(\sum_i x_i\right)^2} \leqslant 6\sqrt{\frac{\sum_i \max_{j\sim i}(z_i - z_j)^2}{\sum_i z_i^2 - \frac{1}{n}\left(\sum_i z_i\right)^2}}.$$

Proof. W.l.o.g we may assume that $|\mathsf{supp}(Z^+)| = |\mathsf{supp}(Z^-)| = \lceil n/2\rceil$ and that $z_1 \geqslant z_2 \geqslant \ldots \geqslant z_n$.

Note that for any $i \in [n]$, we have

$$\max_{j\sim i,\, j<i}(z_j^+ - z_i^+)^2 + \max_{j\sim i,\, j>i}(z_j^- - z_i^-)^2 \leqslant 2\max_{j\sim i}(z_j - z_i)^2\,.$$

Therefore,

$$\begin{aligned}
&\frac{\sum_i \max_{j\sim i}(z_j - z_i)^2}{\sum_i z_i^2} \\
&\geqslant \frac{\sum_i \max_{j\sim i,\, j<i}(z_j^+ - z_i^+)^2 + \sum_i \max_{j\sim i,\, j>i}(z_j^- - z_i^-)^2}{2\left(\sum_{i\in\mathsf{supp}(Z^+)} z_i^2 + \sum_{i\in\mathsf{supp}(Z^-)} z_i^2\right)} \\
&\geqslant \min\left\{\frac{\sum_i \max_{j\sim i,\, j<i}(z_j^+ - z_i^+)^2}{2\sum_{i\in\mathsf{supp}(Z^+)} z_i^2}, \frac{\sum_i \max_{j\sim i,\, j>i}(z_j^- - z_i^-)^2}{2\sum_{i\in\mathsf{supp}(Z^-)} z_i^2}\right\}
\end{aligned}$$

W.l.o.g we may assume that

$$\frac{\sum_i \max_{j\sim i,\, j<i}(z_j^+ - z_i^+)^2}{\sum_{i\in\mathsf{supp}(Z^+)} z_i^2} \leqslant \frac{\sum_i \max_{j\sim i,\, j>i}(z_j^- - z_i^-)^2}{\sum_{i\in\mathsf{supp}(Z^-)} z_i^2}$$

Let $x \stackrel{\text{def}}{=} z^+$. Then we get,

$$\frac{\sum_i \max_{j\sim i}(x_j - x_i)^2}{\sum_i x_i^2} \leqslant 2\frac{\sum_i \max_{j\sim i}(z_j - z_i)^2}{\sum_i z_i^2}$$

We have

$$\begin{aligned}
\sum_i \max_{j\sim i,\, j<i}(x_j^2 - x_i^2) &= \sum_i \max_{j\sim i,\, j<i}(x_j - x_i)(x_j + x_i) \\
&\leqslant \sum_i \max_{j\sim i,\, j<i}\left((x_j - x_i)^2 + 2x_i(x_j - x_i)\right) \\
&\leqslant \sum_i \max_{j\sim i,\, j<i}(x_j - x_i)^2 + 2\sum_i x_i \max_{j\sim i,\, j<i}(x_j - x_i) \\
&\leqslant \sum_i \max_{j\sim i,\, j<i}(x_j - x_i)^2 + 2\sqrt{\sum_i x_i^2}\sqrt{\max_{j\sim i, j<i}(x_j - x_i)^2} \\
&\qquad\qquad \text{(Using the Cauchy-Schwarz inequality)} \\
&\leqslant \lambda_\infty \sum_i x_i^2 + 2\sqrt{\lambda_\infty}\sum_i x_i^2
\end{aligned}$$

Thus we have

$$\frac{\sum_i \max_{j\sim i, j<i}(x_j^2 - x_i^2)}{\sum_i x_i^2} \leqslant 6\sqrt{\frac{\sum_i \max_{j\sim i}(z_j - z_i)^2}{\sum_i z_i^2}}$$

□

We are now ready to complete the proof of Theorem 6.1.3.

Proof of Theorem 6.1.3. Let the x_i's be as defined in Algorithm 6.2.3. W.l.o.g, we may assume that $|\mathsf{supp}(x^+)| < |\mathsf{supp}(x^-)|$. For each $i \in [n]$, we define $y_i = x_i^+$. Using Lemma 6.2.6, we get

$$\frac{\sum_i \max_{j \sim i} \left|y_i^2 - y_j^2\right|}{\sum_i y_i^2 - \frac{1}{n}\left(\sum_i y_i\right)^2} \leqslant 6\sqrt{\frac{\sum_i \max_{j\sim i}(x_i - x_j)^2}{\sum_i x_i^2 - \frac{1}{n}\left(\sum_i x_i\right)^2}}\,.$$

Using Lemma 6.2.5, we get

$$\frac{\sum_i \max_{j \sim i} \left|y_i^2 - y_j^2\right|}{\sum_i y_i^2 - \frac{1}{n}\left(\sum_i y_i\right)^2} \leqslant 576\sqrt{\mathsf{SDPval}\log d}.$$

From Lemma 6.2.4 we get that the set output in Step 3 of Algorithm 6.2.3 has vertex expansion at most $576\sqrt{\mathsf{SDPval}\log d}$. □

6.3 Approximation Algorithms for Hypergraph Eigenvalues

Since L is a non-linear operator, computing its eigenvalues exactly is intractable. In this section we give a $\mathcal{O}(k \log r)$-approximation algorithm for γ_k.

Theorem 6.3.1 (Restatement of Theorem 6.1.5)**.** *There exists a randomized polynomial time algorithm that, given a hypergraph $H = (V, E, w)$ and a parameter $k < |V|$, outputs k orthonormal vectors $u_1, \ldots, u_k$ such that for each $i \in [k]$*

$$\mathcal{R}(u_i) \leqslant \mathcal{O}(i \log r\, \gamma_i)\ .$$

We will prove this theorem inductively. We already know that $\gamma_1 = 0$ and $v_1 = \mathbf{1}/\sqrt{n}$. Now, we assume that we have computed $k-1$ orthonormal vectors $u_1, \ldots, u_{k-1}$ such that $\mathcal{R}(u_i) \leqslant \mathcal{O}(i \log r\, \gamma_i)$. We will now show how to compute u_k.

Our main idea is to show that there exists a unit vector $X \in \mathsf{span}\{\mathsf{v}_1, \ldots, \mathsf{v}_k\}$ which is orthogonal to $\mathsf{span}\{u_1, \ldots, u_{k-1}\}$. We will show that for such an X, $\mathcal{R}(X) \leqslant k\,\gamma_k$ (Proposition 6.3.2). Then we give an SDP relaxation (SDP 6.3.3) and a rounding algorithm (Algorithm 6.3.4, Lemma 6.3.5) to compute an "approximate" X'.

Proposition 6.3.2. *Let $u_1, \ldots, u_{k-1}$ be arbitrary orthonormal vectors. Then*

$$\min_{X \perp u_1, \ldots, u_{k-1}} \mathcal{R}(X) \leqslant k\, \gamma_k \,.$$

Proof. Consider subspaces $S_1 \stackrel{\text{def}}{=} \mathsf{span}\{u_1, \ldots, u_{k-1}\}$ and $S_2 \stackrel{\text{def}}{=} \mathsf{span}\{\mathsf{v}_1, \ldots, \mathsf{v}_k\}$. Since $\mathsf{rank}(S_2) > \mathsf{rank}(S_1)$, there exists $X \in S_2$ such that $X \perp S_1$. We will now show that this X satisfies $\mathcal{R}(X) \leqslant \mathcal{O}(k\, \gamma_k)$, which will finish this proof. Let $X = c_1 \mathsf{v}_1 + \ldots + c_k \mathsf{v}_k$ for scalars $c_i \in \mathbb{R}$ such that $\sum_i c_i^2 = 1$.

Recall that γ_k is defined as

$$\gamma_k \stackrel{\text{def}}{=} \min_{Y \perp \mathsf{v}_1, \ldots, \mathsf{v}_{k-1}} \frac{Y^T L_Y Y}{Y^T Y} \,.$$

We can restate the definition of γ_k as follows,

$$\gamma_k = \min_{Y \perp \mathsf{v}_1, \ldots, \mathsf{v}_{k-1}} \max_{Z \in \mathbb{R}^n} \frac{Y^T L_Z Y}{Y^T Y} \,.$$

Therefore,

$$\gamma_k = \mathsf{v}_k^T L_{\mathsf{v}_k} \mathsf{v}_k \geqslant \mathsf{v}_k^T L_X \mathsf{v}_k \,. \tag{52}$$

The Laplacian matrix L_X, being positive semi-definite, has a Cholesky Decomposition into matrices B_X such that $L_X = B_X B_X^T$.

$$\begin{aligned}
\mathcal{R}(X) = X^T L_X X = \sum_{i,j \in [k]} c_i c_j \mathsf{v}_i^T B_X B_X^T \mathsf{v}_j & \qquad \text{(Cholesky Decomposition of } L_X \text{)} \\
\leqslant \sum_{i,j \in [k]} |c_i c_j| \, \|B_X \mathsf{v}_i\| \cdot \|B_X \mathsf{v}_i\| & \qquad \text{(Cauchy-Schwarz inequality)} \\
= \sum_{i,j \in [k]} |c_i c_j| \sqrt{\mathsf{v}_i^T L_X \mathsf{v}_i} \sqrt{\mathsf{v}_j^T L_X \mathsf{v}_j} \leqslant \sum_{i,j \in [k]} |c_i c_j| \sqrt{\gamma_1 \gamma_j} & \qquad \text{(Using (52))} \\
\leqslant \left(\sum_i |c_i| \right)^2 \max_{i,j} \sqrt{\gamma_i \gamma_j} \leqslant k\, \gamma_k \,. &
\end{aligned}$$

□

Next we present an SDP relaxation (SDP 6.3.3) to compute the vector orthogonal $u_1, \ldots, u_{k-1}$ having the least Rayleigh quotient. The vector $\bar{Y}_i$ is the relaxation of the i^{th} coordinate of the vector u_k that we are trying to compute. The objective function of the SDP and (53) seek to minimize the Rayleigh quotient; Proposition 6.3.2 shows that the objective value of this SDP is at most $k\,\gamma_k$. (54) demands the solution be orthogonal to $u_1, \ldots, u_{k-1}$.

SDP 6.3.3.

$$\mathsf{SDPval} \stackrel{\text{def}}{=} \min \sum_{e \in E} w(e) \max_{i,j \in e} \left\| \bar{Y}_i - \bar{Y}_j \right\|^2 .$$

subject to

$$\sum_{i \in V} \left\| \bar{Y}_i \right\|^2 = 1 \tag{53}$$

$$\sum_{i \in V} u_l(i)\, \bar{Y}_i = 0 \qquad \forall l \in [k-1] \tag{54}$$

Figure 16: SDP Relaxation for for γ_k.

Algorithm 6.3.4 (Rounding Algorithm for Computing Eigenvalues)**.**

1. Solve SDP 6.3.3 on the input hypergraph H with the previously computed $k-1$ vectors $u_1, \ldots, u_{k-1}$.
2. Sample a random Gaussian vector $g \sim \mathcal{N}(0,1)^n$. Set $X_i \stackrel{\text{def}}{=} \left\langle \bar{Y}_i, g \right\rangle$.
3. Output $X / \left\| X \right\|$.

Figure 17: Rounding Algorithm for γ_k.

Lemma 6.3.5. *With constant probability Algorithm 6.3.4 outputs a vector u_k such that*

1. $u_k \perp u_l \; \forall l \in [k-1]$.

2. $\mathcal{R}(u_k) \leqslant 192\,\mathsf{SDPval}\,\log r$.

Proof. We first verify condition (1). For any $l \in [k-1]$, we using (54)

$$\langle X, u_l \rangle = \sum_{i \in V} \langle \bar{Y}_i, g \rangle\, u_l(i) = \left\langle \sum_{i \in V} u_l(i)\, \bar{Y}_i, g \right\rangle = 0\,.$$

We now prove condition (2). To bound $\mathcal{R}(X)$ we need an upper bound on the numerator and a lower bound on the denominator of the $\mathcal{R}(\cdot)$ expression. For the sake of brevity let L denote L_X. Then Using Fact 2.5.5

$$\begin{aligned}\mathbb{E}\left[X^T L X\right] &\leqslant \sum_{e \in E} w(e)\, \mathbb{E}\left[\max_{i,j \in e} (X_i - X_j)^2\right] \leqslant 4 \log r \sum_{e \in E} w(e) \max_{i,j \in e} \left\|\bar{Y}_i - \bar{Y}_j\right\|^2 \\ &= 4\,\mathsf{SDPval}\,\log r\,.\end{aligned}$$

Therefore, by Markov's Inequality,

$$\mathbb{P}\left[X^T L X \leqslant 96\,\mathsf{SDPval}\,\log r\right] \geqslant 1 - \frac{1}{24}\,. \tag{55}$$

For the denominator, using linearity of expectation, we get

$$\mathbb{E}\left[\sum_{i \in V} X_i^2\right] = \sum_i \mathbb{E}\left[\langle \bar{Y}_i, g \rangle^2\right] = \sum_i \left\|\bar{Y}_i\right\|^2 = 1 \qquad \text{(Using (53))}\,.$$

Now applying Lemma 2.5.6 to the denominator we conclude

$$\mathbb{P}\left[\sum_i X_i^2 \geqslant \frac{1}{2}\right] \geqslant \frac{1}{12}\,. \tag{56}$$

Using Union-bound on (55) and (56) we get that

$$\mathbb{P}\left[\mathcal{R}(X) \leqslant 192\,\mathsf{SDPval}\,\log r\right] \geqslant \frac{1}{24}\,.$$

□

We now have all the ingredients to prove Theorem 6.1.5.

Proof of Theorem 6.1.5. We will prove this theorem inductively. For the basis of induction, we have the first eigenvector $u_1 = \mathsf{v}_1 = \mathbf{1}/\sqrt{n}$. We assume that we have computed $u_1, \ldots, u_{k-1}$ satisfying $\mathcal{R}(u_i) \leqslant \mathcal{O}(i \log r\, \gamma_i)$. We now show how to compute u_k.

Proposition 6.3.2 implies that for SDP 6.3.3,

$$\mathsf{SDPval} \leqslant k\, \gamma_k \,.$$

Therefore, from Lemma 6.3.5, we get that Algorithm 6.3.4 will output a unit vector which is orthogonal to all u_i for $i \in [k-1]$ and

$$\mathcal{R}(u_k) \leqslant 192\, k \log r\, \gamma_k \,.$$

□

6.4 *Algorithm for* Hypergraph Balanced Separator

In this section we prove Theorem 6.1.8.

Theorem 6.4.1 (Restatement of Theorem 6.1.8)**.** *There exists a randomized polynomial time algorithm that given $H = (V, E, w)$, an instance of the* c-Hypergraph Balanced Separator *problem, outputs a c'-balanced set $S \subset V$ such that* $\mathsf{sp}(S) = \mathcal{O}\left(\sqrt{\log n}\right)\, \mathsf{OPT}$*, where* OPT *is the least sparsity of a c-balanced set and* $c' \geqslant c/100$.

Proof of Theorem 6.1.8. We prove this theorem by giving an SDP relaxation for this problem (SDP 6.4.2) and a rounding algorithm for it (Algorithm 6.4.4). Firstly, we need a suitable objective function for the relaxation that captures hypergraph expansion. Motivated by Theorem 4.2.11, we can have objective function to be a relaxation of

$$X^T L(X) = \sum_{e \in E} \max_{i,j \in e} (X_i - X_j)^2 \,.$$

We relax the scalar X_u to be a vector $\bar{u}$. Ideally, we would want all X_u to be in the set $\{-1, 1\}$ so that we can identify the cut. Therefore, we add the constraint that all vectors $\bar{u}$ have length 1 (57). Since we want the integral solution to be c-balanced, we add the corresponding constraint for vectors (58) . Finally, we add ℓ_2^2 triangle inequality constraints between all triplets of vertices (59), as all integral solutions of the relaxation will trivially satisfy this.

SDP 6.4.2.

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

subject to

$$\|\bar{u}\|^2 = 1 \qquad \forall u \in V \tag{57}$$

$$\sum_{u,v} \|\bar{u} - \bar{v}\|^2 \geqslant 4c(1-c)\,|V|^2 \tag{58}$$

$$\|\bar{u} - \bar{v}\|^2 + \|\bar{v} - \bar{w}\|^2 \geqslant \|\bar{u} - \bar{w}\|^2 \qquad \forall u, v, w \in V \tag{59}$$

Figure 18: SDP Relaxation for for Hypergraph Balanced Separator.

Our main ingredient is the following theorem due to [13].

Theorem 6.4.3 ([13])**.** *There exists a randomized polynomial time algorithm that given an ℓ_2^2-space on $X = \{\bar{u}\}$ satisfying $\sum_{u,v} \|\bar{u} - \bar{v}\|^2 \geqslant 4c(1-c)n^2$, outputs two sets $S, T \subset X$ such that $|S|, |T| \geqslant c'n$ and*

$$\min_{u \in S, v \in T} \|\bar{u} - \bar{v}\|^2 \geqslant \frac{1}{C\sqrt{\log n}}\,.$$

Here c', C are functions only of c.

By definition X_u, X_v and using (59), it follows that

$$|X_u - X_v| \leqslant \|\bar{u} - \bar{v}\|^2 \qquad \forall u, v \in V\,.$$

Next, using Theorem 6.4.3,

$$\sum_{u,v} |X_u - X_v| \geqslant \sum_{u \in S, v \in T} |X_u - X_v| \geqslant c'^2 n^2 \frac{1}{C\sqrt{\log n}}\,.$$

Algorithm 6.4.4.

1. Solve SDP 6.4.2.
2. Compute sets S, T using Theorem 6.4.3.
3. For each $u \in V$, define $X_u \stackrel{\text{def}}{=} \min_{v \in S} \|\bar{u} - \bar{v}\|^2$. Sort the $\{X_u : u \in V\}$ in increasing order and output the set $A \subset V$ having the least sparsity in this ordering (See Proposition 4.4.2).

Figure 19: Rounding Algorithm for Hypergraph Balanced Separator

Therefore,

$$\begin{aligned}
\frac{\sum_{e \in E} \max_{u,v \in e} |X_u - X_v|}{\sum_{u,v} |X_u - X_v|} &\leqslant \frac{C\sqrt{\log n}}{c'^2} \cdot \frac{\sum_{e \in E} \max_{u,v \in e} \|\bar{u} - \bar{v}\|^2}{n^2} \\
&\leqslant \frac{C\sqrt{\log n}}{c'^2} \cdot \frac{\sum_{e \in E} \max_{u,v \in e} \|\bar{u} - \bar{v}\|^2}{\sum_{u,v} \|\bar{u} - \bar{v}\|^2} \\
&\qquad \left(\text{Since} \sum_{u,v} \|\bar{u} - \bar{v}\|^2 \leqslant n^2\right) \\
&\leqslant \frac{C\sqrt{\log n}}{c'^2} \, \mathsf{OPT} \,.
\end{aligned}$$

Invoking Proposition 4.4.2, we get that the set A output by Algorithm 6.4.4 satisfies $|A| \in [c'n, (1 - c')n]$ and

$$\mathsf{sp}(A) \leqslant \frac{C\sqrt{\log n}}{c'^2} \, \mathsf{OPT} \,.$$

This finishes the proof of the theorem.

□

6.5 Conclusion

In this chapter we gave *optimal* approximation algorithms for vertex expansion and hypergraph expansion via approximation algorithms for λ_∞ and the hypergraph eigenvalues. The approximation factors for λ_∞ and γ_2 are optimal (upto constant factors) under SSE. We get a $\mathcal{O}(k \log r)$-approximation algorithm for γ_k, but we do not know of any hardness result for γ_k other than what follows from the hardness for

γ_2. Closing the gap between the approximation upper bounds and the computational lower bounds for γ_2 is left as an open problem.

Acknowledgements. The results in Section 6.2 were obtained in joint work with Prasad Raghavendra and Santosh Vempala [75].

CHAPTER VII

APPROXIMATION ALGORITHMS FOR SMALL SET EXPANSION PROBLEMS

7.1 Introduction

In this chapter, we study the "small set" versions of the Hypergraph Expansion problem and the Vertex Expansion problem. As in Chapter 5, we can again afford to work with a more general definition of expansion. Given a hypergraph $H = (V, E, w)$ where weight function $w : V \cup E \to \mathbb{R}^+$, the expansion of a set $S \subset V$ is defined as

$$\phi(S) \stackrel{\text{def}}{=} \frac{\sum_{e \in E(S,\bar{S})} w(e)}{\sum_{u \in V} w(u)} .$$

We recall the Hypergraph Small Set Expansion (Problem 2.1.7).

Problem 7.1.1 (Hypergraph Small Set Expansion)**.** Given a hypergraph $H = (V, E, w)$ and a parameter $\delta \in (0, 1/2]$, the Hypergraph Small Set Expansion problem (H-SSE) is to find a set $S \subset V$ of size at most δn that minimizes $\phi(S)$. The value of the optimal solution to H-SSE is called the small set expansion of H. That is, for $\delta \in (0, 1/2]$, the small set expansion $\phi_{H,\delta}$ of a hypergraph $H = (V, E, w)$ is defined as

$$\phi_{H,\delta} = \min_{\substack{S \subset V \\ 0 < |S| \leqslant \delta n}} \phi(S).$$

Note that for $\delta = 1/2$, the Hypergraph Small Set Expansion problem is the Hypergraph Expansion problem.

7.1.1 Summary of Results

Raghavendra, Steurer and Tetali [87] gave an algorithm for Small Set Expansion in graphs that finds a set of size $\mathcal{O}(\delta n)$ with expansion $\mathcal{O}(\sqrt{\mathsf{OPT} \log(1/\delta)})$ (where

OPT is the expansion of the optimal solution). Later Bansal et. al.[16] gave a $\mathcal{O}(\sqrt{\log n \log(1/\delta)})$ approximation algorithm for the problem. We present analogs of the results of Bansal et. al.[16] and Raghavendra et. al.[87] for hypergraphs.

Theorem 7.1.2. *There is a randomized polynomial-time approximation algorithm for the* Hypergraph Small Set Expansion *problem that given a hypergraph* $H = (V, E, w)$*, and parameters* $\varepsilon \in (0,1)$ *and* $\delta \in (0, 1/2)$*, finds a set* $S \subset V$ *of size at most* $(1+\varepsilon)\delta n$ *such that*

$$\phi(S) \leqslant \mathcal{O}_{\varepsilon}\left(\delta^{-1} \log \delta^{-1} \log\log \delta^{-1} \cdot \sqrt{\log n} \cdot \phi_{H,\delta}\right) = \tilde{\mathcal{O}}_{\varepsilon}\left(\delta^{-1}\sqrt{\log n}\, \phi_{H,\delta}\right),$$

(where the constant in the $\mathcal{O}$ *notation depends polynomially on* $1/\varepsilon$*). That is, the algorithm gives* $\mathcal{O}(\sqrt{\log n})$ *approximation when* δ *and* ε *are fixed.*

We state our second result, Theorem 7.1.3, for r-uniform hypergraphs. We present and prove a more general Theorem 7.5.3 that applies to any hypergraph in Section 7.5.

Theorem 7.1.3 (Informal Statement)**.** *There is a randomized polynomial-time algorithm that given an* r*–uniform hypergraph* $H = (V, E, w)$ *with vertex weights* $w(v) = d_v$*, and parameters* $\varepsilon \in (0,1)$ *and* $\delta \in (0, 1/2)$ *finds a set* $S \subset V$ *of size at most* $(1+\varepsilon)\delta n$ *such that*

$$\phi(S) \leqslant \tilde{\mathcal{O}}_{\varepsilon}\left(\delta^{-1}\left(\sqrt{\frac{\log r}{r}\phi_{H,\delta}} + \phi_{H,\delta}\right)\right).$$

Our algorithms for H-SSE are bi-criteria approximation algorithms in that they output a set S of size at most $(1+\varepsilon)\delta n$. We note that this is similar to the algorithm of Bansal et. al.[16] for SSE, which also finds a set of size at most $(1+\varepsilon)\delta n$ rather than a set of size at most δn. The algorithm of [87] finds a set of size $\mathcal{O}(\delta n)$. The approximation factor of our first algorithm does not depend on the size of hyperedges in the input hypergraph. It has the same dependence on n as the algorithm of Bansal et. al.[16] for SSE. However, the dependence on $1/\delta$ is quasi-linear; whereas it is logarithmic in the algorithm of Bansal et. al.[16]. In fact, we show that the integrality gap of the standard SDP relaxation for H-SSE is at least linear in $1/\delta$ (Theorem 7.6.1). The approximation guarantee of our second algorithm is analogous to that of the algorithm of [87].

Small Set Vertex Expansion. Our techniques can also be used to obtain an approximation algorithm for Small Set Vertex Expansion (SSVE) in graphs (Problem 2.1.7).

Problem 7.1.4 (Small Set Vertex Expansion)**.** Given graph $G = (V, E)$ and a parameter $\delta \in (0, 1/2]$, the Small Set Vertex Expansion (SSVE) is to find a set $S \subset V$ of size at most δn that minimizes $\phi^{\mathsf{V}}(S)$. The value of the optimal solution to SSVE is called the small set vertex expansion of G and is denoted by $\phi^{\mathsf{V}}_{G,\delta}$. That is, for $\delta \in (0, 1/2]$, the small set expansion $\phi^{\mathsf{V}}_{G,\delta}$ of a graph $G = (V, E)$ is defined as

$$\phi^{\mathsf{V}}_{G,\delta} = \min_{\substack{S \subset V \\ 0 < |S| \leqslant \delta n}} \phi^{\mathsf{V}}(S).$$

The Small Set Vertex Expansion recently gained interest due to its connection to obtaining subexponential time, constant factor approximation algorithms for many combinatorial problems like Sparsest Cut and Graph Coloring ([9, 74]). Using a reduction from Vertex Expansion in graphs to Hypergraph Expansion (Theorem 7.1.5, similar to Theorem 4.2.19), we can get an approximation algorithm for SSVE having the same approximation guarantee as that for H-SSE.

Theorem 7.1.5 (Extension of Theorem 4.2.19)**.** *There exist absolute constants $c_1, c_2 \in \mathbb{R}^+$ such that for every graph $G = (V, E)$, of maximum degree d, there exists a polynomial time computable hypergraph $H = (V', E')$ having the hyperedges of cardinality at most $d + 1$ such that*

$$c_1 \phi_{H,\delta} \leqslant \phi^{\mathsf{V}}_{G,\delta} \leqslant c_2 \phi_{H,\delta}.$$

Also, $\eta^H_{max} \leqslant \log_2(d_{max} + 1)$, where d_{max} is the maximum degree of G (where η^H_{max} is defined in Definition 7.5.1).

From this theorem, Theorem 7.1.2 and Theorem 7.5.3 we immediately get algorithms for SSVE.

Theorem 7.1.6 (Corollary to Theorem 7.1.2 and Theorem 7.1.5)**.** *There is a randomized polynomial-time approximation algorithm for the* Small Set Vertex Expansion *that given a graph $G = (V, E)$, and parameters $\varepsilon \in (0, 1)$ and $\delta \in (0, 1/2))$ finds a set $S \subset V$ of size at most $(1 + \varepsilon)\delta n$ such that*

$$\phi^{\mathsf{V}}(S) \leqslant \mathcal{O}_\varepsilon \left(\sqrt{\log n}\, \delta^{-1} \log \delta^{-1} \log\log \delta^{-1} \cdot \phi^{\mathsf{V}}_{G,\delta} \right) .$$

Theorem 7.1.7 (Corollary to Theorem 7.5.3 and Theorem 7.1.5)**.** *There is a randomized polynomial-time algorithm for the* Small Set Vertex Expansion *that given a graph $G = (V, E)$ of maximum degree d, parameters $\varepsilon \in (0, 1)$ and $\delta \in (0, 1/2)$ finds a set $S \subset V$ of size at most $(1 + \varepsilon)\delta n$ such that*

$$\begin{aligned} \phi^{\mathsf{V}}(S) &\leqslant \mathcal{O}_\varepsilon \left(\sqrt{\phi^{\mathsf{V}}_{G,\delta} \log d} \cdot \delta^{-1} \log \delta^{-1} \log\log \delta^{-1} \right) \\ &= \tilde{\mathcal{O}}_\varepsilon \left(\delta^{-1} \sqrt{\phi^{\mathsf{V}}_{G,\delta} \log d} \right) . \end{aligned}$$

We note that the Small Set Vertex Expansion for $\delta = 1/2$ is just the Vertex Expansion. In that case, Theorem 7.1.7 gives the same approximation guarantee as the algorithm of Theorem 6.1.3.

7.1.2 Proof Overview

Our general approach to solving H-SSE is similar to the approach of Bansal et. al.[16]. We recall how the algorithm of Bansal et. al.[16] for (graph) SSEworks. The algorithm solves a semidefinite programming relaxation for SSE and gets an SDP solution. The SDP solution assigns a vector $\bar{u}$ to each vertex u. Then the algorithm generates *an orthogonal separator*. An orthogonal separator S (introduced by [31]) with distortion D is a distribution over random subset of vertices such that

(a) If $\bar{u}$ and $\bar{v}$ are close to each other then the probability that u and v are separated by S is small; namely, it is at most $\alpha D\|\bar{u}-\bar{v}\|^2$, where α is a normalization factor such that $\mathbb{P}[u \in S] = \alpha\|\bar{u}\|^2$.

(b) If the angle between $\bar{u}$ and $\bar{v}$ is larger than a certain threshold, then the probability that both u and v are in S is much smaller than the probability that one of them is in S.

Bansal et. al.[16] showed that condition (b) together with SDP constraints implies that S is of size at most $(1+\varepsilon)\delta n$ with sufficiently high probability. Then condition (a) implies that the expected number of cut edges is at most D times the SDP value. That means that S is a D–approximate solution to SSE.

