Diversity Embeddings and the Hypergraph Sparsest Cut

by

Adam Daniel Jozefiak

B.Sc., University of British Columbia, 2020

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Computer Science)

The University of British Columbia
(Vancouver)

August 2022

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

Diversity Embeddings and the Hypergraph Sparsest Cut

submitted by **Adam Daniel Jozefiak** in partial fulfillment of the requirements for the degree of **Master of Science** in **Computer Science**.

Examining Committee:

Bruce Shepherd, Associate Head of Faculty Affairs and Professor, Computer Science, UBC
Supervisor

Nicholas J.A. Harvey, Professor, Computer Science, UBC
Supervisory Committee Member

Abstract

Good approximations have been attained for the sparsest cut problem by rounding solutions to convex relaxations via low-distortion metric embeddings [5, 24]. Recently, Bryant and Tupper showed that this approach extends to the hypergraph setting by formulating a linear program whose solutions are diversities which are rounded via diversity embeddings into ℓ_1 [31]. Diversities are a generalization of metric spaces in which the nonnegative function is defined on all subsets as opposed to only on pairs of elements.

We show that this approach yields an $O(\log n)$-approximation when either the supply or demands are given by a graph. This result improves upon Plotkin et al.'s $O(\log(kn)\log n)$-approximation [28], where k is the number of demands, in the setting where the supply is given by a graph and the demands are given by a hypergraph. Additionally, we provide an $O(\min\{r_G, r_H\}\log(r_H)\log n)$-approximation for when the supply and demands are given by hypergraphs whose hyperedges are bounded in cardinality by r_G and r_H respectively.

To establish these results we provide an $O(\log n)$-distortion ℓ_1 embedding for the class of diversities known as diameter diversities. This improves upon Bryant and Tupper's $O((\log n)\hat{}2)$-distortion embedding [31]. The smallest known distortion with which an arbitrary diversity can be embedded into ℓ_1 is $O(n)$. We show that for any $\varepsilon > 0$ and any $p > 0$, there is a family of diversities which cannot be embedded into ℓ_1 in polynomial time with distortion smaller than $O(n\hat{}(1-\varepsilon))$ based on querying the diversities on sets of cardinality at most $O((\log n)\hat{}p)$, unless $P = NP$. This disproves (an algorithmic refinement) of Bryant and Tupper's conjecture that there exists an $O(\sqrt{n})$-distortion ℓ_1 embedding based off a diversity's induced metric. In addition, we demonstrate via hypergraph cut sparsifiers that it is sufficient to develop a low-distortion embedding for diversities induced by sparse hypergraphs for the purpose of obtaining good approximations for the sparsest cut in hypergraphs.

Lay Summary

Given a network with pairwise capacities and demands, a fundamental problem is to compute its bottleneck or *sparsest cut*. A classic approach is to formulate a convex relaxation whose solutions form a metric on the nodes. Cut information is then extracted via embedding this metric into ℓ_1. In spite of the powerful ability of hypergraphs to model multiway relationships, the sparsest cut has received relatively little attention in this setting. Extending the approach of metric embeddings to the hypergraph sparsest cut yields a convex relaxation whose solutions are nonnegative set-functions that satisfy a certain type of triangle-inequality. These are known as diversities. We show that this approach of diversity embeddings is tractable and guarantees good approximations when either the supply or demand hyperedges are small in cardinality. We also make progress on open questions in the theory of ℓ_1 diversity embeddings.

Preface

This thesis is an original intellectual product of the author, Adam Daniel Jozefiak, with the guidance and mentorship of Bruce Shepherd. In addition, much of this thesis includes the restatement of past results which are necessary for presenting our original contributions. Much of this includes the work of Bryant and Tupper, who introduced and formulated the definition of a diversity, along with other seminal results of Linial, London, and Rabinovich, and Aumann and Rabani, and Bourgain, to name but a few of the giants who have led us to where we have begun from.

Table of Contents

Acknowledgments

First and foremost, I would like to thank my supervisor Bruce Shepherd for his guidance and mentorship, not merely limited to this thesis. I would not be in my current position without his support. During the course of my graduate studies, I have grown and acquired valuable wisdom that I could not foresee, and for this I have the utmost gratitude.

I would also like to thank Nick Harvey, not only for his participation in the examination committee of this thesis, but for his guidance and teaching over the last few years. Together, Bruce, Nick, and Hu Fu are responsible for my training in the fields of algorithms, optimization, and theoretical computer science. This thesis stands as a testament to this education and training.

I am also indebted in gratitude to Jesse Perla for his guidance and mentorship. I am grateful and fortunate for the opportunity to have worked with Jesse on research in computational economics. If it were not for Bruce and Jesse's thoughtfulness and their advice to consider the path of operations research and the business school I would not be embarking on my journey to MIT.

Looking further into the past, thank you Peggy Ament for believing in me at a time when few others could. I would also like to thank my parents for their unconditional love. My family raised me to be who I am today and they continue to support me to this day.

Most of all, I would like to thank Simora for her unconditional love and support. Thank you for bearing with me during the process of writing this thesis. .

Dedication

To my love Simora,

I promise that your sacrifices will be made worthwhile a thousand times over.

Chapter 1

Introduction

1.1 Introduction

The sparsest cut problem is a fundamental problem in theoretical computer science. In this thesis we consider the sparsest cut problem in the general hypergraph setting. That is, the problem is defined by two hypergraphs, a supply hypergraph and a demand hypergraph.

We let $G = (V, E_G, w_G)$ be a hypergraph with node set V, hyperedge set E_G, and nonnegative hyperedge weights $w_G : E_G \to \mathbb{R}_+$ and $H = (V, E_H, w_H)$ be a hypergraph with node set V, hyperedge set E_H, and nonnegative hyperedge weights $w_H : E_H \to \mathbb{R}_+$. We refer to G and H as the *supply hypergraph* and the *demand hypergraph*, respectively. We use the notation r_G for the rank of G and r_H for the rank H where we define the rank of a hypergraph as follows

Definition 1.1.1 (Hypergraph Rank) *Let $H = (V, E)$ be a hypergraph with node set V and hyperedge set E. We say that H has rank k if the maximum cardinality of a hyperedge of H is k. Or, more formally the rank is defined as*

$$\max\{|U| : U \in E\} = k \tag{1.1}$$

A hypergraph of rank two is simply a graph. In settings where G or H is a graph we state explicitly that G is a *supply graph* and H is a *demand graph*, respectively.

Then we define the sparsity of a cut as follows

Definition 1.1.2 *Let $G = (V, E_G, w_G)$ be a supply hypergraph and $H = (V, E_H, w_H)$ be a demand hypergraph. For $A \subseteq V$ such that $A \neq \emptyset, V$, the sparsity of the cut defined by A, which we denote by $\phi(A)$, is defined as*

$$\phi(A) = \frac{\sum_{U \in E_G} w_G(U) \mathbb{1}_{\{U \cap A \neq \emptyset, U\}}}{\sum_{S \in E_H} w_H(S) \mathbb{1}_{\{S \cap A \neq \emptyset, S\}}} \tag{1.2}$$

where $\mathbb{1}$ is the indicator variable.

$$\mathbb{1}_p = \begin{cases} 1 & \text{if } p = \textit{true} \\ 0 & \text{if } p = \textit{false} \end{cases} \tag{1.3}$$

The sparsest cut of G and H is defined as

Definition 1.1.3 *Let $G = (V, E_G, w_G)$ be a supply hypergraph and $H = (V, E_H, w_H)$ be a demand hypergraph. Then the sparsity of the sparsest cut of G and H, denoted by ϕ, is defined as*

$$\phi = \min_{A \subseteq V: A \neq \emptyset, V} \phi(A) = \min_{A \subseteq V: A \neq \emptyset, V} \frac{\sum_{U \in E_G} w_G(U) \mathbb{1}_{\{U \cap A \neq \emptyset, U\}}}{\sum_{S \in E_H} w_H(S) \mathbb{1}_{\{S \cap A \neq \emptyset, S\}}} \tag{1.4}$$

Additionally, we refer to an argmin of (1.4) as the sparsest cut of G and H. If the underlying hypergraphs are ambiguous we may use the notation $\phi_{G,H}$ and $\phi_{G,H}(A)$.

Computing the sparsest cut is NP-hard, even for the case where G and H are graphs [27]. Consequently, there is a rich history of approximation algorithms for the sparsest cut problem in the graph setting which has culminated in an $O(\sqrt{\log n} \log \log n)$-approximation factor for general supply and demand graphs [3]. A common approach to achieve such bounds begins with formulating a convex relaxation, such as a linear program (LP) or a semidefinite program (SDP), and then rounding its optimal solution to obtain an integral, but approximate, sparsest cut. Often, solutions of these convex relaxations form a metric space for the nodes V and the rounding step involves embedding this metric into another metric space such as the ℓ_1 or the ℓ_2 metric.

Extending such metric-relaxations to the hypergraph setting results in a convex relaxation whose solutions are vectors δ which assign nonnegative values to arbitrary subsets as opposed to only pairwise distances $d(u, v)$ from a metric space (V, d). The vector δ does satisfy certain triangle inequalities and the ordered tuples (V, δ) are termed “diversities” according to Bryant and Tupper [10]. Analogous to the approach based on metric embeddings into ℓ_1 [5, 24], one can extract an approximate sparsest cut in the hypergraph setting via diversity embeddings into ℓ_1. The focus of this thesis is largely concerned with low-distortion embeddings of diversities into ℓ_1 and their application to the sparsest cut in hypergraphs.

In Section 1.2 we summarize our contributions to this topic, in Section 1.3 we provide an overview of the history and related work of this problem, in both the graph and hypergraph settings. In Section 1.4 we overview the organization and structure of this thesis.

1.2 Overview of Our Results

In this section we list our contributions as theorem and conjecture statements.

1.2.1 Approximating the Sparsest Cut in Hypergraphs

Our contributions for the sparest cut problem are in the setting where the supply and demand hypergraphs are in general hypergraphs, namely Theorem 1.2.1. This more general hypergraph setting has received relatively little investigation except for recent work in the setting where G is a supply hypergraph and H is a demand graph [19, 25, 26].

Theorem 1.2.1 *Let $G = (V, E_G, w_G)$ be a supply hypergraph with rank r_G and $H = (V, E_H, w_H)$ be a demand hypergraph with rank r_H. Then there is a randomized polynomial-time $O(\min\{r_G, r_H\} \log n \log r_H)$-approximation algorithm for the sparsest cut of G and H.*

Although the approach of Theorem 1.2.1 is introduced by Bryant and Tupper in [31], we present the first randomized polynomial-time implementation of this approach. We next consider the case where the supply is a graph, but the demands arise from a hypergraph.

Theorem 1.2.2 *Let $G = (V, E_G, w_G)$ be a supply graph, that is it has rank $r_G = 2$, and $H = (V, E_H, w_H)$ be a demand hypergraph. Then there is a randomized polynomial-time $O(\log n)$-approximation algorithm for the sparsest cut of G and H.*

As for this setting where G is a supply graph and H is a demand hypergraph, Theorem 1.2.2 is an improvement upon an existing polynomial time $O(\log n \log(|E_H| r_H))$-approximation algorithm due to Plotkin et al. [28]. Specifically, the approximation factor guaranteed by Plotkin et al.'s algorithm has a logarithmic dependence on $|E_H|$, which is exponentially large in general, while Theorem 1.2.2 truly guarantees a polylogarithmic approximation factor.

1.2.2 Diversity Embeddings

We shall see that Theorem 1.2.1 relies critically on the following embedding result. It is also an improvement upon the existing $O(\log^2 n)$-distortion embedding of a diameter diversity into the ℓ_1 diversity, $(\mathbb{R}^{O(\log n)}, \delta_1)$, [31]. See Definition 2.2.7 for the definition of a diameter diversity.

Theorem 1.2.3 *Let (X, δ_{diam}) be a diameter diversity with induced metric (X, d) and with $|X| = n$. Then there exists a randomized polynomial time embedding of (X, δ_{diam}) into the ℓ_1 diversity $(\mathbb{R}^{O(\log^2 n)}, \delta_1)$ with distortion $O(\log n)$.*

In contrast, we also give bad news in terms of embedding a general diversity.

Theorem 1.2.4 *For any $p \geq 0$ and for any $\varepsilon > 0$, there does not exist a polynomial-time diversity ℓ_1 embedding that queries a diversity on sets of cardinality at most $O(\log^p n)$ and achieves a distortion of $O(n^{1-\varepsilon})$, unless P=NP.*

Theorem 1.2.4 provides an inapproximability result for embedding diversities into ℓ_1 based only on their induced metric space. This answers an algorithmic-refinement of Bryant and Tupper's conjecture that there exists an $O(\sqrt{n})$-distortion embedding into ℓ_1 solely using the induced metric of a diversity [9], see Conjecture 3.4.2.

1.2.3 The Minimum Cost Hypergraph Steiner Problem

Theorem 1.2.5 *There exists a polynomial time $O(\log n)$-approximation algorithm for the minimum cost hypergraph Steiner problem. Specifically, for a hypergraph $G = (V, E, w)$ with nonnegative hyperedge weights $w : E \to \mathbb{R}_+$ and for a set of Steiner nodes $S \subseteq V$, $HSP(G, S)$ can be approximated up to a factor of $O(\log |S|)$ in polynomial time.*

To our knowledge, this is the first polynomial-time approximation algorithm for the minimum-cost hypergraph Steiner problem, a generalization of the classic Steiner tree and the minimum-cost spanning subhypergraph problems. For our purposes, this approximation algorithm provides a critical subroutine for Theorem 1.2.1.

1.2.4 A Sparse Diversity Embedding Conjecture

Conjecture 1.2.6 *Let $H = (V, E, w)$ be a hypergraph with node set V, hyperedge set E, and nonnegative hyperedges weights $w : E \to \mathbb{R}_+$, where $|V| = n$ and $|E| = m$. Let $(V, \delta_{\mathscr{H}})$ be the corresponding hypergraph Steiner diversity defined by H. Then for some $p > 0$ there is an $O(\log^p(n+m))$-distortion embedding of $(V, \delta_{\mathscr{H}})$ into the ℓ_1 diversity.*

We conjecture that a diversity embedding into ℓ_1 with polylogarithmic distortion in the number of hyperedges that define the diversity is possible. This conjecture is motivated by the fact that hypergraph cut sparsifiers allow us to replace an instance of the sparsest cut with dense hypergraphs by an instance with sparse hypergraphs in polynomial time and with an arbitrarily small approximation factor. Then the diversity that arises as the solution to the sparsest-cut convex-relaxation is one that is defined by a sparse hypergraph. We believe that this sparse structure may give rise to low-distortion embeddings, hence our Conjecture 1.2.6.

1.3 Related Work and History

A closely related property to the sparsity of a cut is the *expansion* of the cut. This is defined as the ratio of the weight of the (hyper)edges crossing the cut to the number of nodes in the smaller partition of the cut. Formally,

Definition 1.3.1 *Let $G = (V, E, w)$ be a (hyper)graph with nonnegative hyperedge weights $w : E \to \mathbb{R}_+$ and let $A \subseteq V$ such that $A \neq \emptyset, V$. Then the expansion of the cut A, $\Phi(A)$, is defined as*

$$\Phi(A) = \frac{\sum_{U \in E} w(U) \mathbb{1}_{\{U \cap A \neq \emptyset, U\}}}{\min\{|A|, |A^C|\}} \tag{1.5}$$

Then the (hyper)graph expansion of G is defined as

Definition 1.3.2 *Let $G = (V, E)$ be (hyper)graph with nonnegative hyperedge weights $w : E \to \mathbb{R}_+$. Then the expansion of the (hyper)graph G, denoted by Φ, is defined as*

$$\Phi = \min_{A \subseteq V : A \neq \emptyset, V} \Phi(A) = \min_{A \subseteq V : A \neq \emptyset, V} \frac{\sum_{U \in E} w(U) \mathbb{1}_{\{U \cap A \neq \emptyset, U\}}}{\min\{|A|, |A^C|\}} \tag{1.6}$$

Often, when discussing (hyper)graph expansion, the supply (hyper)graph G has unit weight and the expansion of a cut is simply the ratio of the number of (hyper)edges crossing the cut to the number of nodes in the smaller partition of the cut. The expansion of a cut is closely related to the sparsity of the cut when the demand network is a complete graph with unit weights, which we call *uniform demands*. The problem of computing the minimum expansion cut, even with uniform demands, is NP-hard [27].

Definition 1.3.3 *A demand graph $H = (V, E_H, w_H)$ is said to be a uniform demand graph if H is a complete graph with unit weights, that is for each distinct pair of elements $u, v \in V$, $w_H(\{u,v\}) = 1$. Furthermore, it is easy to verify that for $A \subseteq V$ we have that*

$$\sum_{S \in E_H} w_H(S) \mathbb{1}_{\{S \cap A \neq \emptyset, S\}} = |A||A^C| \tag{1.7}$$

Thus for a supply hypergraph G and a uniform demand graph H, we have for each $A \subseteq V$ with $|A| \leq |A^C|$, that $\frac{n}{2}|A| \leq |A||A^C| \leq n|A|$. Hence up to a factor of 2, approximability of the sparsest cut and the minimum expansion cut is equivalent. In particular, existence of an $O(\alpha)$-approximation algorithm for the sparsest cut is also an $O(\alpha)$-approximation algorithm for the (hyper)graph expansion problem and vice-versa. Many of the breakthroughs for the sparsest cut problem are in this specific setting of uniform demands due to this close connection with (hyper)graph expansion. Expander graphs have played a ubiquitous role in theoretical computer science and consequently approximation algorithms for computing the cut with minimal expansion are critical to tasks like verifying whether a graph is an expander and constructing spectral sparsifiers [19].

We begin with discussing the history of the sparsest cut problem in the setting where G and H are both graphs. Leighton and Rao gave the first approximation algorithm for the sparsest cut problem with uniform demands, achieving an $O(\log n)$ approximation factor [23]. Specifically, Leighton and Rao showed that the flow-cut gap for the uniform demand multicommodity flow problem and the sparsest cut problem with uniform demands is $\Theta(\log n)$, allowing algorithms for the multicommodity flow problem to be employed as approximation algorithms for the sparsest cut problem in this setting. Moreover, they showed that the flow-cut gap in this setting is $\Omega(\log n)$ due to an example where the supply graph is a unit-capacity constant-degree expander graph. The Leighton and Rao algorithm is an LP approach and their $\Omega(\log n)$ integrality gap necessitated the use of more powerful tools to improve the approximability of the sparsest cut problem, even for the uniform demands setting (see further below).

Later, Linial, London, and Rabinovich [24] and Aumann and Rabani [5] generalized Leighton and Rao's flow-cut gap result to the setting of an arbitrary demand graph. Specifically, Linial et al. showed that the flow-cut gap is bounded by the minimum distortion of embedding a finite metric space into the ℓ_1 metric. According to Bourgain, the minimum distortion of embedding a finite metric into ℓ_1 is $O(\log n)$. However, before [24] it remained an open question as to whether this distortion factor is tight. Linial et al.'s result coupled with Leighton and Rao's $\Omega(\log n)$ flow-cut gap lower bound definitively answered this question. Furthermore, Linial et al. were the first to give a randomized polynomial time implementation of Bourgain's $O(\log n)$-distortion ℓ_1 embedding. From their geometric analysis of graphs, Linial et al. gave an approximation algorithm for the sparsest cut problem via embedding finite metric spaces into the ℓ_1 metric, producing the first $O(\log n)$ approximation factor for the sparsest cut problem for arbitrary supply and demand graphs.

One can strengthen the metric relaxation of the sparsest cut problem, introduced by Linial et al., by adding negative-type metric constraints through the use of an SDP relaxation. (X, d) is a *metric of the negative-type* if $(X, \sqrt{d})$ is a subset of the ℓ_2 metric space. We use the notation ℓ_2^2 for a metric of the negative-type. The minimum distortion of embedding a finite metric of the negative-type into the ℓ_1 metric space yields the integrality gap for this SDP approach. Furthermore, a polynomial time algorithm that embeds a finite metric of the negative-type into the ℓ_1 metric with low distortion yields a polynomial time approximation algorithm for the sparsest cut problem.

On this front, Arora, Rao, and Vazirani gave a polynomial time approximation algorithm for the sparsest cut problem with uniform demands that attains an approximation factor of $O(\sqrt{\log n})$ [4]. However, Arora et al.'s approach is a randomized rounding scheme of their SDP relaxation as opposed to directly embedding an arbitrary ℓ_2^2 metric into ℓ_1.

Later, Chawla, Gupta, and Räcke proved that a finite ℓ_2^2 metric can indeed be embedded into the ℓ_1 metric with distortion $O(\log^{\frac{3}{4}} n)$ and they gave a polynomial time algorithm for achieving such a metric embedding [13]. For arbitrary demand graphs, this improved the $O(\log n)$ approximation factor for the sparsest cut problem due to Linial et al.'s metric embedding technique to $O(\log^{\frac{3}{4}} n)$. Later, Arora, Lee, and Naor [3] improved this result by showing that a finite ℓ_2^2 metric can be embedded in the ℓ_1 metric space with distortion $O(\sqrt{\log n}\log\log n)$ and in turn they gave a polynomial time approximation algorithm to the generalized sparsest cut problem that achieves this approximation factor [3]. This metric embedding result is tight up to a factor of $O(\log\log n)$.

