NEW ROUNDING TECHNIQUES FOR THE DESIGN AND ANALYSIS OF APPROXIMATION ALGORITHMS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF MANAGEMENT
SCIENCE AND ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Shayan Oveis Gharan
May 2014

Abstract

We study two of the most central classical optimization problems, namely the Traveling Salesman problems and Graph Partitioning problems and develop new approximation algorithms for them. We introduce several new techniques for rounding a fractional solution of a continuous relaxation of these problems into near optimal integral solutions. The two most notable of those are the maximum entropy rounding by sampling method and a novel use of higher eigenvectors of graphs.

To my beloved wife, Farnaz, my mum, my dad, Sheida, Shahram, Shadi, Shahab and all members of my family.

Acknowledgements

Foremost, I would like to express my deepest appreciation to my advisor, Amin Saberi, who introduced me to the field of computing. Amin thought me how to conduct academic research, how to write a mathematical paper and how to present my work effectively to others; the three skills that shaped my career. I would like to express my gratitude to Luca Trevisan who introduced me to the field of Spectral graph theory. I started working with Luca after taking a course on this subject and we have managed to advance this area in several directions since then.

I would like to thank all my teachers at Stanford University for offering advanced courses that have helped me throughout my career. In particular, I would like to thank Persi Diaconis, Ashish Goel, Serge Plotkin, Tim Roughgarden, Ryan Williams and Yinyu Ye. In addition, a thank you to all my coauthors: Arash Asadpour, Michel Goemans, Tsz Chiu Kwok, Bundit Laekhanukit, Lap Chi Lau, James Lee, Yin Tat Lee, Russell Lyons, Aleksander Madry, Vahideh Manshadi, Vahab Mirrokni, Mohit Singh, Jan Vondrak and Morteza Zadimoghaddam; working with each one of them has been a tremendous learning experience.

I am deeply grateful to the Management Science and Engineering department, Stanford School of Engineering, Caroline Pease and Anne Schnoebele for supporting my graduate research during the past five years. Also, I gratefully acknowledge the National Science Foundation for their financial support.

I have greatly benefited from the summer internships at IBM Almaden Research Center, Microsoft Research, New England and Microsoft Research, Redmond. I want to thank my supervisors in those institutes for giving me the opportunity to work and network with researchers from all around the world.

Last but not least, my special thanks goes to my wife, Farnaz, and everyone else in my family: my mum, my dad, Sheida, Shahram, Shadi and Shahab. They have supported me during this PhD by means that words cannot explain.

Contents

List of Figures

List of Algorithms

Chapter 1

Introduction

Optimization problems arise in a wide range of fields including logistics, planning, marketing, advertising, and policy-making. The amount of data collected and stored of our daily activities is growing rapidly and requires better exploitation of our computing power. Better utilization of this data can result in billions of dollars of revenue in advertising, health care, finance, and many other disciplines. However, as of now, computing an optimal solution using simple exhaustive search algorithms would require trillions of years to compute an optimal solution even if they were to use all the computing power ever built by mankind. Hence, we are faced with the challenge to develop mathematical tools that help us design efficient algorithms to determine an optimal or near optimal solution.

The main focus of this thesis is to study the approximability of *classical* NP-hard optimization problems. This area of research arises from the fact that many important problems are known to be NP-hard, i.e., under standard conjectures [For09] they cannot be solved optimally in polynomial time. Instead, one can hope to find an approximate solution, one that is not optimal, but is guaranteed to be within a small factor from the optimal solution.

We say a polynomial time algorithm is an α approximation for a minimization problem if the output of the algorithm is within a factor α of the optimum in the worst case. The most common approach in designing approximation algorithms involves four main steps:

i) Formulate the problem as an integer program.

ii) Relax the integrality constraint and obtain a convex (linear) relaxation of the problem.

iii) Compute an optimal fractional solution to the continuous relaxation.

iv) Round the fractional solution to an integer solution.

There are several ways to write a continuous relaxation for discrete optimization problems. Of these, the most well known are the linear programing relaxations and semi-definite programming relaxations. For most of these methods, the first three steps outlined above can be taken systematically.

but the last step is the most challenging. The main goal of this thesis is to address this difficult last step: specifically, to develop new techniques for rounding fractional solutions of an optimization problem.