We start with an SDP relaxation of the Rayleigh quotient of the hypergraph

$$\mathcal{R}(X) = \frac{X^T L(X)}{X^T X} = \frac{\sum_{e \in E} w(e) \max_{i,j \in e}(X_i - X_j)^2}{d \sum_i X_i^2}$$

together with the "small-set" constraints of Bansal et. al.[16]. If we run this algorithm on an instance of H-SSE, we will still find a set of size at most $(1+\varepsilon)\delta n$, but the cost of the solution might be very high. Indeed, consider a hyperedge e. Even though every two vertices u and v in e are unlikely to be separated by S, at least one pair out of $\binom{|e|}{2}$ pairs of vertices is quite likely to be separated by S; hence, e is quite likely to be cut by S. To deal with this problem, we develop *hypergraph orthogonal separators*. In the definition of a hypergraph orthogonal separator, we strengthen condition (a) by requiring that a hyperedge e is cut by S with small probability if all vertices in e are close to each other. Specifically, we require that

$$\mathbb{P}[e \text{ is cut by } S] \leqslant \alpha D \max_{u,v \in e} \|\bar{u} - \bar{v}\|^2. \tag{60}$$

We show that there is a hypergraph orthogonal separator with distortion proportional to $\sqrt{\log n}$ (the distortion also depends on parameters of the orthogonal separator). Plugging this hypergraph orthogonal separator in the algortihm of Bansal et. al.[16], we get Theorem 7.1.2. We also develop another variant of hypergraph orthogonal separators, ℓ_2–ℓ_2^2 orthogonal separators. An ℓ_2–ℓ_2^2 orthogonal separator with ℓ_2–distortion $D_{\ell_2}(r)$ and ℓ_2^2–distortion $D_{\ell_2^2}$ satisfies the following condition[1]

$$\mathbb{P}\left[e \text{ is cut by } S\right] \leqslant \alpha D_{\ell_2}(|e|) \cdot \min_{w\in E} \|\bar{w}\| \cdot \max_{u,v\in e} \|\bar{u}-\bar{v}\| + \alpha D_{\ell_2^2} \cdot \max_{u,v\in e} \|\bar{u}-\bar{v}\|^2 . \quad (61)$$

We show that there is an ℓ_2-ℓ_2^2 hypergraph orthogonal separator whose ℓ_2 and ℓ_2^2 distortions do not depend on n (in contrast, there is no hypergraph orthogonal separator whose distortion does not depend on n). This result yields Theorem 7.1.3.

We now give a brief conceptual overview of our construction of hypergraph orthogonal separators. We use the framework developed in [31] for (graph) orthogonal separators. For simplicity, we ignore vector normalization steps in this overview; let us assume that all the vectors are unit vectors. (Note, however, that these normalization steps are crucial). We first design a procedure that partitions the hypergraph into two pieces (the procedure labels every vertex with either 0 or 1). In a sense, each set S in the partition is a "very weak" hypergraph orthogonal separator. It satisfies property (60) with $D_0 \sim \sqrt{\log n} \log\log(1/\delta)$ and $\alpha_0 = 1/2$ and a weak variant of property (b): if the angle between vectors $\bar{u}$ and $\bar{v}$ is larger than the threshold then events $u \in S$ and $v \in S$ are "almost" independent. We repeat the procedure $l = \log_2(1/\delta) + \mathcal{O}(1)$ times and obtain a partition of graph into $2^l = \mathcal{O}(1/\delta)$ pieces. Then we randomly choose one set S among them; this set S is our hypergraph orthogonal separator. Note that by running the procedure many times we decrease exponentially in l the probability that two vertices, as in condition (b), belong to S. So condition (b) holds for S. Also, we affect the distortion in (60) in two ways. First, the probability that the edge is cut increases by a factor of l. That is, we get $\mathbb{P}\left[e \text{ is cut by } S\right] \leqslant l \times \alpha_0 D_0 \max_{u,v\in e} \|\bar{u}-\bar{v}\|^2$. Second, the probability that we

[1] It may look strange that we have two terms in the bound. One may expect that we can either have only term $D_{\ell_2^2} \max_{u,v\in e} \|\bar{u}-\bar{v}\|^2$ (as in the previous definition) or only term $D_{\ell_2}(|e|) \cdot \min_{w\in E} \|\bar{w}\| \cdot \max_{u,v\in e} \|\bar{u}-\bar{v}\|$. However, the latter is not possible — there is no ℓ_2–ℓ_2^2 separator with $D_{\ell_2^2} = 0$.

choose a vertex u goes down from $\|\bar{u}\|^2/2$ to $\Omega(\delta)\|\bar{u}\|^2$ since, roughly speaking, we choose one set S among $\mathcal{O}(1/\delta)$ possible sets. That is, the parameter α of S is $\Omega(\delta)$. Therefore, $\mathbb{P}[e \text{ is cut by } S] \leqslant \alpha(\alpha_0 l D_0/\alpha) \max_{u,v\in e} \|\bar{u}-\bar{v}\|^2$. That is, we get a hypergraph orthogonal separator with distortion $(\alpha_0 l D_0/\alpha) \sim \tilde{\mathcal{O}}(\delta^{-1}\sqrt{\log n})$. The construction of ℓ_2^2 orthogonal separators is similar but a bit more technical.

Organization. We present our SDP relaxation and introduce our main technique, hypergraph orthogonal separators, in Section 7.2. We describe our first algorithm for H-SSE in Section 7.2.3, and then describe an algorithm that generates hypergraph orthogonal separators in Section 7.3. We define ℓ_2–ℓ_2^2 hypergraph orthogonal separators, give an algorithm that generates them, and then present our second algorithm for H-SSE in Section 7.4 and Section 7.5. Finally, we show a simple SDP integrality gap for H-SSE in Section 7.6. This integrality gap also gives a lower bound on the quality of m-orthogonal separators. We give a proof of Theorem 7.1.5 in Section 7.7.

7.2 Algorithm for Hypergraph Small Set Expansion

7.2.1 SDP Relaxation for Hypergraph Small Set Expansion

We use the SDP relaxation for H-SSE shown in SDP 7.2.1. There is an SDP variable $\bar{u}$ for every vertex $u \in V$. Every combinatorial solution S (with $|S| \leqslant \delta n$) defines the corresponding (intended) SDP solution:

$$\bar{u} = \begin{cases} \frac{e}{\sqrt{w(S)}} & \text{if } u \in S \\ 0 & \text{otherwise} \end{cases}.$$

where e is a fixed unit vector. It is easy to see that this solution satisfies all SDP constraints. Note that $\max_{u,v\in e} \|\bar{u}-\bar{v}\|^2$ is equal to $1/w(S)$, if e is cut, and to 0, otherwise. Therefore, the objective function equals

$$\sum_{e\in E} w(e) \max_{u,v\in e} \|\bar{u}-\bar{v}\|^2 = \sum_{e\in E(S,\bar{S})} w(e)\frac{1}{w(S)} = \frac{|E(S,\bar{S})|}{w(S)} = \phi(S).$$

Thus our SDP for H-SSE is indeed a relaxation.

SDP 7.2.1.

$$\mathsf{SDPval} \stackrel{\text{def}}{=} \min \sum_{e\in E} w(e) \max_{u,v\in e} \|\bar{u}-\bar{v}\|^2$$

subject to

$$\begin{aligned}
\sum_{v\in V} \langle \bar{u},\bar{v}\rangle &\leqslant \delta n \cdot \|\bar{u}\|^2 \qquad \forall u \in V && (62)\\
\sum_{u\in V} w(u)\,\|\bar{u}\|^2 &= 1 && (63)\\
\|\bar{u}-\bar{v}\|^2 + \|\bar{v}-\bar{w}\|^2 &\geqslant \|\bar{u}-\bar{w}\|^2 \qquad \text{for every } u,v,w \in V && (64)\\
0 \leqslant \langle \bar{u},\bar{v}\rangle &\leqslant \|\bar{u}\|^2 \qquad \text{for every } u,v \in V. && (65)
\end{aligned}$$

Figure 20: SDP relaxation for H-SSE

7.2.2 Hypergraph Orthogonal Separators

The main technical tool for proving Theorem 7.1.2 is *hypergraph orthogonal separators*. In this chapter, we extend the technique of orthogonal separators to hypergraphs thereby introducing hypergraph orthogonal separators. We then use hypergraph orthogonal separators to solve H-SSE. In Section 7.4, we introduce another version of hypergraph orthogonal separators, namely the ℓ_2–ℓ_2^2 hypergraph orthogonal separators, and then use them to prove Theorem 7.1.3 and Theorem 7.5.3.

Definition 7.2.2 (Hypergraph Orthogonal Separators)**.** Let $\{\bar{u} : u \in V\}$ be a set of vectors in the unit ball that satisfy ℓ_2^2–triangle inequalities (64) and (65). We say that a random set $S \subset V$ is a *hypergraph m-orthogonal separator* with distortion D, probability scale $\alpha > 0$, and separation threshold $\beta \in (0,1)$ if it satisfies the following properties.

1. For every $u \in V$,

$$\mathbb{P}[u \in S] = \alpha\|\bar{u}\|^2.$$

2. For every u and v such that $\|\bar{u}-\bar{v}\|^2 \geqslant \beta \min\{\|\bar{u}\|^2, \|\bar{v}\|^2\}$

$$\mathbb{P}\left[u \in S \text{ and } v \in S\right] \leqslant \alpha \frac{\min\{\|\bar{u}\|^2, \|\bar{v}\|^2\}}{m}.$$

3. For every $e \subset V$,

$$\mathbb{P}\left[e \text{ is cut by } S\right] \leqslant \alpha D \max_{u,v \in e} \|\bar{u}-\bar{v}\|^2.$$

The definition of a hypergraph m-orthogonal separator is similar to that of a (graph) m-orthogonal separator: a random set S is an m-orthogonal separator if it satisfies properties 1, 2, and property 3′, which is property 3 restricted to edges e of size 2.

3′. For every $\{u, v\}$,

$$\mathbb{P}\left[e \text{ is cut by } S\right] \leqslant \alpha D \|\bar{u}-\bar{v}\|^2.$$

We design an algorithm that generates a hypergraph m-orthogonal separator with distortion $\mathcal{O}_\beta(\sqrt{\log n} \cdot m \log m \log\log m)$. We note that the distortion of *any* hypergraph orthogonal separator must depend on m at least linearly (see Section 7.6). We remark that there are two constructions of (graph) orthogonal separators, "orthogonal separators via ℓ_1" and "orthogonal separators via ℓ_2", with distortions, $\mathcal{O}_\beta(\sqrt{\log n} \log m)$ and $\mathcal{O}_\beta(\sqrt{\log n \log m})$, respectively (presented in [31]). Our construction of hypergraph orthogonal separators uses the framework of orthogonal separators via ℓ_1. We prove the following theorem in Section 7.3.

Theorem 7.2.3. *There is a polynomial-time randomized algorithm that given a set of vertices V, a set of vectors $\{\bar{u}\}$ satisfying ℓ_2^2–triangle inequalities* (64) *and* (65)*, parameters $m \geqslant 2$ and $\beta \in (0,1)$, generates a hypergraph m-orthogonal separator with probability scale $\alpha \geqslant 1/n$ and distortion $D = \mathcal{O}\left(\beta^{-1} m \log m \log\log m \times \sqrt{\log n}\right)$.*

7.2.3 Rounding Algorithm

In this section, we present our algorithm for Hypergraph Small Set Expansion. Our algorithm uses hypergraph orthogonal separators that we describe in Section 7.3. We use the approach of Bansal et. al.[16]. Suppose that we are given a polynomial-time algorithm that generates hypergraph m-orthogonal separators with distortion $D(m, \beta)$ (with probability scale $\alpha > 1/\mathsf{poly}(n)$). We show how to get a $D^* \stackrel{\text{def}}{=} 4D(4/(\varepsilon\delta), \varepsilon/4)$ approximation for H-SSE.

Theorem 7.2.4. *There is a randomized polynomial-time approximation algorithm for the* Hypergraph Small Set Expansion *that given a hypergraph $H = (V, E)$, and parameters $\varepsilon \in (0, 1)$ and $\delta \in (0, 1/2)$ finds a set $S \subset V$ of size at most $(1 + \varepsilon)\delta n$ such that*

$$\phi(S) \leqslant 4D(4/(\varepsilon\delta), \varepsilon/4) \cdot \phi_{H,\delta}.$$

Proof. We solve the SDP relaxation for H-SSE and obtain an SDP solution $\{\bar{u}\}$. Denote the SDP value by SDPval. Consider a hypergraph orthogonal separator S with $m = 4/(\varepsilon\delta)$ and $\beta = \varepsilon/4$. Define a set S':

$$S' = \begin{cases} S & \text{if } |S| \leqslant (1+\varepsilon)\delta n \\ \emptyset & \text{otherwise} \end{cases} .$$

Clearly, $|S'| \leqslant (1 + \varepsilon)\delta n$. Bansal et. al.[16] showed that

$$\mathbb{P}\left[u \in S'\right] \in \left[\frac{\alpha}{2}\, \|\bar{u}\|^2, \alpha\|\bar{u}\|^2\right] \qquad \text{for every } u \in V\,.$$

Note that

$$\mathbb{P}\left[S' \text{ cuts edge } e\right] \leqslant \mathbb{P}\left[S \text{ cuts edge } e\right] \leqslant \alpha D^* \max_{u,v\in e} \|\bar{u} - \bar{v}\|^2.$$

where $D^* = D(4/(\varepsilon\delta), \varepsilon/4)$ for the sake of brevity. Let

$$Z \stackrel{\text{def}}{=} w(S') - \frac{\sum_{e\in E(S',\bar{S}')} w(e)}{4D^* \cdot \mathsf{SDPval}}\,.$$

We have,

$$\begin{aligned}
\mathbb{E}[Z] &= \mathbb{E}[w(S')] - \frac{\mathbb{E}\left[\sum_{e\in E(S',\bar{S}')} w(e)\right]}{4D^* \cdot \mathsf{SDPval}} \\
&\geqslant \sum_{u\in V}\left(\frac{\alpha}{2}\cdot\|\bar{u}\|^2\right) w(u) - \frac{\sum_{e\in E}\left(\alpha D^* \max_{u,v\in e}\|\bar{u}-\bar{v}\|^2\right) w(e)}{4D^*\cdot \mathsf{SDPval}} \\
&= \frac{\alpha}{2} - \frac{1}{4D^*\cdot\mathsf{SDPval}} \times \alpha D^* \mathsf{SDPval} = \frac{\alpha}{4}.
\end{aligned}$$

Since $Z \leqslant w(S') < n$ (always), by Markov's inequality, we have $\mathbb{P}[Z > 0] \geqslant \alpha/(4n)$ and hence

$$\mathbb{P}[\phi(S) < 4D^* \cdot \mathsf{SDPval}] \geqslant \alpha/(4n)\,.$$

We sample S independently $4n/\alpha$ times and return the first set S' such that $\phi(S) < 4D^* \cdot \mathsf{SDPval}$. This gives a set S' such that $|S'| \leqslant (1+\varepsilon)\delta n$, and $\phi(S') \leqslant 4D^*\phi_{H,\delta}$. The algorithm succeeds (finds such a set S') with a constant probability. By repeating the algorithm n times, we can make the success probability exponentially close to 1. □

In Section 7.3, we describe how to generate an m-hypergraph orthogonal separator with distortion $D = \mathcal{O}\left(\sqrt{\log n} \times \beta^{-1} m \log m \log\log m\right)$. That gives us an algorithm for H-SSE with approximation factor $\mathcal{O}_\varepsilon\left(\delta^{-1}\log\delta^{-1}\log\log\delta^{-1} \times \sqrt{\log n}\right)$.

7.3 Generating Hypergraph Orthogonal Separators

In this section, we present an algorithm that generates a hypergraph m-orthogonal separator. At the high level, the algorithm is similar to the algorithm for generating orthogonal separators by Chlamtac et. al.[31]. We use a different procedure for generating words $W(u)$ (see below) and set parameters differently; also the analysis of our algorithm is different.

In our algorithm, we use a "normalization" map ψ from [31]. Map ψ maps a set $\{\bar{u}\}$ of vectors satisfying ℓ_2^2–triangle inequalities (64) and (65) to $\mathbb{R}^n$. It has the following properties.

1. For all vertices u, v, w,

$$\|\psi(\bar{u}) - \psi(\bar{v})\|_2^2 + \|\psi(\bar{v}) - \psi(\bar{w})\|_2^2 \geqslant \|\psi(\bar{u}) - \psi(\bar{w})\|_2^2 .$$

2. For all vertices u and v,

$$\langle \psi(\bar{u}), \psi(\bar{v}) \rangle = \frac{\langle \bar{u}, \bar{v} \rangle}{\max\{\|\bar{u}\|^2, \|\bar{v}\|^2\}} .$$

   In particular, for every $\bar{u} \neq 0$, $\|\psi(\bar{u})\|_2^2 = \langle \psi(\bar{u}), \psi(\bar{u}) \rangle = 1$. Also, $\psi(0) = 0$.

3. For all non-zero vectors $\bar{u}$ and $\bar{v}$,

$$\|\psi(\bar{u}) - \psi(\bar{v})\|_2^2 \leqslant \frac{2\,\|\bar{u} - \bar{v}\|^2}{\max\{\|\bar{u}\|^2, \|\bar{v}\|^2\}} .$$

We also use the following theorem of Arora, Lee and Naor [11] (See also [12]).

Theorem 7.3.1 ([11]). *There exist constants $C \geqslant 1$ and $p \in (0, 1/4)$ such that for every n unit vectors x_u ($u \in V$), satisfying ℓ_2^2–triangle inequalities* (64), *and every $\Delta > 0$, the following holds. There exists a polynomial time algorithm to sample a random subset U of V such that for every $u, v \in V$ with $\|x_u - x_v\|^2 \geqslant \Delta$,*

$$\mathbb{P}\left[u \in U \text{ and } d(v, U) \geqslant \frac{\Delta}{C\sqrt{\log n}}\right] \geqslant p,$$

where $d(v, U) = \min_{u \in U} \|x_u - x_v\|^2$.

First we describe an algorithm that randomly assigns each vertex u a symbol, either 0 or 1. Then we use this algorithm to generate an orthogonal separator.

Lemma 7.3.2. *There is a randomized polynomial-time algorithm that given a finite set V, unit vectors $\psi(\bar{u})$ for $u \in V$ satisfying ℓ_2^2-triangle inequalities and a parameter $\beta \in (0, 1)$, returns a random assignment $\omega : V \to \{0, 1\}$ that satisfies the following properties.*

- *For every u and v such that $\|\psi(\bar{u}) - \psi(\bar{v})\|^2 \geqslant \beta$,*

$$\mathbb{P}\left[\omega(u) \neq \omega(v)\right] \geqslant 2p,$$

  where $p > 0$ is the constant from Theorem 7.3.1.

- *For every set $e \subset V$ of size at least 2,*

$$\mathbb{P}\left[\omega(u) \neq \omega(v) \text{ for some } u, v \in e\right] \leqslant \mathcal{O}\left(\beta^{-1}\sqrt{\log n} \max_{u,v \in e} \|\psi(\bar{u}) - \psi(\bar{v})\|^2\right).$$

Proof. Let U be the random set from Theorem 7.3.1 for vectors $x_u = \psi(\bar{u})$ and $\Delta = \beta$. Choose $t \sim (0, \beta/(C\sqrt{\log n}))$ uniformly at random. Let

$$\omega(u) = \begin{cases} 0 & \text{if } d(u, U_i) \leqslant t \\ 1 & \text{otherwise} \end{cases}.$$

Consider first vertices u and v such that $\|\psi(\bar{u}) - \psi(\bar{v})\|^2 \geqslant \beta$. By Theorem 7.3.1,

$$\mathbb{P}\left[u \in U \text{ and } d(v, U) \geqslant \frac{\Delta}{C\sqrt{\log n}}\right] \geqslant p$$

and

$$\mathbb{P}\left[v \in U \text{ and } d(u, U) \geqslant \frac{\Delta}{C\sqrt{\log n}}\right] \geqslant p\,.$$

Note that in the former case, when $u \in U$ and $d(v, U) \geqslant \frac{\Delta}{C\sqrt{\log n}}$, we have $\omega(u) = 0$ and $\omega(v) = 1$; in the latter case, when $v \in U$ and $d(u, U) \geqslant \frac{\Delta}{C\sqrt{\log n}}$, we have $\omega(v) = 0$ and $\omega(u) = 1$. Therefore, the probability that $\omega(u) \neq \omega(v)$ is at least $2p$.

Now consider a set $e \subset V$ of size at least 2. Let

$$\tau_m = \min_{w \in e} d(U, \psi(\bar{w})) \quad \text{and} \quad \tau_M = \max_{w \in e} d(U, \psi(\bar{w})).$$

We have, $\tau_M - \tau_m \leqslant \max_{u,v \in e} \|\psi(\bar{u}) - \psi(\bar{v})\|^2$. Note that if $t < \tau_m$ then $\omega(u) = 1$ for all $u \in e$; if $t \geqslant \tau_M$ then $\omega(u) = 0$ for all $u \in e$. Thus $\omega(u) \neq \omega(v)$ for some $u, v \in e$ only if $t \in [\tau_m, \tau_M)$. Since the probability density of the random variable t is at most

$C\sqrt{\log n}$, we get,

$$\mathbb{P}\left[\exists\, u, v \in e : \omega(u) \neq \omega(v)\right] \leqslant \mathbb{P}\left[t \in [\tau_m, \tau_M)\right]$$
$$\leqslant \frac{C\sqrt{\log n}}{\Delta}\,(\tau_M - \tau_m) \leqslant \frac{C\sqrt{\log n}}{\beta} \max_{u,v\in e} \|\bar{u} - \bar{v}\|^2. \quad (66)$$

□

We now amplify the result of Lemma 7.3.2.

Lemma 7.3.3. *There is a randomized polynomial time algorithm that given V, vectors $\psi(\bar{u})$ and $\beta \in (0,1)$ as in Lemma 7.3.2, and a parameter $m \geqslant 2$, returns a random assignment $\omega : V \to \{0,1\}$ such that:*

- *For every u and v such that $\|\psi(\bar{u}) - \psi(\bar{v})\|^2 \geqslant \beta$,*

$$\mathbb{P}\left[\tilde{\omega}(u) \neq \tilde{\omega}(v)\right] \geqslant \frac{1}{2} - \frac{1}{\log_2 m}.$$

- *For every set $e \subset V$ of size at least 2,*

$$\mathbb{P}\left[\tilde{\omega}(u) \neq \tilde{\omega}(v) \text{ for some } u, v \in e\right]$$
$$\leqslant \mathcal{O}\left(\beta^{-1}\sqrt{\log n} \cdot \log\log m \cdot \max_{u,v\in e} \|\psi(\bar{u}) - \psi(\bar{v})\|^2\right).$$

Proof. Let $K = \max\left(\left\lceil \frac{\log_2 \log_2 m}{-\log_2(1-4p)} \right\rceil, 1\right)$. We independently sample K assignments $\omega_1, \ldots, \omega_K$. Let

$$\tilde{\omega}(u) = \omega_1(u) \oplus \cdots \oplus \omega_K(u),$$

where $\oplus$ denotes addition modulo 2. Consider u and v such that $\|\psi(\bar{u}) - \psi(\bar{v})\|^2 \geqslant \beta$. Let

$$\tilde{p} = \mathbb{P}\left[\omega_i(u) \neq \omega_i(v)\right] \geqslant 2p \quad \text{for} \quad i \in \{1, \ldots, K\}$$

(the expression does not depend on the value of i since all ω_i are identically distributed). Note that $\tilde{\omega}(u) \neq \tilde{\omega}(v)$ if and only if $\omega_i(u) \neq \omega_i(v)$ for an odd number of values i.

Therefore,

$$\mathbb{P}[\omega(u) \neq \omega(v)] = \sum_{0 \leqslant k \leqslant K/2} \binom{K}{2k+1} \tilde{p}^{2k+1}(1-\tilde{p})^{K-2k-1} = \frac{1-(1-2\tilde{p})^K}{2}$$
$$\geqslant \frac{1-(1-4p)^K}{2} \geqslant \frac{1}{2} - \frac{1}{\log_2 m}.$$

Now let $e \subset V$ be a subset of size at least 2. We have,

$$\mathbb{P}[\tilde{\omega}(u) \neq \tilde{\omega}(v)] \leqslant \mathbb{P}[\omega_i(u) \neq \omega_i(v) \text{ for some } i]$$
$$\leqslant \mathcal{O}\left(K\beta^{-1}\sqrt{\log n} \max_{u,v \in e} \|\psi(\bar{u}) - \psi(\bar{v})\|^2\right).$$

□

We are now ready to present our algorithm for the hypergraph orthogonal separator (Algorithm 7.3.4).

Algorithm 7.3.4 (Hypergraph Orthgonal Separator)**.**

1. Set $l = \lceil \log_2 m/(1 - \log_2(1 + 2/\log_2 m)) \rceil = \log_2 m + \mathcal{O}(1)$.
2. Sample l independent assignments $\tilde{\omega}_1, \ldots, \tilde{\omega}_l$ using Lemma 7.3.3.
3. For every vertex u, define word $W(u) = \tilde{\omega}_1(u) \ldots \tilde{\omega}_l(u) \in \{0,1\}^l$.
4. If $n \geqslant 2^l$, pick a word $W \in \{0,1\}^l$ uniformly at random. If $n < 2^l$, pick a random word $W \in \{0,1\}^l$ so that $\mathbb{P}_W[W = W(u)] = 1/n$ for every $u \in V$. This is possible since the number of distinct words constructed in step 3 is at most n (we may pick a word W not equal to any $W(u)$).
5. Pick $r \sim (0,1)$ uniformly at random.
6. Let $S = \{u \in V : \|\bar{u}\|^2 \geqslant r \text{ and } W(u) = W\}$.

Figure 21: Hypergraph Orthogonal Separator

Theorem 7.3.5. *Random set S output by Algorithm 7.3.4 is a hypergraph orthogonal separator with distortion*

$$D = \mathcal{O}\left(\sqrt{\log n} \times \frac{m \log m \log\log m}{\beta}\right),$$

probability scale $\alpha \geqslant 1/n$ and separation threshold β.

Proof. We verify that S satisfies properties 1–3 in the definition of a hypergraph m-orthogonal separator with $\alpha = \max\left\{1/2^l, 1/n\right\}$.

Property 1. We compute the probability that $u \in S$. Observe that $u \in S$ if and only if $W(u) = W$ and $r \leqslant \|\bar{u}\|^2$ (these two events are independent). If $n \geqslant 2^l$, the probability that $W = W(u)$ is $1/2^l$ since we choose W uniformly at random from $\{0,1\}^l$; if $n < 2^l$ the probability is $1/n$. That is,

$$\mathbb{P}\left[W = W(u)\right] = \max\left\{1/2^l, 1/n\right\} = \alpha$$

and

$$\mathbb{P}\left[r \leqslant \|\bar{u}\|^2\right] = \|\bar{u}\|^2 .$$

We conclude that property 1 holds.

Property 2. Consider two vertices u and v such that $\|\bar{u} - \bar{v}\|^2 \geqslant \beta \min\left\{\|\bar{u}\|^2, \|\bar{v}\|^2\right\}$. Assume without loss of generality that $\|\bar{u}\|^2 \leqslant \|\bar{v}\|^2$. Note that $u, v \in S$ if and only if $r \leqslant \|\bar{u}\|^2$ and $W = W(u) = W(v)$. We first upper bound the probability that $W(u) = W(v)$. We have,

$$2\left\langle \bar{u}, \bar{v}\right\rangle = \|\bar{u}\|^2 + \|\bar{v}\|^2 - \|\bar{u} - \bar{v}\|^2 \leqslant (1-\beta)\|\bar{u}\|^2 + \|\bar{v}\|^2 \leqslant (2-\beta)\|\bar{v}\|^2 .$$

Therefore, $2\left\langle \bar{u}, \bar{v}\right\rangle / \|\bar{v}\|^2 \leqslant 2 - \beta$. Hence,

$$\|\psi(\bar{u}) - \psi(\bar{v})\|^2 = 2 - 2\left\langle \psi(\bar{u}), \psi(\bar{v})\right\rangle = 2 - \frac{2\left\langle \bar{u}, \bar{v}\right\rangle}{\max\left\{\|\bar{u}\|^2, \|\bar{v}\|^2\right\}} \geqslant \beta = \Delta .$$

From Lemma 7.3.3 we get that

$$\mathbb{P}\left[\tilde{\omega}_i(u) \neq \tilde{\omega}_i(v)\right] \geqslant \frac{1}{2} - \frac{1}{\log_2 m} \qquad \text{for every } i .$$

The probability that $W(u) = W(v)$ is at most $(\frac{1}{2} + \frac{1}{\log_2 m})^l \leqslant 1/m$. We have,

$$\begin{aligned}\mathbb{P}\left[u \in S, v \in S\right] = \mathbb{P}\left[r \leqslant \min\left\{\|\bar{u}\|^2, \|\bar{v}\|^2\right\}\right] \times \quad & \mathbb{P}\left[W(u) = W(v)\right] \\ \times\, & \mathbb{P}\left[W = W(u) = W(v) \mid W(u) = W(v)\right] \\ \leqslant \min\left\{\|\bar{u}\|^2, \|\bar{v}\|^2\right\} \times \alpha \times \frac{1}{m}, &\end{aligned}$$

as required.

Property 3. Let e be an arbitrary subset of V, $|e| \geqslant 2$. Let

$$\rho_m = \min_{w \in e} \|\bar{w}\|^2 \qquad \text{and} \qquad \rho_M = \max_{w \in e} \|\bar{w}\|^2.$$

Note that

$$\rho_M - \rho_m = \|\bar{w}_1\|^2 - \|\bar{w}_2\|^2 \leqslant \|\bar{w}_1 - \bar{w}_2\|^2 \leqslant \max_{u,v \in e} \|\bar{u} - \bar{v}\|^2,$$

for some $w_1, w_2 \in e$. Here we used that SDP constraint (65) implies that $\|\bar{w}_1\|^2 - \|\bar{w}_2\|^2 \leqslant \|\bar{w}_1 - \bar{w}_2\|^2$.

Let $A = \{u \in e : \|\bar{u}\|^2 \geqslant r\}$. Note that $S \cap e = \{u \in A : W(u) = W\}$. Therefore, if e is cut by S then one of the following events happens.

- Event $\mathcal{E}_1$: $A \neq e$ and $S \cap e \neq \emptyset$.
- Event $\mathcal{E}_2$: $A = e$ and $A \cap S \neq \emptyset$, $A \cap S \neq A$.