Next ,we discuss recent work on the sparsest cut problem in the setting where G is a supply hypergraph and H is a demand graph. In this setting, Kapralov et al. in [19] gave a polynomial time approximation algorithm for the hypergraph expansion problem and the sparsest cut problem with uniform demands that yields an $O(\log n)$ approximation factor. Their algorithm is a rounding scheme of an LP relaxation of the hypergraph expansion problem that optimizes over pseudo metrics. We note that Kapralov et al.'s result comes after Louis and Makarychev's $O(\sqrt{\log n})$ approximation factor in [26], however Kapralov et al.'s approximation algorithm is an LP approach whereas Louis and Makarychev's approximation algorithm is an SDP approach. Kapralov et al.'s work is motivated by the construction of spectral sparsifiers for hypergraphs and they require an approximation algorithm for the hypergraph expansion problem in order to extract cuts with approximate minimal expansion.

Louis and Makarychev in [26] obtained a randomized polynomial time approximation algorithm with an approximation factor $O(\sqrt{\log n})$ for the hypergraph expansion problem, and in turn, for the sparsest cut problem with uniform demands. This approximation factor matches the approximation factor for graph expansion in the setting where G is a graph and in turn the sparsest cut with uniform demands, obtained by Arora, Rao, and Vazirani [4]. Louis and Makarychev's approach is a randomized rounding scheme of an SDP relaxation of the sparsest cut problem.

Similar to the progression of [4] to [3], in this setting where G is a supply hypergraph and H is an arbitrary (non-uniform) demand graph, Louis gave a randomized polynomial time approximation algorithm for the sparsest cut problem that obtains an $O(\sqrt{\log r_G \log n}\log\log n)$ approximation factor [25]. Louis' technique is inspired by [3], whereby Louis' algorithm solves an SDP relaxation of the sparsest cut problem, with negative type metric constraints, where its solution, a metric of the negative type, is embedded into ℓ_2 with distortion $O(\sqrt{\log n}\log\log n)$ according to [3]'s metric embedding result. Finally, Louis' algorithm performs randomized rounding to obtain an approximate sparsest cut, incurring an additional $O(\sqrt{\log r_G})$ approximation factor.

As for the setting where G is a supply graph and H is a demand hypergraph, Plotkin et al. provided a polynomial time $O(\log n \log(|E_H| r_H))$-approximation algorithm. Plotkin et al.'s approach rounds a fractional solution of an LP relaxation without the use of metric embeddings. To our knowledge, there is no polynomial-time algorithm in this setting that utilizes metric embeddings. Furthermore, we are not aware of any polynomial-time approach in the more general setting where G and H are arbitrary supply and demand hypergraphs.

Recently introduced by Bryant and Tupper [10], diversities are a generalization of metric spaces where instead of a nonnegative function defined on pairs of elements, it is defined on arbitrary finite sets of elements. Additionally, Bryant and Tupper have developed a substantial theory on diversities [9–11, 31] including the notion of a diversity embedding. They have demonstrated that several types of diversities attain polynomial-time low-distortion embeddings into ℓ_1. Most pertinent to our discussion, in [31] they generalized the work of Linial et al. [24] whereby they show that the flow-cut gap in the hypergraph setting is equivalent to the minimum distortion of embedding some finite diversity into the ℓ_1 diversity. Notably, this work provides an approach to approximating the sparsest cut in hypergraphs. However, Bryant and Tupper did not expand their investigation into an algorithmic direction, leaving open questions of whether this approach is even polynomial-time tractable in the hypergraph setting.

1.4 Organization of this Thesis

In Chapter 2 we introduce the notion of a diversity, present properties of diversities, and define several diversities that are pertinent to our discussion and results in this thesis.

In Chapter 3 we introduce the notion of embedding one diversity into another along with several properties of embedding into ℓ_1. This chapter also includes polynomial-time low-distortion ℓ_1 embeddings of several classes of diversities. Notably, this chapter includes proofs of Theorem 1.2.3, in Section 3.3, and Theorem 1.2.4, in Section 3.4.

In Chapter 4 we present the approach of approximating the hypergraph sparsest cut via diversity embeddings. Specifically, in Section 4.1 we provide an LP relaxation for the sparsest cut and show how we can extract an approximate sparsest cut via diversity embeddings. In Section 4.2 we characterize the optimal solutions of this LP relaxation which is necessary for the proofs of Theorems 1.2.1 and 1.2.2 in the subsequent Section 4.3.

In Chapter 5 we introduce the minimum-cost hypergraph Steiner problem and present a proof of Theorem 1.2.5.

Finally, in Chapter 6 we introduce the notion of a hypergraph cut-sparsifier and show how it can be used to replace a dense instance of the sparsest cut problem with a sparse instance. We conclude this chapter with Conjecture 1.2.6.

Chapter 2

Diversities

2.1 Introduction

Recently introduced by Bryant and Tupper [10], diversities are a generalization of metric spaces where instead of a nonnegative function defined on pairs of elements, it is defined on arbitrary finite sets of elements. In this section we introduce the notion of a diversity along with some of their relevant properties. In the subsequent section we define several diversities that are pertinent to our discussion and results with regards to the sparsest cut problem in hypergraphs. Additionally, in Section 2.3 we show when and how a subadditive set function yields a diversity. This is a necessary technical lemma for the proof of Theorem 3.4.3, our inapproximability result for diversity embeddings into ℓ_1. We begin by formally defining a diversity.

Definition 2.1.1 *Let X be a set, then the collection of finite subsets of X, $\mathscr{P}(X)$, is defined as*

$$\mathscr{P}(X) = \{A \subseteq X : \ |A| \text{ is finite}\}. \tag{2.1}$$

Definition 2.1.2 *A diversity is a pair (X, δ) where X is a set and δ is a real-valued function defined over the finite subsets of X satisfying the following three axioms:*

1. $\forall A \in \mathscr{P}(X), \delta(A) \geq 0$
2. $\delta(A) = 0$ *if and only if* $|A| \leq 1$
3. $\forall A, B, C \in \mathscr{P}(X), C \neq \emptyset \Rightarrow \delta(A \cup B) \leq \delta(A \cup C) + \delta(B \cup C)$

Similar to the notion of a pseudo-metric there is a definition of a pseudo diversity. Formally, (V, δ) is a *pseudo diversity* if (X, δ) satisfies the three axioms of Definition 2.1.2 *with* the second axiom being weakened to

$$\delta(A) = 0 \text{ if } |A| \leq 1 \tag{2.2}$$

We refer to the third axiom of Definition 2.1.2 as the *triangle inequality*, and it is the property that makes (pseudo) diversities a generalization of (pseudo) metrics.

From this point onward, when we refer to (X,δ) as being a "diversity" we mean that (X,δ) is a pseudo diversity (and may not necessarily be a diversity). We make this assumption since in our approach to the sparsest cut the LP relaxations optimize over pseudo diversities.

We note that any (pseudo) diversity (X,δ) yields an induced (pseudo) metric space (X,d) where for any $x,y \in X$ we define d as

$$d(x,y) = \delta(\{x,y\}) \tag{2.3}$$

Conversely, a metric can induce an infinite number of diversities for whom it is the induced metric.

Definition 2.1.3 **($\mathscr{D}_{(X,d)}$)** *For a metric space (X,d), $\mathscr{D}_{(X,d)}$ is the family of diversities for whom (X,d) is their induced metric. For contexts where the set X is unambiguous the notation $\mathscr{D}_d$ may be used.*

A consequence of the triangle inequality is that (pseudo) diversities are monotone increasing set functions.

Proposition 2.1.4 *(Pseudo) diversities are monotone increasing. That is, if (X,δ) is a (pseudo) diversity then for any $A,B \in \mathscr{P}(X)$ we have that*

$$A \subseteq B \Rightarrow \delta(A) \leq \delta(B) \tag{2.4}$$

Proof: We let $A,B \in \mathscr{P}(X)$ be arbitrary and we assume that $A \subseteq B$. Then we let $B \setminus A = \{x_1, x_2, \ldots, x_k\}$. We define $A_0 = A$ and for each $i \in \{1,2,\ldots,k\}$ we define $A_i = A \cup \{x_1,x_2,\ldots,x_i\}$. For any arbitrary $i \in \{0,1,2,\ldots,k-1\}$ we show that $\delta(A_i) \leq \delta(A_{i+1})$.

$$\begin{aligned}
\delta(A_i) &\leq \delta(A_i \cup \{x_{i+1}\}) + \delta(\{x_{i+1}\}) && \text{by the triangle inequality of diversities} && (2.5)\\
&= \delta(A_i \cup \{x_{i+1}\}) && \delta(\{x_{i+1}\}) = 0 \text{ since } |\{x_{i+1}\}| = 1 && (2.6)\\
&= \delta(A_{i+1}) && && (2.7)
\end{aligned}$$

Since $A_k = B$, the above result implies that

$$\delta(A) \leq \delta(A_1) \leq \delta(A_2) \leq \ldots \leq \delta(A_{k-1}) \leq \delta(B) \tag{2.8}$$

thus completing the proof. □

2.2 Examples of Diversities

In this section we provide definitions of several diversities that are relevant to our discussion of and results on the generalized sparsest cut problem in hypergraphs. All of the diversities defined below are in general pseudo diversities, with the exception of the ℓ_1 diversity which is always a diversity. For each diversity that we define below it is easy to verify that it satisfies the axioms of Definitions 2.1.2. We note that these diversities have been previously defined by Bryant and Tupper [9, 10, 31].

We begin with the hypergraph Steiner diversity, a critical diversity to our approach to the sparsest cut problem in hypergraphs. Given a hypergraph $H = (V,E)$, we define $\mathscr{T}_{(H,S)}$, the collection of subsets of hyperedges that correspond to connected subhypergraphs that contain the nodes $S \subseteq V$.

Definition 2.2.1 *Let $H = (V,E)$ be a hypergraph with node set V and hyperedge set E. For a set of nodes $S \subseteq V$, we define, $\mathscr{T}_{(H,S)}$, the collection of subsets of hyperedges that correspond to connected subhypergraphs of H that contain S, as*

$$\mathscr{T}_{(H,S)} = \{t \subseteq E : (V(t),t) \text{ is a connected hypergraph and } S \subseteq V(t)\} \tag{2.9}$$

where we define $V(t)$ as

$$V(t) = \cup_{U \in t} U \tag{2.10}$$

In contexts where the hypergraph H is unambiguous we may use the notation $\mathscr{T}_S$.

A hypergraph Steiner diversity is defined as

Definition 2.2.2 (Hypergraph Steiner Diversity) *Let $H = (V,E,w)$ be a hypergraph with node set V, hyperedge set E, and nonnegative hyperedge weights $w : E \to \mathbb{R}_+$. The hypergraph Steiner diversity $(V, \delta_{\mathscr{H}})$ is defined as*

$$\delta_{\mathscr{H}}(A) = \begin{cases} \min_{t \in \mathscr{T}_A} \sum_{U \in t} w(U) & \text{if } |A| \geq 2 \\ 0 & \text{otherwise} \end{cases} \tag{2.11}$$

where $\mathscr{T}_A$ is the collection of subsets of hyperedges that correspond to connected subhypergraphs of H that contain the nodes A, as defined according to Definition 2.2.1.

Analogous to the fact that any metric space arises from a shortest path metric on some graph, every (pseudo) diversity is a hypergraph Steiner diversity for some hyperedge-weighted hypergraph [9]. Given a diversity (X, δ), one can easily verify that (X, δ) is equivalent to the hypergraph Steiner diversity corresponding to the hyperedge-weighted hypergraph $H = (X, \mathscr{P}(X), w)$ where w is defined as $w(A) = \delta(A)$.

Proposition 2.2.3 (Section 2.3 of [9]) *Let (X, δ) be a (pseudo) diversity. Then (X, δ) is a hypergraph Steiner diversity for some hyperedge-weighted hypergraph $H = (X, E, w)$ where $w : E \to \mathbb{R}_+$.*

2.2.1 The l1 and Cut Diversities

Just as the hypergraph Steiner diversity is a generalization of the shortest path metric of a graph, the ℓ_1 diversity is a natural generalization of the ℓ_1 metric. For context, we first define the ℓ_1 metric.

Definition 2.2.4 *Let $(\mathbb{R}^m, d)$ be a metric space. Then $(\mathbb{R}^m, d)$ is the ℓ_1 metric if for any $x, y \in \mathbb{R}^m$*

$$d(x,y) = \sum_{i=1}^{m} |x_i - y_i| \tag{2.12}$$

Similarly, the ℓ_1 diversity is defined as follows

Definition 2.2.5 (ℓ_1 Diversity) *$(\mathbb{R}^m, \delta_1)$ is an ℓ_1 diversity if for any $A \in \mathscr{P}(\mathbb{R}^m)$*

$$\delta_1(A) = \sum_{i=1}^{m} \max_{a,b \in A} |a_i - b_i| \tag{2.13}$$

Likewise, the cut pseudo metric is generalized by the cut pseudo diversity. A classic property of the ℓ_1 metric is that it can be represented as sum of cut pseudo metrics. Similarly, we present an analogous generalization for the ℓ_1 diversity and cut pseudo diversities as Theorem 3.1.2.

Definition 2.2.6 (Cut Diversity) *Let X be a set and let $U \subseteq X$ be a nonempty subset. Then (X, δ_U) is a cut pseudo diversity (induced by U) if for any $A \in \mathscr{P}(X)$*

$$\delta_U(A) = \begin{cases} 1 & \textit{if } A \cap U \neq \emptyset, A \\ 0 & \textit{otherwise} \end{cases} \tag{2.14}$$

2.2.2 Extremal Diversities

Given a (pseudo) metric space, (X,d), a natural extension of the (pseudo) metric space to a (pseudo) diversity is to define a diversity (X, δ) where $\delta(A)$ is simply the diameter of the set A in the metric space (X,d). Unsurprisingly, we refer to such a diversity as a diameter diversity.

Definition 2.2.7 (Diameter Diversity) *Given a (pseudo) metric space (X,d), a (pseudo) diameter diversity (X, δ_{diam}) satisfies for each $A \in \mathscr{P}(X)$*

$$\delta_{diam}(A) = \max_{x,y \in A} d(x,y) \tag{2.15}$$

Alternatively, one case also extend the (pseudo) metric (X,d) to a (pseudo) diversity by defining a diversity (X, δ) where $\delta(A)$ is simply the minimum cost Steiner tree containing the nodes A in the complete graph with node set X and edge weights d. We refer to such a diversity as a Steiner diversity. Such a diversity is the special case of the hypergraph Steiner diversity in which the hyperedge-weighted hypergraph $H = (V, E, w)$ is simply a graph.

Definition 2.2.8 (Steiner Diversity) *Let $G = (V, E, w)$ be a graph with node set V, edge set E, and non-negative edge weights $w : E \to \mathbb{R}_+$. Then the Steiner diversity $(V, \delta_{Steiner})$ is a Steiner diversity if for any $A \in \mathscr{P}(V)$ we have that $\delta_{Steiner}(A)$ is defined as*

$$\delta_{Steiner}(A) = \begin{cases} \min_{t \in \mathscr{T}_A} \sum_{e \in t} w(e) & \textit{if } |A| \geq 2 \\ 0 & \textit{otherwise} \end{cases} \tag{2.16}$$

where $\mathscr{T}_A$ is the collection of subsets of edges that correspond to connected subgraphs of G that contain the nodes A, as defined according to Definition 2.2.1. Or in other words, $\delta_{Steiner}(A)$ is defined to be the minimum weight of a subtree of G containing the nodes A.

A special case of Steiner diversities and a generalization of tree metrics are the tree diversities. We note that Bryant and Tupper introduced the definition of a tree diversity by the name *phylogenetic diversity*.

Definition 2.2.9 (Tree Diversity) *Let $G = (V,E,w)$ be a graph with node set V, edge set E, and nonnegative edge weights $w : E \to \mathbb{R}_+$. Let $(V, \delta_{Steiner})$ be the corresponding Steiner diversity. If (V,E) is a tree then $(V, \delta_{Steiner})$ is a tree diversity, for which we may use the notation (V, δ_{tree}).*

The following extremal result characterizes the (pseudo) diameter diversity as being the "minimal" (pseudo) diversity and characterizes the (pseudo) Steiner diversity as being the "maximal" (pseudo) diversity among the family of (pseudo) diversities sharing an induced (pseudo) metric.

Theorem 2.2.10 (Bryant and Tupper in [31]) *let (X,d) be a (pseudo) metric space. Let $\mathscr{D}_{(X,d)}$ be the family of diversities for whom (X,d) is their induced (pseudo) metric space. Let $(X, \delta_{diam}) \in \mathscr{D}_{(X,d)}$ be the (pseudo) diameter diversity with induced metric space (X,d). Let $(X, \delta_{Steiner})$ be the (pseudo) Steiner diversity with induced metric space (X,d). Then for any $(X, \delta) \in \mathscr{D}_{(X,d)}$ and any $A \in \mathscr{P}(X)$ it follows that*

$$\delta_{diam}(A) \leq \delta(A) \leq \delta_{Steiner}(A) \tag{2.17}$$

Or in other words (X, δ_{diam}) is the minimal (pseudo) diversity of the family $\mathscr{D}_{(X,d)}$ and $(X, \delta_{Steiner})$ is the maximal (pseudo) diversity of the family $\mathscr{D}_{(X,d)}$.

Proof: Let $(X, \delta) \in \mathscr{D}_{(X,d)}$ and $A \in \mathscr{P}(X)$ be arbitrary. Let $x, y \in A$ be such that

$$d(x,y) = \max_{u,v \in A} d(u,v) \tag{2.18}$$

Let $t \subseteq X \times X$ such that

$$\sum_{(u,v)\in t} d(u,v) = \min_{t' \in \mathscr{T}_A} \sum_{(u,v)\in t'} d(u,v) \tag{2.19}$$

Then it follows that

$$\begin{aligned}
\delta_{\text{diam}}(A) &= \max_{u,v\in A} d(u,v) && \text{by definition of } \delta_{\text{diam}} && (2.20)\\
&= d(x,y) && \text{by choice of } x \text{ and } y && (2.21)\\
&= \delta(\{x,y\}) && \text{by } (X,\delta) \in \mathscr{D}_{(X,d)} && (2.22)\\
&\leq \delta(A) && \text{by } \delta \text{ being increasing, Proposition 2.1.4} && (2.23)\\
&\leq \sum_{(u,v)\in t} d(u,v) && \text{by the triangle inequality, Proposition A.1.1} && (2.24)\\
&= \min_{t'\in\mathscr{T}_A} \sum_{(u,v)\in t'} d(u,v) && \text{by choice of } t && (2.25)\\
&= \delta_{\text{Steiner}}(A) && \text{by definition of } \delta_{\text{Steiner}} && (2.26)
\end{aligned}$$

This completes the proof. $\square$

Finally, we introduce the k-diameter diversity which generalizes the diameter diversity and satisfies a similar extremal result to Theorem 2.2.10. We note that Bryant and Tupper introduced the definition of a k-diameter diversity by the name *truncated diversity*.

Definition 2.2.11 (k-Diameter Diversity) *Given a (pseudo) diversity (X,δ) and a $k \in \mathbb{Z}_{\geq 0}$, we say that $(X,\delta_{k\text{-}diam})$ is the k-diameter (pseudo) diversity of (X,δ) if for any $A \in \mathscr{P}(X)$*

$$\delta_{k\text{-}diam}(A) = \max_{B \subseteq A : |B| \leq k} \delta(B) \tag{2.27}$$

Furthermore, if a (pseudo) diversity satisfies equation (2.27) then it is referred to as a k-diameter (pseudo) diversity.

From the definition of the k-diameter diversity, it is easy to see that a diameter diversity is a 2-diameter diversity.

Fact 2.2.12 *A 2-diameter (pseudo) diversity is a diameter (pseudo) diversity. Furthermore if (X,δ) is a diversity with the induced metric space (X,d) then $\mathscr{D}_{(X,d)} = \mathscr{D}_{(X,\delta,2)}$.*

Additionally, we introduce a generalization of $\mathscr{D}_{(X,d)}$, the family of diversities which have the induced metric space (X,d).

Definition 2.2.13 ($\mathscr{D}_{(X,\delta,k)}$) *For a (pseudo) diversity (X,δ) and $k \in \mathbb{Z}_{\geq 0}$, $\mathscr{D}_{(X,\delta,k)}$ is the family of (pseudo) diversities which are equivalent to δ on subsets of X of cardinality at most k. For contexts where the set X is unambiguous the notation $\mathscr{D}_{(\delta,k)}$ may be used.*

Next, we prove that a k-diameter diversity (X,δ) is the minimal diversity of the family $\mathscr{D}_{(X,\delta,k)}$, generalizing Theorem 2.2.10 due to Bryant and Tupper. We use this extremal property of the k-diameter diversity in order to characterize the optimal solutions of the diversity-relaxation for the sparsest cut, Theorem 4.2.1.