If $\mathcal{E}_1$ happens then $r \in [\rho_m, \rho_M]$ since $A \neq e$ and $A \neq \emptyset$. We have,

$$\mathbb{P}\left[\mathcal{E}_1\right] \leqslant \mathbb{P}\left[r \in (\rho_m, \rho_M]\right] \leqslant |\rho_M - \rho_m| \leqslant \max_{u,v \in e} \|\bar{u} - \bar{v}\|^2.$$

If $\mathcal{E}_2$ happens then (1) $r \leqslant \rho_m$ (since $A = e$) and (2) $W(u) \neq W(v)$ for some $u, v \in e$. The probability that $r \leqslant \rho_m$ is ρ_m. We now upper bound the probability that

$W(u) \neq W(v)$ for some $u, v \in e$. For each $i \in \{1, \ldots, l\}$,

$$
\begin{aligned}
&\mathbb{P}\left[\tilde{\omega}_i(u) \neq \tilde{\omega}_i(v) \text{ for some } u, v \in e\right] \\
&\leqslant \mathcal{O}\left(\beta^{-1}\sqrt{\log n} \cdot \log\log m\right) \max_{u,v\in e} \|\psi(\bar{u}) - \psi(\bar{v})\|^2 \\
&\leqslant \mathcal{O}\left(\beta^{-1}\sqrt{\log n} \cdot \log\log m\right) \max_{u,v\in e} \frac{2\|\bar{u} - \bar{v}\|^2}{\min\left\{\|\bar{u}\|^2, \|\bar{v}\|^2\right\}} \\
&\leqslant \mathcal{O}\left(\beta^{-1}\sqrt{\log n} \cdot \log\log m\right) \times \rho_m^{-1} \times \max_{u,v\in e} \|\bar{u} - \bar{v}\|^2 .
\end{aligned}
$$

By the union bound over $i \in \{1, \ldots, l\}$, the probability that $W(u) \neq W(v)$ for some $u, v \in e$ is at most $\mathcal{O}\left(l \times \beta^{-1}\sqrt{\log n} \cdot \log\log m\right) \times \rho_m^{-1} \times \max_{u,v\in e} \|\bar{u} - \bar{v}\|^2$. Therefore,

$$
\begin{aligned}
\mathbb{P}\left[\mathcal{E}_2\right] &\leqslant \rho_m \times \mathcal{O}\left(l \times \beta^{-1}\sqrt{\log n}\, \log\log m\right) \times \rho_m^{-1} \times \max_{u,v\in e} \|\bar{u} - \bar{v}\|^2 \\
&\leqslant \mathcal{O}\left(\beta^{-1}\sqrt{\log n}\, \log m \log\log m\right) \times \max_{u,v\in e} \|\bar{u} - \bar{v}\|^2.
\end{aligned}
$$

We get that the probability that e is cut by S is at most

$$
\mathbb{P}\left[\mathcal{E}_1\right] + \mathbb{P}\left[\mathcal{E}_2\right] \leqslant \mathcal{O}\left(\beta^{-1}\sqrt{\log n}\, \log m \log\log m\right) \times \max_{u,v\in e} \|\bar{u} - \bar{v}\|^2.
$$

For $D = \mathcal{O}\left(\beta^{-1}\sqrt{\log n}\, \log m \log\log m\right)/\alpha$ we get

$$
\mathbb{P}\left[e \text{ is cut by } S\right] \leqslant \alpha D \max_{u,v\in e} \|\bar{u} - \bar{v}\|^2.
$$

Note that $\alpha \geqslant 1/2^l \geqslant \Omega(1/m)$. Thus

$$
D \leqslant \mathcal{O}\left(\beta^{-1}\sqrt{\log n}\, m \log m \log\log m\right).
$$

□

7.4 ℓ_2–ℓ_2^2 *Hypergraph Orthogonal Separators*

In this section, we present another variant of hypergraph orthogonal separators, which we call ℓ_2–ℓ_2^2 hypergraph orthogonal separators. The advantage of ℓ_2–ℓ_2^2 hypergraph orthogonal separators is that their distortions do not depend on n (the number of vertices). Then in Section 7.5, we use ℓ_2–ℓ_2^2 hypergraph orthogonal separators to prove Theorem 7.5.3 (which, in turn, implies Theorem 7.1.3).

Definition 7.4.1 (ℓ_2–ℓ_2^2 Hypergraph Orthogonal Separator)**.** Let $\{\bar{u} : u \in V\}$ be a set of vectors in the unit ball. We say that a random set $S \subset V$ is *a ℓ_2–ℓ_2^2 hypergraph m-orthogonal separator* with ℓ_2–distortion $D_{\ell_2} : \mathbb{N} \to \mathbb{R}$, ℓ_2^2–distortion $D_{\ell_2^2}$, probability scale $\alpha > 0$, and separation threshold $\beta \in (0, 1)$ if it satisfies the following properties.

1. For every $u \in V$,
$$\mathbb{P}\left[u \in S\right] = \alpha \|\bar{u}\|^2.$$

2. For every u and v such that $\|\bar{u} - \bar{v}\|^2 \geqslant \beta \min\left\{\|\bar{u}\|^2, \|\bar{v}\|^2\right\}$
$$\mathbb{P}\left[u \in S \text{ and } v \in S\right] \leqslant \alpha \frac{\min\left\{\|\bar{u}\|^2, \|\bar{v}\|^2\right\}}{m}.$$

3. For every $e \subset V$,
$$\mathbb{P}\left[e \text{ is cut by } S\right] \leqslant \alpha D_{\ell_2^2} \cdot \max_{u,v \in e} \|\bar{u} - \bar{v}\|^2 + \alpha D_{\ell_2}(|e|) \cdot \min_{w \in e} \|\bar{w}\| \cdot \max_{u,v \in e} \|\bar{u} - \bar{v}\|.$$

(This definition differs from Definition 7.2.2 only in item 3.)

Theorem 7.4.2. *There is a polynomial-time randomized algorithm that given a set of vertices V, a set of vectors $\{\bar{u}\}$ satisfying ℓ_2^2–triangle inequalities, and parameters m and β generates an ℓ_2–ℓ_2^2 hypergraph m-orthogonal separator with probability scale $\alpha \geqslant 1/n$ and distortions:*

$$D_{\ell_2^2} = \mathcal{O}(m),$$
$$D_{\ell_2}(r) = \mathcal{O}\left(\beta^{-1/2} \sqrt{\log r}\, m \log m \log\log m\right).$$

Note that distortions $D_{\ell_2^2}$ and D_{ℓ_2} do not depend on n.

The algorithm and its analysis are very similar to those in the proof of Theorem 7.2.3. The only difference is that we use another procedure to generate random assignments $\omega : V \to \{0, 1\}$. The following lemma is an analog of Lemma 7.3.2.

Lemma 7.4.3. *There is a randomized polynomial time algorithm that given a finite set V, vectors $\psi(\bar{u})$ for $u \in V$, satisfying ℓ_2^2 triangle inequalities, and a parameter $\beta \in (0,1)$, returns a random assignment $\omega : V \to \{0,1\}$ that satisfies the following properties.*

- *For every set $e \subset V$ of size at least 2,*

$$\mathbb{P}\left[\omega(u) \neq \omega(v) \text{ for some } u, v \in e\right] \leqslant \mathcal{O}\left(\beta^{-1/2}\sqrt{\log|e|}\right) \times \max_{u,v\in e} \|\psi(\bar{u}) - \psi(\bar{v})\|.$$

- *For every u and v such that $\|\psi(\bar{u}) - \psi(\bar{v})\|^2 \geqslant \beta$,*

$$\mathbb{P}\left[\omega(u) \neq \omega(v)\right] \geqslant 0.3.$$

Proof. We sample a random Gaussian vector $g \sim \mathcal{N}(0, I_n)$ (each component g_i of g is distributed as $\mathcal{N}(0,1)$, all random variables g_i are mutually independent). Let N be a Poisson process on $\mathbb{R}$ with rate $1/\sqrt{\beta}$. Let

$$\omega(u) = \begin{cases} 1 & \text{if } N(\langle g, \psi(\bar{u})\rangle) \text{ is even} \\ 0 & \text{if } N(\langle g, \psi(\bar{u})\rangle) \text{ is odd} \end{cases}.$$

Note that $\omega(u) = \omega(v)$ if and only if $N(\langle g, \psi(\bar{u})\rangle) - N(\langle g, \psi(\bar{v})\rangle)$ is even.

Consider a set $e \subset V$ of size at least 2. Denote $\mathsf{diam}(e) = \max_{u,v\in e} \|\psi(\bar{u}) - \psi(\bar{v})\|$. Let $\tau_m = \min_{w\in e} \langle g, \psi(\bar{w})\rangle$ and $\tau_M = \max_{w\in e} \langle g, \psi(\bar{w})\rangle$. Note that

$$N(\tau_m) = \min_{w\in e} N(\langle g, \psi(\bar{w})\rangle),$$
$$N(\tau_M) = \max_{w\in e} N(\langle g, \psi(\bar{w})\rangle).$$

If all numbers $N(\langle g, \psi(\bar{u})\rangle)$ are equal then $\omega(u) = \omega(v)$ for all $u, v \in e$. Thus if $\omega(u) \neq \omega(v)$ for some $u, v \in e$ then $N(\langle g, \psi(\bar{u})\rangle) \neq N(\langle g, \psi(\bar{v})\rangle)$ for some $u, v \in e$. In particular, then $N(\tau_M) - N(\tau_m) > 0$. Given g, $N(\tau_M) - N(\tau_m)$ is a Poisson random variable with rate $(\tau_M - \tau_m)/\sqrt{\beta}$. We have,

$$\begin{aligned}\mathbb{P}\left[\omega(u) \neq \omega(v) \text{ for some } u, v \in e \mid g\right] &\leqslant \mathbb{P}\left[N(\tau_M) - N(\tau_m) > 0 \mid g\right] \\ &= 1 - e^{-(\tau_M - \tau_m)/\sqrt{\beta}} \leqslant \beta^{-1/2}(\tau_M - \tau_m).\end{aligned}$$

Let $\xi_{uv} = \langle g, \psi(\bar{u})\rangle - \langle g, \psi(\bar{v})\rangle$ for $u, v \in e$ $(u \neq v)$. Note that ξ_{uv} are Gaussian random variables with mean 0, and

$$\mathsf{Var}\left[\xi_{uv}\right] = \mathsf{Var}\left[\langle g, \psi(\bar{u})\rangle - \langle g, \psi(\bar{v})\rangle\right] = \|\psi(\bar{u}) - \psi(\bar{v})\|^2 \leqslant \mathsf{diam}(e)^2$$

Note that the expectation of the maximum of (not necessarily independent) r Gaussian random variables with standard deviation bounded by σ is $\mathcal{O}\left(\sqrt{\log r}\sigma\right)$ (Fact 2.5.5). We have,

$$\mathbb{E}\left[\tau_M - \tau_m\right] = \mathbb{E}\left[\max_{u,v\in e}(\xi_{uv})\right] = \mathcal{O}\left(\sqrt{\log|e|}\mathsf{diam}(e)\right)$$

since the total number of random variables ξ_{uv} is $|e|(|e|-1)$. Therefore,

$$\begin{aligned}\mathbb{P}\left[\omega(u) \neq \omega(v) \text{ for some } u, v \in e\right] &\leqslant \beta^{-1/2}\ \mathbb{E}\left[\tau_M - \tau_m\right] \\ &= \mathcal{O}\left(\beta^{-1/2}\sqrt{\log|e|}\ \max_{u,v\in e}\|\psi(\bar{u}) - \psi(\bar{v})\|\right). \quad (67)\end{aligned}$$

We proved that ω satisfies the first property. Now we verify that ω satisfies the second condition. Consider two vertices u and v with $\|\psi(\bar{u}) - \psi(\bar{v})\|^2 \geqslant \beta$. Given g, the random variable $Z = N(\langle g, \psi(\bar{u})\rangle) - N(\langle g, \psi(\bar{v})\rangle)$ has Poisson distribution with rate $\lambda = |\langle g, \psi(\bar{u})\rangle) - \langle g, \psi(\bar{v})\rangle|/\sqrt{\beta}$. We have,

$$\mathbb{P}\left[Z \text{ is even} \mid g\right] = \sum_{k=0}^{\infty}\mathbb{P}\left[Z = 2k \mid g\right] = \sum_{k=0}^{\infty}\frac{e^{-\lambda}\lambda^{2k}}{(2k)!} = \frac{1+e^{-2\lambda}}{2}.$$

Note that λ is the absolute value of a Gaussian random variable with mean 0 and standard deviation $\sigma = \|\psi(\bar{u}) - \psi(\bar{v})\|/\sqrt{\beta} \geqslant 1$. Thus

$$\mathbb{P}\left[Z \text{ is even}\right] = \mathbb{E}\left[\frac{1+e^{-2\sigma|\gamma|}}{2}\right],$$

where γ is a standard Gaussian random variable, $\gamma \sim \mathcal{N}(0,1)$. We have,

$$\mathbb{P}\left[\omega(u) \neq \omega(v)\right] = \mathbb{E}\left[\frac{1-e^{-2\sigma|\gamma|}}{2}\right] \geqslant \mathbb{E}\left[\frac{1-e^{-2|\gamma|}}{2}\right] \geqslant 0.3.$$

□

Now we use Algorithm 7.3.4 to obtain ℓ_2–ℓ_2^2 hypergraph orthogonal separators. The only difference is that we use the procedure from Lemma 7.4.3 rather than from Lemma 7.3.2 to generate assignments ω. We obtain a ℓ_2–ℓ_2^2 hypergraph orthogonal separator.

Theorem 7.4.4. *Random set S obtained from Algorithm 7.3.4 using the procedure from Lemma 7.4.3 (instead of Lemma 7.3.2) is a hypergraph m-orthogonal separator with distortion*

$$D_{\ell_2^2} = \mathcal{O}(m),$$
$$D_{\ell_2}(r) = \mathcal{O}\left(\beta^{-1/2}\sqrt{\log r}\, m \log m \log\log m\right),$$

probability scale $\alpha \geqslant 1/n$ and separation threshold $\beta \in (0,1)$.

Proof. The proof of the theorem is almost identical to that of Theorem 7.3.5. We first check conditions 1 and 2 of ℓ_2–ℓ_2^2 hypergraph orthogonal separators in the same way as we checked conditions 1 and 2 of hypergraph orthogonal separators in Theorem 7.3.5. When we verify that property 3 holds, we use bounds from Lemma 7.4.3. The only difference is how we upper bound the probability of the event $\mathcal{E}_2$.

If $\mathcal{E}_2$ happens then (1) $r \leqslant \rho_m$ (since $A = e$) and (2) $W(u) \neq W(v)$ for some $u, v \in e$. The probability that $r \leqslant \rho_m$ is ρ_m. We upper bound the probability that $W(u) \neq W(v)$ for some $u, v \in e$. For each $i \in \{1, \dots, l\}$,

$$\begin{aligned}
&\mathbb{P}\left[\tilde{\omega}_i(u) \neq \tilde{\omega}_i(v) \text{ for some } u, v \in e\right] \\
&\leqslant \mathcal{O}\left(\beta^{-1/2}\sqrt{\log |e|}\, \log\log m\right) \max_{u,v\in e} \|\psi(\bar{u}) - \psi(\bar{v})\| \\
&\leqslant \mathcal{O}\left(\beta^{-1/2}\sqrt{\log |e|}\, \log\log m\right) \max_{u,v\in e} \frac{\|\bar{u} - \bar{v}\|}{\min\{\|\bar{u}\|, \|\bar{v}\|\}} \\
&\leqslant \mathcal{O}\left(\beta^{-1/2}\sqrt{\log |e|}\, \log\log m\right) \times \rho_m^{-1/2} \times \max_{u,v\in e} \|\bar{u} - \bar{v}\|\,.
\end{aligned}$$

By the union bound over $i \in \{1, \dots, l\}$, the probability that $W(u) \neq W(v)$ for some $u, v \in e$ is at most $\mathcal{O}\left(l \times \beta^{-1/2}\sqrt{\log |e|}\, \log\log m\right) \times \rho_m^{-1/2} \times \max_{u,v\in e} \|\bar{u} - \bar{v}\|$.

Therefore,

$$\begin{aligned}
\mathbb{P}\left[\mathcal{E}_2\right] &\leqslant \rho_m \times \mathcal{O}\left(l \times \beta^{-1/2}\sqrt{\log|e|}\,\log\log m\right) \times \rho_m^{-1/2} \times \max_{u,v\in e}\|\bar{u}-\bar{v}\| \\
&\leqslant \mathcal{O}\left(l \times \beta^{-1/2}\sqrt{\log|e|}\,\log\log m\right) \times \rho_m^{1/2} \times \max_{u,v\in e}\|\bar{u}-\bar{v}\|.
\end{aligned}$$

We get that the probability that e is cut by S is at most

$$\begin{aligned}
\mathbb{P}\left[\mathcal{E}_1\right] + \mathbb{P}\left[\mathcal{E}_2\right] &\leqslant \max_{u,v\in e}\|\bar{u}-\bar{v}\|^2 + \mathcal{O}\left(l \times \beta^{-1/2}\sqrt{\log|e|}\,\log\log m\right) \\
&\qquad \times \rho_m^{1/2} \times \max_{u,v\in e}\|\bar{u}-\bar{v}\| \\
&\leqslant \max_{u,v\in e}\|\bar{u}-\bar{v}\|^2 + \mathcal{O}\left(l \times \beta^{-1/2}\sqrt{\log|e|}\,\log\log m\right) \\
&\qquad \times \min_{w\in e}\|\bar{w}\| \times \max_{u,v\in e}\|\bar{u}-\bar{v}\|.
\end{aligned}$$

For $D_{\ell_2^2} = 1/\alpha$ and $D_{\ell_2}(r) = \mathcal{O}\left(\beta^{-1/2}\sqrt{\log r}\,\log m \log\log m\right)/\alpha$, we get

$$\mathbb{P}\left[e \text{ is cut by } S\right] \leqslant \alpha D_{\ell_2^2} \cdot \max_{u,v\in e}\|\bar{u}-\bar{v}\|^2 + \alpha D_{\ell_2}(|e|) \cdot \min_{w\in e}\|\bar{w}\| \cdot \max_{u,v\in e}\|\bar{u}-\bar{v}\|.$$

Note that $\alpha \geqslant 1/2^l \geqslant \Omega(1/m)$. Thus

$$\begin{aligned}
D_{\ell_2^2} &= \mathcal{O}(m), \\
D_{\ell_2}(r) &= \mathcal{O}\left(\beta^{-1/2}\sqrt{\log r}\, m \log m \log\log m\right).
\end{aligned}$$

□

7.5 *Algorithm for Hypergraph Small Set Expansion via ℓ_2–ℓ_2^2 Hypergraph Orthogonal Separators*

In this section, we present another algorithm for Hypergraph Small Set Expansion. The algorithm finds a set with expansion proportional to $\sqrt{\phi_{G,\delta}}$. The proportionality constant depends on degrees of vertices and hyperedge size but not on the graph size. Here, we present our result for arbitrary hypergraphs. The result for uniform hypergraphs (Theorem 7.1.3) stated in the introduction follows from our general result. In order to state our result for arbitrary graphs, we need the following definition.

Definition 7.5.1. Consider a hypergraph $H = (V, E)$. Suppose that for every edge e we are given a non-empty subset $e^\circ \subseteq e$.Let

$$\eta(u) = \sum_{e: u \in e^\circ} \frac{\log_2 |e|}{|e^\circ|},$$

$$\eta_{max} = \max_{u \in V} \eta(u).$$

Finally, let $\eta^H_{\max}$ be the minimum of η_{max} over all possible choices of subsets e°.

Claim 7.5.2. *1.* $\eta^H_{max} \leqslant \max_{u \in V} \sum_{e: u \in e} (\log_2 |e|)/|e|$.

2. If H is a r-uniform graph with maximum degree d_{max} then $\eta^H_{max} \leqslant (d_{max} \log_2 r)/r$.

3. Suppose that we can choose one vertex in every edge so that no vertex is chosen more than once. Then $\eta^H_{max} \leqslant \log_2 r_{max}$, where r_{max} is the size of the largest hyperedge in H.

Proof.

1. Let $e^\circ = e$ for every $e \in E$. We have, $\eta^H_{\max} \leqslant \max_{u \in V} \sum_{e: u \in e} (\log_2 |e|)/|e|$.

2. By item 1,

$$\eta^H_{\max} \leqslant \max_{u \in V} \sum_{e: u \in e} (\log_2 |e|)/|e| = \max_{u \in V} \sum_{e: u \in e} (\log_2 r)/r = (d_{\max} \log_2 r)/r.$$

3. For every edge $e \in E$, let e° be the set that contains the vertex chosen for e. Then $|e^\circ| = 1$ and $|\{e : u \in e^\circ\}| \leqslant 1$ for every u. We have,

$$\eta^H_{\max} \leqslant \max_{u \in V} \sum_{e: u \in e^\circ} \frac{\log_2 |e|}{|e^\circ|} \leqslant \max_{u \in V} \sum_{e: u \in e^\circ} \frac{\log_2 r_{\max}}{1} = \log_2 r_{\max}.$$

□

Theorem 7.5.3. *There is a randomized polynomial-time algorithm that given a hypergraph $H = (V, E)$ with vertex weights $w(v) = d_v$, and parameters $\varepsilon \in (0, 1)$ and*

$\delta \in (0, 1/2]$, finds a set $S \subset V$ of size at most $(1+\varepsilon)\delta n$ such that

$$\begin{aligned}\phi(S) &\leqslant \mathcal{O}_\varepsilon\left(\delta^{-1}\log\delta^{-1}\log\log\delta^{-1}\sqrt{\eta^H_{max}\cdot\phi_{H,\delta}} + \delta^{-1}\phi_{H,\delta}\right)\\ &= \tilde{\mathcal{O}}_\varepsilon\left(\delta^{-1}\left(\sqrt{\eta^H_{max}\phi_{H,\delta}} + \phi_{H,\delta}\right)\right),\end{aligned}$$

In particular, if H is an r-uniform hypergraph then we have,

$$\phi(S) \leqslant \tilde{\mathcal{O}}_\varepsilon\left(\delta^{-1}\left(\sqrt{\frac{\log_2 r}{r}\phi_{H,\delta}} + \phi_{H,\delta}\right)\right)$$

Proof. The proof is similar to that of Theorem 7.2.4. We solve the SDP relaxation for H-SSE and obtain an SDP solution $\{\bar{u}\}$. Denote the SDP value by SDPval. Consider an ℓ_2–ℓ_2^2 hypergraph orthogonal separator S with $m = 4/(\varepsilon\delta)$ and $\beta = \varepsilon/4$. Define a set S':

$$S' = \begin{cases} S & \text{if } |S| \leqslant (1+\varepsilon)\delta n \\ \emptyset & \text{otherwise} \end{cases}.$$

Clearly, $|S'| \leqslant (1+\varepsilon)\delta n$. As in the proof of Theorem 7.2.4,

$$\mathbb{P}\left[u \in S'\right] \in \left[\frac{\alpha}{2}\|\bar{u}\|^2, \alpha\|\bar{u}\|^2\right].$$

Note that

$$\begin{aligned}&\mathbb{P}\left[S' \text{ cuts edge } e\right] \leqslant \mathbb{P}\left[S \text{ cuts edge } e\right]\\ &\qquad \leqslant \alpha D_{\ell_2}\max_{u,v\in e}\|\bar{u}-\bar{v}\|^2 + \alpha D_{\ell_2}(r)\min_{w\in e}\|\bar{w}\|\max_{u,v\in e}\|\bar{u}-\bar{v}\|.\end{aligned}$$

Let

$$\mathcal{C} = \alpha^{-1}\,\mathbb{E}\left[\sum_{e\in E(S',\bar{S}')} w(e)\right] \qquad \text{and} \qquad Z = w(S') - \frac{\sum_{e\in E(S',\bar{S}')} w(e)}{4\mathcal{C}}.$$

We have,

$$\begin{aligned}&\mathbb{E}\left[Z\right] = \mathbb{E}\left[w(S')\right] - \mathbb{E}\left[\frac{\sum_{e\in E(S',\bar{S}')} w(e)}{4\mathcal{C}}\right]\\ &\qquad \geqslant \sum_{u\in V}\left(\frac{\alpha}{2}\cdot\|\bar{u}\|^2\right)w(u) - \frac{\alpha}{4} = \frac{\alpha}{2} - \frac{\alpha}{4} = \frac{\alpha}{4}.\end{aligned}$$

Now we upper bound $\mathcal{C}$.

$$\begin{aligned}\mathcal{C} = \alpha^{-1}\,\mathbb{E}\left[\sum_{e\in E(S',\bar{S}')} w(e)\right] &\leqslant \alpha^{-1}\sum_{e\in E} w(e)\,\mathbb{P}\left[e \text{ is cut by } S\right]\\ &\leqslant D_{\ell_2^2}\sum_{e\in E} w(e)\max\|\bar{u}-\bar{v}\|^2 + \sum_{e\in E} w(e)\,D_{\ell_2}(|e|)\min_{w\in e}\|\bar{w}\|\max_{u,v\in e}\|\bar{u}-\bar{v}\|\\ &\leqslant D_{\ell_2^2}\cdot\mathsf{SDPval} + \sqrt{\sum_{e\in E} w(e)\,D_{\ell_2}(|e|)^2\min_{w\in e}\|\bar{w}\|^2}\sqrt{\sum_{e\in E} w(e)\max_{u,v\in e}\|\bar{u}-\bar{v}\|^2}\\ &\leqslant D_{\ell_2^2}\cdot\mathsf{SDPval} + \sqrt{\sum_{e\in E}\sum_{w\in e}\frac{D_{\ell_2}(|e|)^2}{|e|}\|\bar{w}\|^2}\,\sqrt{\mathsf{SDPval}}\end{aligned}$$

For every vertex w,

$$\sum_{e:w\in e}\frac{D_{\ell_2}(|e|)^2}{|e|} \leqslant \mathcal{O}_\beta(m\log m\log\log m)^2\sum_{e:w\in e}\frac{\log_2|e|}{|e|} \leqslant \mathcal{O}_\beta(m\log m\log\log m)^2\times\eta^H_{\max}.$$

and $\sum_{w\in V} d_u\|\bar{w}\|^2 = \sum_{w\in V} w_u\|\bar{w}\|^2 = 1$. Therefore,

$$\mathcal{C} \leqslant \mathcal{O}_\beta\left(m\mathsf{SDPval} + m\log m\log\log m\,\sqrt{\eta^H_{\max}\cdot\mathsf{SDPval}}\right).$$

By the argument from Theorem 7.2.4, we get that if we sample S' sufficiently many times (i.e., $(4n^2/\alpha)$ times), we will find a set S' such that

$$\phi(S') \leqslant 4\mathcal{C} \leqslant \mathcal{O}_\beta\left(\delta^{-1}\log\delta^{-1}\log\log\delta^{-1}\sqrt{\eta^H_{\max}\cdot\mathsf{SDPval}} + \delta^{-1}\mathsf{SDPval}\right)$$

with probability exponentially close to 1. □

7.6 SDP *Intgrality Gap*

In this section, we present an integrality gap for the SDP relaxation for H-SSE. We also give a lower bound on the distortion of a hypergraph m-orthogonal separator.

Theorem 7.6.1. *For $\delta = 1/r$, the integrality gap of the* SDP *for* H-SSE *is at least $1/(2\delta) = r/2$.*

Proof. Consider a hypergraph $H = (V, E)$ on $n = r$ vertices with one hyperedge $e = V$ (e contains all vertices). Note that the expansion of every set of size $\delta n = 1$ is 1. Thus $\phi_{H,\delta} = 1$.

Consider an SDP solution that assigns vertices mutually orthogonal vectors of length $1/\sqrt{r}$. It is easy to see this is a feasible SDP solution. Its value is $\max_{u,v\in e} \|\bar{u}-\bar{v}\|^2 = 2/r$. Therefore, the SDP integrality gap is at least $r/2$. □

Now we give a lower bound on the distortion of hypergraph m-orthogonal separators.

Lemma 7.6.2. *For every $m > 4$, there is an* SDP *solution such that every hypergraph m-orthogonal separator with separation threshold $\beta \geqslant 0$ has distortion at least $\lceil m \rceil/4$.*

Proof. Consider the SDP solution from Theorem 7.6.1 for $n = r = \lceil m \rceil$. Consider a hypergraph m-orthogonal separator S for this solution. Let D be its distortion. Note that condition (2) from the definition of hypergraph orthogonal separators applies to any pair of distinct vertices (u, v) since $\langle \bar{u}, \bar{v}\rangle = 0$.

By the inclusion–exclusion principle, we have,

$$\begin{aligned}
\mathbb{P}\left[|S| = 1\right] &\geqslant \sum_{u\in S} \mathbb{P}\left[u \in S\right] - \frac{1}{2} \sum_{u,v\in S, u\neq v} \mathbb{P}\left[u\in S, v\in S\right] \\
&\geqslant \sum_{u\in S} \alpha\|\bar{u}\|^2 - \frac{1}{2} \sum_{u,v\in S,u\neq v} \frac{\alpha \min\left\{\|\bar{u}\|^2, \|\bar{v}\|^2\right\}}{m} \\
&= \alpha - \frac{\alpha n(n-1)}{2mr} = \alpha\left(1 - \frac{(n-1)}{2m}\right) \geqslant \alpha/2.
\end{aligned}$$

On the other hand, if $|S| = 1$ then S cuts e. We have,

$$\mathbb{P}\left[|S| = 1\right] \leqslant \mathbb{P}\left[S \text{ cuts } e\right] \leqslant \alpha D \max_{u,v\in e} \|\bar{u}-\bar{v}\|^2 = 2\alpha D/r.$$

We get that $\alpha/2 \leqslant 2\alpha D/r$ and thus $D \geqslant r/4 = \lceil m \rceil/4$. □

7.7 Reduction from Vertex Expansion to Hypergraph Expansion

Theorem 7.7.1 (Restatement of Theorem 7.1.5)**.** *There exist absolute constants $c_1', c_2' \in \mathbb{R}^+$ such that for every graph $G = (V, E)$, of maximum degree d, there exists* *a polynomial time computable hypergraph $H = (V', E')$ having the hyperedges of cardinality at most $d+1$ such that*

$$c_1' \phi_{H,\delta} \leqslant \phi^{\mathsf{V}}_{G,\delta} \leqslant c_2' \phi_{H,\delta},$$

and $\eta^H_{max} \leqslant \log_2 d_{max}$.

Proof. Starting with graph G, we use Theorem 8.3.2 to obtain a graph $G' = (V', E')$ such that

$$c_1 \phi^{\mathsf{V}}_{G,\delta} \leqslant \Phi^{\mathsf{V}}_{G',\delta} \leqslant c_2 \phi^{\mathsf{V}}_{G,\delta}. \tag{68}$$

Next we construct hypergraph $H = (V', E'')$ using the reduction in Theorem 4.2.19. We get that

$$\phi_H(S) = \Phi^{\mathsf{V}}(S) \qquad \text{for every } S' \subset V,$$

and hence

$$\phi_{H,\delta} = \Phi^{V}_{G',\delta}$$

We get from (68),

$$c_1 \phi^{\mathsf{V}}_{G,\delta} \leqslant \phi_{H,\delta} \leqslant c_2 \phi^{\mathsf{V}}_{G,\delta}.$$

Finally, we upper bound $\eta^H_{\max}$. We use part 3 of Claim 7.5.2. We choose vertex v in the hyperedge $\{v\} \cup N^{\mathsf{out}}(\{v\})$. By Claim 7.5.2, $\eta^H_{\max} \leqslant \log_2 r_{\max}$, where $r_{\max}$ is the size of the largest hyperedge. Note that $|\{v\} \cup N^{\mathsf{out}}(\{v\})| = d_v + 1$. Thus $\eta^H_{\max} \leqslant \log_2 r_{\max} \leqslant \log_2(d_{\max} + 1)$ □

7.8 Conclusion

The Small Set Vertex Expansion recently gained interest due to its connection to obtaining sub-exponential-time, constant factor approximation algorithms for many combinatorial problems like Sparsest Cut and Graph Coloring ([9, 74]). To the best of our knowledge, our algorithms are the first approximation algorithms for these problems. Our approximation guarantees are not strong enough to have any implications in obtaining sub-exponential-time, constant factor approximation algorithms for Sparsest Cut and Graph Coloring, etc ([9, 74]).

We do not know of any computational lower bounds for these problems, expect those that follow from computational lower bounds for SMALL SET EXPANSION in graphs $\left(\Omega\left(\sqrt{\mathsf{OPT}\log 1/\delta}\right)\right)$, and for VERTEX EXPANSION $\left(\Omega\left(\sqrt{\mathsf{OPT}\log d}\right)\right)$ and HYPERGRAPH EXPANSION $\left(\Omega\left(\sqrt{\mathsf{OPT}\log r}\right)\right)$. Closing the gap between the approximation upper bounds and the computational lower bounds is left as an open problem.

Acknowledgements. The results in this chapter were obtained in joint work with Yury Makarychev [71].

THE COMPLEXITY OF EXPANSION PROBLEMS

PART III

Computational Lower bounds

CHAPTER VIII

HARDNESS OF VERTEX EXPANSION PARAMETERS

In this chapter, we show a hardness result suggesting that there is no efficient algorithm to recognize vertex expanders. More precisely, our main result is a hardness for the problem of approximating λ_∞ in graphs of bounded degree d. The hardness result shows that the approximability of vertex expansion degrades with the degree, and therefore the problem of recognizing expanders is hard for sufficiently large degree. Furthermore, we exhibit an approximation algorithm for λ_∞ (and hence also for vertex expansion) whose guarantee matches the hardness result up to constant factors.

Formal Statement of Results. Most known hardness results for Vertex Expansion follow from the corresponding hardness results for Edge Expansion. It is natural to ask if one can prove better inapproximability results for Vertex Expansion than those that follow from the inapproximability results for Edge Expansion. Indeed, the best one could hope for would be a lower bound matching the upper bound in Theorem 6.1.3 $\left(\mathcal{O}\left(\sqrt{\mathsf{OPT} \log d}\right)\right)$. Our main result is a reduction from SSE to the problem of distinguishing between the case when vertex expansion of the graph is at most ε and the case when the vertex expansion is at least $\Omega(\sqrt{\varepsilon \log d})$. This immediately implies that it is SSE-hard to find a subset of vertex expansion less than $C\sqrt{\phi^{\mathsf{V}} \log d}$ for some constant C. To the best of our knowledge, our work is the first evidence that vertex expansion might be harder to approximate than edge expansion. More formally, we state our main theorem below.

Theorem 8.0.1. *For every $\eta > 0$, there exists an absolute constant C such that $\forall \varepsilon > 0$ it is* SSE*-hard to distinguish between the following two cases for a given graph*

$G = (V, E)$ with maximum degree $d \geqslant 100/\varepsilon$.

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

$$\phi^{\mathsf{V}}(S) \leqslant \varepsilon\,.$$

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

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

This immediately implies a lower bound for the computation of λ_∞

Theorem 8.0.2 (Corollary to Theorem 8.0.1 and Theorem 6.1.2)**.** *For every $\eta > 0$, there exists an absolute constant C such that $\forall \varepsilon > 0$ it is* SSE*-hard to distinguish between the following two cases for a given graph $G = (V, E)$ with maximum degree $d \geqslant 100/\varepsilon$.*

Yes *: There exists a vector $X \in \mathbb{R}^n$ such that*

$$\frac{\sum_{i\in V} \max_{j\sim i}(X_i - X_j)^2}{\sum_i X_i^2 - \frac{1}{n}(\sum_i X_i)^2} \leqslant \varepsilon$$

No *: For all vectors $X \in \mathbb{R}^n$,*

$$\frac{\sum_{i\in V} \max_{j\sim i}(X_i - X_j)^2}{\sum_i X_i^2 - \frac{1}{n}(\sum_i X_i)^2} \geqslant \min\left\{10^{-10}, C\sqrt{\varepsilon \log d}\right\} - \eta$$

By a suitable choice of parameters in Theorem 8.0.1, we obtain the following.

Theorem 8.0.3. *There exists an absolute constant $\delta_0 > 0$ such that for every constant $\epsilon > 0$ the following holds: Given a graph $G = (V, E)$, it is* SSE*-hard to distinguish between the following two cases:*

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

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

No *: (G is a vertex expander with constant expansion) For all sets* $S \subset V$,

$$\phi^{\mathsf{V}}(S) \geqslant \delta_0$$

In particular, the above result implies that it is SSE-hard to certify that a graph is a vertex expander with constant expansion. This is in contrast to the case of edge expansion, where the Cheeger's inequality can be used to certify that a graph has constant edge expansion.

At a high level, the proof is as follows. We introduce the notion of Balanced Analytic Vertex Expansion for Markov chains. This quantity can be thought of as a CSP on $(d+1)$-tuples of vertices. We show a reduction from Balanced Analytic Vertex Expansion of a Markov chain, say H, to vertex expansion of a graph, say H_1 (Section 8.7). Our reduction is generic and works for any Markov chain H. Surprisingly, the CSP-like nature of Balanced Analytic Vertex Expansion makes it amenable to a reduction from Small Set Expansion (Section 8.6). We construct a gadget for this reduction and study its embedding into the Gaussian graph to analyze its soundness (Section 8.4 and Section 8.5). The gadget involves a sampling procedure to generate a bounded-degree graph.

Hypergraph Expansion Using the reduction from Vertex Expansion to Hypergraph Expansion (Theorem 4.2.19), we get the following hardness results for Hypergraph Expansion and γ_2.

Theorem 8.0.4 (Corollary to Theorem 4.2.19 and Theorem 8.0.1)**.** *For every* $\eta > 0$*, there exists an absolute constant* C *such that* $\forall \varepsilon > 0$ *it is* SSE*-hard to distinguish between the following two cases for a given hypergraph* $H = (V, E, w)$ *with maximum hyperedge size* $r \geqslant 100/\varepsilon$*.*

Yes *: There exists a set* $S \subset V$ *such that*

$$\phi_H(S) \leqslant \varepsilon$$

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

$$\phi_H(S) \geqslant \min\left\{10^{-10}, C\sqrt{\varepsilon \log r}\right\} - \eta$$

Theorem 8.0.5 (Corollary to Theorem 4.2.19 and Theorem 8.0.2)**.** *For every $\eta > 0$, there exists an absolute constant C such that $\forall \varepsilon > 0$ it is SSE-hard to distinguish between the following two cases for a given hypergraph $H = (V, E, w)$ with maximum hyperedge size $r \geqslant 100/\varepsilon$.*

Yes : *There exists an $X \in \mathbb{R}^n$ such that $\langle X, \mu^* \rangle = 0$ and*

$$\mathcal{R}(X) \leqslant \varepsilon$$

No : *For all $X \in \mathbb{R}^n$ such that $\langle X, \mu^* \rangle = 0$,*

$$\mathcal{R}(X) \geqslant \min\left\{10^{-10}, C\varepsilon \log r\right\} - \eta$$

Related Work. An $\mathcal{O}(\log n)$ approximation algorithm for ϕ was obtained by Leighton and Rao [64]. The current best approximation factor for ϕ^{V} is $\mathcal{O}\left(\sqrt{\log n}\right)$ obtained using a convex relaxation by Feige, Lee and Hajiaghayi [39]. Beyond this, the situation is much less clear for the approximability of vertex expansion. Applying Cheeger's method leads to a bound of $\mathcal{O}\left(\sqrt{d\mathsf{OPT}}\right)$ [1] where d is the maximum degree of the input graph. Ambühl, Mastrolilli and Svensson [7] showed that ϕ^{V} and ϕ have no PTAS assuming that SAT does not have sub-exponential time algorithms.

8.1 Proof Overview

Balanced Analytic Vertex Expansion. To exhibit a hardness result, we begin by defining a combinatorial optimization problem related to the problem of approximating vertex expansion in graphs having largest degree d. This problem referred to as Balanced Analytic Vertex Expansion can be motivated as follows.

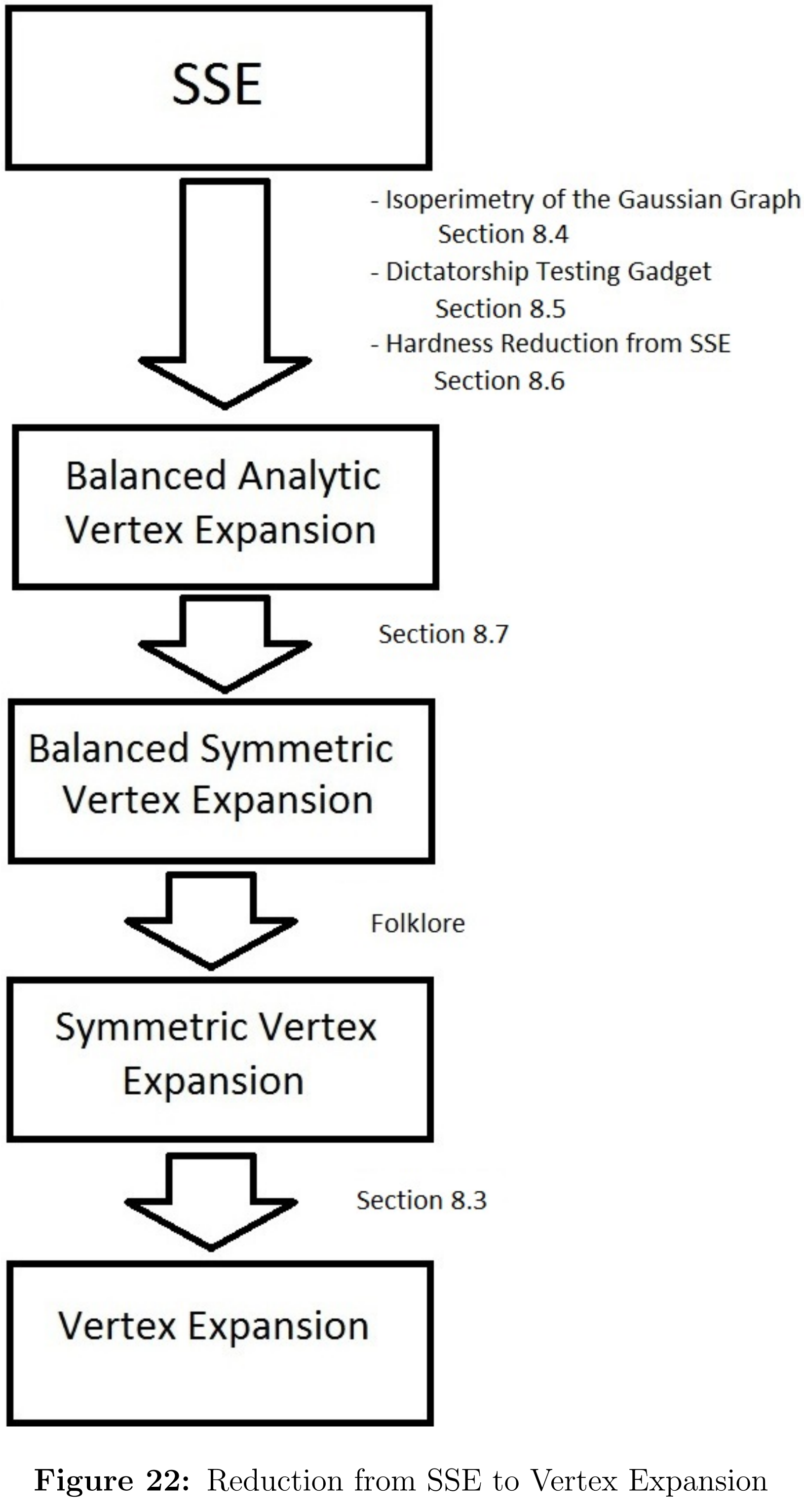

Figure 22: Reduction from SSE to Vertex Expansion

Fix a graph $G = (V, E)$ and a subset of vertices $S \subset V$. For any vertex $v \in V$, v is on the boundary of the set S if and only if $\max_{u \in N(v)} |\mathbb{I}_S[u] - \mathbb{I}_S[v]| = 1$, where $N(v)$ denotes the neighbourhood of vertex v. In particular, the fraction of vertices on the boundary of S is given by $\mathbb{E}_v \max_{u \in N(v)} |\mathbb{I}_S[u] - \mathbb{I}_S[v]|$. The *symmetric* vertex expansion of the set $S \subseteq V$ is given by,

$$n \cdot \frac{|N(S) \cup N(V \backslash S)|}{|S|\,|V \backslash S|} = \frac{\mathbb{E}_v \max_{u \in N(v)} |\mathbb{I}_S[u] - \mathbb{I}_S[v]|}{\mathbb{E}_{u,v} |\mathbb{I}_S[u] - \mathbb{I}_S[v]|}.$$

Note that for a degree d graph, each of the terms in the numerator is maximization over the d edges incident at the vertex. The formal definition of Balanced Analytic Vertex Expansion is as follows.

Definition 8.1.1. An instance of Balanced Analytic Vertex Expansion, denoted by $(V, \mathcal{P})$, consists of a set of variables V and a probability distribution $\mathcal{P}$ over $(d+1)$-tuples in V^{d+1}. The probability distribution $\mathcal{P}$ satisfies the condition that all its $d+1$ marginal distributions are the same (denoted by μ). The goal is to solve the following optimization problem

$$\Phi(V, \mathcal{P}) \stackrel{\text{def}}{=} \min_{F : V \to \{0,1\} | \mathbb{E}_{X,Y \sim \mu} |F(X) - F(Y)| \geqslant \frac{1}{100}} \frac{\mathbb{E}_{(X, Y_1, \ldots, Y_d) \sim \mathcal{P}} \max_i |F(Y_i) - F(X)|}{\mathbb{E}_{X,Y \sim \mu} |F(X) - F(Y)|}$$

For constant d, this could be thought of as a constraint satisfaction problem (CSP) of arity $d+1$. Every d-regular graph G has an associated instance of Balanced Analytic Vertex Expansion whose value corresponds to the vertex expansion of G. Conversely, we exhibit a reduction from Balanced Analytic Vertex Expansion to problem of approximating vertex expansion in a graph of degree $\mathsf{poly}(d)$ (Section 8.7 for details).

Dictatorship Testing Gadget. As with most hardness results obtained via the label cover or the unique games problem, central to our reduction is an appropriate dictatorship testing gadget. Simply put, a dictatorship testing gadget for Balanced Analytic Vertex Expansion is an instance $\mathcal{H}^R$ of the problem such that, on one hand there exists the so-called *dictator* assignments with value ϵ, while every assignment far from every dictator incurs a cost of at least $\Omega(\sqrt{\epsilon \log d})$.

The construction of the dictatorship testing gadget is as follows. Let H be a Markov chain on vertices $\{s, t, t', s'\}$ connected to form a path of length three. The transition probabilities of the Markov chain $\mathcal{H}$ are so chosen to ensure that if μ_H is the stationary distribution of H then $\mu_H(t) = \mu_H(t') = \epsilon/2$ and $\mu_H(s) = \mu_H(s') = (1-\epsilon)/2$. In particular, H has a vertex separator $\{t, t'\}$ whose weight under the stationary distribution is only ϵ.

The dictatorship testing gadget is over the product Markov chain H^R for some large constant R. The constraints $\mathcal{P}$ of the dictatorship testing gadget H^R are given by the following sampling procedure,

- Sample $x \in H^R$ from the stationary distribution of the chain.
- Sample d-neighbours $y_1, \ldots, y_d \in H^R$ of x independently from the transition probabilities of the chain H^R. Output the tuple $(x, y_1, \ldots, y_d)$.

For every $i \in [R]$, the i^{th} dictator solution to the above described gadget is given by the following function,

$$F(x) = \begin{cases} 1 & \text{if } x_i \in \{s, t\} \\ 0 & \text{otherwise} \end{cases}$$

It is easy to see that for each constraint $(x, y_1, \ldots, y_d) \sim \mathcal{P}$, $\max_j |F(x) - F(y_j)| = 0$ unless $x_i = t$ or $x_i = t'$. Since x is sampled from the stationary distribution for μ_H, $x_i \in \{t, t'\}$ happens with probability ϵ. Therefore the expected cost incurred by the i^{th} dictator assignment is at most ϵ.

Soundness Analysis of the Gadget. The soundness property desired of the dictatorship testing gadget can be stated in terms of influences. Specifically, given an assignment $F : V(H)^R \to [0,1]$, the influence of the i^{th} coordinate is given by

$$\mathsf{Inf_i[F]} = \mathop{\mathbb{E}}_{\mathsf{x_{[R]\setminus i}}} \mathop{\mathsf{Var}}_{\mathsf{x_i}}[\mathsf{F(x)}],$$

i.e., the expected variance of the function after fixing all but the i^{th} coordinate randomly. Henceforth, we will refer to a function $F : H^R \to [0,1]$ as *far from every dictator* if the influence of all of its coordinates are small (say $< \tau$).

We show that the dictatorship testing gadget H^R described above satisfies the following soundness – for every function F that is far from every dictator, the cost of F is at least $\Omega(\sqrt{\epsilon \log d})$. To this end, we appeal to the invariance principle to translate the cost incurred to a corresponding isoperimetric problem on the Gaussian space. More precisely, given a function $F : H^R \to [0,1]$, we express it as a polynomial in the eigenfunctions over H. We carefully construct a Gaussian ensemble with the same moments up to order two, as the eigenfunctions at the query points $(x, y_1, \ldots, y_d) \in \mathcal{P}$. By appealing to the invariance principle for low degree polynomials, this translates in to the following isoperimetric question over Gaussian space $\mathcal{G}$.,

Suppose we have a subset $S \subseteq \mathcal{G}$ of the n-dimensional Gaussian space. Consider the following experiment:

- Sample a point $z \in \mathcal{G}$ the Gaussian space.
- Pick d independent perturbations $z'_1, z'_2, \ldots, z'_d$ of the point z by ϵ-noise.
- Output 1 if at least one of the edges (z, z'_i) crosses the cut $(S, \bar{S})$ of the Gaussian space.

Among all subsets S of the Gaussian space with a given volume, which set has the least expected output in the above experiment? The answer to this isoperimetric question corresponds to the soundness of the dictatorship test. A halfspace of volume $\frac{1}{2}$ has an expected output of $\sqrt{\epsilon \log d}$ in the above experiment. We show that among all subsets of constant volume, halfspaces achieve the least expected output value.

This isoperimetric theorem proven in Section 8.4 yields the desired $\Omega(\sqrt{\epsilon \log d})$ bound for the soundness of the dictatorship test constructed via the Markov chain H. Here the noise rate of ϵ arises from the fact that all the eigenfunctions of the Markov chain H have an eigenvalue smaller than $1 - \epsilon$. The details of the argument based on invariance principle is presented in Section 8.5

We show a $\Omega(\sqrt{\epsilon \log d})$ lower bound for the isoperimetric problem on the Gaussian space. The proof of this isoperimetric inequality is included in Section 8.4

We would like to point out here that the traditional noisy cube gadget does not suffice for our application. This is because in the noisy cube gadget while the dictator solutions have an edge expansion of ϵ they have a vertex expansion of ϵd, yielding a much worse value than the soundness.

Reduction from Small Set Expansion problem. Gadget reductions from the UNIQUE GAMES problem cannot be used towards proving a hardness result for edge or vertex expansion problems. This is because if the underlying instance of UNIQUE GAMES has a small vertex separator, then the graph produced via a gadget reduction would also have small vertex expansion. Therefore, we appeal to a reduction from the SMALL SET EXPANSION problem (Section 8.6 for details).

Raghavendra, Steurer and Tulsiani [88] show optimal inapproximability results for the Balanced separator problem using a reduction from the SMALL SET EXPANSION problem. While the overall approach of our reduction is similar to theirs, the details are subtle. Unlike hardness reductions from unique games, the reductions for expansion-type problems starting from SMALL SET EXPANSION are not very well understood. For instance, the work of Raghavendra and Tan [89] gives a dictatorship testing gadget for the Max-Bisection problem, but a SMALL SET EXPANSION based hardness for Max-Bisection still remains open.

8.1.1 Organization

We begin with some definitions and the statements of the SSE hypotheses in Section 8.2. In Section 8.3, we show that the computation of Vertex Expansion and Symmetric Vertex Expansion is equivalent up to constant factors. We prove a new Gaussian isoperimetry results in Section 8.4 that we use in our soundness analysis. In Section 8.5 we show the construction of our main gadget and analyze its soundness and completeness using Balanced Analytic Vertex Expansion as the test function. We show a reduction from a reduction from Balanced Analytic Vertex Expansion to vertex expansion in Section 8.7. In Section 8.6, we use this gadget to show a reduction SSE to Balanced Analytic Vertex Expansion. Finally, in Section 8.8, we show how to put all the reductions together to get optimal SSE-hardness for vertex expansion.

8.2 Preliminaries

Analytic Vertex Expansion Our reduction from SSE to vertex expansion goes via an intermediate problem that we call d-Balanced Analytic Vertex Expansion. We recall the definition of d-Balanced Analytic Vertex Expansion here.

Definition 8.2.1. An instance of d-Balanced Analytic Vertex Expansion, denoted by $(V, \mathcal{P})$, consists of a set of variables V and a probability distribution $\mathcal{P}$ over $(d+1)$-tuples in V^{d+1}. The probability distribution $\mathcal{P}$ satisfies the condition that all its $d+1$ marginal distributions are the same (denoted by μ). The d-Balanced Analytic Vertex Expansion under a function $F : V \to \{0,1\}$ is defined as

$$\Phi(V,\mathcal{P})(F) \stackrel{\text{def}}{=} \frac{\mathbb{E}_{(X,Y_1,\ldots,Y_d)\sim\mathcal{P}} \max_i |F(Y_i) - F(X)|}{\mathbb{E}_{X,Y\sim\mu} |F(X) - F(Y)|} .$$

The d-Balanced Analytic Vertex Expansion of $(V, \mathcal{P})$ is defined as

$$\Phi(V,\mathcal{P}) \stackrel{\text{def}}{=} \min_{F:V\to\{0,1\} \mid \mathbb{E}_{X,Y\sim\mu}|F(X)-F(Y)| \geqslant \frac{1}{100}} \Phi(V,\mathcal{P})(F).$$

We drop the degree d from the notation, when it is clear from the context.

For an instance $(V, \mathcal{P})$ of Balanced Analytic Vertex Expansion and an assignment $F : V \to \{0, 1\}$ define

$$\mathsf{VAL}_{\mathcal{P}}(F) = \mathop{\mathbb{E}}_{(X,Y_1,\ldots,Y_d)\sim\mathcal{P}} \max_i |F(Y_i) - F(X)| \,.$$

Gaussian Graph. Recall that two standard normal random variables X, Y are said to be α-correlated if there exists an independent standard normal random variable Z such that $Y = \alpha X + \sqrt{1-\alpha^2} Z$.

Definition 8.2.2. The *Gaussian Graph* $\mathcal{G}_{\Lambda,\Sigma}$ is a complete weighted graph on the vertex set $V(\mathcal{G}_{\Lambda,\Sigma}) = \mathbb{R}^n$. The weight of the edge between two vertices $u, v \in V(\mathcal{G}_{\Lambda,\Sigma})$ is given by

$$w(\{u, v\}) = \mathbb{P}[X = u \text{ and } Y = v]$$

where $Y \sim \mathcal{N}(\Lambda X, \Sigma)$, where Λ is a diagonal matrix such that $\|\Lambda\| \leqslant 1$ and $\Sigma \succeq \varepsilon I$ is a diagonal matrix.

Remark 8.2.3. Note that for any two non-empty disjoint sets $S_1, S_2 \subset V(\mathcal{G}_{\Lambda,\Sigma})$, the total weight of the edges between S_1 and S_2 can be non-zero even though every single edge in the $\mathcal{G}_{\Lambda,\Sigma}$ has weight zero.

Definition 8.2.4. We say that a family of graphs $\mathcal{G}_d$ is $\Theta(d)$-regular, if there exist absolute constants $c_1, c_2 \in \mathbb{R}^+$ such that for every $G \in \mathcal{G}_d$, all vertices $i \in V(G)$ have $c_1 d \leqslant d_i \leqslant c_2 d$.

We now formalize our notion of hardness.

Definition 8.2.5. A constrained minimization problem $\mathcal{A}$ with its optimal value denoted by $\mathsf{VAL}(\mathcal{A})$ is said to be c-vs-s hard if it is SSE-hard to distinguish between the following two cases.

Yes:

$$\mathsf{VAL}(\mathcal{A}) \leqslant c \,.$$

No:

$$\mathsf{VAL}(\mathcal{A}) \geqslant s\,.$$

Variance. For a random variable X, define the variance and ℓ_1-variance as follows,

$$\mathsf{Var}\,[X] = \mathop{\mathbb{E}}_{X_1,X_2}[(X_1 - X_2)^2]$$

$$\mathop{\mathsf{Var}}_1[X] = \mathop{\mathbb{E}}_{X_1,X_2}[|X_1 - X_2|]$$

where X_1, X_2 are two independent samples of X.

Small-Set Expansion Hypothesis We recall the definition of Small-Set Expansion Hypothesis.

Problem 8.2.6 (Small Set Expansion (γ, δ))**.** Given a regular graph $G = (V, E)$, distinguish between the following two cases:

Yes: There exists a non-expanding set $S \subset V$ with

$$\mu(S) = \delta \qquad \text{and} \qquad \Phi_G(S) \leqslant \gamma\,.$$

No: All sets $S \subset V$ with $\mu(S) = \delta$ are highly expanding having

$$\Phi_G(S) \geqslant 1 - \gamma\,.$$

Hypothesis 8.2.7 (Hardness of approximating Small Set Expansion)**.** *For all $\gamma > 0$, there exists $\delta > 0$ such that the promise problem* Small Set Expansion *(γ, δ) is* NP*-hard.*

For the proofs, we will use the following version of the Small Set Expansion problem, in which we high expansion is guaranteed not only for sets of measure δ, but also within an arbitrary multiplicative factor of δ.

Problem 8.2.8 (Small Set Expansion (γ, δ, M))**.** Given a regular graph $G = (V, E)$, distinguish between the following two cases:

Yes: There exists a non-expanding set $S \subset V$ with

$$\mu(S) = \delta \qquad \text{and} \qquad \Phi_G(S) \leqslant \gamma\,.$$

No: All sets $S \subset V$ with $\mu(S) \in \left(\frac{\delta}{M}, M\delta\right)$ have

$$\Phi_G(S) \geqslant 1 - \gamma\,.$$

The following stronger hypothesis was shown to be equivalent to Small-Set Expansion Hypothesis in [88].

Hypothesis 8.2.9 (Hardness of approximating Small Set Expansion)**.** *For all $\gamma > 0$ and $M \geqslant 1$, there exists $\delta > 0$ such that the promise problem* Small Set Expansion *(γ, δ, M) is* NP*-hard.*

8.3 Reduction between Vertex Expansion and Symmetric Vertex Expansion

In this section we show that the computation of the Vertex Expansion is essentially equivalent to the computation of Symmetric Vertex Expansion. Formally, we prove the following theorems.

Theorem 8.3.1. *Given a graph $G = (V, E)$, there exists a graph H such that* $\max_{i \in V(H)} d_i \leqslant \left(\max_{i \in V(G)} d_i\right)^2 + \max_{i \in V(G)} d_i$

$$\Phi^{\mathsf{V}}(G) \leqslant \phi^{\mathsf{V}}(H) \leqslant \frac{\Phi^{\mathsf{V}}(G)}{1 - \Phi^{\mathsf{V}}(G)}\,.$$

Proof. Let G^2 denote the graph on $V(G)$ that corresponds to two hops in the graph G. Formally,

$$\{u, v\} \in E(G^2) \iff \exists w \in V(G), (u, w) \in E(G) \text{ and } (w, v) \in E(G)\,.$$

Let $H = G \cup G^2$, i.e., $V(H) = V(G)$ and $E(H) = E(G) \cup E(G^2)$.

Let $S \subset V(G)$ be a set with small symmetric vertex expansion $\Phi^{\mathsf{V}}(S) = \varepsilon$. Let $S' = S - N_G(\bar{S})$ be the set of vertices obtained from S by deleting it's internal boundary. It is easy to see that

$$N_H(S') = N_G(S) \cup N_G(\bar{S}) .$$

Moreover, since $N_G(\bar{S}) \leqslant \Phi^{\mathsf{V}}(S) w(S)$ we have $w(S') \geqslant w(S)(1 - \Phi_G^{\mathsf{V}}(S))$. Hence the vertex expansion of the set S' is upper-bounded by,

$$\phi_H^{\mathsf{V}}(S') \leqslant \frac{\Phi_G^{\mathsf{V}}(S)}{1 - \Phi_G^{\mathsf{V}}(S)} .$$

Conversely, suppose $T \subset V(H)$ be a set with small vertex expansion $\phi_H^{\mathsf{V}}(T) = \varepsilon$. Consider the set $T' = T \cup N_G(T)$. Observe that the internal boundary of T' in the graph G is given by $N_G(\bar{T}') = N_G(T)$. Further the external boundary of T' is given by $N_G(T') = N_G(N_G(T)) = N_{G^2}(T)$. Therefore, we have

$$N_G(T') \cup N_G(\bar{T}') = N_G(T) \cup N_{G^2}(T) = N_H(T) .$$

Further since $w(T') \geqslant w(T)$, we have $\Phi_G^{\mathsf{V}}(T') \leqslant \phi_H^{\mathsf{V}}(T)$.

This completes the proof of the Theorem.

□

Theorem 8.3.2. *Given a graph G, there exists a graph G' such that*

$$\max_{i \in V(G)} d_i = \max_{i \in V(G')} d_i \qquad \text{and} \qquad \phi^{\mathsf{V}}(G) = \Theta(\Phi^{\mathsf{V}}(G')) .$$

Moreover, such a G' can be computed in time polynomial in the size of G.