Theorem 2.2.14 *Given a (pseudo) diversity (X,δ) and $k \in \mathbb{Z}_{\geq 0}$ we let $(X,\delta_{k\text{-}diam})$ be the k-diameter (pseudo) diversity of (X,δ). Then for any $(X,\delta') \in \mathscr{D}_{(X,\delta,k)}$ and any $A \in \mathscr{P}(X)$ it follows that*

$$\delta_{k\text{-}diam}(A) \leq \delta'(A) \tag{2.28}$$

Or in other words $(X,\delta_{k\text{-}diam})$ is the minimal (pseudo) diversity of the family $\mathscr{D}_{(X,\delta,k)}$.

Proof: For an arbitrary $k \in \mathbb{Z}_{\geq 0}$, we let $(X,\delta_{k\text{-diam}})$ be the k-diameter diversity of (X,δ). First, we argue that $(X,\delta_{k\text{-diam}})$ is in fact a member of $\mathscr{D}_{(X,\delta,k)}$. We consider an arbitrary $A \in \mathscr{P}(X)$ such that $|A| \leq k$. Then it follows that

$$\delta_{k\text{-diam}}(A) = \max_{B \subseteq A : |B| \le k} \delta(B) \quad \text{by definition of a } k\text{-diameter diversity} \tag{2.29}$$

$$= \delta(A) \quad \text{by } |A| \le k \text{ and diversities being increasing, Proposition 2.1.4} \tag{2.30}$$

Hence, $\forall A \in \mathscr{P}(X)$ such that $|A| \le k$ it follows that $\delta_{k\text{-diam}}(A) = \delta(A)$ and so $(X, \delta_{k\text{-diam}}) \in \mathscr{D}_{(X,\delta,k)}$. Next, we establish the minimality of $(X, \delta_{k\text{-diam}})$ with regards to this family of diversities. We let $(X, \delta') \in \mathscr{D}_{(X,\delta,k)}$ and $A \in \mathscr{P}(X)$ be arbitrary. Then it follows that

$$\delta_{k\text{-diam}}(A) = \delta(B) \quad \text{for some } B \subseteq A \text{ such that } |B| \le k \tag{2.31}$$

$$= \delta'(B) \quad \text{by definition of } (X, \delta') \in \mathscr{D}_{(X,\delta,k)} \tag{2.32}$$

$$\le \delta'(A) \quad \text{by diversities being increasing, Proposition 2.1.4} \tag{2.33}$$

This completes the proof.

□

2.3 Subadditive Set Functions as Diversities

In this section we seek to answer two questions. First, when is a subadditive set function a diversity? Secondly, when can a subadditive set function be modified into a diversity? We proceed with the definition of a subadditive set function.

Definition 2.3.1 (Subadditive Set Function) *Given a set X, a set function $f : 2^X \to \mathbb{R}$ is subadditive if for any $A, B \subseteq X$, f satisfies*

$$f(A \cup B) \le f(A) + f(B) \tag{2.34}$$

Subadditive functions are closely related to diversities if we consider the third axiom of diversities, the triangle inequality. Specifically, in the case where $\emptyset \neq C \subseteq A \cap B$, where (X, δ) is a diversity and $A, B \in \mathscr{P}(X)$. Then we observe that

$$\delta(A \cup B) \le \delta(A \cup C) + \delta(B \cup C) = \delta(A) + \delta(B) \tag{2.35}$$

which can be interpreted as δ being subadditive for sets $A, B \in \mathscr{P}(X)$ with nonempty intersection. One might try to redefine the third axiom of diversities using subadditivity (for subsets with nonempty intersection) instead of a generalized triangle inequality, but a few other nuances would be required in this new definition. Nonetheless, this close relationship between diversities and subadditive set functions motivates this section.

Returning to our motivating questions, we recall the first axiom of a diversity, nonnegativity, and Proposition 2.1.4, a diversity is an increasing set function. The combination of these properties implies that in order for a subadditive set function to be, or even "resemble", a diversity it must be nonnegative and increasing. This intuition is reflected in the lemma below, which answers our motivating questions for this section. We remark that this lemma is also critical to the proof of our inapproximability result for diversity embeddings into ℓ_1, Theorem 3.4.3. Specifically, as a key technical lemma which is used to establish our notion of an independent set diversities, Definition 3.4.5, as in fact being diversities.

Lemma 2.3.2 *Let X be a set and let $f : 2^X \to \mathbb{R}$ be a nonnegative, increasing, and subadditive set function. Then (X,δ) is a pseudo-diversity where δ is defined as*

$$\delta(A) = \begin{cases} f(A) & \text{if } |A| \geq 2 \\ 0 & \text{otherwise} \end{cases} \tag{2.36}$$

Proof: Let δ be defined as in the lemma statement. We begin by showing that (X,δ) satisfies the first two axioms of a pseudo-diversity. By the nonnegativity of f, δ is likewise nonnegative by its construction, thus (X,δ) satisfies the first axiom of pseudo-diversities. Likewise by construction, $\delta(A) = 0$ if $|A| \leq 1$ and thus (X,δ) satisfies the second axiom.

In order to prove that (X,δ) satisfies the third axiom of a pseudo-diversity we first establish that δ is an increasing set function. Let $A \subseteq B \in \mathscr{P}(X)$ be arbitrary. For the case where $|A| \leq 1$ then

$$\delta(A) = 0 \leq \delta(B) \tag{2.37}$$

by the nonnegativity of δ. As for the case where $|A| \geq 2$ then

$$\delta(A) = f(A) \leq f(A \cup B) = \delta(A \cup B) \tag{2.38}$$

where the two equalities follow by the fact that $|A|, |A \cup B| \geq 2$ and the inequality follows by the fact that f is increasing. This concludes the proof that δ is an increasing set function.

We now prove that (X,δ) satisfies the third axiom of pseudo-diversities. Let $A,B,C \subseteq X$ be arbitrary where $C \neq \emptyset$. We consider several cases.

1. The first case is where one of A and B is empty. Without loss of generality, we assume that $A = \emptyset$. Then,

$$\begin{aligned} \delta(A \cup B) &= \delta(B) && && (2.39) \\ &\leq \delta(B \cup C) && \text{by } \delta \text{ being increasing} && (2.40) \\ &\leq \delta(A \cup C) + \delta(B \cup C) && \text{by } \delta \text{ being nonnegative} && (2.41) \end{aligned}$$

2. The second case that we consider is where $|A| = |B| = 1$. There are two subcases that we consider. The first is where $A = B$ in which case $|A \cup B| = 1$ and so

$$\delta(A \cup B) = 0 \qquad \text{by } |A \cup B| = 0 \tag{2.42}$$
$$\leq \delta(A \cup C) + \delta(B \cup C) \qquad \text{by } \delta \text{ being nonnegative} \tag{2.43}$$

Then the second subcase is where $A \neq B$ and so $|A \cup B| \geq 2$. Additionally, we divide this subcase into two more sub-subcases. The first sub-subcase is where $A \cap C = \emptyset$ and $B \cap C = \emptyset$, in which case $|A \cup C|, |B \cup C| \geq 2$ and so

$$\delta(A \cup B) = f(A \cup B) \qquad \text{by } |A \cup B| \geq 2 \tag{2.44}$$
$$\leq f(A) + f(B) \qquad \text{by subadditivity of } f \tag{2.45}$$
$$\leq f(A \cup C) + f(B \cup C) \qquad \text{by } f \text{ being increasing} \tag{2.46}$$
$$= \delta(A \cup C) + \delta(B \cup C) \qquad \text{by } |A \cup C|, |B \cup C| \geq 2 \tag{2.47}$$

The second sub-subcase is where, without loss of generality, $B \subseteq C$, and so

$$\delta(A \cup B) \leq \delta(A \cup C) \qquad \text{by } \delta \text{ being increasing} \tag{2.48}$$
$$\leq \delta(A \cup C) + \delta(B \cup C) \qquad \text{by } \delta \text{ being nonnegative} \tag{2.49}$$

3. The final case is where, without loss of generality, $|A| \geq 2$ and $|B| \geq 1$. Furthermore, we consider two subcases. The first is where $|B \cup C| = 1$ in which case $B = C$ and so

$$\delta(A \cup B) = f(A \cup B) \qquad \text{by } |A \cup B| \geq 2 \tag{2.50}$$
$$= f(A \cup C) \qquad \text{by } B = C \tag{2.51}$$
$$= \delta(A \cup C) \qquad \text{by } |A \cup C| \geq 2 \tag{2.52}$$
$$\leq \delta(A \cup C) + \delta(B \cup C) \qquad \text{by } \delta \text{ being nonnegative} \tag{2.53}$$

Then the second subcase is where $|B \cup C| \geq 2$. Then,

$$\delta(A \cup B) = f(A \cup B) \qquad \text{by } |A \cup B| \geq 2 \tag{2.54}$$
$$\leq f(A) + f(B) \qquad \text{by } f \text{ being subadditive} \tag{2.55}$$
$$\leq f(A \cup C) + f(B \cup C) \qquad \text{by } f \text{ being increasing} \tag{2.56}$$
$$= \delta(A \cup C) + \delta(B \cup C) \qquad \text{by } |A \cup C|, |B \cup C| \geq 2 \tag{2.57}$$

This completes the proof. $\square$

Chapter 3

Diversity Embeddings

3.1 Introduction

In this chapter we discuss diversity embeddings. Analogous to the notion of embedding a metric space into another metric space, one can define the notion of embedding a diversity into another diversity. In this section we introduce such a notion of diversity embeddings along with some properties and known upper and lower bounds on the minimum distortion of embedding into the ℓ_1 diversity. In Section 3.2 we present an $O(\log n)$-distortion embedding of a Steiner diversity into the ℓ_1 diversity. In that section, we also present a corollary of the aforementioned result, an $O(k \log n)$-distortion embedding of a hypergraph Steiner diversity corresponding to a rank-k hypergraph into the ℓ_1 diversity. In Section 3.3 we present an $O(\log n)$-distortion embedding of a diameter diversity into the ℓ_1 diversity. This embedding improves upon an earlier result of Bryant and Tupper who gave an $O(\log^2 n)$-distortion embedding. We also give an $O(k \log n)$-distortion embedding of a k-diameter diversity into the ℓ_1 diversity, as a corollary of our $O(\log n)$-distortion diameter diversity embedding. Finally, in Section 3.4 we present an $\Omega(n)$-inapproximability result for embedding an arbitrary diversity into the ℓ_1 diversity in polynomial-time while solely using its induced metric. We begin by formally defining an embedding of one diversity into another.

Definition 3.1.1 (Diversity Embedding) *Let (X, δ_X) and (Y, δ_Y) be two diversities. Let f be a map from X to Y. We say that f is an embedding of the diversity (X, δ_X) into the diversity (Y, δ_Y) with distortion c if there are constants $c_1, c_2 \geq 1$ such that $c = c_1 c_2$ and for any $A \in \mathscr{P}(X)$ we have that*

$$\frac{1}{c_1}\delta_X(A) \leq \delta_Y(f(A)) \leq c_2 \delta_X(A) \tag{3.1}$$

For convenience, when referring to such an embedding we may use the notation $(X, \delta_X) \to (Y, \delta_Y)$ and $(X, \delta_X) \xrightarrow{c} (Y, \delta_Y)$, where the latter specifies the distortion of the embedding. We may also precede such notation with "f :" in order to specify the map.

If we restrict Equation 3.1 in Definition 3.1.1 to sets A of cardinality two, then we recover the exact definition of a metric embedding. This is significant since facts about metric embeddings can and often do extend to facts about diversity embeddings since each diversity "encodes" an induced metric space.

Our approach to the sparsest cut involves embedding into the ℓ_1 diversity due to the fact that ℓ_1-embeddable diversities are a sum of cut pseudo-diversities. We use the term ℓ_1-*embeddable* to mean embedding into ℓ_1 isometrically, or in other words, with distortion $c = 1$. This fact, due to Bryant and Tupper [31], is a generalization of the analogous fact that ℓ_1-embeddable metrics are a sum of cut pseudo-metrics.

Theorem 3.1.2 *Let (X, δ) be a diversity where $|X| = n < \infty$. Then the following are equivalent.*

1. *(X, δ) is embeddable into ℓ_1^m.*
2. *(X, δ) is a nonnegative combination of $O(nm)$ cut diversities.*

In Chapter 4 we show how Theorem 3.1.2 is utilized in our approach to the sparsest cut problem in hypergraphs, and furthermore, how the distortion of embedding a diversity into the ℓ_1 diversity yields a multiplicative approximation factor. Henceforth, the remainder of this chapter is concerned with the embedding of diversities, both arbitrary diversities and particular families of diversities, into the ℓ_1 diversity with low distortion. This investigation can be formulated by the following question posed by Bryant and Tupper in [9].

Problem 3.1.3 *Let (X, δ) be an arbitrary diversity where $|X| = n$. What is the minimum distortion with which (X, δ) can be embedded into an ℓ_1 diversity?*

This is an existential question in that it seeks a minimum distortion embedding into ℓ_1 without any *algorithmic* or *computational* constraints. Specifically, by "algorithmic" or "computational" we mean constraints on the embedding, the map f as in Definition 3.1.1, being computable in polynomial-time with respect to the size of encoding (X, δ) (often simply the cardinality of X). Implicit in this constraint of polynomial-time computability is that the dimension m of the ℓ_1 diversity $(\mathbb{R}^m, \delta_1)$ must also be bounded by a polynomial factor. We formalize this refinement of Problem 3.1.3 as the following problem.

Problem 3.1.4 *Let (X, δ) be an arbitrary diversity where $|X| = n$. What is the minimum distortion with which (X, δ) can be embedded into an ℓ_1 diversity in polynomial-time.*

Bryant and Tupper have shown in [9] that an arbitrary diversity can be embedded into the ℓ_1 diversity with distortion $O(n)$. In fact, this embedding is solely based off the induced metric of a diversity. Moreover, this embedding is computable in polynomial-time provided that the induced metric of the diversity can be queried in polynomial-time. Below, we restate this result with proof.

Theorem 3.1.5 (Restatement of Theorem 1 in [9]) *Let (X, δ) be an arbitrary diversity where $|X| = n$. Then there is an embedding of (X, δ) into the ℓ_1 diversity, $(\mathbb{R}^n, \delta_1)$, with distortion n.*

Proof: We consider an arbitrary diversity (X, δ) and we let (X, d) be its induced metric space. We enumerate the elements of X as $X = \{x_1, x_2, \ldots, x_n\}$. Then we define our embedding $f : X \to \mathbb{R}^n$ as

$$f(x) = (d(x_1, x), d(x_2, x), \ldots, d(x_n, x)) \tag{3.2}$$

We consider an arbitrary $A \in \mathscr{P}(X)$. It suffices to show that

$$\delta(A) \leq \delta_1(f(A)) \leq n\delta(A) \tag{3.3}$$

We now prove the first inequality. First, we consider an arbitrary $a_0 \in A$.

$$\begin{aligned}
\delta(A) &\leq \sum_{x \in A} d(x, a_0) && \text{by the triangle inequality, Proposition A.1.2} && (3.4)\\
&\leq \sum_{x \in A} \max_{a \in A} d(x,a) && && (3.5)\\
&= \sum_{x \in A} \max_{a,b \in A} d(x,a) - d(x,b) && \text{since } \max_{b \in A} -d(x,b) = -d(x,x) = 0 && (3.6)\\
&= \sum_{x \in A} \max_{a,b \in A} |d(x,a) - d(x,b)| && \text{since } \max_{a,b \in A} d(x,a) - d(x,b) \geq 0 && (3.7)\\
&\leq \sum_{i=1}^{n} \max_{a,b \in A} |d(x_i,a) - d(x_i,b)| && && (3.8)\\
&= \sum_{i=1}^{n} \max_{a,b \in A} |f_i(a) - f_i(b)| && && (3.9)\\
&= \delta_1(f(A)) && \text{by definition of the } \ell_1 \text{ diversity} && (3.10)
\end{aligned}$$

As for the second inequality, we choose for each $i \in \{1, 2, \ldots, n\}$ $a_i, b_i \in A$ such that

$$\max_{a,b \in A} |d(x_i,a) - d(x_i,b)| = d(x_i,a_i) - d(x_i,b_i) \tag{3.11}$$

Then it follows that

$$
\begin{aligned}
\delta_1(f(A)) &= \sum_{i=1}^{n} \max_{a,b\in A} |f_i(a) - f_i(b)| && (3.12)\\
&= \sum_{i=1}^{n} \max_{a,b\in A} |d(x_i,a) - d(x_i,b)| && (3.13)\\
&= \sum_{i=1}^{n} d(x_i,a_i) - d(x_i,b_i) && \text{by choice of } a_i,b_i \in A \quad (3.14)\\
&\leq \sum_{i=1}^{n} d(a_i,b_i) + (x_i,b_i) - d(x_i,b_i) && \text{by the triangle inequality} \quad (3.15)\\
&= \sum_{i=1}^{n} d(a_i,b_i) && (3.16)\\
&= \sum_{i=1}^{n} \delta(\{a_i,b_i\}) && (3.17)\\
&\leq \sum_{i=1}^{n} \delta(A) && \text{by diversities being increasing, Proposition 2.1.4} \quad (3.18)\\
&= n\delta(A) && (3.19)
\end{aligned}
$$

This completes the proof. □

As for a lower bound to Problem 3.1.3, it is evident that some diversities cannot be embedded into the ℓ_1 diversity with distortion smaller than $\Omega(\log n)$. This is due to the fact that each diversity encodes an induced metric space and there are finite metric spaces which cannot be embedded into the ℓ_1 metric with distortion smaller than $\Omega(\log n)$. Specifically, Linial, London, and Rabinovich showed that the flow-cut gap in a graph is equivalent to the minimum distortion necessary to embed some metric space into the ℓ_1 metric. Given that the flow-cut gap for a constant degree expander graph is $\Omega(\log n)$, they concluded that there exist metric spaces which cannot be embedded into ℓ_1 with distortion smaller than $\Omega(\log n)$ [24]. Interestingly, this result answered the open question of what is the minimum distortion to embed an n-point metric into the ℓ_1 metric, the metric variant of Problem 3.1.3. This lower bound matched the $O(\log n)$ upper bound due to Bourgain [8].

Theorem 3.1.6 (Proposition 4.2 in [24]) *There exists a family of metric spaces $\{(X_n,d_n)\}_{n=1}^{\infty}$, where $|X_n| = n$, which cannot be embedded into the ℓ_1 metric with distortion smaller than $\Omega(\log n)$.*

Theorem 3.1.7 (Restatement of Bryant and Tupper's observation in [9]) *There exists a family of diversities $\{(X_n,\delta_n)\}_{n=1}^{\infty}$, where $|X_n| = n$, which cannot be embedded into the ℓ_1 diversity with distortion smaller than $\Omega(\log n)$.*

Proof: We let (X,δ) be an arbitrary diversity with induced metric space (X,d). We recall that any embedding of (X,δ) into the ℓ_1 diversity includes an embedding of (X,d) into the ℓ_1 metric space. Thus we can find a family of diversities, according to theorem 3.1.6, that require distortion $\Omega(\log n)$ for embedding into ℓ_1. □

3.2 A Steiner Diversity Embedding

In this section we present a proof of Bryant and Tupper's $O(\log n)$-distortion embedding of a Steiner diversity into the ℓ_1 diversity. In Subsection 3.2.1 we present a corollary of this result which is an $O(k \log n)$-distortion embedding of a hypergraph Steiner diversity corresponding to a rank-k hypergraph into the ℓ_1 diversity. We say that a hypergraph has rank k if the cardinalities of its hyperedges are bounded by k.

We note that Bryant and Tupper's embedding utilizes Fakcharoenphol, Rao, and Talwar's probabilistic embedding of n-node metric spaces into dominating tree metrics with distortion $O(\log n)$ in expectation [16], which we refer to as the *FRT algorithm* for brevity. Given that tree metrics are isometrically embeddable into the ℓ_1 metric, the FRT algorithm provides an alternate $O(\log n)$-distortion metric embedding into ℓ_1.

Theorem 3.2.1 (The FRT Algorithm) *Let (X,d) be a metric space where $|X| = n$. Then there is a randomized polynomial-time algorithm that embeds (X,d) into a dominating tree metric with distortion $O(\log n)$ in expectation. Specifically, this randomized polynomial-time algorithm produces a tree $T = (X,E,w)$ with nonnegative edge weights w such that the corresponding shortest path metric, which happens to be a tree metric, (X,d_T) satisfies for every $x,y \in X$*

$$d(x,y) \leq d_T(x,y) \tag{3.20}$$

$$\mathbb{E}[d_T(x,y)] \leq O(\log n) d(x,y) \tag{3.21}$$

In addition to the FRT algorithm, Bryant and Tupper's embedding utilizes the fact that tree diversities are isometrically embeddable into the ℓ_1 diversity [10]. This embedding follows by the fact that, like a tree metric, a tree diversity is a sum of cut pseudo-diversities which correspond to the edges of the tree that defines the diversity.

Theorem 3.2.2 *Let (X,δ_{tree}) be a tree diversity. Then there exists an embedding of (X,δ_{tree}) into the ℓ_1 diversity with distortion 1. Moreover, this embedding is computable in polynomial-time with respect to $|X|$.*

Given Theorems 3.2.1 and 3.2.2, we present a proof of Bryant and Tupper's Steiner diversity embedding from [9].