Proof. Given graph G, we construct G' as follows. We start with $V(G') = V(G) \cup E(G)$, i.e., G' has a vertex for each vertex in G and for each edge in G. For each edge $\{u, v\} \in E(G)$, we add edges $\{u, \{u, v\}\}$ and $\{v, \{u, v\}\}$ in G'. For a vertex $i \in V(G) \cap V(G')$, we set its weight to be $w(i)$. For a vertex $\{u, v\} \in E(G) \cap V(G')$, we set its weight to be $\min \{w(u)/d_u, w(v)/d_v\}$.

It is easy to see that G' can be computed in time polynomial in the size of G, and that $\max_{i \in V(G)} d_i = \max_{i \in V(G')} d_i$.

We first show that $\phi^{\mathsf{V}}(G) \geqslant \Phi^{\mathsf{V}}(G')/2$. Let $S \subset V(G)$ be the set having the least vertex expansion in G. Let

$$S' = S \cup \{\{u,v\} \mid \{u,v\} \in E(G) \text{ and } u \in S \text{ or } v \in S\} \, .$$

By construction, we have $w(S) \leqslant w(S')$, $N_G(S) = N_{G'}(S')$ and

$$w(N_{G'}(\bar{S}')) \leqslant \sum_{u \in N_{G'}(S')} d_u \frac{w(u)}{d_u} \leqslant w(N_{G'}(S')) \, .$$

Therefore,

$$\Phi^{\mathsf{V}}(G') \leqslant \Phi^{\mathsf{V}}_{G'}(S') = \frac{w(N_{G'}(S')) + w(N_{G'}(\bar{S}'))}{w(S')} \leqslant \frac{2w(N_G(S))}{w(S)} = 2\phi^{\mathsf{V}}_G(S) = 2\phi^{\mathsf{V}}(G) \, .$$

Now, let $S' \subset V(G')$ be the set having the least value of $\Phi^{\mathsf{V}}_{G'}(S')$ and let $\varepsilon = \Phi^{\mathsf{V}}_{G'}(S')$. We construct the set S as follows. We let $S_1 = S' \backslash N_{G'}(\bar{S}')$, i.e. we obtain S_1 from S' by deleting it's internal boundary. Next we set $S = S_1 \cap V(G)$. More formally, we let S be the following set.

$$S = \left\{ v \in S' \cap V(G) | v \notin N_{G'}(\bar{S}') \right\} \, .$$

By construction, we get that $N_G(S) \subseteq N_{G'}(S') \cup N_{G'}(\bar{S}')$. Now, the internal boundary of S' has weight at most $\varepsilon w(S')$. Therefore, we have

$$w(S_1) \geqslant (1 - \varepsilon) w(S') \, .$$

We need a lower bound on the weight of the set S we constructed. To this end, we make the following observation. For each vertex $\{u,v\} \in S_1 \cap E(G)$, u or v also has to be in S_1 (If not, then deleting $\{u,v\}$ from S' will result in a decrease in the vertex expansion thereby contradicting the optimality of the choice of the set S'). Therefore, we have the following

$$\sum_{\{u,v\} \in S_1 \cap E(G)} w(\{u,v\}) = \sum_{\{u,v\} \in S_1 \cap E(G)} \min \left\{ \frac{w(u)}{d_u}, \frac{w(u)}{d_u} \right\} \leqslant \sum_{u \in S_1 \cap V(G)} w(u) = w(S) \, .$$

Therefore,

$$w(S) \geqslant \frac{w(S_1)}{2} \geqslant (1-\varepsilon)\frac{w(S')}{2}$$

Therefore, we have

$$\phi^{\mathsf{V}}(G) \leqslant \phi^{\mathsf{V}}_G(S) = \frac{w(N_G(S))}{w(S)} \leqslant \frac{w(N_{G'}(S') \cup N_{G'}(\bar{S}'))}{(1-\varepsilon)w(S')/2} = 4\Phi^{\mathsf{V}}_{G'}(S') = 4\Phi^{\mathsf{V}}(G')\,.$$

Putting these two together, we have

$$\frac{\phi^{\mathsf{V}}(G)}{2} \leqslant \Phi^{\mathsf{V}}(G') \leqslant 4\phi^{\mathsf{V}}(G)\,.$$

□

8.4 Isoperimetry of the Gaussian Graph

In this section we bound the Balanced Analytic Vertex Expansion of the Gaussian graph. For the Gaussian Graph, we define the canonical probability distribution on V^{d+1} as follows. The marginal distribution along any component X or Y_i is the standard Gaussian distribution in $\mathbb{R}^n$, denoted here by $\mu = \mathcal{N}(0,1)^n$.

$$\mathcal{P}_{\mathcal{G}_{\Lambda,\Sigma}}(X, Y_1, \ldots, Y_d) = \frac{\Pi_{i=1}^d w(X, Y_i)}{\mu(X)^{d-1}} = \mu(X)\Pi_{i=1}^d\, \mathbb{P}\left[Y = Y_i\right].$$

Here, random variable Y is sampled from $\mathcal{N}(\Lambda X, \Sigma)$.

Theorem 8.4.1. *For any closed set $S \subset of V(\mathcal{G}_{\Lambda,\Sigma})$ with Λ a diagonal matrix satisfying $\|\Lambda\| \leqslant 1$, and Σ a diagonal matrix satisfying $\Sigma \succeq \varepsilon I$, we have*

$$\frac{\mathbb{E}_{(X,Y_1,\ldots,Y_d)\sim\mathcal{P}_{\mathcal{G}_{\Lambda,\Sigma}}} \max_i \left|\mathbb{I}_S\left[X\right] - \mathbb{I}_S\left[Y_i\right]\right|}{\mathbb{E}_{X,Y\sim\mu}\left|\mathbb{I}_S\left[X\right] - \mathbb{I}_S\left[Y\right]\right|} = \frac{\mathbb{E}_{X\sim\mu}\,\mathbb{E}_{Y_1,\ldots Y_d\sim\mathcal{N}(\Lambda X,\Sigma)} \max_i \left|\mathbb{I}_S\left[X\right] - \mathbb{I}_S\left[Y_i\right]\right|}{\mathbb{E}_{X,Y\sim\mu}\left|\mathbb{I}_S\left[X\right] - \mathbb{I}_S\left[Y\right]\right|} \geqslant c\sqrt{\varepsilon \log d}$$

for some absolute constant c.

Lemma 8.4.2. *Let $u, v \in \mathbb{R}^n$ satisfy $|u - v| \leqslant \sqrt{\varepsilon \log d}$. Let Λ be a diagonal matrix satisfying $\|\Lambda\| \leqslant 1$, and let Σ a diagonal matrix satisfying $\Sigma \succeq \varepsilon I$. Let P_u, P_v be the distributions $\mathcal{N}(\Lambda u, \Sigma)$ and $\mathcal{N}(\Lambda v, \Sigma)$ respectively. Then,*

$$d_{\mathsf{TV}}(P_u, P_v) \leqslant 1 - \frac{1}{d}.$$

Proof. First, we note that that for the purpose of estimating their total variation distance, we can view P_u, P_v as one-dimensional Gaussian random variables along the line $\Lambda u - \Lambda v$. Since $\|\Lambda\| \leqslant 1$,

$$\|\Lambda u - \Lambda v\| \leqslant \|u - v\| \leqslant \sqrt{\varepsilon \log d}\,.$$

W.l.o.g., we may take $\Lambda u = 0$ and $\Lambda v = \sqrt{\varepsilon \log d}$. Next, by the definition of total variation distance,

$$\begin{aligned}
d_{\mathsf{TV}}(P_u, P_v) &= \int_{x: P_v(x) \geqslant P_u(x)} |P_v(x) - P_u(x)| dx \\
&= \int_{\Lambda v/2}^{\infty} (P_v(x) - P_u(x)) dx \\
&= \frac{1}{\sqrt{2\pi\varepsilon}} \int_{\Lambda v/2}^{\infty} e^{-\frac{\|x - \Lambda v\|^2}{2\varepsilon}}\, dx - \frac{1}{\sqrt{2\pi\varepsilon}} \int_{\Lambda v/2}^{\infty} e^{-\frac{\|x\|^2}{2\varepsilon}}\, dx \\
&= \frac{1}{\sqrt{2\pi\varepsilon}} \int_{-\Lambda v/2}^{\Lambda v/2} e^{-\frac{\|x\|^2}{2\varepsilon}}\, dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{\log d}/2}^{\sqrt{\log d}/2} e^{-\frac{\|x\|^2}{2}}\, dx \\
&= 1 - 2 \cdot \frac{1}{\sqrt{2\pi}} \int_{\sqrt{\log d}/2}^{\infty} e^{-\frac{\|x\|^2}{2}}\, dx \\
&< 1 - \frac{1}{d}.
\end{aligned}$$

where the last step uses a standard bound on the Gaussian tail.

□

Proof of Theorem 8.4.1. Let μ_X denote the Gaussian distribution $\mathcal{N}(\Lambda X, \Sigma)$. Then the LHS is:

$$\int_{\mathbb{R}^n \setminus S} \left(1 - (1 - \mu_X(S))^d\right)\, d\mu(X) + \int_S \left(1 - (1 - \mu_X(\mathbb{R}^n \setminus S))^d\right)\, d\mu(X).$$

To bound this, we will restrict ourselves to points X for which the μ_X measure of the complementary set is at least $1/d$. Roughly speaking, these will be points near the boundary of S. Define:

$$S_1 = \left\{ x \in S \,:\, \mu_X(\mathbb{R}^n \setminus S) < \frac{1}{2d} \right\},\ S_2 = \left\{ x \in \mathbb{R}^n \setminus S \,:\, \mu_X(S) < \frac{1}{2d} \right\}$$

and

$$S_3 = \mathbb{R}^n \setminus S_1 \setminus S_2.$$

For $u \in \mathbb{R}^n$, let P_u be the distribution $\mathcal{N}(\Lambda u, \Sigma)$. For any $u \in S_1, v \in S_2$, we have

$$d_{\mathsf{TV}}(P_u, P_v) > 1 - \frac{1}{2d} - \frac{1}{2d} = 1 - \frac{1}{d}.$$

Therefore, by Lemma 8.4.2, $\|u - v\| > \sqrt{\varepsilon \log d}$, i.e., $d(S_1, S_2) > \sqrt{\varepsilon \log d}$. Next we bound the measure of S_3. We can assume w.l.o.g. that $\mu(S) \leqslant \mu(\mathbb{R}^n \setminus S)$ and $\mu(S_1) \geqslant \mu(S)/2$ (else $\mu(S_3) \geqslant \mu(S)/2$ and we are done). Applying the isoperimetric inequality for Gaussian space [22, 100], for subsets at this distance,

$$\mu(S_3) \geqslant \sqrt{\frac{2}{\pi}} \sqrt{\varepsilon \log d} \cdot \mu(S_1)\mu(S_2) \geqslant \sqrt{\frac{\varepsilon \log d}{2\pi}} \cdot \mu(S)\mu(\mathbb{R}^n \setminus S).$$

We are now ready to complete the proof.

$$\begin{aligned}
& \frac{1}{2}\left(\int_{\mathbb{R}^n \setminus S} (1 - (1 - \mu_X(S))^d) \, d\mu(X) + \int_S (1 - (1 - \mu_X(\mathbb{R}^n \setminus S)) \, d\mu(X)\right) \\
\geqslant \; & \frac{1}{2}\left(\int_{X \in \mathbb{R}^n \setminus S, \mu_X(S) \geqslant 1/d} (1 - (1 - \mu_X(S))^d) \, d\mu(X)\right. \\
& \quad \left. + \int_{X \in S, \mu_X(\mathbb{R}^n \setminus S) \geqslant 1/d} (1 - (1 - \mu_X(\mathbb{R}^n \setminus S)) \, d\mu(X)\right) \\
\geqslant \; & \frac{e-1}{2e}\left(\int_{X \in \mathbb{R}^n \setminus S, \mu_X(S) \geqslant 1/d} d\mu(X) + \int_{X \in S, \mu_X(\mathbb{R}^n \setminus X) \geqslant 1/d} d\mu(X)\right) \\
\geqslant \; & \frac{e-1}{2e} \mu(S_3) \\
\geqslant \; & c\sqrt{\varepsilon \log d} \cdot \mu(S)\mu(\mathbb{R}^n \setminus S).
\end{aligned}$$

□

We prove the following Theorem which helps us to bound the isoperimetry of the Gaussian graph for over all functions over the range $[0, 1]$.

Theorem 8.4.3. *Given an instance $(V, \mathcal{P})$ and a function $F : V \to [0, 1]$, there exists a function $F' : V \to \{0, 1\}$, such that*

$$\frac{\mathbb{E}_{(X, Y_1, \ldots, Y_d) \sim \mathcal{P}} \max_i |F(X) - F(Y_i)|}{\mathbb{E}_{X, Y \sim \mu} |F(X) - F(Y)|} \geqslant \frac{\mathbb{E}_{(X, Y_1, \ldots, Y_d) \sim \mathcal{P}} \max_i |F'(X) - F'(Y_i)|}{\mathbb{E}_{X, Y \sim \mu} |F'(X) - F'(Y)|}$$

Proof. For every $r \in [0,1]$, we define $F_r : V \to \{0,1\}$ as follows.

$$F_r(X) = \begin{cases} 1 & F(X) \geqslant r \\ 0 & F(X) < r \end{cases}$$

Clearly,

$$F(X) = \int_0^1 F_r(X) dr \,.$$

Now, observe that if $F(X) - F(Y) \geqslant 0$ then $F_r(X) - F_r(Y) \geqslant 0 \; \forall r \in [0,1]$ and similarly, if $F(X) - F(Y) < 0$ then $F_r(X) - F_r(Y) \leqslant 0 \; \forall r \in [0,1]$. Therefore,

$$|F(X) - F(Y)| = \left| \int_0^1 (F_r(X) - F_r(Y))\, dr \right| = \int_0^1 |F_r(X) - F_r(Y)|\, dr \,.$$

Also, observe that if $|F(X) - F(Y_1)| \geqslant |F(Y_i) - F(X)|$ then

$$|F_r(X) - F_r(Y_1)| \geqslant |F_r(Y_i) - F_r(X)| \;\; \forall r \in [0,1]$$

Therefore,

$$\begin{aligned} \frac{\mathbb{E}_{(X,Y_1,\ldots,Y_d)\sim\mathcal{P}} \max_i |F(X) - F(Y_i)|}{\mathbb{E}_{X,Y\sim\mu} |F(X) - F(Y)|} &= \frac{\mathbb{E}_{(X,Y_1,\ldots,Y_d)\sim\mathcal{P}} \max_i \int_0^1 |F_r(X) - F_r(Y_i)|\, dr}{\mathbb{E}_{X,Y\sim\mu} \int_0^1 |F_r(X) - F_r(Y)|\, dr} \\ &= \frac{\int_0^1 \left(\mathbb{E}_{(X,Y_1,\ldots,Y_d)\sim\mathcal{P}} \max_i |F_r(X) - F_r(Y_i)|\right) dr}{\int_0^1 \left(\mathbb{E}_{X,Y\sim\mu} |F_r(X) - F_r(Y)|\right) dr} \\ &\geqslant \min_{r\in[0,1]} \frac{\mathbb{E}_{(X,Y_1,\ldots,Y_d)\sim\mathcal{P}} \max_i |F_r(X) - F_r(Y_i)|}{\mathbb{E}_{X,Y\sim\mu} |F_r(X) - F_r(Y)|} \end{aligned}$$

Let r' be the value of r which minimizes the expression above. Taking F' to be $F_{r'}$ finishes the proof.

□

Corollary 8.4.4 (Corollary to Theorem 8.4.1 and Theorem 8.4.3)**.** *Let* $F : V(\mathcal{G}_{\Lambda,\Sigma}) \to [0,1]$ *be any function. Then, for some absolute constant* c*,*

$$\frac{\mathbb{E}_{(X,Y_1,\ldots,Y_d)\sim\mathcal{P}_{\mathcal{G}_{\Lambda,\Sigma}}} \max_i |F(X) - F(Y_i)|}{\mathbb{E}_{X,Y\sim\mu} |F(X) - F(Y)|} \geqslant c\sqrt{\varepsilon \log d}\,.$$

8.5 Dictatorship Testing Gadget

In this section we initiate the construction of the dictatorship testing gadget for reduction from SSE.

The dictatorship testing gadget is obtained by picking an appropriately chosen constant sized Markov-chain H, and considering the product Markov chain H^R. Given a Markov chain H, define an instance of Balanced Analytic Vertex Expansion with vertices as V_H and the constraints given by the following canonical probability distribution over V_H^{d+1}.

- Sample $X \sim \mu_H$, the stationary distribution of the Markov chain V_H.
- Sample $Y_1, \ldots, Y_d$ independently from the neighbours of X in V_H

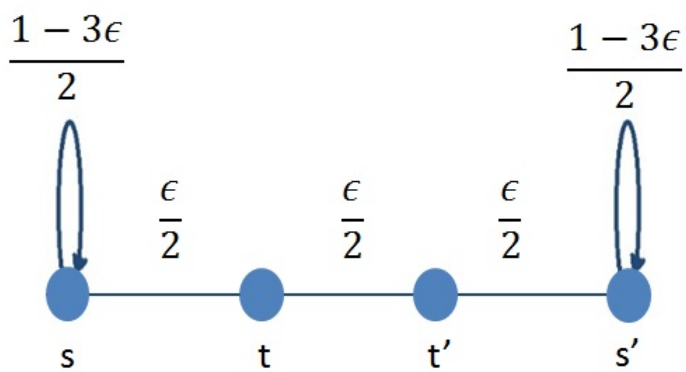

Figure 23: Gadget for the reduction

For our application, we use a specific Markov chain H on four vertices. Define a Markov chain H on $V_H = \{s, t, t', s'\}$ as follows (Figure 23),$p(s|s) = p(s'|s') = 1 - \frac{\epsilon}{1-2\epsilon}$, $p(t|s) = p(t'|s') = \frac{\epsilon}{1-2\epsilon}$, $p(s|t) = p(s'|t') = \frac{1}{2}$ and $p(t'|t) = p(t|t') = \frac{1}{2}$. It is easy to see that the stationary distribution of the Markov chain H over V_H is given by,

$$\mu_H(s) = \mu_H(s') = \frac{1}{2} - \varepsilon \qquad\qquad \mu_H(t) = \mu_H(t') = \varepsilon$$

From this Markov chain, construct a dictatorship testing gadget $(V_H^R, \mathcal{P}_H^R)$ as described above. We begin by showing that this dictatorship testing gadget has small vertex separators corresponding to dictator functions.

Proposition 8.5.1 (Completeness)**.** *For each $i \in [R]$, the i^{th}-dictator set defined as $F(x) = 1$ if $x_i \in \{s,t\}$ and 0 otherwise satisfies,*

$$\underset{1}{\mathsf{Var}}[F] = \frac{1}{2} \qquad \text{and} \qquad \mathsf{VAL}_{\mathcal{P}_{H^R}}(F) \leqslant 2\epsilon$$

Proof. Clearly,

$$\mathop{\mathbb{E}}_{X,Y \sim \mu_H} |F(X) - F(Y)| = \frac{1}{2}$$

Observe that for any choice of $(X, Y_1, \ldots, Y_d) \sim \mathcal{P}_{H^R}$, $\max_i |F(X) - F(Y_i)|$ is non-zero if and only if either $x_i = t$ or $x_i = t'$. Therefore we have,

$$\mathop{\mathbb{E}}_{(X,Y_1,\ldots,Y_d)\sim\mathcal{P}_H} \max_i |F(X) - F(Y_i)| \leqslant \mathop{\mathbb{P}}_{[}[x]_i \in \{t, t'\}]) = 2\varepsilon\,,$$

which concludes the proof. □

8.5.1 Soundness

We will show a general soundness claim that holds for the dictatorship testing gadgets $\big(V(H^R), \mathcal{P}_{H^R}\big)$ constructed out of arbitrary Markov chains H with a given spectral gap. Towards formally stating the soundness claim, we recall some background and notation about polynomials over the product Markov chain H^R.

8.5.2 Polynomials over H^R

In this section, we recall how functions over the product Markov chain H^R can be written as multi-linear polynomials over the eigenfunctions of H.

Let $e_0, e_1, \ldots, e_n : V(H) \to \mathbb{R}$ be an orthonormal basis of eigenvectors of H and let $\lambda_0, \ldots, \lambda_n$ be the corresponding eigenvalues. Here $e_0 = 1$ is the constant function whose eigenvalue $\lambda_0 = 1$. Clearly $e_0, \ldots, e_n$ form an orthonormal basis for the vector space of functions from $V(H)$ to $\mathbb{R}$.

It is easy to see that the eigenvectors of the product chain H^R are given by products of $e_0, \ldots, e_n$. Specifically, the eigenvectors of H^R are indexed by $\sigma \in [n]^R$ as follows,

$$e_\sigma(x) = \prod_{i=1}^{R} e_{\sigma_i}(x_i)$$

Every function $f : H^R \to \mathbb{R}$ can be written in this orthonormal basis $f(x) = \sum_{\sigma\in[n]^R} \hat{f}_\sigma e_\sigma(x)$. For a multi-index $\sigma \in [n]^R$, the function e_σ is a monomial of degree $|\sigma| = |\{i|\sigma_i \neq 0\}|$.

For a polynomial $Q = \sum_\sigma \hat{Q}_\sigma e_\sigma$, the polynomial $Q^{>p}$ denotes the projection on to degrees higher than p, i.e., $Q^{>p} = \sum_{\sigma,|\sigma|>p} \hat{Q}_\sigma e_\sigma$. The influences of a polynomial $Q = \sum_\sigma \hat{Q}_\sigma$ are defined as,

$$\mathsf{Inf_i(Q)} = \sum_{\sigma:\sigma_i\neq 0} \hat{\mathsf{Q}}_\sigma^2$$

The above notions can be naturally extended to vectors of multi-linear polynomials $Q = (Q_0, Q_1, \ldots, Q_d)$.

Note that every real-valued function on the vertices $V(H)$ of a Markov chain H can be thought of as a random variable. For each $i > 0$, the random variable $e_i(x)$ has mean zero and variance 1. The same holds for all $e_\sigma(x)$ for all $|\sigma| \neq 0$. For a function $Q : V(H^R) \to \mathbb{R}$ (or equivalently a polynomial), $\mathsf{Var}[Q]$ denotes the variance of the random variable $Q(x)$ for a random x from stationary distribution of H^R. It is an easy computation to check that this is given by,

$$\mathsf{Var}[Q] = \sum_{\sigma:|\sigma|\neq 0} \hat{Q}_\sigma^2$$

We will make use of the following Invariance Principle due to Isaksson and Mossel [49].

Theorem 8.5.2 ([49]). *Let $X = (X_1, \ldots, X_n)$ be an independent sequence of ensembles, such that $\mathbb{P}[X_i = x] \geqslant \alpha > 0, \forall i, x$. Let Q be a d-dimensional multi-linear polynomial such that*

$$\mathsf{Var}[Q_j(X)] \leqslant 1 \qquad \mathsf{Var}\left[Q_j^{>p}\right] \leqslant (1-\varepsilon\eta)^{2p} \qquad \text{and} \qquad \mathsf{Inf_i(Q_j)} \leqslant \tau$$

where $p = \frac{1}{18}\log(1/\tau)/\log(1/\alpha)$. Finally, let $\psi : \mathbb{R}^k \to \mathbb{R}$ be Lipschitz continuous. Then,

$$|\mathbb{E}[\psi(Q(X))] - \mathbb{E}[\psi(Q(Z))]| = \mathcal{O}\left(\tau^{\frac{\varepsilon\eta}{18}/\log\frac{1}{\alpha}}\right)$$

where Z is an independent sequence of Gaussian ensembles with the same covariance structure as X.

8.5.3 Noise Operator

We define a noise operator $\Gamma_{1-\eta}$ on functions on the Markov chain H as follows :

$$\Gamma_{1-\eta}F(X) \stackrel{\text{def}}{=} (1-\eta)F(X) + \eta \mathop{\mathbb{E}}_{Y\sim X} F(Y)$$

for every function $F : H \to \mathbb{R}$. Similarly, one can define the noise operator $\Gamma_{1-\eta}$ on functions over H^R.

Applying the noise operator $\Gamma_{1-\eta}$ on a function F, smoothens the function or makes it closer to a low-degree polynomial. This resulting function $\Gamma_{1-\eta}F$ is close to a *low-degree polynomial*, and therefore is amenable to applying an invariance principle. Formally, one can show the following decay of coefficients of high degree for $\Gamma_{1-\eta}F$.

Lemma 8.5.3. *(Decay of High degree Coefficients) Let Q_j be the multi-linear polynomial representation of $\Gamma_{1-\eta}F(X)$, and let ε be the spectral gap of the Markov chain H. Then,*

$$\mathsf{Var}\left[Q_j^{>p}\right] \leqslant (1-\varepsilon\eta)^{2p}$$

Proof. The Fourier expansion of the function F is $F = \sum_\sigma \hat{f}_\sigma e_\sigma$ where $\{e_\sigma\}$ is the set of eigenvectors of H^k. It is easy to see that

$$e_\sigma = e_{\sigma_1} \otimes \ldots \otimes e_{\sigma_k},$$

where the $\{e_{\sigma_i}\}$ are the eigenvectors of H.

$$\begin{aligned}
\Gamma_{1-\eta}F(X) &= (1-\eta)F(X) + \eta \mathop{\mathbb{E}}_{Y\sim X} F(Y) \\
&= \sum_\sigma \hat{f}_\sigma \mathbb{E}\left[e_\sigma(X) + \mathop{\mathbb{E}}_{Y\sim X} F(Y)\right] \\
&= \sum_\sigma \hat{f}_\sigma \Pi_{i\in\sigma}\left((1-\eta)e_{\sigma_i}(X_i) + \mathop{\mathbb{E}}_{Y_i\sim X_i} e_{\sigma_i}(Y_i)\right)
\end{aligned}$$

We bound the second moment of $\Gamma_{1-\eta}F$ as follows

$$
\begin{aligned}
\mathop{\mathbb{E}}_X \left(\Gamma_{1-\eta}F(X)\right)^2 &= \sum_\sigma \hat{f}_\sigma^2 \mathop{\mathbb{E}}_X \Pi_{i\in\sigma} \left((1-\eta)e_{\sigma_i}(X_i) + \eta \mathop{\mathbb{E}}_{Y_i\sim X_i} e_{\sigma_i}(Y_i)\right)^2 \\
&= \sum_\sigma \hat{f}_\sigma^2 \Pi_{i\in\sigma} \left((1-\eta)^2 \mathop{\mathbb{E}}_{X_i} e_{\sigma_i}(X_i)^2 + \eta^2 \mathop{\mathbb{E}}_{X_i} \left(\mathop{\mathbb{E}}_{Y_i\sim X_i} e_{\sigma_i}(Y_i)\right)^2 \right. \\
&\qquad \left. +2\eta(1-\eta) \mathop{\mathbb{E}}_{X_i} \mathop{\mathbb{E}}_{Y_i\sim X_i} e_{\sigma_i}(X_i)e_{\sigma_i}(Y_i)\right)^2 \\
&= \sum_\sigma \hat{f}_\sigma^2 \Pi_{i\in\sigma} \left((1-\eta)^2 + \eta^2\lambda_i^2 + 2\eta(1-\eta)\lambda_i\right) \\
&= \sum_\sigma \hat{f}_\sigma^2 \Pi_{i\in\sigma} \left(1-\eta+\eta\lambda_i\right)^2
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathsf{Var}\left[Q_j^{>p}\right] &\leqslant 4 \sum_{\sigma:|\sigma|>p} \hat{f}_\sigma^2 \Pi_{i\in\sigma} \left(1-\eta+\eta\lambda_i\right)^2 \\
&\leqslant \sum_{\sigma:|\sigma|>p} \hat{f}_\sigma^2 \left(1-\varepsilon\eta\right)^{2|\sigma|} \\
&\leqslant \left(1-\varepsilon\eta\right)^{2p}
\end{aligned}
$$

Here the second inequality follows from the fact that all non-trivial eigenvalues of H are at most $1-\varepsilon$ and the third inequality follows Parseval's indentity. □

Furthermore, on applying the noise operator $\Gamma_{1-\eta}$, the resulting function $\Gamma_{1-\eta}F$ can have a bounded number of influential coordinates as shown by the following lemma.

Lemma 8.5.4. *(Sum of Influences Lemma) If the spectral gap of a Markov chain is at least ε then for any function $F : V_H^R \to \mathbb{R}$,*

$$\sum_{i\in[R]} \mathsf{Inf}_\mathsf{i}(\Gamma_{1-\eta}\mathsf{F}) \leqslant \frac{1}{\eta\epsilon} \mathsf{Var}\left[\mathsf{F}\right]$$

Proof. By suitable normalization, we may assume without loss of generality that $\mathsf{Var}[F] = 1$. If Q denotes the multi-linear representation of $\Gamma_{1-\eta}F$, then the sum of influences can be written as,

$$\sum_{i\in[R]} \mathsf{Inf_i}(\Gamma_{1-\eta}\mathsf{F}) \leqslant \sum_{|\sigma|\neq 0} |\sigma|\hat{Q}_\sigma^2$$
$$\leqslant \sum_{|\sigma|\neq 0} |\sigma|(1-\eta\epsilon)^{2|\sigma|}\hat{F}_\sigma^2$$
$$\leqslant \left(\max_{k\in\mathbb{Z}_{\geqslant 0}} k(1-\eta\epsilon)^{2k}\right)\sum_{|\sigma|\neq 0} \hat{F}_\sigma^2 < \frac{1}{\eta\epsilon}$$

where we used the fact that the function $h(t) = t(1-\eta\epsilon)^{2t}$ achieves its maximum value at $t = -\frac{1}{2}\ln(1-\eta\epsilon)$. □

8.5.4 Soundness Claim

Now we are ready to formally state our soundness claim for a dictatorship test gadget constructed out of a Markov chain.

Proposition 8.5.5 (Soundness)**.** *For all $\epsilon, \eta, \alpha, \tau > 0$ the following holds. Let H be a finite Markov-chain with a spectral gap of at least ε, and the probability of every state under stationary distribution is $\geqslant \alpha$. Let $F : V(H^R) \to \{0,1\}$ be a function such that* $\max_{i\in[R]} \mathsf{Inf_i}(\Gamma_{1-\eta}\mathsf{F}) \leqslant \tau$. *Then we have*

$$\mathop{\mathbb{E}}_{(X,Y_1,\ldots,Y_d)\sim\mathcal{P}_{H^R}}[\max_i |F(Y_i) - F(X)|]$$
$$\geqslant \Omega(\sqrt{\varepsilon\log d}) \mathop{\mathbb{E}}_{X,Y\sim\mu_{H^R}} |F(X) - F(Y)| - O(\eta) - \tau^{\Omega(\varepsilon\eta/\log(1/\alpha))}\,.$$

For the sake of brevity, we define $\mathsf{soundness}(V(H^R), \mathcal{P}_{H^R})$ to be the following :

Definition 8.5.6.

$$\mathsf{soundness}(V(H^R), \mathcal{P}_{H^R}) \stackrel{\text{def}}{=} \min_{F:\max_{i\in[R]}\mathsf{Inf_i(F)}\leqslant\tau} \frac{\mathbb{E}_{(X,Y_1,\ldots,Y_d)\sim\mathcal{P}_{H^R}}[\max_i |F(Y_i) - F(X)|]}{\mathbb{E}_{X,Y\sim\mu_{H^R})} |F(X) - F(Y)|}$$

In the rest of the section, we will present a proof of Proposition 8.5.5. First, we construct Gaussian random variables with moments matching the eigenvectors of the chain H.

Gaussian Ensembles. Let $Q = (Q_0, Q_1, \ldots, Q_d)$ be the multi-linear polynomial representation of the vector-valued function $(\Gamma_{1-\eta}F(X), \Gamma_{1-\eta}F(Y_1), \ldots, \Gamma_{1-\eta}F(Y_d))$. Let E denote the ensemble of nd random variables

$$(e_0(X), e_1(X), \ldots, e_n(X)), (e_0(Y_1), \ldots, e_n(Y_1)), \ldots, (e_0(Y_d), \ldots, e_n(Y_d)) .$$

Let $E_1, \ldots, E_R$ be R independent copies of the ensemble E. Clearly, the polynomial Q can be thought of as a polynomial over $E_1, \ldots, E_R$. For each random variable x in $E_1, \ldots, E_R$ and a value β in its support, $\mathbb{P}[x = \beta]$ is at least the minimum probability of a vertex in H under its stationary distribution.

This polynomial Q satisfies the requirements of Theorem 8.5.2 because on the one hand, the influences of F are $\leqslant \tau$ and on the other by Lemma 8.5.3, $\mathsf{Var}[Q^{\geqslant p}] \leqslant (1 - \varepsilon\eta)^{2p}$. Now we will apply the invariance principle to relate the soundness to the corresponding quantity on the Gaussian graph, and then appeal to the isoperimetric result on the Gaussian graph (Theorem 8.4.1).

The invariance principle translates the polynomial $(Q_0(X), Q_1(Y_1), \ldots Q_d(Y_d))$ on the sequence of independent ensembles $E_1, \ldots, E_R$, to a polynomial on a corresponding sequence of Gaussian ensembles with the same moments up to degree two.

Consider the ensemble E. For each $i \neq 0$,

$$\mathbb{E}[e_i(X)] = \mathbb{E}[e_i(Y_1) = 0] = \ldots \mathbb{E}[e_i(Y_d)] = 0 .$$

For each $i \neq j$, it is easy to see that,

$$\mathbb{E}[e_i(X)e_j(X)] = \mathbb{E}[e_i(Y_1)e_j(Y_1)] = \ldots \mathbb{E}[e_i(Y_d)e_j(Y_d] = 0 .$$

Moreover,

$$\mathbb{E}[e_i(X)e_j(Y_a)] = \mathbb{E}[e_i(Y_a)e_j(Y_b)] = 0$$

whenever $i \neq j$ and all $a, b \in \{1, \ldots d\}$. The only non-trivial correlations are $\mathbb{E}[e_i(X)e_i(Y_a)]$ and $\mathbb{E}[e_i(Y_a)e_i(Y_b)]$ for all $i \in [n]$ and $a, b \in [d]$. It is easy to check that

$$\mathbb{E}[e_i(X)e_i(Y_a)] = \lambda_i \qquad \mathbb{E}[e_i(Y_a)e_i(Y_b)] = \lambda_i^2$$

From the above discussion, we see that the Gaussian ensemble $z = (z_X, z_{Y_1}, \ldots, z_{Y_d})$ has the same covariance as the ensemble E.

1. Sample z_X and n-dimensional Gaussian random vector.

2. Sample $z_{Y_1}, \ldots, z_{Y_d} \in \mathbb{R}^n$ i.i.d as follows : The i^{th} coordinate of each z_{Y_a} is sampled from $\lambda_i z_X(i) + \sqrt{1-\lambda_i^2}\xi_{a,i}$ where $\xi_{a,i}$ is a Gaussian random variable independent of z_X and all other $\xi_{a,i}$.

Let $Z_X, Z_{Y_1}, \ldots, Z_{Y_d} \in \mathbb{R}^{nR}$ be the ensemble obtained by R independent samples from $z_X, z_{Y_1}, \ldots, z_{Y_d}$.

Let Σ denote the $nR \times nR$ diagonal matrix whose entries are $1-\lambda_1^2, \ldots, 1-\lambda_n^2$ repeated R times. Since the spectral gap of H is ε, we have that

$$1 - \lambda_i^2 \geqslant 2\varepsilon - \epsilon^2 > \epsilon$$

for all $i \in \{1, \ldots, n\}$. Therefore, we have $\Sigma \succeq \epsilon I$.

Proof of soundness. Now we return to the proof of the main soundness claim for the dictatorship testing gadget $(V(H^R), \mathcal{P}_{\mathcal{H}^R})$ constructed out an arbitrary Markov chain.

Proof of Proposition 8.5.5. Let $Q = (Q_0, Q_1, \ldots, Q_d)$ be the multi-linear polynomial representation of the vector-valued function $(\Gamma_{1-\eta}F(X), \Gamma_{1-\eta}F(Y_1), \ldots, \Gamma_{1-\eta}F(Y_d))$.

Define a function $s : \mathbb{R} \to \mathbb{R}$ as follows

$$s(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \in [0,1] \\ 1 & \text{if } x > 1 \end{cases}$$

Define a function $\Psi : \mathbb{R}^{d+1} \to \mathbb{R}$ as,

$$\Psi(x, y_1, \ldots, y_d) = \max_i |s(y_i) - s(x)| .$$

Clearly, Ψ is a Lipschitz function with a constant of 1.

Using the fact that F is bounded in $[0, 1]$,

$$\mathop{\mathbb{E}}_{(X,Y_1,\ldots,Y_d)\sim\mathcal{P}_{H^R}} \max_a |F(X) - F(Y_a)|$$

$$\geqslant \mathop{\mathbb{E}}_{(X,Y_1,\ldots,Y_d)\sim\mathcal{P}_{H^R}} \max_a |\Gamma_{1-\eta}F(X) - \Gamma_{1-\eta}F(Y_a)| - 2\eta \quad (69)$$

Furthermore, since $\Gamma_{1-\eta}F$ is also bounded in $[0, 1]$, we have $s(\Gamma_{1-\eta}F) = \Gamma_{1-\eta}F$. Therefore,

$$\mathop{\mathbb{E}}_{(X,Y_1,\ldots,Y_d)\sim\mathcal{P}_{H^R}} \max_a |\Gamma_{1-\eta}F(X) - \Gamma_{1-\eta}F(Y_a)|$$

$$= \mathop{\mathbb{E}}_{(X,Y_1,\ldots,Y_d)\sim\mathcal{P}_{H^R}} \max_a |s(\Gamma_{1-\eta}F(X)) - s(\Gamma_{1-\eta}F(Y_a))| \quad (70)$$

Apply the invariance principle to the polynomial $Q = (\Gamma_{1-\eta}F, \Gamma_{1-\eta}F, \ldots, \Gamma_{1-\eta}F)$ and Lipschitz function Ψ. By invariance principle Theorem 8.5.2, we get

$$\mathop{\mathbb{E}}_{(X,Y_1,\ldots,Y_d)\sim\mathcal{P}_{H^R}} \max_a |s(\Gamma_{1-\eta}F(X)) - s(\Gamma_{1-\eta}F(Y_a))|$$

$$\geqslant \mathop{\mathbb{E}}_{(Z_X,Z_{Y_1},\ldots,Z_{Y_d})\sim\mathcal{P}_{\mathcal{G}_{\Lambda,\Sigma}}} \max_a |s(\Gamma_{1-\eta}F(Z_X)) - s(\Gamma_{1-\eta}F(Z_{Y_a}))| - \tau^{\Omega(\epsilon\eta/\log(1/\alpha))} \quad (71)$$

Observe that $s \circ (\Gamma_{1-\eta}F)$ is bounded in $[0, 1]$ even over the Gaussian space. Hence, by using the isoperimetric result on Gaussian graphs (Corollary 8.4.4), we know that

$$\mathop{\mathbb{E}}_{(Z_X,Z_{Y_1},\ldots,Z_{Y_d})\sim\mathcal{P}_{\mathcal{G}_{\Lambda,\Sigma}}} \max_a |s(\Gamma_{1-\eta}F(Z_X)) - s(\Gamma_{1-\eta}F(Z_{Y_a}))|$$

$$\geqslant c\sqrt{\varepsilon \log d} \mathop{\mathbb{E}}_{Z_X,Z_Y\sim\mu_{\mathcal{G}_{\Lambda,\Sigma}}} |s(\Gamma_{1-\eta}F(Z_X)) - s(\Gamma_{1-\eta}F(Z_Y))| \quad (72)$$

Now we apply the invariance principle on the polynomial $(\Gamma_{1-\eta}F, \Gamma_{1-\eta}F)$ and the functional $\Psi : \mathbb{R}^2 \to \mathbb{R}$ given by

$$\Psi(a, b) = |s(a) - s(b)|\,.$$

This yields,

$$\mathop{\mathbb{E}}_{Z_X,Z_Y\sim\mu_{\mathcal{G}_{\Lambda,\Sigma}}} |s(\Gamma_{1-\eta}F(Z_X)) - s(\Gamma_{1-\eta}F(Z_Y))|$$

$$\geqslant \mathop{\mathbb{E}}_{X,Y\sim\mu(H^R)} |s(\Gamma_{1-\eta}F(X)) - s(\Gamma_{1-\eta}F(Y))| - \tau^{\Omega(\epsilon\eta/\log(1/\alpha))} \quad (73)$$

Over H^R, the function $\Gamma_{1-\eta}F$ is bounded in $[0,1]$, which implies that $s(\Gamma_{1-\eta}F(X)) = \Gamma_{1-\eta}F(X)$ and $\Gamma_{1-\eta}F(X) \geqslant F(X) - \eta$.

$$\mathop{\mathbb{E}}_{X,Y\sim\mu(H^R)} |s(\Gamma_{1-\eta}F(X)) - s(\Gamma_{1-\eta}F(Y))| \geqslant \mathbb{E}_{X,Y\sim\mu(H^R)} |F(X) - F(Y)| - 2\eta \quad (74)$$

From equations (69) to (74) we get,

$$\mathop{\mathbb{E}}_{(X,Y_1,\ldots,Y_d)\sim\mathcal{P}_{H^R}} \max_a |F(X) - F(Y_a)| \geqslant \Omega(\sqrt{\epsilon \log d})\, \mathbb{E}_{X,Y\sim\mu(H^R)} |F(X) - F(Y)| \\ -4\eta - \tau^{\Omega(\epsilon\eta/\log(1/\alpha))}$$

□

8.6 *Hardness Reduction from* SSE

In this section we will present a reduction from Small Set Expansion problem to Balanced Analytic Vertex Expansion problem. Let $G = (V, E)$ be an instance of Small Set Expansion (γ, δ, M). Starting with the instance $G = (V, E)$ of Small Set Expansion(γ, δ, M), our reduction produces an instance $(\mathcal{V}', \mathcal{P}')$ of Balanced Analytic Vertex Expansion.

To describe our reduction, let us fix some notation. For a set A, let $A^{\{R\}}$ denote the set of all multi-sets with R elements from A. Let $G_\eta = (1-\eta)G + \eta K_V$ where K_V denotes the complete graph on the set of vertices V. For an integer R, define $G_\eta^{\otimes R}$ to be the product graph $G_\eta^{\otimes R} = (G_\eta)^R$.

Define a Markov chain H on $V_H = \{s, t, t', s'\}$ as follows ((Figure 23)),

$$p(s|s) = p(s'|s') = 1 - \frac{\epsilon}{1-2\epsilon}, \qquad p(t|s) = p(t'|s') = \frac{\epsilon}{1-2\epsilon},$$
$$p(s|t) = p(s'|t') = \frac{1}{2}, \qquad p(t'|t) = p(t|t') = \frac{1}{2}\,.$$

It is easy to see that the stationary distribution of the Markov chain H over V_H is given by,

$$\mu_H(s) = \mu_H(s') = \frac{1}{2} - \epsilon \qquad\qquad \mu_H(t) = \mu_H(t') = \epsilon$$

The reduction consists of two steps. First, we construct an "unfolded" instance $(\mathcal{V}, \mathcal{P})$ of the BALANCED ANALYTIC VERTEX EXPANSION, then we merge vertices of $(\mathcal{V}, \mathcal{P})$ to create the final output instance $(\mathcal{V}', \mathcal{P}')$. The details of the reduction are presented below (Figure 24).

Reduction
Input: A graph $G = (V, E)$ - an instance of SMALL SET EXPANSION(γ, δ, M).
Parameters: $R = \frac{1}{\delta}$, ϵ
Unfolded instance $(\mathcal{V}, \mathcal{P})$
Set $\mathcal{V} = (V \times V_H)^R$. The probability distribution μ on $\mathcal{V}$ is given by $(\mu_V \times \mu_H)^R$. The probability distribution $\mathcal{P}$ is given by the following sampling procedure.

1. Sample a random vertex $A \in V^R$.
2. Sample $d+1$ random neighbors $B, C_1, \ldots, C_d \sim G_\eta^{\otimes R}(A)$ of the vertex A in the tensor-product graph $G_\eta^{\otimes R}$.
3. Sample $x \in V_H^R$ from the product distribution μ^R.
4. Independently sample d neighbours $y^{(1)}, \ldots, y^{(d)}$ of x in the Markov chain H^R, i.e., $y^{(i)} \sim \mu_H^R(x)$.
5. Output $((B, x), (C_1, y_1), \ldots, (C_d, y_d))$

Folded Instance $(\mathcal{V}', \mathcal{P}')$
Fix $\mathcal{V}' = (V \times \{s, t\})^{\{R\}}$. Define a projection map $\Pi : \mathcal{V} \to \mathcal{V}'$ as follows:

$$\Pi(A, x) = \{(a_i, x_i) | x_i \in \{s, t\}\}$$

for each $(A, x) = ((a_1, x_1), (a_2, x_2), \ldots, (a_R, x_R))$ in $(V \times \{s, t\})^{\{R\}}$.
Let μ' be the probability distribution on $\mathcal{V}'$ obtained by projection of probability distribution μ on $\mathcal{V}$. Similarly, the probability distribution $\mathcal{P}'$ on $(\mathcal{V}')^{d+1}$ by applying the projection Π to the probability distribution $\mathcal{P}$.

Figure 24: Hardness Reduction

Observe that each of the queries $\Pi(B, x)$ and $\{\Pi(C_i, y_i)\}_{i=1}^d$ are distributed according to μ' on $\mathcal{V}'$. Let $F' \colon \mathcal{V}' \to \{0, 1\}$ denote the indicator function of a subset for the instance. Let us suppose that

$$\mathop{\mathbb{E}}_{X,Y \sim \mathcal{V}} [|F'(X) - F'(Y)|] \geqslant \frac{1}{10}$$

For the whole reduction, we fix $\eta = \varepsilon/(100d)$. We will restrict $\gamma < \varepsilon/(100d)$. We will fix its value later.

Theorem 8.6.1. *(Completeness) Suppose there exists a set $S \subset V$ such that $\mathsf{vol}(S) = \delta$ and $\Phi(S) \leqslant \gamma$ then there exists $F' : \mathcal{V}' \to \{0,1\}$ such that,*

$$\mathop{\mathbb{E}}_{X,Y \sim \mathcal{V}'} \left[|F'(X) - F'(Y)| \right] \geqslant \frac{1}{10}$$

and,

$$\mathop{\mathbb{E}}_{X,Y_1,\ldots,Y_d \sim \mathcal{P}} \left[\max_i |F'(X) - F'(Y_i)| \right] \leqslant 2\epsilon + \mathcal{O}\left(d(\eta + \gamma)\right) \leqslant 4\varepsilon$$

Proof. Define $F : \mathcal{V} \to \{0,1\}$ as follows:

$$F(A,x) = \begin{cases} 1 & \text{if } |\Pi(A,x) \cap (S \times \{s,t\})| = 1 \\ 0 & \text{otherwise} \end{cases}$$

Observe that by definition of F, the value of $F(A,x)$ only depends on $\Pi(A,x)$. So the function F naturally defines a map $F' : \mathcal{V}' \to \{0,1\}$. Therefore we can write,

$$\begin{aligned} \mathbb{P}\left[F(A,x) = 1\right] &= \sum_{i \in [R]} \mathbb{P}\left[x_i \in \{s,t\}\right] \mathbb{P}\left[\{a_1,\ldots,a_R\} \cap S = \{a_i\} | x_i \in \{s,t\}\right] \\ &\geqslant R \cdot \frac{1}{2} \cdot \frac{1}{R} \cdot \left(1 - \frac{1}{R}\right)^{R-1} \geqslant \frac{1}{10} \end{aligned}$$

and,

$$\begin{aligned} \mathbb{P}\left[F(A,x) = 1\right] &= \mathbb{P}\left[|\Pi(A,x) \cap (S \times \{s,t\})| = 1\right] \\ &\leqslant \mathop{\mathbb{E}}_{(A,x) \sim \mathcal{V}} \left[|\Pi(A,x) \cap (S \times \{s,t\})|\right] \\ &= R \cdot \frac{1}{2} \cdot \frac{|S|}{|V|} \leqslant \frac{1}{2} \end{aligned}$$

The above bounds on $\mathbb{P}\left[F(A,x) = 1\right]$ along with the fact that F takes values only in $\{0,1\}$, we get that

$$\mathop{\mathbb{E}}_{X,Y \sim \mathcal{V}'} |F'(X) - F'(Y)| = \mathop{\mathbb{E}}_{(A,x),(B,y) \sim \mathcal{V}} |F(A,x) - F(B,y)| \geqslant \frac{1}{10}$$

Suppose we sample $A \in V^R$ and $B, C_1, \ldots, C_d$ independently from $G_\eta^{\otimes R}(A)$. Let us denote $A = (a_1, \ldots, a_R)$, $B = (b_1, \ldots, b_R)$, $C_i = (c_{i1}, \ldots, c_{iR})$ for all $i \in [d]$. Note that,

$$\begin{aligned}
&\mathbb{P}\left[\exists i \in [R] \text{ such that } |\{a_i, b_i\} \cap S| = 1\right] \\
&\leqslant \textstyle\sum_{i\in[R]}(1-\eta)\,\mathbb{P}\left[(a_i, b_i) \in E[S, \bar{S}]\right] + \eta\,\mathbb{P}\left[(a_i, b_i) \in S \times \bar{S}\right] \\
&\leqslant R(\mathsf{vol}(S)\Phi(S) + 2\eta\mathsf{vol}(S)) \leqslant 2(\gamma + \eta)\,.
\end{aligned}$$

Similarly, for each $j \in [d]$,

$$\mathbb{P}\left[\exists i \in [R] || \{a_i, c_{ji}\} \cap S| = 1\right] \leqslant \sum_{i\in[R]} \mathbb{P}\left[(a_i, c_{ji}) \in E[S, \bar{S}]\right] \leqslant R\mathsf{vol}(S)\Phi(S) \leqslant 2(\gamma+\eta)\,.$$

By a union bound, with probability at least $1 - 2(d+1)(\gamma+\eta)$ we have that none of the edges $\{(a_i, b_i)\}_{i\in[R]}$ and $\{(a_i, c_{ji})\}_{j\in[d], i\in[R]}$ cross the cut $(S, \bar{S})$.

Conditioned on the above event, we claim that if $(B, x) \cap (S \times \{t, t'\}) = \emptyset$ then

$$\max_i |F(B, x) - F(C_i, y_i)| = 0\,.$$

First, if $(B, x) \cap (S \times \{t, t'\}) = \emptyset$ then for each $b_i \in S$ the corresponding $x_i \in \{s, s'\}$. In particular, this implies that for each $b_i \in S$, either all of the pairs (b_i, x_i), $\{(c_{ji}, y_{ji})\}_{j\in[d]}$ are either in $S \times \{s, t\}$ or $S \times \{s', t'\}$, thereby ensuring that $\max_i |F(B, x) - F(C_i, y_i)| = 0$.

From the above discussion we conclude,

$$\begin{aligned}
&\mathbb{E}_{(B,x),(C_1,y_1),\ldots,(C_d,y_d)\sim\mathcal{P}}\left[\max_i |F(B, x) - F(C_i, y_i)|\right] \\
&\leqslant \mathbb{P}\left[|(B, x) \cap (S \times \{t, t'\})\,| \geqslant 1\right] + 2(d+1)(\gamma+\eta) \\
&\leqslant \mathbb{E}\left[|(B, x) \cap (S \times \{t, t'\})\,|\right] + 2(d+1)(\gamma+\eta) \\
&= R \cdot \mathsf{vol}(S) \cdot \epsilon + 2(d+1)(\gamma+\eta) = \epsilon + 2(d+1)(\gamma+\eta)
\end{aligned}$$

□

Let $F' : \mathcal{V}' \to \{0,1\}$ be a subset of the instance $(\mathcal{V}', \mathcal{P}')$. Let us define the following notation.

$$\mathsf{VAL}_{\mathcal{P}'}(F') \stackrel{\text{def}}{=} \mathop{\mathbb{E}}_{(X,Y_1,\ldots,Y_d)\sim\mathcal{P}'} \left[\max_{i\in[d]} |F'(X) - F'(Y_i)|\right]$$

and

$$\mathop{\mathsf{Var}}_1[F'] \stackrel{\text{def}}{=} \mathop{\mathbb{E}}_{X,Y\sim\mathcal{V}'} |F'(X) - F'(Y)| \ .$$

We define the functions $F : \mathcal{V} \to [0,1]$ and $f_A, g_A : V_H^R \to [0,1]$ for each $A \in V^R$ as follows.

$$F(A,x) \stackrel{\text{def}}{=} F'(\Pi(A,x)) \qquad f_A(x) \stackrel{\text{def}}{=} F(A,x) \qquad g_A(x) \stackrel{\text{def}}{=} \mathop{\mathbb{E}}_{B\sim G_\eta^{\otimes R}(A)} F(B,x)$$

Lemma 8.6.2.

$$\mathsf{VAL}_{\mathcal{P}'}(F') \geqslant \mathop{\mathbb{E}}_{A\in V^R} \mathsf{VAL}_{\mu_H^R}(g_A)$$

Proof.

$$\begin{aligned}
\mathsf{VAL}_{\mathcal{P}'}(F') &= \mathsf{VAL}_{\mathcal{P}}(F) \\
&= \mathop{\mathbb{E}}_{A\sim V^R} \mathop{\mathbb{E}}_{x\sim\mu_H^R} \mathop{\mathbb{E}}_{y_1,\ldots,y_d\sim\mu_H^R(x)} \mathop{\mathbb{E}}_{B,C_1,\ldots,C_d\sim G_\gamma^{\otimes R}(A)} \max_i |F(B,x) - F(C_i,y_i)| \\
&\geqslant \mathop{\mathbb{E}}_{A\sim V^R} \mathop{\mathbb{E}}_{x\sim\mu_H^R} \mathop{\mathbb{E}}_{y_1,\ldots,y_d\sim\mu_H^R(x)} \max_i \left| \mathop{\mathbb{E}}_{B\sim G_\gamma^{\otimes R}(A)} F(B,x) - \mathop{\mathbb{E}}_{C_i\sim G_\gamma^{\otimes R}(A)} F(C_i,y_i) \right| \\
&\geqslant \mathop{\mathbb{E}}_{A\sim V^R} \mathop{\mathbb{E}}_{x\sim\mu_H^R} \mathop{\mathbb{E}}_{y_1,\ldots,y_d\sim\mu_H^R(x)} \max_i |g_A(x) - g_A(y_i)| \\
&= \mathop{\mathbb{E}}_{A\in V^R} \mathsf{VAL}_{\mu_H^R}(g_A)
\end{aligned}$$

□

Lemma 8.6.3.

$$\mathop{\mathbb{E}}_{A\sim V^R} \mathop{\mathbb{E}}_{x\sim\mu_H^R} g_A(x)^2 \geqslant \mathop{\mathbb{E}}_{(A,x)\sim\mathcal{V}} F^2(A,x) - \mathsf{VAL}_{\mathcal{P}'}(F')$$

Proof.

$$\mathop{\mathbb{E}}_{A\sim V^R}\mathop{\mathbb{E}}_{x\sim\mu_H^R} g_A(x)^2 = \mathop{\mathbb{E}}_{A\sim V^R}\mathop{\mathbb{E}}_{x\sim\mu_H^R}\mathop{\mathbb{E}}_{B,C\sim G_\eta^{\otimes R}(A)} F(B,x)F(C,x)$$

$$= \frac{1}{2}\mathop{\mathbb{E}}_{A\sim V^R}\mathop{\mathbb{E}}_{x\sim\mu_H^R}\mathop{\mathbb{E}}_{B,C\sim G_\eta^{\otimes R}(A)} F^2(B,x)+F^2(C,x)-(F(B,x)-F(C,x))^2$$

$$= \mathop{\mathbb{E}}_{A\sim V^R}\mathop{\mathbb{E}}_{x\sim\mu_H^R} F^2(A,x) - \frac{1}{2}\mathop{\mathbb{E}}_{A\sim V^R}\mathop{\mathbb{E}}_{x\sim\mu_H^R}\mathop{\mathbb{E}}_{B,C\sim G_\eta^{\otimes R}(A)}(F(B,x)-F(C,x))^2 \tag{75}$$

where in the last step we used the fact that B, C have the same distribution as $A \sim V^R$.

Since the function F is bounded in $[0,1]$, we have

$$\mathop{\mathbb{E}}_{A\sim V^R}\mathop{\mathbb{E}}_{x\sim\mu_H^R}\mathop{\mathbb{E}}_{B,C\sim G_\eta^{\otimes R}(A)}(F(B,x)-F(C,x))^2$$

$$\leqslant \mathop{\mathbb{E}}_{A\sim V^R}\mathop{\mathbb{E}}_{x\sim\mu_H^R}\mathop{\mathbb{E}}_{B,C\sim G_\eta^{\otimes R}(A)}|F(B,x)-F(C,x)| \tag{76}$$

$$\mathop{\mathbb{E}}_{A\sim V^R}\mathop{\mathbb{E}}_{x\sim\mu_H^R}\mathop{\mathbb{E}}_{B,C\sim G_\eta^{\otimes R}(A)}|F(B,x)-F(C,x)|$$

$$\leqslant \mathop{\mathbb{E}}_{A\sim V^R}\mathop{\mathbb{E}}_{x\sim\mu_H^R}\mathop{\mathbb{E}}_{y\sim\mu_H^R(x)}\mathop{\mathbb{E}}_{B,C,D\sim G_\eta^{\otimes R}(A)}|F(B,x)-F(D,y)|+|F(C,x)-F(D,y)|$$

$$= 2\mathop{\mathbb{E}}_{A\sim V^R}\mathop{\mathbb{E}}_{x\sim\mu_H^R}\mathop{\mathbb{E}}_{y\sim\mu_H^R(x)}\mathop{\mathbb{E}}_{B,D\sim G_\eta^{\otimes R}(A)}|F(B,x)-F(D,y)| \tag{77}$$

(because (B,D), (C,D) have same distribution)

$$\leqslant 2\mathop{\mathbb{E}}_{A\sim V^R}\mathop{\mathbb{E}}_{x\sim\mu_H^R}\mathop{\mathbb{E}}_{y_1,\ldots,y_d\sim\mu_H^R(x)}\mathop{\mathbb{E}}_{B,D_1,\ldots,D_d\sim G_\eta^{\otimes R}(A)}\max_i|F(B,x)-F(D_i,y_i)|$$

$$= 2\mathsf{VAL}_{\mathcal{P}}(F) = 2\mathsf{VAL}_{\mathcal{P}'}(F') \tag{78}$$

Equations (75), (76) and (78) yield the desired result. □

Lemma 8.6.4.

$$\mathop{\mathbb{E}}_{A\sim V^R}\mathop{\mathsf{Var}}_1[g_A] = \mathop{\mathbb{E}}_{A\sim V^R}\mathop{\mathbb{E}}_{x,y\in\mu_H^R}|g_A(x)-g_A(y)| \geqslant \frac{1}{2}(\mathop{\mathsf{Var}}_1[F])^2 - \mathsf{VAL}_{\mathcal{P}'}(F')$$

Proof. Since the function g_A is bounded in $[0,1]$ we can write

$$\mathop{\mathbb{E}}_{A\sim V^R}\mathop{\mathbb{E}}_{x,y\in\mu_H^R}|g_A(x)-g_A(y)| \geqslant \mathop{\mathbb{E}}_{A\sim V^R}\mathop{\mathbb{E}}_{x,y\in\mu_H^R}(g_A(x)-g_A(y))^2$$

$$\geqslant \mathop{\mathbb{E}}_{A\sim V^R}\mathop{\mathbb{E}}_{x\in\mu_H^R} g_A^2(x) - \mathop{\mathbb{E}}_{A}\mathop{\mathbb{E}}_{x,y\in\mu_H^R} g_A(x)g_A(y) \tag{79}$$

In the above expression there are two terms. From Lemma 8.6.3, we already know that

$$\mathop{\mathbb{E}}_{A \sim V^R} \mathop{\mathbb{E}}_{x \in \mu_H^R} g_A^2(x) \geqslant \mathop{\mathbb{E}}_{(A,x)\sim \mathcal{V}} F^2(A,x) - \mathsf{VAL}_{\mathcal{P}'}(F') \tag{80}$$

Let us expand out the other term in the expression.

$$\mathop{\mathbb{E}}_{A} \mathop{\mathbb{E}}_{x,y \in \mu_H^R} g_A(x) g_A(y) = \mathop{\mathbb{E}}_{A} \mathop{\mathbb{E}}_{B,C \sim G_\eta^{\otimes R}(A)} \mathop{\mathbb{E}}_{x,y \in \mu_H^R} F'(\Pi(B,x)) F'(\Pi(C,y)) \tag{81}$$

Now consider the following graph $\mathcal{H}$ on $\mathcal{V}'$ defined by the following edge sampling procedure.