Theorem 3.2.3 (Restatement of Theorem 2 in [9]) *Let $(X,\delta_{Steiner})$ be a Steiner diversity where $|X| = n$. Then there is a randomized polynomial-time algorithm that embeds $(X,\delta_{Steiner})$ into the ℓ_1 diversity with $O(\log n)$ distortion.*

Proof: Given Theorem 3.2.2, it suffices to prove that there is a randomized polynomial-time algorithm that embeds $(X,\delta_{\text{Steiner}})$ into a dominating tree diversity with distortion $O(\log n)$ in expectation. That is, we show that there is a randomized polynomial-time embedding into a random tree diversity (X,δ_{tree}) such that for each $A \in \mathscr{P}(X)$ the embedding satisfies

$$\delta_{\text{Steiner}}(A) \leq \delta_{\text{tree}}(A) \tag{3.22}$$

$$\mathbb{E}[\delta_{\text{tree}}(A)] \leq \delta_{\text{Steiner}}(A) \tag{3.23}$$

We let (X,d) be the induced metric space of $(X,\delta_{\text{Steiner}})$. We then apply the FRT algorithm to (X,d), according to Theorem 3.2.1, and we let (X,d_T) be the resulting random tree metric satisfying (3.20) and (3.21). We then define (X,δ_{tree}) to be a random tree diversity, that is induced by the tree metric (X,d_T). That is, for an arbitrary $A \in \mathscr{P}(X)$ we define (X,δ_{tree}) as

$$\delta_{\text{tree}}(A) = \min_{t \in \mathscr{T}_{(T,A)}} \sum_{(u,v) \in t} d_T(u,v) \tag{3.24}$$

where $T = (X,E)$ is the spanning tree of the nodes X that is defined by the tree metric (X,d_T). We recall that $\mathscr{T}_{(T,A)}$ is the collection of subsets of edges of T that correspond to subgraphs of T that contain the nodes A.

We recall that the FRT algorithm has randomized polynomial-time complexity. Since we can extend a tree metric into a tree diversity trivially in polynomial-time then this embedding of $(X,\delta_{\text{Steiner}})$ into the random tree diversity (X,δ_{tree}) runs in randomized polynomial-time. Then we let $A \in \mathscr{P}(X)$ be arbitrary and we choose $t \in \mathscr{T}_{(T,A)}$ such that

$$\delta_{\text{tree}}(A) = \sum_{(u,v) \in t} d_T(u,v) \tag{3.25}$$

Then (3.22) is established with the following argument

$$\begin{aligned}
\delta_{\text{Steiner}}(A) &\leq \sum_{(u,v) \in t} d(u,v) && \text{by the triangle inequality} && (3.26)\\
&\leq \sum_{(u,v) \in t} d_T(u,v) && \text{by } (X,d_T) \text{ dominating } (X,d), \text{(3.20)} && (3.27)\\
&= \delta_{\text{tree}}(A) && \text{by choice of } t && (3.28)
\end{aligned}$$

As for (3.23), we choose $t' \subseteq X \times X$ such that

$$\delta_{\text{Steiner}}(A) = \sum_{(u,v) \in t'} d(u,v) \tag{3.29}$$

Then we argue that

$$
\begin{aligned}
\mathbb{E}[\delta_{\text{tree}}(A)] &\leq \mathbb{E}[\sum_{(u,v)\in t'} d_T(u,v)] && \text{by the triangle inequality} && (3.30)\\
&= \sum_{(u,v)\in t'} \mathbb{E}[d_T(u,v)] && \text{by linearity of expectation} && (3.31)\\
&\leq \sum_{(u,v)\in t'} O(\log n) d(u,v) && \text{by (3.21)} && (3.32)\\
&= O(\log n)\delta_{\text{Steiner}}(A) && \text{by choice of } t' && (3.33)
\end{aligned}
$$

This completes the proof. □

3.2.1 A Hypergraph Steiner Diversity Embedding

In this section we provide a proof, due to Bryant and Tupper [9], of the fact that a hypergraph Steiner diversity corresponding to a rank k hypergraph can be embedded into the ℓ_1 diversity with distortion $O(k \log n)$.

At a high level, this $O(k \log n)$-distortion embedding follows by approximating a hyperedge-weighted rank-k hypergraph with an edge-weighted graph while incurring an $O(k)$ approximation factor. Or in other words, embedding a hypergraph Steiner diversity into a Steiner diversity with distortion $O(k)$. Then, the additional $O(\log n)$ factor is incurred by embedding this Steiner diversity into the ℓ_1 diversity.

Theorem 3.2.4 *Let $H = (V, E, w)$ be a rank k hypergraph with node set V, edge set E, and nonnegative hyperedge weights $w : E \to \mathbb{R}_+$. Let $(V, \delta_{\mathscr{H}})$ be the corresponding hypergraph Steiner diversity. Then there is a Steiner diversity $(V, \delta_{Steiner})$ into which $(V, \delta_{\mathscr{H}})$ can be polynomial-time embedded with distortion $O(k)$.*

Proof: We let (X, d) be the induced metric space of $(V, \delta_{\mathscr{H}})$. We first note that for any $u, v \in V$ and any $U \in E$ such that $u, v \in U$ then

$$d(u,v) = \delta_{\mathscr{H}}(\{u,v\}) \leq w(U) \tag{3.34}$$

by definition of a hypergraph Steiner diversity and the fact that the hyperedge U is a connected subhypergraph of H that contains the nodes u and v.

Then we let $(V, \delta_{\text{Steiner}}) \in \mathscr{D}_{(V,d)}$ be the corresponding Steiner diversity with induced metric space (V, d). We remark that $(V, \delta_{\text{Steiner}})$ can be represented by the metric space (V, d) or by the complete edge-weighted graph $G = (V, E', d)$. Moreover, (V, d) is the shortest path metric for the hyperedge-weighted hypergraph H and can be computed in polynomial-time, for instance, by reducing the hypergraph to a bipartite graph in which each hyperedge is replaced by a corresponding node and then computing the shortest path in said graph. Hence, this embedding of $(V, \delta_{\mathscr{H}})$ into $(V, \delta_{\text{Steiner}})$ can be computed in polynomial-time.

It remains to show that this embedding of $(V, \delta_{\mathscr{H}})$ into $(V, \delta_{\text{Steiner}})$ has distortion $O(k)$. That is, for any $A \in \mathscr{P}(V)$ it follows that

$$\delta_{\mathscr{H}}(A) \leq \delta_{\text{Steiner}}(A) \leq (k-1)\delta_{\mathscr{H}}(A) \tag{3.35}$$

We let $A \in \mathscr{P}(V)$ be arbitrary and we choose $t \in \mathscr{T}_{(H,A)}$ to be such that

$$\delta_{\mathscr{H}}(A) = \sum_{U \in t} w(U) \tag{3.36}$$

Then the first inequality of (3.35),

$$\delta_{\mathscr{H}}(A) \leq \delta_{\text{Steiner}}(A) \tag{3.37}$$

follows by the fact that $(V, \delta_{\mathscr{H}}), (V, \delta_{\text{Steiner}}) \in \mathscr{D}_{(V,d)}$ and $(V, \delta_{\text{Steiner}})$ is the maximal diversity of this family, Theorem 2.2.10.

As for the second inequality,

$$\begin{aligned}
\delta_{\mathscr{H}}(A) &= \sum_{U \in t} w(U) && \text{by the choice of } t, \text{(3.36)} && (3.38)\\
&\geq \sum_{U \in t} \max_{u,v \in U} d(u,v) && \text{by (3.34)} && (3.39)
\end{aligned}$$

For each $U \in t$ we define $T(U) \subseteq E'$ to be a subtree of the complete graph G' that spans the nodes U and only contains the nodes U. Therefore, $|T(U)| = k-1$ since $T(U)$ is a spanning tree. Then we continue as follows

$$\begin{aligned}
\sum_{U \in t} \max_{u,v \in U} d(u,v) &= \sum_{U \in t, (u',v') \in T(U)} \frac{1}{k-1} \max_{u,v \in U} d(u,v) && \text{by } |T(U)| = k-1 && (3.40)\\
&\geq \frac{1}{k-1} \sum_{U \in t, (u',v') \in T(U)} d(u',v') && && (3.41)\\
&\geq \frac{1}{k-1} \sum_{(u',v') \in \cup_{U \in t} T(U)} d(u',v') && \text{by not double-counting edges} && (3.42)\\
&\geq \delta_{\text{Steiner}}(A) && && (3.43)
\end{aligned}$$

where the last inequality follows by the fact that

$$\cup_{U \in t} T(U) \in \mathscr{T}_{(G,A)} \tag{3.44}$$

or in other words, $\cup_{U \in t} T(U)$ is a connected subgraph of G that contains the nodes A. This completes the proof.

□

Then the main result of this section follows as a corollary of Theorems 3.2.3 and 3.2.4.

Corollary 3.2.5 *Let $H = (V, E, w)$ be a rank-k hypergraph with node set V, edge set E, and nonnegative hyperedge weights $w : E \to \mathbb{R}_+$. Let $(V, \delta_{\mathscr{H}})$ be the corresponding hypergraph Steiner diversity. Then $(V, \delta_{\mathscr{H}})$ can be embedded into the ℓ_1 diversity with distortion $O(k \log n)$, in randomized polynomial-time.*

Proof: We let $f : (V, \delta_{\mathscr{H}}) \to (V, \delta_{\text{Steiner}})$ be a polynomial-time $O(k)$-distortion embedding of $(V, \delta_{\mathscr{H}})$ into some Steiner diversity $(V, \delta_{\text{Steiner}})$, due to Theorem 3.2.4. We let $g : (V, \delta_{\text{Steiner}}) \to (\mathbb{R}^m, \delta_1)$ be a randomized polynomial-time $O(\log n)$-distortion embedding of $(V, \delta_{\text{Steiner}})$ into the ℓ_1 diversity for some dimension m, due to Theorem 3.2.3. Then the map $g \cdot f$ is an embedding of $(V, \delta_{\mathscr{H}})$ into the ℓ_1 diversity with distortion $O(k \log n)$. Moreover, this embedding is computable in randomized polynomial-time. This completes the proof. $\square$

3.3 Diameter Diversity Embeddings

In this section we provide two polynomial-time low-distortion embeddings of the diameter diversity into ℓ_1. The first is an $O(\log^2 n)$-distortion embedding due to Bryant and Tupper [31], presented as Theorem 3.3.7 in Section 3.3.2. The second is an $O(\log n)$-distortion embedding, Theorem 1.2.3, restated as Theorem 3.3.8 in Section 3.3.3. The distortion achieved by the latter embedding is asymptotically optimal given the $\Omega(\log n)$ lower bound on the distortion of embedding arbitrary diversities into ℓ_1, Theorem 3.1.7.

Although our analysis reduces the distortion by a logarithmic factor from $O(\log^2 n)$ to $O(\log n)$, we provide both embeddings to juxtapose their associated proofs. Specifically, we include Lemma 3.3.6, which states that the diameter diversity, $(\mathbb{R}^d, \delta_{\text{diam}})$, whose induced metric is the ℓ_1 metric is within a factor of $O(d)$ of the ℓ_1 diversity, $(\mathbb{R}^d, \delta_1)$. This result is significant in that it essentially states that the ℓ_1 diversity defined over a low-dimensional space behaves similarly to a diameter diversity. This is a potentially relevant observation for tackling the remaining unsolved questions in the realm of diversity embeddings into ℓ_1.

3.3.1 Fréchet Embeddings

Both ℓ_1 embeddings of the diameter diversity are based off polynomial-time implementations of Bourgain's original $O(\log n)$-distortion metric embedding into ℓ_1 [8].

Theorem 3.3.1 (Restatement of Proposition 1 of [8]) *Let (X, d) be a finite metric space where $|X| = n$. There exists an embedding of (X, d) into the ℓ_1 metric $(\mathbb{R}^k, d_1)$, where $k \in O(2^n)$, with distortion $O(\log n)$.*

We remark that Bourgain's embedding is a scaled Fréchet embedding, which we define below.

Definition 3.3.2 *Let (X, d) be a metric space. A Fréchet embedding is a map $f : (X, d) \to (\mathbb{R}^k, d')$ where each coordinate, $f_i : X \to \mathbb{R}$, of the embedding is defined as*

$$f_i(x) = d(x, A_i) = \min_{y \in A_i} d(x, y) \tag{3.45}$$

for some nonempty $A_i \subseteq X$.

Due to the triangle inequality of metrics, a Fréchet embedding is coordinate-wise non-expansive, which we define and prove below.

Proposition 3.3.3 *Let (X, d) be a metric space and let $f : (X, d) \to (\mathbb{R}^k, d')$ be a Fréchet embedding. Then for any coordinate, $i \in \{1, 2, \ldots, k\}$ and any $x, y \in X$, it follows that*

$$|f_i(x) - f_i(y)| \leq d(x,y) \tag{3.46}$$

Proof: Let $x, y \in X$ and let $i \in \{1, 2, \ldots, k\}$ be arbitrary. Then it follows that

$$\begin{aligned}
|f_i(x) - f_i(y)| &= |d(x,A_i) - d(y,A_i)| && \text{for some nonempty } A_i \subseteq X && (3.47)\\
&= |d(x,u) - d(y,u')| && \text{for some } u, u' \in A_i && (3.48)
\end{aligned}$$

Without loss of generality, we assume that $d(x,u) - d(y,u') \geq 0$. Then we have that

$$\begin{aligned}
|d(x,u) - d(y,u')| &= d(x,u) - d(y,u') && && (3.49)\\
&\leq d(x,u') - d(y,u') && \text{since } d(x,u) = \min_{v \in A_i} d(x,v) && (3.50)\\
&\leq d(x,y) + d(y,u') - d(y,u') && \text{by the triangle inequality} && (3.51)\\
&= d(x,y) && && (3.52)
\end{aligned}$$

This completes the proof. □

The original embedding due to Bourgain is an existential result that is algorithmically intractable. It was later that Linial, London, and Rabinovich [24] who provided randomized polynomial-time implementations of Theorem 3.3.1. This is achieved by sampling a relatively small subset of the collection of coordinate maps

$$\{f_i(x) = d(x,A_i)\}_{A_i \subseteq X} \tag{3.53}$$

in order to achieve a randomized polynomial-time complexity.

The following implementation of Bourgain's embedding according to Linial et al. [24] forms the basis of Bryant and Tupper's $O(\log^2 n)$-distortion embedding.

Theorem 3.3.4 *Let (X,d) be a metric space with $|X| = n$. Then there exists an embedding, $f : X \to \mathbb{R}^{O(\log n)}$, of (X,d) into the ℓ_1 metric $(\mathbb{R}^{O(\log n)}, d_1)$ with distortion $O(\log n)$. That is,*

$$d(x,y) \leq \|f(x) - f(y)\|_1 \leq O(\log n) d(x,y) \tag{3.54}$$

The fact that Bourgain's embedding is a scaled Fréchet embedding is key to Theorem 3.3.8. Specifically, Theorem 3.3.8 is based off the following implementation of Bourgain's embedding.

Theorem 3.3.5 (Lemma 3 in [5]) *Let (X,d) be a metric space with $|X| = n$. Then there exists an embedding, $f : X \to \mathbb{R}^{O(\log^2 n)}$, of (X,d) into the ℓ_1 metric $(\mathbb{R}^{O(\log^2 n)}, d_1)$ with distortion $O(\log n)$. That is,*

$$\frac{1}{O(\log n)} d(x,y) \leq \|f(x) - f(y)\|_1 \leq d(x,y) \tag{3.55}$$

Furthermore, the embedding f is a scaled Fréchet embedding where for an arbitrary coordinate $i \in \{1,2,\ldots,O(\log^2 n)\}$, f_i is defined as

$$f_i(x) = \frac{1}{O(\log^2 n)} d(x, A_i) \tag{3.56}$$

where $A_i \subseteq X$.

Scaling the Fréchet embedding, in the above theorem, is a necessary step. In fact, the original embedding of Bourgain is also scaled proportionally to the exponentially large dimension of the ℓ_1 metric space being embedded into.

3.3.2 An $O(\log^2 n)$-Distortion Embedding

In this section we present the proof of Theorem 3.3.7, closely following Bryant and Tupper's original proof. First, we prove their technical lemma.

Lemma 3.3.6 (Restatement of Lemma 1 of [31]) *Let $(\mathbb{R}^k, d_1)$ be a metric space where d_1 is the ℓ_1 metric, let $(\mathbb{R}^k, \delta^{(1)}_{diam})$ be the diameter diversity whose induced metric space is $(\mathbb{R}^k, d_1)$, and let $(\mathbb{R}^k, \delta_1)$ be the ℓ_1^k diversity. Then for any $A \in \mathscr{P}(\mathbb{R}^k)$ it follows that*

$$\delta^{(1)}_{diam}(A) \leq \delta_1(A) \leq k\delta^{(1)}_{diam}(A) \tag{3.57}$$

Proof: Let $A \in \mathscr{P}(\mathbb{R}^k)$ be arbitrary and let $x, y \in A$ be such that $d(x,y) = \max_{a,b\in A} d(a,b)$. Then it follows that

$$\begin{aligned}
\delta^{(1)}_{\text{diam}}(A) &= \max_{a,b\in A} d_1(a,b) && (3.58)\\
&= d_1(x,y) && \text{by choice of } x \text{ and } y \quad (3.59)\\
&= \sum_{i=1}^{k} |x_i - y_i| && \text{by definition of the } \ell_1 \text{ metric} \quad (3.60)\\
&\leq \sum_{i=1}^{k} \max_{a,b\in A} |a_i - b_i| && (3.61)\\
&\leq \sum_{i=1}^{k} \max_{a,b\in A} \|a-b\|_1 && (3.62)\\
&= kd_1(x,y) && (3.63)\\
&= k\delta^{(1)}_{\text{diam}}(A) && \text{by definition of the diameter diversity} \quad (3.64)
\end{aligned}$$

Given that $\delta_1(A) = \sum_{i=1}^{k} \max_{a,b\in A} |a_i - b_i|$, as in Equation (3.61) above, this completes the proof. □

Given Lemma 3.3.6 and Theorem 3.3.4, we present and prove the following result of Bryant and Tupper.

Theorem 3.3.7 *Let (X, δ_{diam}) be a diameter diversity with $|X| = n$. Then there exists a randomized polynomial-time embedding of (X, δ_{diam}) into the ℓ_1 diversity $(\mathbb{R}^{O(\log n)}, \delta_1)$ with distortion $O(\log^2 n)$.*

Proof: As according to Theorem 3.3.4, we let $f : X \to \mathbb{R}^{O(\log n)}$ be an embedding of (X,d) into the ℓ_1 metric $(\mathbb{R}^{O(\log n)}, d_1)$ with distortion $O(\log n)$. We then define our diversity embedding from $(X, \delta_{\text{diam}})$ to the ℓ_1 diversity $(\mathbb{R}^{O(\log n)}, \delta_1)$ to be simply the map f. Given that Theorem 3.3.4 guarantees that f is computable in randomized polynomial-time, it remains to verify that the corresponding diversity embedding attains a distortion of $O(\log^2 n)$. That is, for any $A \in \mathscr{P}(X)$, we show that

$$\delta_{\text{diam}}(A) \leq \delta_1(f(A)) \leq O(\log^2 n)\delta_{\text{diam}}(A) \tag{3.65}$$

We let $(\mathbb{R}^{O(\log n)}, \delta^{(1)}_{\text{diam}})$ be the diameter diversity with the induced metric space $(\mathbb{R}^{O(\log n)}, d_1)$. We choose $x, y \in A$ such that

$$\delta_{\text{diam}}(A) = \max_{a,b \in A} d(a,b) = d(x,y) \tag{3.66}$$

and we choose $u, v \in A$ such that

$$\delta^{(1)}_{\text{diam}}(f(A)) = \max_{a,b \in A} d_1(f(a), f(b)) = d_1(f(u), f(v)) = \|f(u) - f(v)\|_1 \tag{3.67}$$

Then it follows that

$$\begin{aligned}
\delta_{\text{diam}}(A) &= d(x,y) && \text{by definition of the diameter diversity and choice of } x,y && (3.68)\\
&\leq \|f(x) - f(y)\|_1 && \text{by (3.54), the metric embedding } f && (3.69)\\
&\leq \max_{a,b \in A} \|f(a) - f(b)\|_1 && && (3.70)\\
&= \delta^{(1)}_{\text{diam}}(f(A)) && \text{by definition of the diameter diversity} && (3.71)\\
&\leq \delta_1(f(A)) && \text{by Lemma 3.3.6} && (3.72)\\
&\leq O(\log n)\delta^{(1)}_{\text{diam}}(f(A)) && \text{by Lemma 3.3.6} && (3.73)\\
&= O(\log n)\|f(u) - f(v)\|_1 && \text{by definition of the diameter diversity and choice of } u,v && (3.74)\\
&\leq O(\log^2 n) d(u,v) && \text{by (3.54), the metric embedding } f && (3.75)\\
&\leq O(\log^2 n) \max_{a,b \in A} d(a,b) && && (3.76)\\
&= O(\log^2 n)\delta_{\text{diam}}(A) && \text{by definition of the diameter diversity} && (3.77)
\end{aligned}$$

This completes the proof. □

3.3.3 An $O(\log n)$-Distortion Embedding

In this section we present an $O(\log n)$-distortion ℓ_1 embedding for diameter diversities.

Theorem 3.3.8 (Restatement of Theorem 1.2.3) *Let (X,δ_{diam}) be a diameter diversity with $|X| = n$. Then there exists a randomized polynomial-time embedding of (X,δ_{diam}) into the ℓ_1 diversity $(\mathbb{R}^{O(\log^2 n)},\delta_1)$ with distortion $O(\log n)$.*

Proof: As according to Theorem 3.3.5, we let $f : X \to \mathbb{R}^{O(\log^2 n)}$ be a scaled Fréchet embedding of (X,d) into the ℓ_1 metric $(\mathbb{R}^{O(\log^2 n)},d_1)$ with distortion $O(\log n)$. We then define our diversity embedding from (X,δ_{diam}) to the ℓ_1 diversity $(\mathbb{R}^{O(\log^2 n)},\delta_1)$ to be simply the map f. Given that Theorem 3.3.5 guarantees that f is computable in randomized polynomial-time, it remains to verify that the corresponding diversity embedding attains a distortion of $O(\log n)$. That is, for any $A \in \mathscr{P}(X)$, we show that

$$\frac{1}{O(\log n)}\delta_{\text{diam}}(A) \leq \delta_1(f(A)) \leq \delta_{\text{diam}}(A) \tag{3.78}$$

We choose $x,y \in A$ such that

$$\delta_{\text{diam}}(A) = \max_{a,b\in A} d(a,b) = d(x,y) \tag{3.79}$$

We begin with the first inequality.

$$\begin{aligned}
\frac{1}{O(\log n)}\delta_{\text{diam}}(A) &= \frac{1}{O(\log n)}\max_{u,v\in A} d(u,v) && \text{by definition of a diameter diversity} && (3.80)\\
&= \frac{1}{O(\log n)} d(x,y) && \text{by choice of } x,y && (3.81)\\
&\leq \|f(x)-f(y)\|_1 && \text{by (3.55), the metric embedding } f && (3.82)\\
&= \sum_{i=1}^{O(\log^2 n)} |f_i(x)-f_i(y)| && \text{by definition of the } \ell_1 \text{ metric} && (3.83)\\
&\leq \sum_{i=1}^{O(\log^2 n)} \max_{a,b\in A} |f_i(a)-f_i(b)| && && (3.84)\\
&= \delta_1(f(A)) && \text{by definition of the } \ell_1 \text{ diversity} && (3.85)
\end{aligned}$$

This completes the first inequality. Then, for each $i \in \{1,2,\ldots,O(\log^2 n)\}$ we let $a_i, b_i \in A$ be chosen such that

$$|d(a_i,A_i) - d(b_i,A_i)| = \max_{a,b\in A} |d(a,A_i) - d(b,A_i)| \tag{3.86}$$

We continue with the second inequality.

$$\begin{aligned}
\delta_1(f(A)) &= \sum_{i=1}^{O(\log^2 n)} \max_{a,b\in A} |f_i(a) - f_i(b)| && \text{by definition of the } \ell_1 \text{ diversity} \quad (3.87)\\
&= \frac{1}{O(\log^2 n)} \sum_{i=1}^{O(\log^2 n)} \max_{a,b\in A} |d(a,A_i) - d(b,A_i)| && \text{by (3.56), definition of } f \quad (3.88)\\
&= \frac{1}{O(\log^2 n)} \sum_{i=1}^{O(\log^2 n)} |d(a_i,A_i) - d(b_i,A_i)| && \text{by choice of } a_i, b_i\text{'s} \quad (3.89)\\
&\le \frac{1}{O(\log^2 n)} \sum_{i=1}^{O(\log^2 n)} d(a_i,b_i) && \text{by Proposition 3.3.3} \quad (3.90)\\
&\le \frac{1}{O(\log^2 n)} \sum_{i=1}^{O(\log^2 n)} \max_{a,b\in A} d(a,b) && \text{by } a_i, b_i \in A \quad (3.91)\\
&= \max_{a,b\in A} d(a,b) && (3.92)\\
&= \delta_{\text{diam}}(A) && \text{by definition of the diameter diversity} \quad (3.93)
\end{aligned}$$

This completes the proof □

Theorem 3.3.9 *The diameter diversity embedding into the ℓ_1 diversity from Theorem 3.3.8 achieves an asymptotically optimal distortion.*

Proof: This result follows immediately by Theorem 3.1.7, the ℓ_1-diversity embedding $\Omega(\log n)$-distortion lower bound. □

3.3.4 A k-Diameter Diversity Embedding

In this section we provide a proof of the fact that a k-diameter diversity can be embedded into the ℓ_1 diversity with distortion $O(k \log n)$.

At a high level, this $O(k \log n)$-distortion embedding follows by approximating a k-diameter diversity with a diameter diversity, incurring an $O(k)$ approximation factor. Or in other words, embedding a k-diameter diversity into a diameter diversity with distortion $O(k)$. Then, the additional $O(\log n)$ factor is incurred by embedding this diameter diversity into the ℓ_1 diversity.

Theorem 3.3.10 *Let $(X, \delta_{k\text{-diam}})$ be a k-diameter diversity, where $k \in \mathbb{Z}_{\geq 0}$. Then there is a diameter diversity into which $(X, \delta_{k\text{-diam}})$ can be embedded with distortion $O(k)$, in polynomial-time.*

Proof: We let (X,d) be the induced metric space of $(X, \delta_{k\text{-diam}})$. We let $(X, \delta_{\text{diam}}) \in \mathscr{D}_{(X,d)}$ be the corresponding diameter diversity whose induced metric space is (X,d). It suffices to show that $(X, \delta_{k\text{-diam}})$ embeds into $(X, \delta_{\text{diam}})$ with distortion k. That is, for any $A \in \mathscr{P}(X)$ it suffices to show that

$$\delta_{\text{diam}}(A) \leq \delta_{k\text{-diam}} \leq k\delta_{\text{diam}} \tag{3.94}$$

We let $A \in \mathscr{P}(X)$ be arbitrary and we choose $B \subseteq A$ where $|B| \leq k$ and

$$\delta_{k\text{-diam}}(A) = \delta_{k\text{-diam}}(B) \tag{3.95}$$

Without loss of generality we enumerate the elements of B as

$$B = \{v_1, v_2, \ldots, v_j\} \tag{3.96}$$

where $j \leq k$. Then it follows that

$$\begin{aligned}
\delta_{\text{diam}}(A) &\leq \delta_{k\text{-diam}}(A) && \text{by minimality of the diameter diversity, Theorem 2.2.10} && (3.97)\\
&= \delta_{k\text{-diam}}(B) && \text{by choice of } B && (3.98)\\
&\leq \sum_{i=2}^{j} d(v_1, v_i) && \text{by the triangle inequality, Proposition A.1.2} && (3.99)\\
&\leq \sum_{i=2}^{j} \max_{u,v \in B} d(u,v) && && (3.100)\\
&\leq k \max_{u,v \in B} d(u,v) && \text{by } j \leq k && (3.101)\\
&= k\delta_{\text{diam}}(B) && \text{by definition of the diameter diversity} && (3.102)\\
&\leq k\delta_{\text{diam}}(A) && \text{by } B \subseteq A \text{ and diversities being increasing, Proposition 2.1.4} && (3.103)
\end{aligned}$$

The induced metric space of (X,d) of $(X, \delta_{k\text{-diam}})$ can be computed in polynomial-time. Then the corresponding diameter diversity $(X, \delta_{\text{diam}})$ can be computed in polynomial-time given the (X,d). Hence, this embedding is polynomial-time computable. This completes the proof. □

Then the main result of this section follows as a corollary of Theorems 3.3.10 and 3.3.8.

Corollary 3.3.11 *Let $(X, \delta_{k\text{-diam}})$ be a k-diameter diversity, where $k \in \mathbb{Z}_{\geq 0}$. Then $(X, \delta_{k\text{-diam}})$ can be embedded into the ℓ_1 diversity with distortion $O(k \log n)$, in randomized polynomial-time.*

Proof: We let $f : (X, \delta_{k\text{-diam}}) \to (X, \delta_{\text{diam}})$ be a polynomial-time $O(k)$-distortion embedding of $(X, \delta_{k\text{-diam}})$ into some diameter diversity $(X, \delta_{\text{diam}})$, due to Theorem 3.3.10. We let $g : (X, \delta_{diam}) \to (\mathbb{R}^m, \delta_1)$ be a randomized polynomial-time $O(\log n)$-distortion embedding of $(X, \delta_{\text{diam}})$ into the ℓ_1 diversity for some dimension m, due to Theorem 3.3.8. Then the map $g \cdot f$ is an embedding of $(X, \delta_{k\text{-diam}})$ into the ℓ_1 diversity with distortion $O(k \log n)$. Moreover, this embedding is computable in randomized polynomial-time. This completes the proof. □

3.4 An Inapproximability Result for Diversity Embeddings

In this section we provide an inapproximability result for diversity embeddings into ℓ_1. In short, our result states that there are diversities which cannot be embedded into ℓ_1 with distortion smaller than $\Omega(n)$ by a polynomial-time algorithm that only queries the induced metric of the diversity, unless P=NP. A consequence of this result is that the $O(n)$-distortion ℓ_1 embedding, Theorem 3.1.5, is asymptotically optimal among algorithms that query only the induced metric of a diversity. We start by stating a conjecture of Bryant and Tupper [9] where its algorithmic refinement is disproved by our inapproximability result.

Conjecture 3.4.1 (Restatement of Bryant and Tupper's Conjecture in [9]) *There exists an $O(\sqrt{n})$-distortion diversity embedding into ℓ_1 that is based solely on the induced metric of a diversity.*

Bryant and Tupper conjecture the existence of a diversity embedding that attains low-distortion, specifically $O(\sqrt{n})$, and that this embedding utilizes only a diversity's induced metric. However, Bryant and Tupper do not insist on any algorithmic requirements, notably time complexity. We provide the following algorithmic refinement of their conjecture.

Conjecture 3.4.2 (Algorithmic Refinement of Conjecture 3.4.1) *There exists a polynomial-time $O(\sqrt{n})$-distortion diversity embedding into ℓ_1 that is based solely on the induced metric of a diversity.*

We refute this refinement with the following inapproximability result.

Theorem 3.4.3 (Restatement of 1.2.4) *For any $p \geq 0$ and for any $\varepsilon > 0$, there does not exist a polynomial-time diversity ℓ_1 embedding that queries a diversity on sets of cardinality at most $O(\log^p n)$ and achieves a distortion of $O(n^{1-\varepsilon})$, unless P=NP.*

This theorem statement is quite cumbersome and so we provide the following corollary which more clearly disproves Conjecture 3.4.2.

Corollary 3.4.4 *For any $\varepsilon > 0$, there does not exist a polynomial-time diversity ℓ_1 embedding that is based solely on the induced metric of a diversity with a distortion of $O(n^{1-\varepsilon})$.*

Proof: Setting $p = 0$ in Theorem 3.4.3, we have that there does not exist a polynomial-time diversity ℓ_1 embedding that queries a diversity on sets of cardinality at most $O(1)$ and attains a distortion of $O(n^{1-\varepsilon})$, unless P=NP. Specifically, sets of cardinality $O(1)$ include sets of cardinality two, or in other words, the induced metric of the diversity. □

Therefore, Conjecture 3.4.2 is disproved. An interesting observation is that existing diversity embeddings into ℓ_1 are both computable in polynomial-time and are computed solely using the induced metric of a diversity. Notably, these include Bryant and Tupper's $O(n)$-distortion embedding of an arbitrary diversity into ℓ_1, Theorem 3.1.5, and the two $O(\log n)$-distortion embeddings of the diameter and Steiner diversities into ℓ_1, Theorems 3.2.3 and 3.3.8. This naturally posits the observation that if one were to improve upon the

$O(n)$-distortion of embedding an arbitrary diversity into ℓ_1, say for a hypergraph Steiner diversity, one must construct an algorithm that utilizes the value of the diversity on sets of arbitrary size.

The proof of Theorem 3.4.3 rests upon a reduction from the notorious independent set problem to the problem of embedding a diversity into ℓ_1. In Section 3.4.1 we introduce the independent set problem, state an inapproximability result for it, and define the independent set diversity. In Section 3.4.2 we give our proof of Theorem 3.4.3.

3.4.1 The Independent Set Diversity

In this section we introduce the independent set problem, state an inapproximability result for it, and conclude with defining the independent set diversity.

Definition 3.4.5 (Independent Set) *Given a graph $G = (V,E)$, a subset of nodes $S \subseteq V$ is an independent set of G if for all $u, v \in S$ there does not exist an edge $(u,v) \in E$.*

Thus, the independent set problem is simply the problem of computing the maximum cardinality of an independent set of a graph.

Definition 3.4.6 (Independent Set Problem) *Given a graph $G = (V,E)$, the independent set problem, $ISP(G)$, is defined as*

$$ISP(G) = \max\{|S| : \ S \textit{ is an independent set of } G\} \tag{3.104}$$

For any $\varepsilon > 0$, the independent set problem is inapproximable up to a factor of $O(n^{1-\varepsilon})$, unless $P = NP$ [2, 18].

Theorem 3.4.7 (Inapproximability of the Independent Set Problem) *For any $\varepsilon > 0$, there does not exist a polynomial-time approximation algorithm for the independent set problem with an approximation factor smaller than $O(n^{1-\varepsilon})$, unless P=NP.*

In order to construct the independent set diversity we require the following technical lemma that characterizes an independent set function as being nonnegative, increasing, and subadditive.

Lemma 3.4.8 *Let $G = (V,E)$ be a graph. We define the independent set function $f_{IS} : 2^V \to \mathbb{Z}_{\geq 0}$ as*

$$f_{IS}(A) = \max\{|S| : S \subseteq A, \ S \textit{ is an independent set of } G\} \tag{3.105}$$

Then the set function f_{IS} is nonnegative, increasing, and subadditive.

Proof: By construction, f_{IS} is clearly nonnegative. As for the increasing property of f_{IS}, it suffices to show that for any any arbitrary $A \subseteq B \subseteq V$, $f_{IS}(A) \leq f_{IS}(B)$. We let $A \subseteq B \subseteq V$ be arbitrary and we suppose that $S \subseteq A$ is an independent set of G such that $f_{IS}(A) = |S|$. Since $S \subseteq A \subseteq B$, it also follows that $|S| \leq f_{IS}(B)$. Hence $f_{IS}(A) \leq f_{IS}(B)$ and thus f_{IS} is increasing.

Next, to prove that f_{IS} is subadditive we first argue that independent sets of G are downwards closed. That is, if $S \subseteq V$ is an independent of G, then for any $S' \subseteq S$ it follows that S' is an independent set of G. Let $u, v \in S'$ be arbitrary, since $u, v \in S' \subseteq S$ and S is an independent set of G then $(u, v) \notin E$ and so S' is also an independent set of G.

Now we proceed to prove that f_{IS} is subadditive. Let $A, B \subseteq V$ be arbitrary. It suffices to show that $f_{IS}(A \cup B) \leq f_{IS}(A) + f_{IS}(B)$. We define $A' = A \setminus B$ and $B' = B$. Then $A' \cap B' = \emptyset$ and $A' \cup B' = A \cup B$. Hence,

$$f(A \cup B) = f(A' \cup B') \tag{3.106}$$

We let $S_{A' \cup B'} \subseteq A' \cup B', S_{A'} \subseteq A', S_{B'} \subseteq B'$ be independent sets of G such that

$$f_{IS}(A' \cup B') = |S_{A' \cup B'}|, \; f_{IS}(A') = |S_{A'}|, \; f_{IS}(B') = |S_{B'}| \tag{3.107}$$

Since $S_{A' \cup B'} \subseteq A' \cup B'$ then

$$S_{A' \cup B'} = (S_{A' \cup B'} \cap A') \cup (S_{A' \cup B'} \cap B') \tag{3.108}$$

Moreover, since $A' \cap B' = \emptyset$ then

$$|S_{A' \cup B'}| = |S_{A' \cup B'} \cap A'| + |S_{A' \cup B'} \cap B'| \tag{3.109}$$

We also note that by the downwards closed property of independent sets, $S_{A' \cup B'} \cap A'$ and $S_{A' \cup B'} \cap B'$ are independent sets of G. Furthermore, this implies that

$$|S_{A' \cup B'} \cap A'| \leq f_{IS}(A') \text{ and } |S_{A' \cup B'} \cap B'| \leq f_{IS}(B') \tag{3.110}$$

Putting everything together, we have that

$$\begin{aligned}
f_{IS}(A \cup B) &= f_{IS}(A' \cup B') && \text{by (3.106)} && (3.111) \\
&= |S_{A' \cup B}| && \text{by (3.107)} && (3.112) \\
&= |S_{A' \cup B'} \cap A'| + |S_{A' \cup B'} \cap B'| && \text{by (3.109)} && (3.113) \\
&\leq f_{IS}(A') + f_{IS}(B') && \text{by (3.110)} && (3.114) \\
&\leq f_{IS}(A) + f_{IS}(B) && \text{by the fact that } f_{IS} \text{ is increasing} && (3.115)
\end{aligned}$$

This completes the proof. □

Next, we define the independent set diversity and according to Lemma 2.3.2 and Lemma 3.4.8 it follows that our construction is in fact a diversity.

Definition 3.4.9 (Independent Set Diversity) *Let $G = (V, E)$ be a graph and let f_{IS} be defined as it is in Lemma 3.4.8. Then we define the independent set diversity, $(V, \delta_I S)$, as*

$$\delta_{IS}(A) = \begin{cases} f_{IS}(A) & \textit{if } |A| \geq 2 \\ 0 & \textit{otherwise} \end{cases} \tag{3.116}$$

Proof: We prove that (V, δ_{IS}) is in fact a diversity. According to Lemma 3.4.8, f_{IS} is nonnegative, increasing, and subadditive set function over the ground set V. Then, according to Lemma 2.3.2 and the definition of (V, δ_{IS}) it follows that (V, δ) is a (pseudo)-diversity. In particular, (V, δ_{IS}) is a diversity since for any $A \subseteq V$ where $|A| \geq 2$ it follows that $\delta_{IS}(A) = f_{IS}(A) \geq 1 > 0$. □

3.4.2 Proof of Theorem 3.4.3

Theorem 3.4.10 (Restatement of Theorem 3.4.3) *For any $p \geq 0$ and for any $\varepsilon > 0$, there does not exist a polynomial-time diversity ℓ_1 embedding that queries a diversity on sets of cardinality at most $O(\log^p n)$ with a distortion of $O(n^{1-\varepsilon})$, unless P=NP.*

Proof: For this proof, we assume that P≠NP, and for the sake of contradiction, we assume the negation of the theorem statement. That is, we suppose that for some $p \geq 0$ and for some $\varepsilon > 0$ there is a polynomial-time diversity embedding into ℓ_1 that queries the diversity on sets of cardinality at most $O(\log^p n)$ with a distortion of $O(n^{1-\varepsilon})$.

We let $G = (V, E)$ be an arbitrary graph with $|V| = n$. We let (V, δ_{IS}) be the corresponding independent set diversity as defined by Definition 3.4.5. (V, δ_{IS}) is defined implicitly by the graph G and, at this point, we are not insisting on being able to explicitly compute $\delta_{IS}(A)$ for any $A \subseteq V$. We note that by definition of (V, δ_{IS}) we have that

$$\delta_{IS}(V) = \max\{|S| : S \subseteq V, S \text{ is an independent set of } G\} = \text{ISP}(G) \tag{3.117}$$

Therefore, any approximation of $\delta_{IS}(V)$ yields an approximation of the instance of the independent set problem, ISP(G). In particular, we show how an algorithm that embeds a diversity into ℓ_1 yields an approximation algorithm for $\delta_{IS}(V)$ and in turn for ISP(G).

We let $f : V \to \mathbb{R}^d$, for some $d \in \mathbb{Z}_{\geq 0}$, be the diversity embedding of (V, δ_{IS}) into $(\mathbb{R}^d, \delta_1)$ given by our algorithm at the beginning of this proof. By assumption we can compute f in polynomial-time, provided we can query $\delta_{IS}(A)$ for sets $A \subseteq V$ where $|A| \in O(\log^p n)$ in polynomial-time. Although, (V, δ_{IS}) is defined implicitly by the graph G, we can compute $\delta_{IS}(A)$ for every $A \subseteq V$ where $|A| \in O(\log^p n)$ in polynomial-time, specifically $O(n^p)$, using a brute-force enumeration of all independent sets of G that are subsets of A. Therefore, f is computable in polynomial-time.

Furthermore, the distortion of the embedding f is of the factor $O(n^{1-\varepsilon})$. Specifically, we have the following guarantee on the value of $\delta_{IS}(V)$ with respect to $\delta_1(f(V))$,

$$\frac{1}{c_1}\delta_{IS}(V) \leq \delta_1(f(V)) \leq c_2\delta_{IS}(V) \tag{3.118}$$

where $c_1, c_2 > 0$ and $c_1 c_2 = O(n^{1-\varepsilon})$. Given that $\delta_{IS}(V) = \text{ISP}(G)$ and that f is computable in polynomial-time, this implies that we have a polynomial-time approximation algorithm for the independent set problem with an approximation factor of $O(n^{1-\varepsilon})$. However, this contradicts Theorem 3.4.7. This completes the proof.