- Sample $A \in V^R$, and $x, y \in \mu_H^R$.
- Sample independently $B \sim G_\eta^{\otimes R}(A)$ and $C \sim G_\eta^{\otimes R}(A)$
- Output the edge $\Pi(B,x)$ and $\Pi(C,y)$

Let λ denote the second eigenvalue of the adjacency matrix of the graph $\mathcal{H}$.

$$\begin{aligned}
&\mathop{\mathbb{E}}_{A} \mathop{\mathbb{E}}_{B,C \sim G_\eta^{\otimes R}(A)} \mathop{\mathbb{E}}_{x,y \in \mu_H^R} F'(\Pi(B,x)) F'(\Pi(C,y)) = \langle F', \mathcal{H} F' \rangle \\
&\quad \leqslant \left(\mathop{\mathbb{E}}_{(A,x)\sim\mathcal{V}} F'(\Pi(A,x)) \right)^2 + \lambda \left(\mathop{\mathbb{E}}_{(A,x)\sim\mathcal{V}} (F'(\Pi(A,x)))^2 - (\mathop{\mathbb{E}}_{(A,x)\sim\mathcal{V}} F'(\Pi(A,x)))^2 \right) \\
&\quad = \lambda \mathop{\mathbb{E}}_{(A,x)\sim\mathcal{V}} F(A,x)^2 + (1-\lambda)(\mathop{\mathbb{E}}_{(A,x)\sim\mathcal{V}} F(A,x))^2 \\
&\qquad\qquad (\text{because } F'(\Pi(A,x)) = F(A,x))
\end{aligned}$$

Using the above inequality with equations (79), (80), (81) we can derive the following,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{A \sim V^R} \mathop{\mathbb{E}}_{x,y \in \mu_H^R} |g_A(x) - g_A(y)| \\
&\geqslant \mathop{\mathbb{E}}_{A \sim V^R} \mathop{\mathbb{E}}_{x \in \mu_H^R} g_A^2(x) - \mathop{\mathbb{E}}_{A} \mathop{\mathbb{E}}_{x,y \in \mu_H^R} g_A(x) g_A(y) \\
&\geqslant (1-\lambda) \left[\mathop{\mathbb{E}}_{(A,x)\sim\mathcal{V}} F^2(A,x) - (\mathop{\mathbb{E}}_{(A,x)\sim\mathcal{V}} F(A,x))^2 \right] - \mathsf{VAL}_{\mathcal{P}'}(F') \\
&\geqslant (1-\lambda) \mathsf{Var}[F] - \mathsf{VAL}_{\mathcal{P}'}(F') \\
&\geqslant (1-\lambda) (\mathop{\mathsf{Var}}_1[F])^2 - \mathsf{VAL}_{\mathcal{P}'}(F') \\
&\qquad \left(\text{because } \mathsf{Var}[F] > \mathop{\mathsf{Var}}_1[F]^2 \text{ for all } F \right).
\end{aligned}$$

To finish the argument, we need to bound the second eigenvalue λ for the graph $\mathcal{H}$. Here we will present a simple argument showing that the second eigenvalue λ for the graph $\mathcal{H}$ is strictly less than $\frac{1}{2}$. Let us restate the procedure to sample edges from $\mathcal{H}$ slightly differently.

- Define a map $\mathcal{M} : V \times V_H \to (V \cup \perp) \times (V_H \cup \{\perp\})$ as follows, $\mathcal{M}(b, x) = (b, x)$ if $x \in \{s, t\}$ and $\mathcal{M}(b, x) = (\perp, \perp)$ otherwise. Let $\Pi' : ((V \cup \perp) \times (V_H \cup \perp))^R \to (V \times \{s, t\})^{\{R\}}$ denote the following map.

$$\Pi'(B', x') = \{(b'_i, x'_i) | x_i \in \{s, t\}\}$$

- Sample $A \in V^R$ and $x, y \in \mu_H^R$
- Sample independently $B = (b_1, \ldots, b_R) \sim G_\eta^{\otimes R}(A)$ and $C = (c_1, \ldots, c_R) \sim G_\eta^{\otimes R}(A)$.
- Let $\mathcal{M}(B, x), \mathcal{M}(C, y) \in ((V \cup \{\perp\}) \times (V_H \cup \{\perp\}))^R$ be obtained by applying $\mathcal{M}$ to each coordinate of (B, x) and (C, y).
- Output an edge between $(\Pi'(\mathcal{M}(B, x)), \Pi'(\mathcal{M}(C, y)))$.

It is easy to see that the above procedure also samples the edges of $\mathcal{H}$ from the same distribution as earlier. Note that Π' is a projection from $((V \cup \perp) \times (V_H \cup \perp))^R$ to $(V \times \{s, t\})^{\{R\}}$. Therefore, the second eigenvalue of the graph $\mathcal{H}$ is upper bounded by the second eigenvalue of the graph on $((V \cup \perp) \times (V_H \cup \{\perp\}))^R$ defined by $\mathcal{M}(B, x) \sim \mathcal{M}(C, y)$. Let $\mathcal{H}_1$ denote the graph defined by the edges $\mathcal{M}(B, x) \sim \mathcal{M}(C, y)$. Observe that the coordinates of $\mathcal{H}_1$ are independent, i.e., $\mathcal{H}_1 = \mathcal{H}_2^R$ for a graph $\mathcal{H}_2$ corresponding to each coordinate of $\mathcal{M}(B, x)$ and $\mathcal{M}(C, y)$. Therefore, the second eigenvalue of $\mathcal{H}_1$ is at most the second eigenvalue of $\mathcal{H}_2$. The Markov chain $\mathcal{H}_2$ on $(V \cup \{\perp\}) \times (V_H \cup \perp)$ is defined as follows,

- Sample $a \in V$ and two neighbors $b \sim G_\eta(a)$ and $c \sim G_\eta(a)$.

- Sample $x, y \in V_H$ independently from the distribution μ_H.
- Output an edge between $\mathcal{M}(b,x)$ $\mathcal{M}(c,y)$.

Notice that in the Markov chain $\mathcal{H}_2$, for every choice of $\mathcal{M}(b,x)$ in $(V \cup \{\perp\}) \times (V_H \cup \perp)$, with probability at least $\frac{1}{2}$, the other endpoint $\mathcal{M}(c,y) = (\perp, \perp)$. Therefore, the second eigenvalue of $\mathcal{H}_2$ is at most $\frac{1}{2}$, giving a bound of $\frac{1}{2}$ on the second eigenvalue of $\mathcal{H}$. □

Now we restate a claim from [88] that will be useful for our our soundness proof.

Theorem 8.6.5. *(Restatement of Lemma 6.11 from [88]) Let G be a graph with a vertex set V. Let a distribution on pairs of tuples (A, B) be defined by $A \sim V^R$, $B \sim G_\eta^{\otimes R}(A)$. Let $\ell : V^R \to [R]$ be a labelling such that over the choice of random tuples and two random permutations π_A, π_B*

$$\mathop{\mathbb{P}}_{A \sim V^R, B \sim G_\eta^{\otimes R}(A)} \mathop{\mathbb{P}}_{\pi_A, \pi_B} \left\{ \pi_A^{-1}\left(\ell(\pi_A(A))\right) = \pi_B^{-1}\left(\ell(\pi_B(B))\right) \right\} \geqslant \zeta$$

Then there exists a set $S \subset V$ with $\mathsf{vol}(S) \in \left[\frac{\zeta}{16R}, \frac{3}{\eta R}\right]$ satisfying $\Phi(S) \leqslant 1 - \zeta/16$.

The following lemma asserts that if the graph G is a *NO*-instance of Small Set Expansion (γ, δ, M) then for almost all $A \in V^R$ the functions have no influential coordinates.

Lemma 8.6.6. *Fix $\delta = 1/R$. Suppose for all sets $S \subset V$ with $\mathsf{vol}(S) \in (\delta/M, M\delta)$, $\Phi(S) \geqslant 1 - \gamma$ then for all $\tau > 0$,*

$$\mathop{\mathbb{P}}_{A \sim V^R} \left[\exists i \mid \mathsf{Inf_i}[\mathsf{\Gamma_{1-\eta} g_A}] \geqslant \tau\right] \leqslant \frac{1000}{\tau^3 \varepsilon^2 \eta^2} \cdot \max(1/M, \gamma)$$

Proof. For each $A \in V^R$, let

$$L_A = \{i \in [R] \mid \mathsf{Inf_i}(\mathsf{\Gamma_{1-\eta} f_A}) > \tau/2\}$$

and

$$L'_A = \{i \in [R] \mid \mathsf{Inf_i}(\mathsf{\Gamma_{1-\eta} g_A}) > \tau\}.$$

Call a vertex $A \in V^R$ to be *good* if $L'_A \neq \emptyset$. By Lemma 8.5.4, the sum of influences of $\Gamma_{1-\eta} g_A$ is at most $\frac{1}{\epsilon\eta} \mathsf{Var}[g_A] \leqslant \frac{1}{\epsilon\eta}$. Therefore, the cardinality of L'_A is upper bounded by $|L'_A| \leqslant \frac{2}{\tau\epsilon\eta}$. Similarly, the cardinality of L_A is upper bounded by $|L_A| \leqslant \frac{1}{\tau\epsilon\eta}$.

The lemma asserts that at most a $\frac{1000}{\tau^3\eta^2\epsilon^2} \cdot \max(1/M, \gamma)$ fraction of vertices are *good*. For the sake of contradiction, assume that $\mathbb{P}_{A \in V^R}[L'_A \neq \emptyset] \geqslant 1000 \max(1/M, \gamma)/\tau^2\epsilon^2\eta^2$.

Define a labelling $\ell : V^R \to [R]$ as follows: for each $A \in V^R$, with probability $\frac{1}{2}$ choose a random coordinate in L_A and with probability $1/2$, choose a random coordinate in L'_A. If the sets L_A, L'_A are empty, then we choose a uniformly random coordinate in $[R]$.

Observe that for each $A \in V^R$, the function g_A is the average over bounded functions $f_B \colon V_H^R \to [0,1]$, where $B \sim G_\eta^R(A)$. Fix a vertex $A \in V^R$ such that $L'_A \neq \emptyset$ and a coordinate $i \in L'_A$. In particular, we have that $\mathsf{Inf_i}[\Gamma_{1-\eta} \mathsf{g_A}] \geqslant \tau$. Using convexity of influences, this implies that,

$$E_{B \sim G_\eta^{\otimes R}(A)} \mathsf{Inf_i}[\Gamma_{1-\eta} \mathsf{f_B}] \geqslant \tau\,.$$

Specifically, this implies that for at least a $\frac{\tau}{2}$ fraction of the neighbours $B \sim G_\eta^R(A)$, the influence of the i^{th} coordinate on f_B is at least $\frac{\tau}{2}$. Hence, if $L'_A \neq \emptyset$ then for at least a $\tau/2$ fraction of neighbours $B \sim G_\eta^{\otimes R}(A)$ we have $L'_A \cap L_B \neq \emptyset$.

By definition of the functions f_A, g_A, it is clear that for every permutation $\pi : [R] \to [R]$, $f_A(\pi(x)) = f_{\pi(A)}(x)$ and $g_A(\pi(x)) = g_{\pi(A)}(x)$. Therefore, for every permutation $\pi : [R] \to [R]$ and $A \in V^R$,

$$L_A = \pi^{-1}(L_{\pi(A)}) \qquad \text{and } L'_A = \pi^{-1}(L'_{\pi(A)})$$

From the above discussion, for every *good* vertex $A \in V^R$, for at least a $\tau/2$ fraction of the vertices $B \sim G_\eta^{\otimes R}(A)$, and every pair of permutations $\pi_A, \pi_B : [R] \to [R]$, we have $\pi_A^{-1}(L'_{\pi_A(A)}) \cap \pi_B^{-1}(L_{\pi_B(B)}) \neq \emptyset$. This implies that,

$$\begin{aligned}
&\mathop{\mathbb{P}}_{A\sim V^R, B\sim G_\eta^{\otimes R}(A)} \mathop{\mathbb{P}}_{\pi_A,\pi_B} \left\{\pi_A^{-1}\left(\ell(\pi_A(A))\right) = \pi_B^{-1}\left(\ell(\pi_B(B))\right)\right\} \\
&\geqslant \mathop{\mathbb{P}}_{A\sim V^R}[L'_A \neq \emptyset] \cdot \mathop{\mathbb{P}}_{B\sim G_\eta^{\otimes R}(A)}[L'_A \cap L_B \neq \emptyset | L'_A \neq \emptyset] \\
&\qquad \cdot \mathbb{P}\left[\pi_A^{-1}(\ell(\pi_A(A))) = \pi_B^{-1}(\ell(\pi_B(B))) \mid L'_A \cap L_B \neq \emptyset\right] \\
&\geqslant \mathop{\mathbb{P}}_{A\sim V^R}[L'_A \neq \emptyset] \cdot \left(\frac{\tau}{2}\right) \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{|L'_A|}\frac{1}{|L_B|} \\
&\geqslant \mathop{\mathbb{P}}_{A\sim V^R}[L'_A \neq \emptyset] \cdot \left(\frac{\tau}{2}\right) \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \left(\frac{\tau\eta\epsilon}{2}\right)^2 \\
&\geqslant 16\max(1/M, \gamma)
\end{aligned}$$

By Theorem 8.6.5, this implies that there exists a set $S \subset V$ with $\mathsf{vol}(S) \in [\frac{1}{MR}, \frac{3}{\eta R}]$ satisfying $\Phi(S) \leqslant 1-\gamma$. A contradiction. □

Finally, we are ready to show the soundness of the reduction.

Theorem 8.6.7. *(Soundness) For all ϵ, d there exists choice of M and γ, η such that the following holds. Suppose for all sets $S \subset V$ with $\mathsf{vol}(S) \in (\delta/M, M\delta)$, $\Phi(S) \geqslant 1-\eta$, then for all $F' : \mathcal{V}' \to [0,1]$ such that $\mathsf{Var}_1[F'] \geqslant \frac{1}{10}$, we have $\mathsf{VAL}_{\mathcal{P}'}(F') \geqslant \Omega(\sqrt{\epsilon \log d})$*

Proof. Recall that we had fixed $\eta = \varepsilon/(100d)$. We will choose τ to small enough so that the error term in the soundness of dictatorship test (Proposition 8.5.5) is smaller than ϵ. Since the least probability of any vertex in Markov chain H is ϵ, setting $\tau = \epsilon^{1/\epsilon^3}$ would suffice.

First, we know that if G is a *NO*-instance of Small Set Expansion (γ, δ, M) then for almost all $A \in V^R$, the function g_A has no influential coordinates. Formally, by Lemma 8.6.6, we will have

$$\mathop{\mathbb{P}}_{A\sim V^R}\left[\exists i \mid \mathsf{Inf_i}[\Gamma_{1-\eta}\mathsf{g_A}] \geqslant \tau\right] \leqslant \frac{1000}{\tau^3\eta^2} \cdot \max(1/M, \gamma)\,.$$

For an appropriate choice of M, γ, the above inequality implies that for all but an ϵ-fraction of vertices $A \in V^R$, the function g_A will have no influential coordinates.

Without loss of generality, we may assume that $\mathsf{VAL}_{\mathcal{P}'}(F') \leqslant \sqrt{\epsilon \log d}$, else we would be done. Applying Lemma 8.6.4, we get that $\mathbb{E}_{A \in V^R} \mathsf{Var}_1[g_A] \geqslant (\mathsf{Var}_1[F])^2 - \mathsf{VAL}_{\mathcal{P}'}(F') \geqslant \frac{1}{200}$. This implies that for at least a $\frac{1}{400}$ fraction of $A \in V^R$, $\mathsf{Var}_1[g_A] \geqslant 1/400$. Hence for at least an $1/400 - \epsilon$ fraction of vertices $A \in V^R$ we have,

$$\underset{1}{\mathsf{Var}}[g_A] \geqslant \frac{1}{400} \qquad \text{and} \qquad \max_i \mathsf{Inf_i}(\mathsf{\Gamma_{1-\eta}}(\mathsf{g_A})) \leqslant \tau$$

By appealing to the soundness of the gadget (Proposition 8.5.5), for every such vertex $A \in V^R$, $\mathsf{VAL}_{\mu_H^R}(g_A) \geqslant \Omega(\sqrt{\epsilon \log d}) - O(\epsilon) = \Omega(\sqrt{\epsilon \log d})$. Finally, by applying Lemma 8.6.2, we get the desired conclusion.

$$\mathsf{VAL}_{\mathcal{P}'}(F') \geqslant \mathop{\mathbb{E}}_{A \in V^R} \mathsf{VAL}_{\mu_H^R}(g_A) \geqslant \Omega(\sqrt{\epsilon \log d})$$

□

8.7 Reduction from Analytic d-Vertex Expansion to Vertex Expansion

Theorem 8.7.1. *A* c-*vs*-s *hardness for* d-Balanced Analytic Vertex Expansion *implies a* $4\ \mathsf{c}$-*vs*-$\mathsf{s}/16$ *hardness for* Balanced Symmetric Vertex Expansion *on graphs of degree at most* D*, where* $D = \max\{100d/s, 2\log(1/c)\}$.

At a high level, the proof of Theorem 8.7.1 has two steps.

1. We show that a c-vs-s hardness for Balanced Analytic Vertex Expansion implies a $2\ \mathsf{c}$-vs-$\mathsf{s}/4$ hardness for instances of Balanced Analytic Vertex Expansion having uniform distribution (Proposition 8.7.2).

2. We show that a c-vs-s hardness for instances of d-Balanced Analytic Vertex Expansion having uniform stationary distribution implies a $2\ \mathsf{c}$-vs-$\mathsf{s}/2$ hardness for Balanced Symmetric Vertex Expansion on $\Theta(D)$-regular graphs. (Proposition 8.7.5).

Proposition 8.7.2. *A* c*-vs-*s *hardness for* Balanced Analytic Vertex Expansion *implies a* 2 c*-vs-*s/4 *hardness for instances of* Balanced Analytic Vertex Expansion *having uniform distribution.*

Proof. Let $(V, \mathcal{P})$ be an instance of Balanced Analytic Vertex Expansion. We construct an instance $(V', \mathcal{P}')$ as follows. Let $T = 2n^2$. We first delete all vertices i from V which have $\mu(i) < 1/2n^2$, i.e. $V \leftarrow V \setminus \{i \in V : \mu(i) < 1/2n^2\}$. Note that after this operation, the total weight of the remaining vertices is still at least $1 - 1/2n$ and the Balanced Analytic Vertex Expansion can increase or decrease by at most a factor of 2. Next for each i, we introduce introduce $\lceil \mu(i)T \rceil$ copies of vertex i. We will call these vertices the cloud for vertex i and index them as (i, a) for $a \in [\mu(i)T]$.

We set the probability mass of each $(d+1)$-tuple $((i, a), (j_1, b_1) \ldots, (j_d, b_d))$ as follows :

$$\mathcal{P}'((i, a), (j_1, b_1) \ldots, (j_d, b_d)) = \frac{\mathcal{P}(i, j_1, \ldots, j_d)}{(\mu(i)T) \cdot \Pi_{\ell=1}^{d} (\mu(j_\ell)T)}$$

It is easy to see that $\mu'(i, a) = 1/T$ for all vertices $(i, a) \in V'$. The analytic d-vertex expansion under a function F is given by,

$$\frac{\mathbb{E}_{((i,a),(j_1,b_1)\ldots,(j_d,b_d)) \sim \mathcal{P}'} \max_\ell |F(i, a) - F(j_\ell, b_\ell)|}{\mathbb{E}_{(i,a),(j,b) \sim \mu'} |F(i, a) - F(j, b)|}$$

where $X = (i, a)$ and $Y_\ell = (j, b)$ which are sampled as follows:

1. Sample a $(d+1)$-tuple $(i, j_1, \ldots, j_d)$ from $\mathcal{P}$.

2. Sample a uniformly at random from $1, \ldots, \mu(i)T$.

3. Sample b_ℓ uniformly at random from $\{1, \ldots, \mu(j_\ell)T\}$ for each $\ell \in [d]$.

Completeness Suppose, $\Phi(V, \mathcal{P}) \leqslant c$. Let f be the corresponding cut function. The function $f : V \to \{0, 1\}$ can be trivially extended to a function $F : V' \to \{0, 1\}$ thereby certifying that $\Phi(V', \mathcal{P}') \leqslant 2c$.

Soundness Suppose $\Phi(V,\mathcal{P}) \geqslant s$. Let $F : V' \to \{0,1\}$ be any balanced function. By convexity of absolute value function we get

$$\mathop{\mathbb{E}}_{((i,a),(j_1,b_1),\ldots,(j_d,b_d))\sim\mathcal{P}'} \max_{\ell} |F(i,a) - F(j_{\ell,b_\ell})| \geqslant \mathop{\mathbb{E}}_{(i,j_1,\ldots,j_d)\sim\mathcal{P}} \max_{\ell} \left| \mathop{\mathbb{E}}_{a} F(i,a) - \mathop{\mathbb{E}}_{\ell} F(j_\ell, b_\ell) \right| .$$

So if we define $f(i) = E_a F(i,a)$, the numerator for analytic d-vertex expansion in $(V,\mathcal{P})$ for f is only lower than the corresponding numerator for F in $(V',\mathcal{P}')$. We need to lower bound the denominator, $\mathbb{E}_{i,j\sim\mu} |f(i) - f(j)|$. The requisite lower bound follows from the following two lemmas.

Lemma 8.7.3.

$$\mathop{\mathbb{E}}_{i,j\sim\mu} |f(i) - f(j)| \geqslant \mathop{\mathbb{E}}_{(i,a),(j,b)\sim\mu'} |F(i,a) - F(j,b)| - \mathop{\mathbb{E}}_{(i,a),(i,b)\sim\mu'} |F(i,a) - F(i,b)|$$

Proof. The Lemma follows directly from the following two inequalities.

$$\mathop{\mathbb{E}}_{(i,a),(j,b)} |F(i,a) - F(j,b)| \leqslant \mathop{\mathbb{E}}_{(i,a)} |F(i,a) - f(i)| + \mathop{\mathbb{E}}_{(j,b)} |F(j,b) - f(j)| + \mathop{\mathbb{E}}_{i,j} |f(i) - f(j)|$$

which follows from Triangle Inequality, and

$$\mathop{\mathbb{E}}_{i,a} |F(i,a) - f(i)| \leqslant \mathop{\mathbb{E}}_{i,a,b} |F(i,a) - F(i,b)|$$

□

Lemma 8.7.4.

$$\mathop{\mathbb{E}}_{i,a,b} |F(i,a) - F(i,b)| \leqslant 2\mathsf{VAL}_{\mathcal{P}'}(F) = 2 \mathop{\mathbb{E}}_{(i,a),(j_1,c_1),\ldots(j_d,c_d)\sim\mathcal{P}'} \max_{\ell} |F(i,a) - F(j_\ell, c_\ell)|$$

Proof. Sample $(i, j_1, \ldots, j_d) \sim \mathcal{P}$. For any neighbour (j,c) of $(i,a),(i,b)$, using the Triangle Inequality we have

$$|F(i,a) - F(i,b)| \leqslant |F(i,a) - F(j,c)| + |F(j,c) - F(i,b)|$$

Therefore,

$$\begin{aligned} |F(i,a) - F(i,b)| &\leqslant \frac{\sum_\ell |F(i,a) - F(j_\ell, c_\ell)| + \sum_\ell |F(i,b) - F(j_\ell, c_\ell)|}{d} \\ &\leqslant \max_\ell |F(i,a) - F(j_\ell, c_\ell)| + \max_\ell |F(i,b) - F(j_\ell, c_\ell)| \end{aligned}$$

Taking expectations over the uniformly random choice of a and b from the cloud of i,

$$\mathop{\mathbb{E}}_{(i,a),(i,b)} |F(i,a) - F(i,b)| \leqslant 2 \mathop{\mathbb{E}}_{((i,a),(j_1,b_1),\ldots,(j_d,b_d))\sim\mathcal{P}'} \max_\ell |F(i,a) - F(j_\ell, c_\ell)|$$

□

Lemma 8.7.3 and Lemma 8.7.4 together show that

$$\mathop{\mathbb{E}}_{i,j} |f(i) - f(j)| \geqslant \frac{\mathbb{E}_{(i,a),(j,b)} |F(i,a) - F(j,b)|}{2}.$$

as long as the value $\mathsf{VAL}_{\mathcal{P}'}(F) < \mathsf{Var}_1[F]/4$. Therefore, for any $F : V' \to \{0,1\}$,

$$\frac{\mathbb{E}_{((i,a),(j_1,b_1)\ldots,(j_d,b_d))\sim\mathcal{P}'} \max_\ell |F(i,a) - F(j_\ell, b_\ell)|}{\mathbb{E}_{(i,a),(j,b)\sim\mu'} |F(i,a) - F(j,b)|} \geqslant \frac{s}{4}.$$

Theorem 8.4.3 shows that the minimum value of Balanced Analytic Vertex Expansion is obtained by boolean functions. Therefore, $\Phi(V', \mathcal{P}') \geqslant s/4$. □

Proposition 8.7.5. *A* c*-vs-*s *hardness for instances of* d-Balanced Analytic Vertex Expansion *having uniform stationary distribution implies a* $2\,\mathsf{c}$*-vs-*$\mathsf{s}/4$ *hardness for* Balanced Symmetric Vertex Expansion *on* $\Theta(D)$*-regular graphs. Here* $D \geqslant \max\{100d/s, 2\log(1/c)\}$.

Proof. Let $(V', \mathcal{P}')$ be an instance of d-Balanced Analytic Vertex Expansion. We construct a graph G from $(V', \mathcal{P}')$ as follows. We initially set $V(G) = V'$. For each vertex X we pick D neighbors by sampling D/d tuples from the marginal distribution of $\mathcal{P}'$ on tuples containing X in the first coordinate.

Let d_i denote the degree of vertex i, i.e. the number of vertices adjacent to vertex i in G. It is easy to see that $d_i \geqslant D$ and $\mathbb{E}[d_i] = 2D\ \forall i \in V(G)$. Let $L = \{i \in V(G) | d_i > 4D\}$. Using Hoeffding's Inequality, we get a tight concentration for d_i around $2D$.

$$\mathbb{P}[d_i > 4D] \leqslant e^{-D}.$$

Therefore, $\mathbb{E}[|L|] < n/e^D$. We delete these vertices from G, i.e. $V(G) \leftarrow V(G) \backslash L$. With constant probability, all remaining vertices will have their degrees in the range $[D/2, 4D]$. Also, the vertex expansion of every set will decrease by at most an additive $1/e^D$.

Completeness Let $\Phi(V', \mathcal{P}') \leqslant c$ and let $F : V' \to \{0,1\}$ be the function corresponding to $\Phi(V', \mathcal{P}')$. Let the set S be the support of the function F. Clearly, the set S is balanced. Therefore, with constant probability, we have

$$\Phi^{\mathsf{V}}(G) \leqslant \Phi^{\mathsf{V}}_G(S) \leqslant \Phi(V', \mathcal{P}') + 1/e^D \leqslant 2c.$$

Soundness Suppose $\Phi(V', \mathcal{P}') \geqslant s$. Let $F : V' \to \{0,1\}$ be any balanced function.

Since the max is larger than the average, we get

$$\mathop{\mathbb{E}}_X \max_{Y_i \in N_G(X)} |F(X) - F(Y_i)| \geqslant \frac{d}{D} \sum_{j=1}^{D/d} \mathop{\mathbb{E}}_{(X, Y_1, \ldots, Y_d) \sim \mathcal{P}} \max_i |F(X) - F(Y_i)|$$

By Hoeffding's inequality (Fact 2.5.3), we get

$$\begin{aligned} &\mathbb{P}\left[\left(\mathop{\mathbb{E}}_X \max_{Y_i \in N(X)} |F(X) - F(Y_i)|\right) < s/4\right] \\ &\leqslant \mathbb{P}\left[\left(\frac{d}{D} \sum_{j=1}^{D/d} \mathop{\mathbb{E}}_{(X, Y_1, \ldots, Y_d) \sim \mathcal{P}} \max_i |F(X) - F(Y_i)|\right) < s/4\right] \\ &\leqslant \exp\left(-n(sD/d)^2\right) \end{aligned}$$

Here, the last inequality follows from Hoeffding's inequality over the index X. There are at most 2^n boolean functions on V. Therefore, using a union bound on all those functions we get,

$$\mathbb{P}\left[\Phi^{\mathsf{V}}(G) \geqslant s/4\right] \geqslant 1 - 2^n \exp\left(-n(sD/d)^2\right).$$

Since $D > d/s$, we get that with probability $1 - o(1)$, $\Phi^{\mathsf{V}}(G) \geqslant s/4$.

□

Proof of Theorem 8.7.1. Theorem 8.7.1 follows directly from Proposition 8.7.2 and Proposition 8.7.5. □

8.8 Hardness of Vertex Expansion

We are now ready to prove Theorem 8.0.1. We restate the Theorem below.

Theorem 8.8.1. *For every $\eta > 0$, there exists an absolute constant C such that $\forall \varepsilon > 0$ it is* SSE*-hard to distinguish between the following two cases for a given graph $G = (V, E)$ with maximum degree $d \geqslant 100/\varepsilon$.*

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

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

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

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

Proof. From Theorem 8.6.1 and Theorem 8.6.7 we get that for an instance of Balanced Analytic Vertex Expansion $(V, \mathcal{P})$, it is SSE-hard to distinguish between the following two cases cases:

Yes :

$$\Phi(V, \mathcal{P}) \leqslant \varepsilon$$

No :

$$\Phi(V, \mathcal{P}) \geqslant \min\left\{10^{-4}, c_1\sqrt{\varepsilon \log d}\right\} - \eta$$

Now from Theorem 8.7.1 we get that for a graph G, it is SSE-hard to distinguish between the following two cases cases:

Yes :

$$\Phi^{\mathsf{V,bal}} \leqslant \varepsilon$$

No :

$$\Phi^{\mathsf{V,bal}} \geqslant \min\left\{10^{-6}, c_2\sqrt{\varepsilon \log d}\right\} - \eta$$

We use a standard reduction from Balanced Symmetric Vertex Expansion to Symmetric Vertex Expansion. A c-vs-s hardness for b-Balanced Symmetric Vertex Expansion implies a $2\ \mathsf{c}$-vs-$\mathsf{s}/2$ hardness for Symmetric Vertex Expansion. This can be seen as follows. Fix a graph $G = (V, E)$.

Completeness If G has Balanced-symmetric vertex expansion at most c, then clearly its symmetric vertex expansion is also at most c.