$\square$

Chapter 4

Approximating the Sparsest Cut in Hypergraphs

4.1 Approximating the Sparsest Cut via Diversity Embeddings

In this chapter we present approximation algorithms for the sparsest cut problem in hypergraphs via diversity embeddings. Our approach follows the seminal technique of utilizing metric embeddings for the sparsest cut problem in the setting where the both supply and demand hypergraphs are simply graphs. This is approach is due to Linial, London, and Rabinovich [24] and Aumann and Rabani [5]. They formulate an LP relaxation for the sparsest cut that optimizes over pseudo-metric spaces and then embed the metric space corresponding to the optimal LP solution into the ℓ_1 metric with low distortion. An approximate sparsest cut is then extracted due to the fact that ℓ_1 metrics are a nonnegative sum of cut pseudo-metrics. The approach that we follow is due to Bryant and Tupper [31] where they show that the flow-cut gap of the sparsest cut and the maximum concurrent multicommodity flow is bounded above by the minimum distortion of embedding a diversity into ℓ_1. This generalizes the classical graph setting.

In this section, we present this work of Bryant and Tupper. Specifically, we present an LP relaxation for the sparsest cut in hypergraphs that optimizes over pseudo diversities. Then, we show how embedding a diversity, corresponding to an optimal solution, into ℓ_1 yields an approximate sparsest cut in hypergraphs. We culminate this discussion with a randomized polynomial-time approximation algorithm for the sparsest cut in hypergraphs, Theorem 4.1.1. Bryant and Tupper did not specifically address the tractability of their approach. Namely, how to solve the LP relaxation and then extract cut information in polynomial time.

In Section 4.2 we characterize the optimal solutions of the LP relaxation of the sparsest cut. In particular we show that if either the supply or demand hypergraphs is a graph, then the resulting optimal solution is a Steiner or a diameter diversity, which both have $O(\log n)$-distortion embeddings into the ℓ_1 diversity (see Theorems 3.2.3 and 3.3.8. Then, in Section 4.3 we provide algorithmic details which show that there is a randomized polynomial-time $O(\log n)$-approximation algorithm for the sparsest cut problem in the setting where either the supply or demand hypergraphs is a graph.

Throughout this chapter we consider an instance of the sparsest cut problem defined by a supply hypergraph $G = (V, E_G, w_G)$ with rank r_G and a demand hypergraph $H = (V, E_H, w_H)$ with rank r_H, as defined in Section 1.1 and Definition 1.1.1.

4.1.1 A Linear Programming Relaxation

According to the definition of the sparsest cut ϕ, Definition 1.1.3, and the definition of a cut pseudo diversity, Definition 2.2.6, we have that the sparsest cut ϕ is equivalent to

$$\phi = \min_{A \subseteq V} \frac{\sum_{U \in E_G} w_G(U)\delta_A(U)}{\sum_{S \in E_H} w_H(S)\delta_A(S)} \tag{4.1}$$

Furthermore, we can relax this optimization over cut pseudo diversities δ_A by the following optimization over all pseudo diversities

$$\min_{(V,\delta) \text{ is a pseudo-diversity}} \frac{\sum_{U \in E_G} w_G(U)\delta(U)}{\sum_{S \in E_H} w_H(S)\delta(S)} \tag{4.2}$$

We call (4.2) the *sparsest cut diversity-relaxation*. Moreover, since (4.2) is a relaxation of (4.1) we have that

$$\phi \geq \min_{(V,\delta) \text{ is a pseudo-diversity}} \frac{\sum_{U \in E_G} w_G(U)\delta(U)}{\sum_{S \in E_H} w_H(S)\delta(S)} \tag{4.3}$$

It can be shown that (4.2) is equivalent to the following optimization problem

$$\begin{aligned} \min \sum_{U \in E_G} & w_G(U)\delta(U) \\ \text{s.t.} \sum_{S \in E_H} & w_H(S)\delta(S) \geq 1 \\ & (V, \delta) \text{ is a pseudo-diversity} \end{aligned} \tag{4.4}$$

We recall Definition 2.2.1, $\mathscr{T}_{(G,S)}$, the collection of subsets of hyperedges of G that correspond to connected subhypergraphs that contain the nodes S. Then (4.4) can be shown to be equivalent to the following LP

$$\begin{aligned} \min \sum_{U \in E_G} & w(U)d_U \\ \text{s.t.} \sum_{S \in E_H} & w_H(S)y_S \geq 1 \\ \sum_{U \in t} & d_U \geq y_S && \forall S \in E_H, t \in \mathscr{T}_{(G,S)} \\ & d_U \geq 0 && \forall U \in E_G \\ & y_S \geq 0 && \forall S \in E_H \end{aligned} \tag{4.5}$$

We can take a feasible solution of (4.5), $\{d_U\}_{U \in E_G}$ and $\{y_S\}_{S \in E_H}$, and define a corresponding feasible solution to (4.4) with an equivalent objective value. This corresponding solution is a hypergraph Steiner diversity

$$(V, \delta) \text{ that corresponds to the hypergraph } (V, E_G, w) \tag{4.6}$$

where $w : E_G \to \mathbb{R}_+$ is defined as $w(U) = d_U$.

We remark that this LP relaxation has polynomially[1] many variables and in general it has exponentially many constraints. Specifically, for an $S \in E_H$ it could be the case $\mathscr{T}_{(G,S)}$ contains exponentially many subsets of hyperedges. Hence, the set of constraints

$$\{\sum_{U \in t} d_U \geq y_S\}_{t \in \mathscr{T}_{(G,S)}} \tag{4.7}$$

is exponentially large in general. Thus, we can solve this LP (approximately) if there is a polynomial time (approximate) separation oracle for

$$\min_{t \in \mathscr{T}_{(G,S)}} \sum_{U \in t} d_U \geq y_S \tag{4.8}$$

for each $S \in E_H$. This is the hypergraph Steiner problem (HSP) it is NP-complete [6]. We do have an $O(\log r_H)$-approximate separation oracle however. We defer discussion of HSP to Chapter 5.

4.1.2 Rounding the Linear Programming Relaxation

Given a solution to LP Relaxation (4.5) we immediately have a solution to the sparsest cut diversity-relaxation (4.2), namely (V,δ) from (4.6). We show how to extract a cut from this diversity that approximates the sparsest cut. We let $f : V \to \mathbb{R}^m$ be an embedding from (V,δ) to the ℓ_1 diversity $(\mathbb{R}^m, \delta_1)$ with distortion $c \geq 1$. This yields the following inequalities

$$\frac{\sum_{U \in E_G} w_G(U)\delta_1(f(U))}{\sum_{S \in E_H} w_H(S)\delta_1(f(S))} \leq c\frac{\sum_{U \in E_G} w_G(U)\delta(U)}{\sum_{S \in E_H} w_H(S)\delta(S)} \leq c\phi \tag{4.9}$$

where the first inequality follows by the fact that f is an embedding with distortion c and the second inequality follows by the fact that (V,δ) is a feasible solution to (4.2) which is a relaxation of ϕ according to Equation (4.3).

According to Theorem 3.1.2, δ_1 is a nonnegative sum of $O(nm)$ cut pseudo-diversities. Therefore, there exists some collection of subsets of $2^{f(V)}$, $\mathscr{F} \subseteq 2^{f(V)}$ where $|\mathscr{F}| \in O(nm)$, such that for any $A \subseteq V$

$$\delta_1(f(A)) = \sum_{B \in \mathscr{F}} \alpha_B \delta_B(f(A)) \tag{4.10}$$

where $\{\alpha_B\}_{B \in \mathscr{F}}$ are positive scalars and $\{(f(V),\delta_B)\}_{B \in \mathscr{F}}$ is a collection of cut pseudo-diversities. This immediately yields that

$$\frac{\sum_{U \in E_G} w_G(U)\delta_1(f(U))}{\sum_{S \in E_H} w_H(S)\delta_1(f(S))} = \frac{\sum_{U \in E_G} w_G(U)\left[\sum_{B \in \mathscr{F}} \alpha_B \delta_B(f(U))\right]}{\sum_{S \in E_H} w_H(S)\left[\sum_{B \in \mathscr{F}} \alpha_B \delta_B(f(S))\right]} \tag{4.11}$$

Next, rearranging the order of the summations, we have that (4.11) is equivalent to

$$\frac{\sum_{B \in \mathscr{F}} \alpha_B \left[\sum_{U \in E_G} w_G(U)\delta_B(f(U))\right]}{\sum_{B \in \mathscr{F}} \alpha_B \left[\sum_{S \in E_H} w_H(S)\delta_B(f(S))\right]} \tag{4.12}$$

[1] Unless G and H are not defined explicitly.

Finally, it can be shown that there exists some cut $B_0 \subseteq V$ where $f(B_0) \in \mathscr{F}$ such that

$$\frac{\sum_{U\in E_G} w_G(U)\delta_{f(B_0)}(f(U))}{\sum_{S\in E_H} w_H(S)\delta_{f(B_0)}(f(S))} \leq \frac{\sum_{B\in\mathscr{F}} \alpha_B\left[\sum_{U\in E_G} w_G(U)\delta_B(f(U))\right]}{\sum_{B\in\mathscr{F}} \alpha_B\left[\sum_{S\in E_H} w_H(S)\delta_B(f(S))\right]} \tag{4.13}$$

We note that the left side of the inequality (4.13) is simply

$$\phi(B_0) = \frac{\sum_{U\in E_G} w_G(U)\delta_{B_0}(U)}{\sum_{S\in E_H} w_H(S)\delta_{B_0}(S)} = \frac{\sum_{U\in E_G} w_G(U)\delta_{f(B_0)}(f(U))}{\sum_{S\in E_H} w_H(S)\delta_{f(B_0)}(f(S))} \tag{4.14}$$

where the second equality follows by the fact that cuts are preserved under the map $f: V \to \mathbb{R}^m$.

Hence putting together (4.9), (4.11), (4.12), (4.13), (4.14), and the fact that $\phi \leq \phi(B_0)$ we have that

$$\phi \leq \phi(B_0) \leq c\phi \tag{4.15}$$

4.1.3 An Approximation Algorithm for the Sparsest Cut

Theorem 4.1.1 *Let $G = (V, E_G, w_G)$ be a supply hypergraph with rank r_G and $H = (V, E_H, w_H)$ be a demand hypergraph with rank r_H.*

1. *Let α_{LP} be the approximation factor for some (randomized) polynomial-time algorithm that (approximately) solves LP Relaxation (4.5).*

2. *Let α_{div} be the approximation factor for some (randomized) polynomial-time algorithm that computes the hypergraph Steiner diversity corresponding to the optimal solution of LP Relaxation 4.5, as defined according to Equation (4.6).*

3. *Let α_f be the distortion of embedding the aforementioned diversity into the ℓ_1 diversity for some (randomized) polynomial time algorithm.*

Then the approach outlined in this section forms a (randomized) polynomial-time $O(\alpha_{LP}\alpha_{div}\alpha_f)$-approximation algorithm for the sparsest cut problem in hypergraphs.

4.2 Characterizing the Optimal Solutions of the Sparsest Cut Diversity-Relaxation

In this section we provide our contributions, the characterization of the optimal solutions of the sparsest cut diversity-relaxation, Equation (4.2), or equivalently, of LP Relaxation (4.5). We show that when the supply hypergraph is a graph then the optimal diversity is a Steiner diversity and when the demand hypergraph is a graph then the optimal diversity is a diameter diversity. Notably, both of these diversities have $O(\log n)$-distortion polynomial-time embeddings into ℓ_1.

Theorem 4.2.1 *Let $G = (V, E_G, w_G)$ be a supply hypergraph and let $H = (V, E_H, w_H)$ be a demand hypergraph with rank r_H. Let (V, δ) be a (pseudo) diversity attaining the optimal objective value to the sparsest* *cut diversity-relaxation, Equation (4.2). Let $(V, \delta_{r_H\text{-diam}})$ be the r_H-diameter diversity of (V, δ). Then (V, δ) can be asummed to be $(V, \delta_{r_H\text{-diam}})$.*

Proof: According to Theorem 2.2.14, (V, δ) and $(V, \delta_{r_H\text{-diam}})$ are both members of $\mathscr{D}_{(V,\delta,r_H)}$, and moreover, $(V, \delta_{r_H\text{-diam}})$ is the minimal diversity of this family. By definition of $\mathscr{D}_{(V,\delta,r_H)}$ and by the fact that $\forall S \in E_H, |S| \leq r_H$, it follows that

$$\sum_{S \in E_H} w_H(S)\delta(S) = \sum_{S \in E_H} w_H(S)\delta_{r_H\text{-diam}}(S) \tag{4.16}$$

By the minimality, of $(V, \delta_{r_H\text{-diam}})$ among the family $\mathscr{D}_{(V,\delta,r_H)}$ it follows that

$$\sum_{U \in E_G} w_G(U)\delta_{r_H\text{-diam}}(U) \leq \sum_{U \in E_G} w_G(U)\delta(U) \tag{4.17}$$

Then (4.16) and (4.17) imply that

$$\frac{\sum_{U \in E_G} w_G(U)\delta_{r_H\text{-diam}}(U)}{\sum_{S \in E_H} w_H(S)\delta_{r_H\text{-diam}}(S)} \leq \frac{\sum_{U \in E_G} w_G(U)\delta(U)}{\sum_{S \in E_H} w_H(S)\delta(S)} \tag{4.18}$$

Therefore, the optimal diversity for the sparsest cut diversity-relaxation, (V, δ), can be assumed to be its the r_H-diameter diversity $(V, \delta_{r_H\text{-diam}})$, thus completing the proof. □

Corollary 4.2.2 *Let $G = (V, E_G, w_G)$ be a supply hypergraph and let $H = (V, E_H, w_H)$ be a demand graph, that is it has rank $r_H = 2$. Let (V, δ) be a (pseudo) diversity attaining the optimal objective value to the sparsest cut diversity-relaxation, Equation (4.2). Let (V, d) be the induced metric space of (V, δ) and let $(V, \delta_{diam}) \in \mathscr{D}_{(V,d)}$ be the diameter diversity whose induced metric space is (V, d). Then (V, δ) can be assumed to be (V, δ_{diam}), a diameter diversity.*

Proof: This corollary follows by Fact 2.2.12 and Theorem 4.2.1. □

Theorem 4.2.3 *Let $G = (V, E_G, w_G)$ be a supply graph, that is it has rank $r_G = 2$, and let $H = (V, E_H, w_H)$ be a demand hypergraph. Let (V, δ) be a (pseudo) diversity attaining the optimal objective value to the sparsest cut diversity-relaxation, Equation (4.2). Let (V, d) be the induced metric space of (V, δ) and let $(V, \delta_{Steiner}) \in \mathscr{D}_{(V,d)}$ be the Steiner diversity whose induced metric space is (V, d). Then (V, δ) can be assumed to be $(V, \delta_{Steiner})$, a Steiner diversity.*

Proof: Let $U \in E_G$ be arbitrary. Since $r_G = 2$ then $|U| = 2$ and we let $U = \{u, v\}$. Then it follows that

$$\delta(U) = d(u, v) = \delta_{\text{Steiner}}(U) \tag{4.19}$$

where the two equalities follow by the fact that $(V, \delta), (V \delta_{\text{Steiner}}) \in \mathscr{D}_{(V,d)}$. From this we have that

$$\sum_{U \in E_G} w_G(U)\delta(U) = \sum_{U \in E_G} w_H(S)\delta_{\text{Steiner}}(U) \tag{4.20}$$

By theorem 2.2.10, $(V, \delta_{\text{Steiner}})$ is the maximal diversity of the family $\mathscr{D}_{(V,d)}$ and so it follows that

$$\sum_{S \in E_H} w_H(S)\delta(S) \leq \sum_{S \in E_H} w_H(S)\delta_{\text{Steiner}}(S) \tag{4.21}$$

Then (4.20) and (4.21) imply that

$$\frac{\sum_{U \in E_G} w_G(U)\delta_{\text{Steiner}}(U)}{\sum_{S \in E_H} W_H(S)\delta_{\text{Steiner}}(S)} \leq \frac{\sum_{U \in E_G} w_G(U)\delta(U)}{\sum_{S \in E_H} w_H(S)\delta(S)} \tag{4.22}$$

Therefore, the optimal diversity for the sparsest cut diversity-relaxation (V, δ) can be assumed to be the Steiner diversity $(V, \delta_{\text{Steiner}})$. This completes the proof. □

4.3 Algorithmic Implications

Based off our approach, we first present a polynomial-time approximation algorithm for the case where the supply and demand hypergraphs are arbitrary hypergraphs with ranks r_G and r_H, respectively.

Theorem 4.3.1 (Restatement of 1.2.1) *Let $G = (V, E_G, w_G)$ be a supply hypergraph with rank r_G and let $H = (V, E_H, w_H)$ be a demand hypergraph with rank r_H. Then there is a randomized polynomial-time $O(\min\{r_G, r_H\} \log n \log r_H)$-approximation algorithm for the sparsest cut of G and H.*

Proof: According to Theorem 4.1.1 there is a polynomial-time $O(\alpha_{\text{LP}}\alpha_{\text{div}}\alpha_f)$-approximation algorithm for the sparsest cut of G and H, where α_{LP} α_{div}, and α_f are as defined in Theorem 4.1.1.

According to Corollary 5.3.2, there is an $O(\log r_H)$-approximation algorithm for LP Relaxation (4.5), hence $\alpha_{\text{LP}} = O(\log r_H)$.

We let (V, δ) be an optimal solution to the sparsest cut diversity-relaxation (4.2). We note that (V, δ) is a hypergraph Steiner diversity corresponding to a rank r_G hyperedge-weighted hypergraph, namely (V, E_G, w) where $w(U) = d_U$ and $\{d_U\}_{U \in E_G}$ are from an optimal solution to LP Relaxation (4.5). Then according to Corollary 3.2.5 there exists a randomized polynomial-time $O(r_G \log n)$-distortion embedding of (V, δ) into the ℓ_1 diversity.

Alternatively, according to Theorem 4.2.1 (V, δ) is a r_H-diameter diversity. Then according to Corollary 3.3.11 there is a randomized polynomial-time $O(r_H \log n)$-distortion embedding of (V, δ) into the ℓ_1 diversity. Thus, we can choose whether to embed (V, δ) into ℓ_1 as a hypergraph Steiner diversity or a r_H-diameter diversity based on whether r_G or r_H is smaller. Therefore, $\alpha_f = O(\min\{r_G, r_H\} \log n)$.

Since the two embeddings, Corollary 3.2.5 and 3.3.11 only require the induced metric space of (V, δ) then we only need to compute $\delta(A)$ for $A \in \mathscr{P}(V)$ where $|A| = 2$. Thus, by Corollary 5.3.1 we have that $\alpha = O(\log 2) = O(1)$.

Hence, we have a randomized polynomial-time $O(\min\{r_G, r_H\} \log n \log r_H)$-approximation algorithm for the sparsest cut problem in G and H. This completes the proof □

This is the first randomized polynomial-time approximation algorithm for the setting where G and H are arbitrary hypergraphs. An immediate corollary of this result is an $O(\log n)$-approximation algorithm for the case where the demand hypergraph is simply a graph

Corollary 4.3.2 *Let $G = (V, E_G, w_G)$ be a supply hypergraph and $H = (V, E_H, w_H)$ be a demand graph, that is it has rank $r_H = 2$. Then there is a randomized polynomial-time $O(\log n)$-approximation algorithm for the sparsest cut of G and H.*

Proof: This corollary follows immediately by Theorem 4.3.1 and the fact that $r_H = 2$. $\square$

For this setting where G is a hypergraph and H is a graph, there is a randomized polynomial-time $O(\sqrt{\log r_G \log n} \log\log n)$-approximation algorithm due to Louis [25] which is based off an SDP relaxation. However, among LP-based approaches our algorithm is the first to attain an $O(\log n)$-approximation when H is an arbitrary graph. Specifically, Kapralov et al. [19] attain an $O(\log n)$-approximation when H has uniform demands.

We note that this corollary can be proven alternatively by our characterization of the optimal diversity to the sparsest cut diversity-relaxation (4.2) being a diameter diversity for the case when the demand hypergraph is a graph, Corollary 4.2.2. Similarly, our characterization Theorem 4.2.3 underpins a randomized polynomial-time $O(\log n)$-approximation algorithm for the case where the supply hypergraph is a graph.

Theorem 4.3.3 (Restatement of 1.2.2) *Let $G = (V, E_G, w_G)$ be a supply graph, that is it has rank $r_G = 2$, and $H = (V, E_H, w_H)$ be a demand hypergraph. Then there is a randomized polynomial-time $O(\log n)$-approximation algorithm for the sparsest cut of G and H.*

Proof: According to Theorem 4.1.1 there is a polynomial-time $O(\alpha_{\text{LP}} \alpha_{\text{div}} \alpha_f)$-approximation algorithm for the sparsest cut of G and H, where α_{LP} α_{div}, and α_f are as defined in Theorem 4.1.1.

We recall that for each $S \in E_H$ the LP Relaxation (4.5) may have exponentially many constraints of the form

$$\{\sum_{U \in t} d_U \geq y_S\}_{t \in \mathscr{T}_{(G,S)}} \tag{4.23}$$

Approximately separating over these constraints amounts to approximating the minimum cost Steiner tree for the nodes S. Therefore, we can approximately separate over these constraints in polynomial time using a polynomial-time $O(1)$-approximation algorithm for the minimum-cost Steiner tree problem [12, 22, 30, 32]. Hence $\alpha_{\text{LP}} = O(1)$.

We let (V, δ) be the optimal diversity of the sparsest cut diversity-relaxation. Then according to Theorem 4.2.3 (V, δ) is a Steiner diversity, and moreover, we can compute $\delta(A)$ for any $A \subseteq V$ in polynomial time up to a factor of $O(1)$, again, by a polynomial-time $O(1)$-approximation algorithm for the Steiner tree problem. Therefore, $\alpha_{\text{div}} = O(1)$.

Finally, according to Theorem 3.2.3, there is a randomized polynomial-time $O(\log n)$-distortion embedding of (V, δ), a Steiner diversity, into ℓ_1. Hence $\alpha_f = O(\log n)$.

Hence, we have a randomized polynomial-time $O(\log n)$-approximation algorithm for the sparsest cut problem in G and H. This completes the proof.

$\square$

The previous state-of-the-art algorithm for the setting where G is a graph and H is a hypergraph is a polynomial-time $O(\log n \log(|E_H| r_H)$-approximation algorithm due to Plotkin et al. [28]. Our $O(\log n)$-approximation is a notable improvement due to the fact that $|E_H|$ may be exponentially large.

Chapter 5

The Minimum Cost Hypergraph Steiner Problem

In this chapter we give an asymptotically optimal approximation algorithm for the minimum cost hypergraph Steiner problem, a key problem that emerges in the computation of diversities and in their application to the generalized sparsest cut problem in hypergraphs. Not only is our algorithm optimal but it is also the first approximation algorithm for this problem. In Section 5.1 we define and introduce the problem, commenting on its relevance to the computation of diversities and to their application to the sparsest cut problem in hypergraphs. We also our new results, Theorems 5.1.2 and 5.1.3. In Section 5.2 we provide proofs of these two theorems. In Section 5.3 we provide applications of our approximation algorithm to the minimum-cost hypergraph Steiner problem.

5.1 The Minimum Cost Hypergraph Steiner Problem

We begin by defining the the minimum cost hypergraph Steiner problem.

Definition 5.1.1 (Minimum Cost Hypergraph Steiner Problem (HSP)) *Let $G = (V, E, w)$ be a hypergraph with nonnegative hyperedge weights $w : E \to \mathbb{R}_+$. For a set of Steiner nodes $S \subseteq V$, we define $\mathscr{T}_S$ to be the collection of connected subhypergraphs of G that contain the Steiner nodes S. Then the minimum cost hypergraph Steiner Problem, $HSP(G, S)$ is defined as*

$$HSP(G,S) = \min_{t \in \mathscr{T}_S} \sum_{U \in t} w(U) \tag{5.1}$$

For convenience we may refer to the minimum cost hypergraph Steiner problem as simply the hypergraph Steiner problem, hence the use of the abbreviation HSP.

In the context of diversities and the generalized sparsest cut problem in hypergraphs, the minimum cost hypergraph Steiner problem emerges in the following two computational problems.

1. Given a hypergraph Steiner diversity $(V, \delta_{\mathscr{H}})$ defined by a hypergraph $H = (V, E, w)$ with nonnegative hyperedge weights $w : E \to \mathbb{R}_+$, for any $S \subseteq V$, the diversity value $\delta_{\mathscr{H}}(S)$ is precisely defined

as $\delta(S) = HSP(H,S)$. If one is given $(V, \delta_{\mathscr{H}})$ by the hypergraph H then to (approximately) query $\delta_{\mathscr{H}}(S)$ one must (approximately) solve $HSP(H,S)$. For the application to sparsest cut in hypergraphs, the hypergraph Steiner diversity that emerges from the sparsest cut LP Relaxation (4.5) is defined implicitly by a hyperedge-weighted hypergraph. Thus, to (approximately) solve the sparsest cut problem in hypergraphs via diversity embeddings it is necessary to (approximately) solve the minimum cost hypergraph Steiner problem.

2. Additionally, the sparsest cut LP Relaxation (4.5) has polynomially many variables but exponentially many constraints. Specifically, for each demand $S \in E_H$, there may be an exponential number of constraints of the form

$$\sum_{U \in t} d_U \geq y_S, \forall t \in \mathscr{T}_{(G,S)} \tag{5.2}$$

which can be (approximately) separated over by (approximately) solving an instance of the hypergaph Steiner problem. Hence, computing a polynomial time (approximate) separation oracle, and in turn, obtaining a polynomial time (approximation) algorithm for the sparsest cut LP relaxation amounts to obtaining a polynomial time (approximation) algorithm for the minimum cost hypergraph Steiner problem.

Thus, it is evident that the minimum cost hypergraph Steiner problem is a computational bottleneck for our approach of utilizing diversity embeddings for the sparsest cut problem in hypergraphs. This problem is a natural generalization of both the Steiner tree problem in graphs and the minimum cost spanning sub-hypergraph problem (MSSP) [6], and yet surprisingly, it has not been previously investigated. On this note, we state our contribution which is the first polynomial time approximation algorithm for the minimum cost hypergraph Steiner problem, and moreover, this algorithm also happens to be optimal unless P=NP.

Theorem 5.1.2 (Restatement of Theorem 1.2.5) *There exists a polynomial time $O(\log n)$-approximation algorithm for the minimum cost hypergraph Steiner problem. Specifically, for a hypergraph $G = (V, E, w)$ with nonnegative hyperedge weights $w : E \to \mathbb{R}_+$ and for a set of Steiner nodes $S \subseteq V$, $HSP(G,S)$ can be approximated up to a factor of $O(\log |S|)$ in polynomial time.*

Theorem 5.1.3 *The minimum cost hypergraph Steiner problem cannot be approximated to a factor smaller than $\Omega(\log n)$ in polynomial time unless P=NP. Specifically, for a hypergraph $G = (V, E, w)$ with nonnegative hyperedge weights $w : E \to \mathbb{R}_+$ and for a set of Steiner nodes $S \subseteq V$, $HSP(G,S)$ cannot be approximated up to a factor smaller than $\Omega(\log |S|)$ in polynomial time unless P=NP.*

In the subsequent section we provide proofs of Theorem 5.1.2 and Theorem 5.1.3. Respectively, these theorems rest upon a reduction to the minimum cost node-weighted Steiner tree problem and a reduction from the set cover problem.

5.2 An Optimal Algorithm for the Minimum Cost Hypergraph Steiner Problem

In this section we provide a proof of Theorem 5.1.2 followed by a proof of Theorem 5.1.3.

5.2.1 Proof of Theorem 5.1.2

We proceed by defining the minimum cost node-weighted Steiner tree problem.

Definition 5.2.1 *Let $G = (V, E, w)$ be a graph with nonnegative node weights $w : V \to \mathbb{R}_+$. For a set of Steiner nodes $S \subseteq V$ we define $\mathscr{T}_S$ to be the set of minimally connected subgraphs of G containing the Steiner nodes S, or simply subtrees of G spanning S, in accordance with Definition 2.2.1. Then the minimum cost node-weighted Steiner tree problem (NSTP), $NSTP(G,S)$, is defined as*

$$NSTP(G,S) = \min_{t \in \mathscr{T}_S} \sum_{v \in \cup_{e \in t} e} w(v) \tag{5.3}$$

According to Klein and Ravi [21] the minimum cost node-weighted Steiner tree problem has a polynomial time $O(\log n)$-approximation algorithm, which we state as the following theorem.

Theorem 5.2.2 (Restatement of Theorem 1.1 of [21]) *Let $G = (V, E, w)$ be a graph with nonnegative node weights $w : V \to \mathbb{R}_+$. For a set of Steiner nodes $S \subseteq V$ there is a polynomial time $O(\log |S|)$-approximation algorithm for the minimum-cost node-weighted Steiner tree problem $NSTP(G,S)$.*

As noted earlier, our algorithm for the minimum-cost hypergraph Steiner problem follows by a reduction to the minimum-cost node-weighted Steiner tree problem.

Lemma 5.2.3 *Let $G = (V, E, w)$ be a hypergraph with nonnegative hyperedge weighted $w : E \to \mathbb{R}_+$ and let $S \subseteq V$ be a set of Steiner nodes. Then the minimum-cost hypergraph Steiner problem $HSP(G,S)$ can be reduced to an instance of the minimum-cost node-weighted Steiner tree problem in polynomial time.*

Proof: We first construct the instance of the minimum-cost node-weighted Steiner tree problem to which we are reducing from $\mathrm{HSP}(G,S)$. For each hyperedge $U \in E$ we define an associated node v_U and we denote these nodes by V_E as defined below.

$$V_E = \{v_U : U \in E\} \tag{5.4}$$

We then define $V' = V \cup V_E$. Next, for each $v_U \in V_E$ we create an edge between v_U and each $v \in U$. Specifically, we define this collection of edges E' as

$$E' = \{\{v, v_U\} : v \in U\} \tag{5.5}$$

Finally, we define the nonnegative node-weights $w' : V' \to \mathbb{R}_+$ as

$$w'(v) = \begin{cases} 0 & \text{if } v \in V \\ w(U) & \text{if } v = v_U \text{ for some } v_U \in V_E \end{cases} \tag{5.6}$$

Thus, our node-weighted graph is defined to be $G' = (V', E', w')$ and the instance of the minimum-cost node-weighted Steiner tree problem that we are reducing to is NSTP(G', S). It is easy to see that this reduction can be computed in polynomial time with respect to G. It remains to prove the correctness of this reduction.

We first prove that HSP$(G, S) \geq$ NSTP(G', s). Let t be subset of hyperedges of G that correspond to a connected subhypergraph of G that contains the Steiner nodes S. We define feasible solution to NSTP(G', S) as

$$t' = \{\{v, v_U\} : v \in U, U \in t\} \tag{5.7}$$

Then the objective value of this solution to NSTP(G', S) is

$$\sum_{v \in \cup_{e \in t'} e} w'(v) = \sum_{U \in t} w'(v_U) = \sum_{U \in t} w(U) \tag{5.8}$$

where the first equality follows by the fact that $w(v) = 0$ for all $v \in V$ and the second equality follows by the fact that $w'(v_U) = w(U)$ for all $v_U \in V_E$. Thus, the objective value of t' to NSTP(G', S) is at least as small as that of t to HSP(G, S). It remains to show that t' is a feasible solution to NSTP(G', S), that is t' is connected and t' contains the nodes S.

We first prove the latter condition on t'. Let $v \in S$ be arbitrary. Since t is subhypergraph of G that contains the nodes S there is some $U \in t$ such that $v \in U$. By construction of t', there is an edge $(v, v_U) \in t'$ and so t' contains v. Since v is arbitrary, it follows that t' contains the Steiner nodes S.

As for the former condition on t', we let $v, v' \in \cup_{e \in t'} e$ be two arbitrary but distinct nodes of the graph defined by t'. Without loss of generality, we can assume that $v, v' \in V$ since we can always form a path from any $v_U \in V_E$ to some node in V, specifically, by considering an edge from v_U to one of the nodes $u \in U$. Since t is a connected subhypergraph of G, there is a sequence of hyperedges $U_1, U_2, \ldots, U_k$ such that $v \in U_1$, $v' \in U_k$, and for each $i \in \{1, 2, \ldots, k-1\}$, $U_i \cap U_{i+1} \neq \emptyset$. By the last property of $U_1, U_2, \ldots, U_k$, we can create a sequence of nodes $v_1, v_2, \ldots, v_k$ such that for each $i \in \{1, 2, \ldots, k-1\}, v_i \in U_i \cap U_{i+1}$. Then, by the construction of t', it follows that the sequence

$$v, U_1, v_1, U_2, v_2, \ldots, v_{k-1}, U_k, v' \tag{5.9}$$

of nodes of G' forms a path in the subgraph t' of G'. Hence HSP$(G, S) \geq$ NSTP(G', s).

Lastly, we prove the remaining inequality, HSP$(G, S) \leq$ NSTP(G', s). Let t' be a connected subhypergraph of G' containing the Steiner nodes S, or in other words t' is a feasible solution to NSTP(G', S). We define a feasible solution to HSP(G, S) as

$$t = \{U : v_U \in V_E \cap t'\} \tag{5.10}$$

Then the objective value of this solution to HSP(G,S) is

$$\sum_{U \in t} w(U) = \sum_{v_U \in V_E \cap t'} w'(v_U) = \sum_{v \in t'} w'(v) \tag{5.11}$$

where the first equality follows by the fact that $w(U) = w'(v_U)$ for all $v_U \in V_E \cap t'$ and the second equality follows by the fact that $w'(v) = 0$ for all $v \in V$. Thus, the objective value of t to HSP(G,S) is at least as small as that of t' to NSTP(G',S). It remains to show that t is a feasible solution to HSP(G,S), that is, t is connected and t contains the Steiner nodes S.

We first prove the latter condition on t. Let $v \in S$ be arbitrary. Due to the fact that t' contains the Steiner nodes S, and in particular v, and that t' is connected there must be some edge $(v,v') \in t'$. Furthermore, G' is a bipartite graph by construction, with the partitions $V \subseteq V'$ and $V_E \subseteq V'$, and since $v \in S \subseteq V$ then $v' \in V_E$. Therefore, $v' = v_U$ for some $v_U \in V_E$ where U is a hyperedge of G that contains the node v. Finally, $U \in t$ by construction of t and so the subhypergraph t contains the node v. Since v is arbitrary, it follows that t contains the Steiner nodes S.

As for the former condition on t, we let $v, v' \in \cup_{U \in t} U$ be two arbitrary but distinct nodes of the subhypergraph defined by t. Since t' is a connected bipartite subgraph of G' there is a sequence of nodes $v_1, v_{U_1}, v_2, v_{U_2}, \ldots, v_{U_k-1}, v_k$ where $v = v_1$, $v' = v_k$ and for each $i \in \{1,2,\ldots,k-1\}$ there are edges $(v_i, v_{U_i}), (v_{U_i}, v_{U_{i+1}}) \in E'$. Given that such edges were created when constructing t', for each $i \in \{2,3,\ldots,k-1\}$ it follows that $v_i \in U_i \cap U_{i+1}$. Or in other words, the sequence of hyperedge $U_1, U_2, \ldots, U_{k-1}$ forms a path from $v \in U_1$ to $v \in U_{k-1}$, where the membership of $v \in U_1$ and $v \in U_{k-1}$ follows by construction of the edges of G'. Hence it is shown that t is connected which completes the proof that t is a feasible solution of HSP(G,S). Hence, HSP$(G,S) \leq$ NSTP(G',S) as claimed.

□

Having proven Lemma 5.2.3 it is easy to see that Theorem 5.1.2 follows as a corollary.

Corollary 5.2.4 (Restatement of Theorem 5.1.2) *There exists a polynomial time $O(\log n)$-approximation algorithm for the minimum cost hypergraph Steiner problem. Specifically, for a hypergraph $G = (V,E,w)$ with nonnegative hyperedge weights $w : E \to \mathbb{R}_+$ and for a set of Steiner nodes $S \subseteq V$, HSP(G,S) can be approximated up to a factor of $O(\log |S|)$ in polynomial time.*

Proof: The proof follows immediately as a corollary of Theorem 5.2.2 and Lemma 5.2.3. □

5.2.2 Proof of Theorem 5.1.3

As with the previous subsection, we proceed by defining the set cover problem.

Definition 5.2.5 *Let V be a universal set of n nodes and let $E = \{U_1, U_2, \ldots, U_k\}$ be a collection of subsets of V where $\cup_{i=1}^k U_i = V$. Then the set cover problem, SCP(V,E), is defined as the minimum number of sets $U_i \in E$ whose their union contains V. More formally, SCP(V,E) is defined as*

$$SCP(V,E) = \min_{\mathscr{A} \subseteq E : V \subseteq \cup_{U \in \mathscr{A}} U} |\mathscr{A}| \tag{5.12}$$

The set cover problem is a well-known NP-complete problem [20] cannot be approximated with a factor smaller than $O(\log n)$ in polynomial time [1, 15, 29]. Specifically, Feige gave a lower bound of $(1-o(1))\ln n$ for polynomial-time approximation unless NP has slightly superpolynomial time algorithms [17].

Theorem 5.2.6 *The set cover problem cannot be approximated to a factor smaller than $(1-o(1)\ln n$ in polynomial time (unless NP has slightly superpolynomial time algorithms).*

As noted earlier, our inapproximability result for the minimum-cost hypergraph Steiner problem follows by a reduction from the set cover problem.

Lemma 5.2.7 *Let V be a universal set of n nodes and let $E = \{U_1, U_2, \ldots, U_k\}$ be a collection of subsets of V where $\cup_{i=1}^{k} U_i = V$. Then the set cover problem $SCP(V,E)$ can be reduced to an instance of the minimum-cost hypergraph Steiner problem in polynomial time.*

Proof: Given an instance of the set cover problem, w create a new node u and a node set

$$V' = V \cup \{u\} \tag{5.13}$$

For each subset $U_i \in E$ of V, we create a corresponding hyperedge $U_i' = U_i \cup \{u\}$ and we define the set of hyperedges as

$$E' = \{U_i' : U_i \in E\} \tag{5.14}$$

As for the nonnegative hyperedge weights, we define $w : V' \to \mathbb{R}_+$ as

$$w(U_i') = 1, \quad \forall U_i' \in E' \tag{5.15}$$

Thus, out hyperedge-weighted hypergraph is defined to be $G = (V', E', w)$ and the instance of the minimum-cost hypergraph Steiner problem that we are reducing to is HSP(G,V). It is easy to see that this reduction can be computed in polynomial time with respect to the set cover instance SCP(V,E). It remains to prove the correctness of the reduction.

As usual we let $\mathscr{T}_V$ be the collection of connected subhypergraphs of G containing the nodes V. We create a map $f : \{\mathscr{A} \subseteq E : V \subseteq \cup_{U_i \in \mathscr{A}} U_i\} \to \mathscr{T}_V$ whereby

$$f(\mathscr{A}) = \{U_i' : U_i \in \mathscr{A}\} \tag{5.16}$$

and

$$f^{-1}(t) = \{U_i : U_i' \in t\} \tag{5.17}$$

We first prove the correctness of the map, that is, we show that $f(\mathscr{A}) \in \mathscr{T}_V$ and $f^{-1}(t) \in \{\mathscr{A} \subseteq E : V \subseteq \cup_{U_i \in \mathscr{A}} U_i\}$. Let $\mathscr{A} \subseteq E$ such that $V \subseteq \cup_{U_i \in \mathscr{A}} U_i$ be arbitrary. By construction it follows that each $U_i \subseteq U_i'$

and so $V \subseteq \cup_{U_i \inf(\mathscr{A})} U_i'$. As for the connectivity of $f(\mathscr{A})$, since each U_i' contains the node u, then $f(\mathscr{A})$ is connected so $f(\mathscr{A}) \in \mathscr{T}_V$.

As for the other direction, let $t \in \mathscr{T}_V$ be arbitrary. By definition of $\mathscr{T}_V$ we have that $V \subseteq \cup_{U_i' \in t} U_i'$, and since $u \notin V$ and each $U_i = U_i \setminus \{u\}$, then $V \subseteq \cup_{U_i \in f^{-1}(t)} U_i$. Hence $f^{-1}(t) \in \{\mathscr{A} \subseteq E : V \subseteq \cup_{U_i \in \mathscr{A}} U_i\}$.

Hence, f and f^{-1} map between solutions of SCP(V,E) and HSP(G,V). It remains to argue that objective value is preserved under f and f^{-1}. By the fact that $w(U_i') = 1$ for each $U_i' \in E'$, it follows that for each $t \in \mathscr{T}_V$ the objective value of the feasible solution t is equal to $\sum_{U_i' \in t} w(U_i') = |t|$. Then for any aribtrary $\mathscr{A} \in \{\mathscr{A} \subseteq E : V \subseteq \cup_{U_i \in \mathscr{A}} U_i\}$ we have that

$$|\mathscr{A}| = |f(\mathscr{A})| = \sum_{U_i' \in f(\mathscr{A})} w(U_i') \tag{5.18}$$

where the first equality follows by the definition of f and the second follows by the above observation.

Likewise, for an arbitrary $t \in \mathscr{T}_V$ we have that

$$\sum_{U_i' \in t} w(U_i') = |t| = |f^{-1}(t)| \tag{5.19}$$

where the first equality follows by the above observation and the second follows by the definition of f^{-1}.

Therefore, any feasible solution of our instance of SCP(V,E) can be mapped to a feasible solution of our instance of HSP(G',V) with an equivalent object value, and vice-versa. This completes the proof.

□

Having proven Lemma 5.2.7 it is easy to see that Theorem 5.1.3 follows as a corollary.

Corollary 5.2.8 (Restatement of Theorem 5.1.3) *The minimum cost hypergraph Steiner problem cannot be approximated to a factor smaller than $(1-o(1))\ln n$ in polynomial time (unless NP has slightly superpolynomial time algorithms). Specifically, for a hypergraph $G = (V,E,w)$ with nonnegative hyperedge weights $w : E \to \mathbb{R}_+$ and for a set of Steiner nodes $S \subseteq V$, HSP(G,S) cannot be approximated up to a factor smaller than $(1-o(1)\ln|S|$ in polynomial time (unless NP has slightly superpolynomial time algorithms)*

Proof: The proof follows immediately as a corollary of Theorem 5.2.6 and Lemma 5.2.7. □

5.3 Applications

In this section we provide two applications of our approximation algorithm for the minimum cost hypergraph Steiner problem, Theorem 5.1.2. The first application is that of approximately computing a hypergraph Steiner diversity given a hyperedge-weighted hypergraph.