Soundness Suppose we have a polynomial time algorithm that outputs a set S having $\phi^{\mathsf{V}}(S) \leqslant s$ whenever G has a set S' having $\phi^{\mathsf{V}}(S') \leqslant 2c$. Then this algorithm can be used as an oracle to find a balanced set having symmetric vertex expansion at most s. This would contradict the hardness of Balanced Symmetric Vertex Expansion.

First we find a set, say T, having $\phi^{\mathsf{V}}(T) \leqslant s$. If we are unable to find such a T, we stop. If we find such a set T and T has balance at least b, then we stop. Else, we delete the vertices in T from G and repeat. We continue until the number of deleted vertices first exceeds a $b/2$ fraction of the vertices.

If the process deletes less than $b/2$ fraction of the vertices, then the remaining graph (which has at least $(1 - b/2)n$ vertices) has conductance $2c$, and thus in the original graph every b-balanced cut has conductance at least c. This is a contradiction !

If the process deletes between $b/2$ and $1/2$ of the nodes, then the union of the deleted sets gives a set T' with $\phi^{\mathsf{V}}(T') \leqslant s$ and balance of T' at least $b/2$.

Using this we get that for a graph G, it is SSE-hard to distinguish between the following two cases cases:

Yes :

$$\Phi^{\mathsf{V}} \leqslant \varepsilon$$

No :

$$\Phi^{\mathsf{V}} \geqslant \min\left\{10^{-8}, c_3\sqrt{\varepsilon \log d}\right\} - \eta$$

Finally, using the computational equivalence of VERTEX EXPANSION and SYMMETRIC VERTEX EXPANSION (Theorem 8.3.1), we get that for a graph G, it is SSE-hard to distinguish between the following two cases cases:

Yes :

$$\phi^{\mathsf{V}} \leqslant \varepsilon$$

No :

$$\phi^{\mathsf{V}} \geqslant \min\left\{10^{-10}, C\sqrt{\varepsilon \log d}\right\} - \eta$$

This completes the proof of the theorem. □

8.9 Conclusion

In this chapter we showed that any polynomial time algorithm algorithm that outputs a set having vertex expansion less than $C\sqrt{\phi^{\mathsf{V}} \log d}$, for some absolute constant C, will disprove the SSE hypothesis; alternatively, to improve on the bound of $\mathcal{O}\left(\sqrt{\phi^{\mathsf{V}} \log d}\right)$, one has to disprove the SSE hypothesis. From an algorithmic standpoint, we believe that Theorem 8.0.3 exposes a clean algorithmic challenge of recognizing a vertex expander – a challenging problem that is not only interesting on its own right, but whose resolution would probably lead to a significant advance in approximation algorithms.

Acknowledgements. The results in this chapter were obtained in joint work with Prasad Raghavendra and Santosh Vempala.

CHAPTER IX

CONCLUSION

In this thesis we studied three notions of expansion, namely edge expansion in graphs, vertex expansion in graphs and hypergraph expansion. We showed that the notion of Laplacian eigenvalues and Cheeger Inequalities cuts across these three problems. We studied higher orders of these expansion quantities and gave optimal higher order Cheeger Inequalities for edge expansion in graphs, and made partial progress towards establishing optimal higher order Cheeger Inequalities for vertex expansion in graphs and hypergraph expansion.

We summarize the contributions of this thesis in Table 1, Table 2, Table 3 and Table 4, below.

Table 1: Higher Order Cheeger Inequalities

	Edge Expansion in graphs	Vertex Expansion and Hypergraph Expansion
Cheegers Inequality	$\frac{\lambda_2}{2} \leqslant \phi_G \leqslant \sqrt{2\lambda_2}$ [3, 1]	$\frac{\gamma_2}{2} \leqslant \phi_H \leqslant \sqrt{2\gamma_2}$ and $\frac{\lambda_\infty}{2} \leqslant \phi_G^{\mathsf{V}} \leqslant \sqrt{2\lambda_\infty}$ for Vertex Expansion [21].
Small Set Expansion	$\mathcal{O}\left(\sqrt{\lambda_k \log k}\right)$ $\Omega(\sqrt{\lambda_k \log k})$ for Noisy hypercube graph.	$\mathcal{O}\left(\sqrt{r\,\gamma_k \log k}\right)$ $\tilde{\mathcal{O}}\left(k\sqrt{\gamma_k \log r}\right)$
K SPARSE-CUTS	$\frac{\lambda_k}{2} \leqslant \phi_G^k \leqslant \mathcal{O}\left(\sqrt{\lambda_{2k} \log k}\right)$ Lower bound tight for hypercube, upper bound tight for Noisy hypercube.	$\frac{\gamma_k}{2} \leqslant \phi_k \leqslant \mathcal{O}\left(k^3\sqrt{\gamma_k \log r}\right)$ Lower bound tight for hypercube.

Table 2: Approximation Algorithms

	Edge Expansion in graphs	Vertex Expansion and Hypergraph Expansion
Sparsest Cut	$\mathcal{O}\left(\sqrt{\mathsf{OPT}}\right)$ [3, 1] $\mathcal{O}\left(\sqrt{\log n}\right)\mathsf{OPT}$ [13]	$\mathcal{O}\left(\sqrt{\mathsf{OPT}\log r}\right)$ $\mathcal{O}\left(\sqrt{\log n}\right)\mathsf{OPT}$ and $\mathcal{O}\left(\sqrt{\log n}\right)\mathsf{OPT}$ for Vertex Expansion [39].
Small Set Expansion	$\mathcal{O}\left(\sqrt{\mathsf{OPT}\log k}\right)$ [87] $\mathcal{O}\left(\sqrt{\log n \log k}\right)\mathsf{OPT}$ [16]	$\tilde{\mathcal{O}}\left(k\sqrt{\mathsf{OPT}\log r}\right)$ $\tilde{\mathcal{O}}\left(k\sqrt{\log n}\right)\mathsf{OPT}$
Sparsest k-partition	$\mathcal{O}\left(\sqrt{\mathsf{OPT}\log k}\right)$ $\mathcal{O}\left(\sqrt{\log n \log k}\right)\mathsf{OPT}$	–

Table 3: Computing Eigenvalues

	Adjacency Matrix	Hypergraph Markov Operator
Upper bound	Exact computation for all eigenvalues.	$\mathcal{O}(k\log r)$-approximation.
Lower bound	Exact computation for all eigenvalues.	$\Omega(\log r)$ hardness under SSE.

9.1 Future Directions

Partitioning into Expanders. Similar to the k sparse-cuts problem and the Sparsest k-partition problem is the problem of partitioning a graph into expanders while minimizing the largest expansion among the parts. More formally, given a graph $G = (V, E, w)$ and a parameter $\alpha \in \mathbb{R}^+$, the problem asks to compute a partition of the vertex set V into sets $S_1, S_2 \ldots$ (the number of sets is not specified) such that the graph induced on each S_i has expansion at least α while minimizing $\max_i \phi_G(S_i)$. This problem, whilst being of interest in its own right, could also have numerous practical

Table 4: Mixing Time Bounds

	Random-walks on graphs	Hypergraph Dispersion Process
Upper bound	$\mathcal{O}\left(\frac{\log n}{\lambda_2}\right)$ (folklore)	$\mathcal{O}\left(\frac{\log n}{\gamma_2}\right)$
Lower bound	$\Omega\left(\frac{1}{\lambda_2}\right)$ (folklore)	$\Omega\left(\frac{1}{\gamma_2}\right)$

applications. Tanaka [101] and Gharan and Trevisan [41] study a slight variant of this problem, and show that if there is a sufficiently large gap between ϕ_G^k and ϕ_G^{k+1}, then graph can be partitioned into k pieces while ensuring that expansion of the graph induced on each part is lower bounded by a function of λ_k and the expansion of each part in G is upper bounded by a function of λ_{k+1}. Kannan et. al.[51] study a variant of this problem which asks to compute a partition of the graph into expanders while minimizing the total fraction of edges cut; they give a bi-criteria approximation algorithm for this problem. We showed that their algorithm can produce a partition of the graph into a set of expanders, say $S_1, S_2, \ldots$, with $\max_i \phi_G(S_i)$ being unbounded. Computing a partition of the graph into expanders while minimizing $\max_i \phi_G(S_i)$ seems to be a challenging open problem.

Hypergraph Markov Operators. In Chapter 4, we introduced a new (non-linear) hypergraph Markov Operator and studied its eigenvalues. We showed that there exists no linear operator whose eigenvalues can be used to estimate hypergraph expansion in a *Cheeger*-like manner. We also argued that our operator is the "best" operator for studying hypergraph expansion parameters. We ask, what properties of graphs and random walks generalize to hypergraphs and this Markov operator? We also ask if there are any other hypergraph Markov operators whose eigenvalues can be used to estimate hypergraph parameters.

Approximation Algorithms. Multiplicative approximation algorithms for all the expansion problems studied in this thesis rely on the ARV structure theorem [13]. Since all these problems also generalize the SPARSEST CUT problem, improving the approximation guarantee for the SPARSEST CUT problem is a common barrier for improving the approximation guarantee for any of them. Arora, Rao and Vazirani gave a $\mathcal{O}\left(\sqrt{\log n}\right)$-approximation algorithm, based on a rounding algorithm for the standard SDP relaxation for SPARSEST CUT augmented with the *triangle inequality constraints*. This problem has resisted attempts to find polynomial-time approximation algorithms for them that match their known lower bounds. A natural question to ask is whether we can get better approximation guarantees if we allow ourselves super-polynomial computational time? Towards this end, one could try to understand the power of the LP and SDP hierarchies for combinatorial optimization problems.

The hierarchies of linear and semidefinite programs, such as the ones defined by Lovasz and Schrijver [76], Sherali and Adams [93] and Lasserre [61], provide sequences of increasingly powerful convex relaxations starting from a basic integer program. They form a powerful computational model, where one can solve the program at r^{th} level (also called round) in time $n^{\mathcal{O}(r)}$. The gap between integral and fractional solutions decreases with r, and reaches zero at the n^{th} level, where the program is guaranteed to find an optimal integral solution. Most known LP/SDP based algorithms can be derived by 2 or 3 rounds of one of these hierarchies.

Recently, Barak, Raghavendra and Steurer [17] showed a new method to analyze and round SDP hierarchies. At a high level, they use global correlations inside the high-dimensional SDP solution combined with the hierarchy constraints to obtain a better rounding of this solution into an integral one. They show that the surprising subexponential time algorithm for UNIQUE GAMES, due to Arora, Barak and Steurer [8], can be rederived by using these techniques.

Arora et. al.'s [8] showed that if a graph has "many" eigenvalues of its normalized

Laplacian matrix which are "small", then the graph has a small set having small expansion. They used this in obtaining a subexponential time algorithm for the Unique Games problem. We (in joint work with Prasad Raghavendra and Santosh Vempala) conjecture that if a graph is a vertex expander (i.e. its vertex expansion is $\Omega(1)$), then there exists a way to assign non-negative weights to its edges such that the (weighted) degree of each vertex is equal to one, and the number of "small" eigenvalues of the resulting normalized Laplacian matrix is "small". We show that this conjecture, togethor with the tools of Barak et. al.[17] implies a subexponential time constant factor approximation algorithm for Sparsest Cut, graph coloring, etc.

Understanding the power of the hierarchy for these problems is a very challenging questions, and will have numerous applications in other areas of theory as well.

Hardness. Our computational lower bounds for Vertex Expansion and Hypergraph Expansion rely on the SSE hypothesis. An interesting open problem is whether we can prove analogous lower bounds based on the Unique Games Conjecture. The Unique Games Conjecture is currently not known to imply hardness results for problems closely tied to graph expansion such as Sparsest Cut, Vertex Expansion, etc., the reason being that the *hard* instances of these problems are required to have certain global structure, namely expansion. Gadget reductions from a Unique Games instance preserve the global properties of the Unique Games instance such as lack of expansion. Therefore, showing Unique Games hardness for expansion problems requires a new way of reducing a Unique Games instance to the expansion problem.

Currently, we do not know of any hardness results for Hypergraph Small Set Expansion and Small Set Vertex Expansion other than those that follow from the hardneses results for Small Set Expansion in graphs. Closing the gap between the upper and lower bounds for these problems is also an interesting open problem.

Obtaining optimal NP-hardness results for expansion problems is a fundamental open problem on which there has been very little progress. Surprisingly, determining whether the exact computation of λ_∞ (and γ_2) is NP-hard also remains open.

REFERENCES

[1] ALON, N., "Eigenvalues and expanders," *Combinatorica*, vol. 6, no. 2, pp. 83–96, 1986.

[2] ALON, N. and CHUNG, F. R., "Explicit construction of linear sized tolerant networks," *Annals of Discrete Mathematics*, vol. 38, pp. 15–19, 1988.

[3] ALON, N. and MILMAN, V. D., "λ_1, isoperimetric inequalities for graphs, and superconcentrators," *J. Comb. Theory, Ser. B*, vol. 38, no. 1, pp. 73–88, 1985.

[4] ALPERT, C. J. and KAHNG, A. B., "Recent directions in netlist partitioning: A survey," *Integration: The VLSI Journal*, vol. 19, pp. 1–81, 1995.

[5] ALPERT, C. J. and KAHNG, A. B., "Recent directions in netlist partitioning: a survey," *Integration, the VLSI journal*, vol. 19, no. 1, pp. 1–81, 1995.

[6] AMBÜHL, C., MASTROLILLI, M., and SVENSSON, O., "Inapproximability results for maximum edge biclique, minimum linear arrangement, and sparsest cut," *SIAM Journal on Computing*, vol. 40, no. 2, pp. 567–596, 2011.

[7] AMBÜHL, C., MASTROLILLI, M., and SVENSSON, O., "Inapproximability results for maximum edge biclique, minimum linear arrangement, and sparsest cut," *SIAM Journal on Computing*, vol. 40, no. 2, pp. 567–596, 2011.

[8] ARORA, S., BARAK, B., and STEURER, D., "Subexponential algorithms for unique games and related problems," in *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pp. 563–572, IEEE, 2010.

[9] ARORA, S. and GE, R., "New tools for graph coloring," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pp. 1–12, Springer, 2011.

[10] ARORA, S., KHOT, S. A., KOLLA, A., STEURER, D., TULSIANI, M., and VISHNOI, N. K., "Unique games on expanding constraint graphs are easy," in *Proceedings of the 40th annual ACM symposium on Theory of computing*, pp. 21–28, ACM, 2008.

[11] ARORA, S., LEE, J., and NAOR, A., "Euclidean distortion and the sparsest cut," *Journal of the American Mathematical Society*, vol. 21, no. 1, pp. 1–21, 2008.

[12] ARORA, S., RAO, S., and VAZIRANI, U., "Expander flows, geometric embeddings and graph partitioning," *Journal of the ACM (JACM)*, vol. 56, no. 2, p. 5, 2009.

[13] Arora, S., Rao, S., and Vazirani, U., "Expander flows, geometric embeddings and graph partitioning," *Journal of the ACM (JACM)*, vol. 56, no. 2, p. 5, 2009.

[14] Aumann, Y. and Rabani, Y., "An o($\log k$) approximate min-cut max-flow theorem and approximation algorithm," *SIAM Journal on Computing*, vol. 27, no. 1, pp. 291–301, 1998.

[15] Avis, D., "Diameter partitioning," *Discrete & Computational Geometry*, vol. 1, no. 1, pp. 265–276, 1986.

[16] Bansal, N., Feige, U., Krauthgamer, R., Makarychev, K., Nagarajan, V., Naor, J., and Schwartz, R., "Min-max graph partitioning and small set expansion," in *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pp. 17–26, IEEE, 2011.

[17] Barak, B., Raghavendra, P., and Steurer, D., "Rounding semidefinite programming hierarchies via global correlation," in *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pp. 472–481, IEEE, 2011.

[18] Barnard, S. T. and Simon, H. D., "Fast multilevel implementation of recursive spectral bisection for partitioning unstructured problems," *Concurrency: Practice and Experience*, vol. 6, no. 2, pp. 101–117, 1994.

[19] Ben-Tal, A. and Nemirovski, A., *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*, vol. 2. Siam, 2001.

[20] Biswal, P., Lee, J. R., and Rao, S., "Eigenvalue bounds, spectral partitioning, and metrical deformations via flows," *Journal of the ACM (JACM)*, vol. 57, no. 3, p. 13, 2010.

[21] Bobkov, S., Houdré, C., and Tetali, P., "λ_∞ vertex isoperimetry and concentration," *Combinatorica*, vol. 20, no. 2, pp. 153–172, 2000.

[22] Borell, C., "The brunn-minkowski inequality in gauss space," *Inventiones Mathematicae*, vol. 30, no. 2, pp. 207–216, 1975.

[23] Borell, C., "Geometric bounds on the ornstein-uhlenbeck velocity process," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 70, no. 1, pp. 1–13, 1985.

[24] Brubaker, S. C. and Vempala, S. S., "Random tensors and planted cliques," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pp. 406–419, Springer, 2009.

[25] Catalyurek, U. V. and Aykanat, C., "Hypergraph-partitioning-based decomposition for parallel sparse-matrix vector multiplication," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 10, no. 7, pp. 673–693, 1999.

[26] Celis, L. E., Devanur, N. R., and Peres, Y., "Local dynamics in bargaining networks via random-turn games," in *Internet and Network Economics*, pp. 133–144, Springer, 2010.

[27] Charikar, M., Makarychev, K., and Makarychev, Y., "Near-optimal algorithms for unique games," in *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pp. 205–214, ACM, 2006.

[28] Chawla, S., Gupta, A., and Räcke, H., "Embeddings of negative-type metrics and an improved approximation to generalized sparsest cut," in *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pp. 102–111, Society for Industrial and Applied Mathematics, 2005.

[29] Cheeger, J., "A lower bound for the smallest eigenvalue of the laplacian," *Problems in analysis*, vol. 625, pp. 195–199, 1970.

[30] Cheeger, J., Kleiner, B., and Naor, A., "A $(log n)^{\Omega(1)}$ integrality gap for the sparsest cut sdp," in *Foundations of Computer Science, 2009. FOCS'09. 50th Annual IEEE Symposium on*, pp. 555–564, IEEE, 2009.

[31] Chlamtac, E., Makarychev, K., and Makarychev, Y., "How to play unique games using embeddings," in *Foundations of Computer Science, 2006. FOCS'06. 47th Annual IEEE Symposium on*, pp. 687–696, IEEE, 2006.

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

[33] Chung, F., *Spectral Graph Theory.* American Mathematical Society, 1997.

[34] Cooper, J. and Dutle, A., "Spectra of uniform hypergraphs," *Linear Algebra and its Applications*, vol. 436, no. 9, pp. 3268–3292, 2012.

[35] Devanur, N. R., Khot, S. A., Saket, R., and Vishnoi, N. K., "Integrality gaps for sparsest cut and minimum linear arrangement problems," in *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pp. 537–546, ACM, 2006.

[36] Devine, K. D., Boman, E. G., Heaphy, R. T., Bisseling, R. H., and Catalyurek, U. V., "Parallel hypergraph partitioning for scientific computing," in *Parallel and Distributed Processing Symposium, 2006. IPDPS 2006. 20th International*, pp. 10–pp, IEEE, 2006.

[37] Dhillon, I. S., "Co-clustering documents and words using bipartite spectral graph partitioning," in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 269–274, ACM, 2001.

[38] Dinur, I., "The pcp theorem by gap amplification," *Journal of the ACM (JACM)*, vol. 54, no. 3, p. 12, 2007.

[39] Feige, U., Hajiaghayi, M., and Lee, J. R., "Improved approximation algorithms for minimum weight vertex separators," *SIAM Journal on Computing*, vol. 38, no. 2, pp. 629–657, 2008.

[40] Friedman, J. and Wigderson, A., "On the second eigenvalue of hypergraphs," *Combinatorica*, vol. 15, no. 1, pp. 43–65, 1995.

[41] Gharan, S. O. and Trevisan, L., "Partitioning into expanders.," in *SODA*, pp. 1256–1266, SIAM, 2014.

[42] Girard, P., Guiller, L., Landrault, C., and Pravossoudovitch, S., "Low power bist design by hypergraph partitioning: methodology and architectures," in *Test Conference, 2000. Proceedings. International*, pp. 652–661, IEEE, 2000.

[43] Goldreich, O., "Candidate one-way functions based on expander graphs.," *IACR Cryptology ePrint Archive*, vol. 2000, p. 63, 2000.

[44] Hendrickson, B. and Leland, R., "An improved spectral graph partitioning algorithm for mapping parallel computations," *SIAM Journal on Scientific Computing*, vol. 16, no. 2, pp. 452–469, 1995.

[45] Hillar, C. and Lim, L.-H., "Most tensor problems are np-hard," *arXiv preprint arXiv:0911.1393*, 2009.

[46] Hoory, S., Linial, N., and Wigderson, A., "Expander graphs and their applications," *Bulletin of the American Mathematical Society*, vol. 43, no. 4, pp. 439–561, 2006.

[47] Hu, S. and Qi, L., "The laplacian of a uniform hypergraph," *Journal of Combinatorial Optimization*, pp. 1–36, 2013.

[48] Hu, S. and Qi, L., "The eigenvectors associated with the zero eigenvalues of the laplacian and signless laplacian tensors of a uniform hypergraph," *Discrete Applied Mathematics*, 2014.

[49] Isaksson, M. and Mossel, E., "Maximally stable gaussian partitions with discrete applications," *Israel Journal of Mathematics*, vol. 189, no. 1, pp. 347–396, 2012.

[50] Kane, D. M. and Meka, R., "A prg for lipschitz functions of polynomials with applications to sparsest cut," in *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pp. 1–10, ACM, 2013.

[51] Kannan, R., Vempala, S., and Vetta, A., "On clusterings: Good, bad and spectral," *Journal of the ACM (JACM)*, vol. 51, no. 3, pp. 497–515, 2004.

[52] Karypis, G., Aggarwal, R., Kumar, V., and Shekhar, S., "Multilevel hypergraph partitioning: applications in vlsi domain," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 7, no. 1, pp. 69–79, 1999.

[53] Kelner, J. A., "Spectral partitioning, eigenvalue bounds, and circle packings for graphs of bounded genus," *SIAM Journal on Computing*, vol. 35, no. 4, pp. 882–902, 2006.

[54] Khot, S., "On the power of unique 2-prover 1-round games," in *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pp. 767–775, ACM, 2002.

[55] Khot, S., "Ruling out ptas for graph min-bisection, dense k-subgraph, and bipartite clique," *SIAM Journal on Computing*, vol. 36, no. 4, pp. 1025–1071, 2006.

[56] Khot, S. A. and Vishnoi, N. K., "The unique games conjecture, integrality gap for cut problems and embeddability of negative type metrics into l_1," in *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on*, pp. 53–62, IEEE, 2005.

[57] Krauthgamer, R., Naor, J. S., and Schwartz, R., "Partitioning graphs into balanced components," in *Proceedings of the twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 942–949, Society for Industrial and Applied Mathematics, 2009.

[58] Krauthgamer, R. and Rabani, Y., "Improved lower bounds for embeddings into l_1," *SIAM Journal on Computing*, vol. 38, no. 6, pp. 2487–2498, 2009.

[59] Kuehn, A. A. and Hamburger, M. J., "A heuristic program for locating warehouses," *Management science*, vol. 9, no. 4, pp. 643–666, 1963.

[60] Kumar, V., Grama, A., Gupta, A., and Karypis, G., *Introduction to parallel computing*, vol. 110. Benjamin/Cummings Redwood City, 1994.

[61] Lasserre, J. B., "An explicit exact sdp relaxation for nonlinear 0-1 programs," in *IPCO 2001*, pp. 293–303, 2001.

[62] Ledoux, M. and Talagrand, M., *Probability in Banach Spaces*. Springer, 1991.

[63] Lee, J. R., OveisGharan, S., and Trevisan, L., "Multi-way spectral partitioning and higher-order cheeger inequalities," in *Proceedings of the 44th symposium on Theory of Computing*, pp. 1117–1130, ACM, 2012.

[64] Leighton, T. and Rao, S., "Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms," *Journal of the ACM (JACM)*, vol. 46, no. 6, pp. 787–832, 1999.

[65] Lenz, J. and Mubayi, D., "Eigenvalues and quasirandom hypergraphs," *arXiv preprint arXiv:1208.4863*, 2012.

[66] Lenz, J. and Mubayi, D., "Eigenvalues and linear quasirandom hypergraphs," 2013.

[67] Lenz, J. and Mubayi, D., "Eigenvalues of non-regular linear quasirandom hypergraphs," *arXiv preprint arXiv:1309.3584*, 2013.

[68] Linial, N., London, E., and Rabinovich, Y., "The geometry of graphs and some of its algorithmic applications," *Combinatorica*, vol. 15, no. 2, pp. 215–245, 1995.

[69] Louis, A., "Hypergraph markov operators, eigenvalues and approximation algorithms," *Manuscript*, 2014.

[70] Louis, A. and Makarychev, K., "Approximation algorithm for sparsest k-partitioning," in *SODA 2014*, vol. 12, 2014.

[71] Louis, A. and Makarychev, Y., "Approximation algorithms for hypergraph small set expansion and small set vertex expansion," *arXiv preprint arXiv:1404.4575*, 2014.

[72] Louis, A., Raghavendra, P., Tetali, P., and Vempala, S., "Algorithmic extensions of cheegers inequality to higher eigenvalues and partitions," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pp. 315–326, Springer, 2011.

[73] Louis, A., Raghavendra, P., Tetali, P., and Vempala, S., "Many sparse cuts via higher eigenvalues," in *Proceedings of the 44th symposium on Theory of Computing*, pp. 1131–1140, ACM, 2012.

[74] Louis, A., Raghavendra, P., and Vempala, S., "Personal communication," 2012.

[75] Louis, A., Raghavendra, P., and Vempala, S., "The complexity of approximating vertex expansion," in *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pp. 360–369, IEEE, 2013.

[76] Lovasz, L. and Schrijver, A., "Cones of matrices and set-functions and 0-1 optimization," in *SIAM J. on Optimization 1(12)*, pp. 166–190, 1991.

[77] Makarychev, K. and Makarychev, Y., "Approximation algorithm for non-boolean max k-csp," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pp. 254–265, Springer, 2012.

[78] Margulis, G. A., "Explicit group-theoretical constructions of combinatorial schemes and their application to the design of expanders and concentrators," *Problemy peredachi informatsii*, vol. 24, no. 1, pp. 51–60, 1988.

[79] Matula, D. W. and Shahrokhi, F., "Sparsest cuts and bottlenecks in graphs," *Discrete Applied Mathematics*, vol. 27, no. 1, pp. 113–123, 1990.

[80] Mossel, E., O'Donnell, R., and Oleszkiewicz, K., "Noise stability of functions with low influences: invariance and optimality," *FOCS 2005*, vol. 171, no. 1, pp. 295–341, 2005.

[81] Nemirovski, A. and ËIĕUdin, D., *Problem complexity and method efficiency in optimization*. Wiley (Chichester and New York), 1983.

[82] Parzanchevski, O., "Mixing in high-dimensional expanders," *arXiv preprint arXiv:1310.6477*, 2013.

[83] Parzanchevski, O. and Rosenthal, R., "Simplicial complexes: spectrum, homology and random walks," *arXiv preprint arXiv:1211.6775*, 2012.

[84] Parzanchevski, O., Rosenthal, R., and Tessler, R. J., "Isoperimetric inequalities in simplicial complexes," *arXiv preprint arXiv:1207.0638*, 2012.

[85] Peres, Y., Schramm, O., Sheffield, S., and Wilson, D., "Tug-of-war and the infinity laplacian," *Journal of the American Mathematical Society*, vol. 22, no. 1, pp. 167–210, 2009.

[86] Raghavendra, P. and Steurer, D., "Graph expansion and the unique games conjecture," in *Proceedings of the 42nd ACM symposium on Theory of computing*, pp. 755–764, ACM, 2010.

[87] Raghavendra, P., Steurer, D., and Tetali, P., "Approximations for the isoperimetric and spectral profile of graphs and related parameters," in *Proceedings of the 42nd ACM symposium on Theory of computing*, pp. 631–640, ACM, 2010.

[88] Raghavendra, P., Steurer, D., and Tulsiani, M., "Reductions between expansion problems," in *Computational Complexity (CCC), 2012 IEEE 27th Annual Conference on*, pp. 64–73, IEEE, 2012.

[89] Raghavendra, P. and Tan, N., "Approximating csps with global cardinality constraints using sdp hierarchies," in *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 373–387, SIAM, 2012.

[90] Rodrguez, J., "Laplacian eigenvalues and partition problems in hypergraphs," *Applied Mathematics Letters*, vol. 22, no. 6, pp. 916 – 921, 2009.

[91] Saran, H. and Vazirani, V. V., "Finding k cuts within twice the optimal," *SIAM Journal on Computing*, vol. 24, no. 1, pp. 101–108, 1995.

[92] Shalev-Shwartz, S., "Online learning and online convex optimization," *Foundations and Trends in Machine Learning*, vol. 4, no. 2, pp. 107–194, 2011.

[93] Sherali, H. D. and Adams, W. P., "A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems," in *SIAM J. Discrete Math. 3(3)*, pp. 411–430, 1990.

[94] Shi, J. and Malik, J., "Normalized cuts and image segmentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 8, pp. 888–905, 2000.

[95] Sinclair, A. and Jerrum, M., "Approximate counting, uniform generation and rapidly mixing markov chains," *Information and Computation*, vol. 82, no. 1, pp. 93–133, 1989.

[96] Sinclair, A. and Jerrum, M., "Approximate counting, uniform generation and rapidly mixing markov chains," *Information and Computation*, vol. 82, no. 1, pp. 93–133, 1989.

[97] Sipser, M. and Spielman, D. A., "Expander codes," *IEEE Transactions on Information Theory*, vol. 42, pp. 1710–1722, 1996.

[98] Spielmat, D. A. and Teng, S.-H., "Spectral partitioning works: planar graphs and finite element meshes," in *Foundations of Computer Science, 1996. Proceedings., 37th Annual Symposium on*, pp. 96–105, IEEE, 1996.

[99] Steenbergen, J., Klivans, C., and Mukherjee, S., "A cheeger-type inequality on simplicial complexes," *arXiv preprint arXiv:1209.5091*, 2012.

[100] Sudakov, V. N. and Tsirel'son, B. S., "Extremal properties of half-spaces for spherically invariant measures," *Journal of Mathematical Sciences*, vol. 9, no. 1, pp. 9–18, 1978.

[101] Tanaka, M., "Higher eigenvalues and partitions of a graph," tech. rep., 2011.

[102] Trevisan, L., "Approximation algorithms for unique games," in *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on*, pp. 197–205, IEEE, 2005.