Corollary 5.3.1 *Let $H = (V,E,w)$ be a hypergraph with node set V, hyperedge set E, and nonnegative hyperedge weights $w : E \to \mathbb{R}_+$. Let $(V, \delta_{\mathscr{H}})$ be the corresponding hypergraph Steiner diversity defined by H. Then for any $A \in \mathscr{P}(V)$, $\delta(A)$ can be computed up to a factor of $O(\log|A|)$ in polynomial time.*

Proof: This corollary follows immediately from Theorem 5.1.2 and by the fact that for any $A \in \mathscr{P}(V)$ we have that

$$\delta(A) = \min_{t \in \mathscr{T}_{(H,A)}} \sum_{U \in t} w(U) = HSP(H,A) \tag{5.20}$$

and so computing $\delta(A)$ is equivalent to solving $HSP(H,A)$. $\square$

The second application of Theorem 5.1.2 is that LP Relaxation (4.5) can be approximated up to an approximation factor of $O(\log r_H)$ in polynomial time.

Corollary 5.3.2 *Let $G = (V, E_G, w_G)$ be a supply hypergraph and let $H = (V, E_H, w_H)$ be a demand hypergraph with rank r_H. Then the sparsest cut LP Relaxation (4.5) can be be approximated up to a factor of $O(\log r_H)$ in polynomial time.*

Proof: LP Relaxation (4.5) has polynomially many variables and, in general, exponentially many constraints. Specifically, for each $S \in E_H$ the following set of constraints may be exponentially large

$$\{\sum_{U \in t} d_U \geq y_S\}_{t \in \mathscr{T}_{(G,S)}} \tag{5.21}$$

Separating over these constraints amounts to the decision problem

$$HSP(G,S) = \min_{t \in \mathscr{T}_{(G,S)}} \sum_{U \in t} d_U \geq y_S \tag{5.22}$$

We can Theorem 5.1.2 to approximate HSP(G,S) up to an $O(\log r_H)$ factor, as $|S| \leq r_H$. Hence, up to a factor of $O(\log r_H)$, we can verify whether the set of constraints (5.21) are approximately satisfied, and if not, we can find an approximate separating hyperplane. Then by the ellipsoid algorithm we can solve the LP Relaxation (4.5) up to an $O(\log r_H)$ approximation factor in polynomial time. $\square$

Chapter 6

Cut Sparsifiers and Sparse Diversities

6.1 Sparsification of the Sparsest Cut Problem

During this project a two-part question unfolded. Can we characterize the optimal diversity of the sparsest cut relaxation, Equation (4.2), with regards to properties such as the rank and sparsity of the supply and demand hypergraphs? Can we embed said classes of diversities into ℓ_1 with low distortion? One natural class of diversities to investigate are hypergraph Steiner diversities induced by a sparse hypergraph. We recall that the optimal diversity from the sparsest cut diversity relaxation is a hypergraph Steiner diversity induced by a reweighting of the supply hypergraph, see (4.6). Consequently, a low distortion embedding for a diversity induced by a sparse hypergraph would imply that the sparsest cut problem in hypergraphs is "easier" to solve via diversity embeddings when the supply and demand hypergraphs are sparse.

In fact, the existence of hypergraph cut sparsifiers implies that any instance of the sparsest cut can be approximated by a sparse instance up to an arbitrarily small $(1 \pm \varepsilon)$-approximation factor. Consequently, we can restrict our investigation of diversity embeddings to those that are induced by sparse hypergraphs. In this section we formalize these ideas, but first, we define a hypergraph cut sparsifier; note the definition is analogous for graph cut sparsifiers, [7, 14].

Definition 6.1.1 (Hypergraph Cut Sparsifier) *Let $G = (V,E,w)$ be a hypergraph with node set V, hyperedge set E, and hyperedge weights $w : E \to \mathbb{R}_+$. Let $\varepsilon > 0$ and $G' = (V,E',w')$ where $E' \subseteq E$ is a hypergraph with hyperedge weights $w' : E' \to \mathbb{R}_+$, then G' is a $(1 \pm \varepsilon)$-approximate cut sparsifier of G if for any cut $A \subseteq V$, where $A \neq \emptyset, V$, its weight in G' is within a multiplicative factor of $(1 \pm \varepsilon)$ of its weight in G:*

$$(1-\varepsilon) \sum_{U \in E : U \cap A \neq \emptyset, U} w(U) \leq \sum_{U \in E' : U \cap A \neq \emptyset, U} w'(U) \leq (1+\varepsilon) \sum_{U \in E : U \cap A \neq \emptyset, U} w(U) \tag{6.1}$$

(Hyper)graph cut sparsifiers are of practical importance due to the computational and memory benefits when working with sparse (hyper)graphs. Consequently, there is a rich history of work on cut sparsifiers in graphs. In particular, the seminal work of Benczur and Karger [7] showed that for an n-node and m-edge graph and any $\varepsilon \in (0,1)$, there is a near linear-time algorithm that computes a $(1 \pm \varepsilon)$-approximate cut sparsifier with a near linear number of edges, specifically $O(\frac{n \log n}{\varepsilon^2})$ edges. In recent years, there has been a series of work on hypergraph cut sparsifiers, and of particular relevance to our investigation, we utilize the near linear-size hypergraph cut sparsifier of Chen, Khanna, and Nagda [14].

Theorem 6.1.2 *Let $H = (V, E, w)$ by a hypergraph with $|V| = n$, $|E| = m$, and hyperedge weights $w : E \to \mathbb{R}_+$. Then for any $\varepsilon \in (0,1)$, there is a randomized polynomial-time algorithm that constructs a $(1 \pm \varepsilon)$-approximate cut sparsifier of H with $O(\frac{n \log n}{\varepsilon^2})$ hyperedges in $O(mn + \frac{n^{10}}{\varepsilon^7})$-time with high probability.*

Using the notion of a hypergraph cut sparsifier, we prove the following lemma.

Lemma 6.1.3 *Let $G = (V, E_G, w_G)$ be a supply hypergraph and let $H = (V, E_H, w_H)$ be a demand hypergraph. Let $G' = (V, E_{G'}, w_{G'})$ be a $(1 \pm \varepsilon_1)$-approximate cut sparsifier of G and let $H' = (V, E_{H'}, w_{H'})$ be a $(1 \pm \varepsilon_2)$-approximate cut sparsifier of H, where $\varepsilon_1, \varepsilon_2 \in (0,1)$. Then $\phi_{G,H}$ is within a factor of $\frac{(1+\varepsilon_1)(1+\varepsilon_2)}{(1-\varepsilon_1)(1-\varepsilon_2)}$ of $\phi_{G',H'}$.*

Proof: Let $A \subseteq V$ where $A \neq \emptyset, V$ be arbitrary. It suffices to show that

$$\frac{(1-\varepsilon_1)}{(1+\varepsilon_2)} \phi_{G,H}(A) \leq \phi_{G',H'}(A) \leq \frac{(1+\varepsilon_1)}{(1-\varepsilon_2)} \phi_{G,H}(A) \tag{6.2}$$

Since G' is a $(1 \pm \varepsilon_1)$-approximate cut sparsifier of G we have the following inequalities

$$(1-\varepsilon_1) \sum_{U \in E_G : U \cap A \neq \emptyset, V} w_G(U) \leq \sum_{U \in E_{G'} : U \cap A \neq \emptyset, V} w_{G'}(U) \leq (1+\varepsilon_1) \sum_{U \in E_G : U \cap A \neq \emptyset, V} w_G(U) \tag{6.3}$$

Likewise, since H' is a $(1 \pm \varepsilon_2)$-approximate cut sparsifier of H we have the following inequalities

$$(1-\varepsilon_2) \sum_{S \in E_H : S \cap A \neq \emptyset, V} w_H(S) \leq \sum_{S \in E_{H'} : S \cap A \neq \emptyset, V} w_{H'}(S) \leq (1+\varepsilon_2) \sum_{S \in E_H : S \cap A \neq \emptyset, V} w_H(S) \tag{6.4}$$

Since all weights w_H and w'_H are nonnegative, then taking the inverse of (6.4) we have the following inequalities

$$\frac{1}{(1+\varepsilon_2) \sum_{S \in E_H : S \cap A \neq \emptyset, V} w_H(S)} \leq \frac{1}{\sum_{S \in E_{H'} : S \cap A \neq \emptyset, V} w_{H'}(S)} \leq \frac{1}{(1-\varepsilon_2) \sum_{S \in E_H : S \cap A \neq \emptyset, V} w_H(S)} \tag{6.5}$$

Then putting together (6.3) and (6.5) we have that

$$\frac{(1-\varepsilon_1)}{(1+\varepsilon_2)}\phi_{G,H}(A) = \frac{(1-\varepsilon_1)\sum_{U\in E_G:U\cap A\neq\emptyset,U} w_G(U)}{(1+\varepsilon_2)\sum_{S\in E_H:S\cap A\neq\emptyset,S} w_H(S)} \tag{6.6}$$

$$\leq \frac{\sum_{U\in E_{G'}:U\cap A\neq\emptyset,U} w_{G'}(U)}{\sum_{S\in E_{H'}:S\cap A\neq\emptyset,S} w_{H'}(S)} \quad \text{by (6.3) and (6.5)} \tag{6.7}$$

$$\leq \frac{(1+\varepsilon_1)\sum_{U\in E_G:U\cap A\neq\emptyset,U} w_G(U)}{(1-\varepsilon_2)\sum_{S\in E_H:S\cap A\neq\emptyset,S} w_H(S)} \quad \text{by (6.3) and (6.5)} \tag{6.8}$$

$$= \frac{(1+\varepsilon_1)}{(1-\varepsilon_2)}\phi_{G,H}(A) \tag{6.9}$$

Since

$$\phi_{G,H}(A) = \frac{\sum_{U\in E_G:U\cap A\neq\emptyset,U} w_G(U)}{\sum_{S\in E_H:S\cap A\neq\emptyset,S} w_H(S)} \tag{6.10}$$

this completes the proof. □

With Theorem 6.1.2 and Lemma 6.1.3, we obtain the following corollary.

Corollary 6.1.4 *Let $G=(V,E_G,w_G)$ be a supply hypergraph and $H=(V,E_H,w_H)$ be a demand hypergraph where $|V|=n$, $E_G=m_G$, and $E_H=m_H$. Let $\varepsilon\in(0,1)$, then computing the sparsest cut of G and H can be reduced up to a multiplicative approximation factor of $\frac{(1+\varepsilon)^2}{(1-\varepsilon)^2}$, in randomized polynomial-time with respect to n, m_G, m_H, and $\frac{1}{\varepsilon}$, to an instance of the sparsest cut in which the supply and demand hypergraphs each have $O(\frac{n\log n}{\varepsilon^2})$ hyperedges.*

Proof: The result follows by Theorem 6.1.2 and Lemma 6.1.3. □

The implication of Corollary 6.1.4 is that, fixing a constant approximation factor, one can reduce the sparsest cut problem to an instance where the number of supply and demand hyperedges is nearly linear, of order $O(n\log n)$ to be precise. The consequences of this are two-fold. Firstly, the LP relaxation (4.6) for the sparsified instance has fewer variables and constraints and could potentially be solved more quickly and with a smaller approximation factor. But more importantly, the diversities from solutions to this LP relaxation are induced by sparse hypergraphs. This fact motivates the investigation of low distortion embeddings for diversities induced by sparse hypergraphs. On this note, we conclude this thesis with the following conjecture.

Conjecture 6.1.5 (Restatement of Conjecture 1.2.6) *Let $H=(V,E,w)$ be a hypergraph with node set V, hyperedge set E, and nonnegative hyperedges weights $w: E\to\mathbb{R}_+$, where $|V|=n$ and $|E|=m$. Let $(V,\delta_{\mathscr{H}})$ be the corresponding hypergraph Steiner diversity defined by H. Then for some $p>0$ there is an $O(\log^p(n+m))$-distortion embedding of $(V,\delta_{\mathscr{H}})$ into the ℓ_1 diversity.*

Bibliography

[1] N. Alon, D. Moshkovitz, and S. Safra. Algorithmic construction of sets for k-restrictions. *ACM Transactions on Algorithms (TALG)*, 2(2):153–177, 2006. → page 51

[2] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM (JACM)*, 45(3):501–555, 1998. → page 34

[3] S. Arora, J. Lee, and A. Naor. Euclidean distortion and the sparsest cut. *Journal of the American Mathematical Society*, 21(1):1–21, 2008. → pages 2, 6

[4] S. Arora, S. Rao, and U. Vazirani. Expander flows, geometric embeddings and graph partitioning. *Journal of the ACM (JACM)*, 56(2):1–37, 2009. → page 6

[5] Y. Aumann and Y. Rabani. An O(log k) approximate min-cut max-flow theorem and approximation algorithm. *SIAM Journal on Computing*, 27(1):291–301, 1998. → pages iii, 2, 5, 27, 38

[6] G. Baudis, C. Gröpl, S. Hougardy, T. Nierhoff, and H. J. Prömel. Approximating minimum spanning sets in hypergraphs and polymatroids. In *Proc. ICALP*. Citeseer, 2000. → pages 40, 47

[7] A. A. Benczúr and D. R. Karger. Randomized approximation schemes for cuts and flows in capacitated graphs. *SIAM Journal on Computing*, 44(2):290–319, 2015. → page 54

[8] J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel Journal of Mathematics*, 52(1):46–52, 1985. → pages 21, 26

[9] D. Bryant and P. Tupper. Open problem statement: Minimal distortion embeddings of diversities in ℓ_1. *arXiv preprint arXiv:1712.01960*, 2017. → pages 3, 7, 9, 10, 19, 21, 22, 24, 33

[10] D. Bryant and P. F. Tupper. Hyperconvexity and tight-span theory for diversities. *Advances in Mathematics*, 231(6):3172?3198, Dec 2012. ISSN 0001-8708. doi:10.1016/j.aim.2012.08.008. URL http://dx.doi.org/10.1016/j.aim.2012.08.008. → pages 2, 7, 8, 9, 22

[11] D. Bryant and P. F. Tupper. Constant distortion embeddings of symmetric diversities. *Analysis and Geometry in Metric Spaces*, 4(1), 2016. → page 7

[12] J. Byrka, F. Grandoni, T. Rothvoß, and L. Sanita. An improved LP-based approximation for Steiner tree. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 583–592, 2010. → page 44

[13] S. Chawla, A. Gupta, and H. Racke. Embeddings of negative-type metrics and an improved approximation to generalized sparsest cut. 2005. → page 6

[14] Y. Chen, S. Khanna, and A. Nagda. Near-linear size hypergraph cut sparsifiers. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 61–72. IEEE, 2020. → pages 54, 55

[15] I. Dinur and D. Steurer. Analytical approach to parallel repetition. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 624–633, 2014. → page 51

[16] J. Fakcharoenphol, S. Rao, and K. Talwar. A tight bound on approximating arbitrary metrics by tree metrics. *Journal of Computer and System Sciences*, 69(3):485–497, 2004. → page 22

[17] U. Feige. A threshold of ln n for approximating set cover. *Journal of the ACM (JACM)*, 45(4): 634–652, 1998. → page 51

[18] J. Hastad. Clique is hard to approximate within $n^{1-\varepsilon}$. In *Proceedings of 37th Conference on Foundations of Computer Science*, pages 627–636. IEEE, 1996. → page 34

[19] M. Kapralov, R. Krauthgamer, J. Tardos, and Y. Yoshida. Towards tight bounds for spectral sparsification of hypergraphs. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 598–611, 2021. → pages 2, 5, 6, 44

[20] R. M. Karp. Reducibility among combinatorial problems. In *Complexity of computer computations*, pages 85–103. Springer, 1972. → page 51

[21] P. Klein and R. Ravi. A nearly best-possible approximation algorithm for node-weighted Steiner trees. *Journal of Algorithms*, 19(1):104–115, 1995. → page 48

[22] L. Kou, G. Markowsky, and L. Berman. A fast algorithm for Steiner trees. *Acta informatica*, 15(2): 141–145, 1981. → page 44

[23] T. Leighton and S. Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *Journal of the ACM (JACM)*, 46(6):787–832, 1999. → page 5

[24] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995. → pages iii, 2, 5, 7, 21, 27, 38

[25] A. Louis. Hypergraph markov operators, eigenvalues and approximation algorithms. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 713–722, 2015. → pages 2, 6, 44

[26] A. Louis and Y. Makarychev. Approximation algorithms for hypergraph small set expansion and small set vertex expansion. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2014)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2014. → pages 2, 6

[27] D. W. Matula and F. Shahrokhi. Sparsest cuts and bottlenecks in graphs. *Discrete Applied Mathematics*, 27(1-2):113–123, 1990. → pages 2, 4

[28] P. N. K. S. A. Plotkin, S. Rao, and E. Tardos. Bounds on the max-flow min-cut ratio for directed multicommodity flows. 1993. → pages iii, 3, 45

[29] R. Raz and S. Safra. A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, pages 475–484, 1997. → page 51

[30] H. Takahashi. An approximate solution for the Steiner problem in graphs. *Math. Japonica.*, 6: 573–577, 1990. → page 44

[31] P. Tupper and D. Bryant. Diversities and the geometry of hypergraphs. *Discrete Mathematics & Theoretical Computer Science*, 16, 2014. → pages iii, 3, 7, 9, 12, 19, 26, 28, 38

[32] Y.-F. Wu, P. Widmayer, and C.-K. Wong. A faster approximation algorithm for the Steiner problem in graphs. *Acta informatica*, 23(2):223–229, 1986. → page 44

Appendix A

Supporting Materials

A.1 Application of the Triangle Inequality

Proposition A.1.1 *Let (X,δ) be a (pseudo) diversity and let (X,d) be its (pseudo) metric space. Let $A = \{v_1, v_2, \dots, v_k\} \in \mathscr{P}(X)$. Then it follows that*

$$\delta(A) \leq \sum_{i=1}^{k-1} d(v_i, v_{i+1}) \tag{A.1}$$

Proof: For each $i \in \{1, 2, \dots, k-1\}$ we define

$$A_i = \{v_1, v_2, \dots, v_{i+1}\} \tag{A.2}$$

Then for any arbitrary $i \in \{2, 3, \dots, k-1\}$ we have that

$$\delta(A_i) \leq \delta(A_{i-1}) + \delta(\{v_i, v_{i+1}\}) = \delta(A_{i-1}) + d(v_i, v_{i+1}) \tag{A.3}$$

where the inequality follows by the triangle inequality and the fact that $A_{i-1} \cap \{v_i, v_{i+1}\} = \{v_i\} \neq \emptyset$. Then we have that

$$\begin{aligned}
\delta(A) &= \delta(A_{k-1}) && \text{by } A_{k-1} = A && \text{(A.4)}\\
&\leq \delta(A_{k-2}) + d(v_{k-1}, v_k) && \text{by A.3} && \text{(A.5)}\\
&\leq \delta(A_{k-3}) + d(v_{k-2}, v_{k-1}) + d(v_{k-1}, v_k) && \text{by A.3} && \text{(A.6)}\\
& && && \text{(A.7)}\\
&\vdots && && \text{(A.8)}\\
&\leq \sum_{i=1}^{k-1} d(v_i, v_{i+1}) && \text{by repeated application of A.3} && \text{(A.9)}\\
& && && \text{(A.10)}
\end{aligned}$$

This completes the proof. □

Proposition A.1.2 *Let (X,δ) be a (pseudo) diversity and let (X,d) be its (pseudo) metric space. Let $A \in \mathscr{P}(X)$ and $a \in X$ be arbitrary. Then it follows that*

$$\delta(A) \leq \sum_{v \in A} d(v,a) \tag{A.11}$$

Proof: First, we enumerate the elements of A as $A = \{v_1, v_2, \ldots, v_k\}$. For each $i \in \{1,2,\ldots,k\}$ we define

$$A_i = \{a, v_1, v_2, \ldots, v_i\} \tag{A.12}$$

Then for any arbitrary $i \in \{2,3,\ldots,k\}$ we have that

$$\delta(A_i) \leq \delta(A_{i-1}) + \delta(\{v_i, a\}) = \delta(A_{i-1}) + d(v_i, a) \tag{A.13}$$

where the inequality follows by the triangle inequality and the fact that $A_{i-1} \cap \{v_i, a\} = \{a\} \neq \emptyset$. Then we have that

$$\begin{aligned}
\delta(A) &\leq \delta(A \cup \{a\}) && \text{by diversities being increasing, Proposition 2.1.4} && \text{(A.14)}\\
&= \delta(A_k) && \text{by } A \cup \{a\} = A_k && \text{(A.15)}\\
&\leq \delta(A_{k-1}) + d(v_k, a) && \text{by A.13} && \text{(A.16)}\\
&\leq \delta(A_{k-2}) + d(v_{k-1}, a) + d(v_k, a) && \text{by A.13} && \text{(A.17)}\\
&\vdots && && \text{(A.18)}\\
&\leq \sum_{v \in A} d(v,a) && \text{by repeated application of A.13} && \text{(A.19)}\\
& && && \text{(A.20)}
\end{aligned}$$

This completes the proof. □