Over the last thirty years, several methods have been proposed for rounding a fractional solution of a linear or a semidefinite programming relaxation of discrete optimization problems. One of the first was Raghavan and Thompsons randomized rounding approach [RT87] (see Section 3.1 for a brief overview). More recent ones include the iterative rounding method [Jai01, LRS11] and the hyperplane rounding method [GW95, ARV09].

In the first part of this thesis we study several variants of the well-known Traveling Salesman Problem. We propose a new rounding method, called maximum entropy rounding by sampling method, to round a fractional solution of the Linear Programming relaxation of TSP into a near optimal integral solution. This method has also been used in several other contexts in the last couple of years.

In the second part of this thesis we focus on spectral algorithms. The existence of efficient algorithms to compute the eigenvectors and eigenvalues of graphs (see Section 7.6) supplies a useful tool for the design of efficient graph algorithms. Eigenvectors of a graph are optimizers of a continuous relaxation of graph partitioning problems (see Subsection 7.7.1). In fact, one can write a semi-definite programming relaxation of multiway partitioning problems such that the optimizers are the eigenvectors of the graph. Cheeger's inequality [AM85, Alo86] graph coloring algorithms [AK97] or maximum cut algorithms [Tre09] provide a threshold rounding algorithm to round the second or the last eigenvector of a graph into an integral cut (see Section 7.8 for more details). But there are no generalizations of these algorithms to higher eigenvectors of graphs.

In the second half of this thesis we design new rounding algorithms that find an integral k-way partitioning of a graph using the first k eigenvectors. Our rounding algorithms provide a rigorous justification for several practical spectral algorithms that use these eigenvectors. Furthermore, using our knowledge of higher eigenvalues we manage to improve Cheeger's inequality, we design faster spectral graph algorithms and provide new graph partitioning algorithms with better quality solutions.

In Section 1.1 we provide an overview of Part I and in Section 1.2 we provide an overview of Part II.

1.1 New Approximation Algorithms to the Traveling Salesman Problem

The Traveling Salesman Problem (TSP) is perhaps the most well known problem in the areas of approximation algorithms and combinatorial optimization. Today, TSP has applications in planning, scheduling, manufacturing of microchips, and genome sequencing (see [ABCC07] for details), but

it has actually been of keen interest since the very first studies in the fields of Combinatorics and Graph Theory. In the 18^{th} century Euler introduced The Knight's Tour problem on a chessboard, and in the 19^{th} century Hamilton and Kirkman studied Hamiltonian paths in various classes of graphs [Big81].

In the field of computing, many significant developments have been sparked by TSP. Just a few examples: solving an instance of TSP led to one of the first applications of linear and integer programming techniques [DFJ54, DFJ59]; one of the first approximation algorithms ever developed was the 3/2 approximation algorithm of Christofides for TSP [Chr76]; one of the first average case analyses was the work of Karp on TSP [Kar77, Kar79]. TSP was also one of the first problems proved to be NP-complete [Kar72].

In an instance of TSP we are given a set V of n cities with their pairwise symmetric distances. The goal is to find the shortest tour that visits each city at least once. It is NP-hard to approximate TSP with a factor better than 185/184 [Lam12]. Christofides designed a 3/2 approximation algorithm for TSP in 1976 [Chr76], and subsequently no one has succeeded in beating the 3/2 factor despite the fact that many researchers have tried [Wol80, SW90, BP91, Goe95, CV00, GLS05, BEM10, BC11, SWvZ12]. It remains one of the central open problems in the field of computing. One of the major achievements of this thesis is the first significant advance in solving this problem in over 35 years

Before describing our ideas, we must first describe previous work on this problem. Observe that any Hamiltonian cycle is in the intersection of the *spanning tree polytope*[1] and the *perfect matching matching polytope* of the input graph. It turns out that both of these polytopes are integral and it is easy to optimize any function over them (see Subsection 2.4.2 for background). So the main question is how to optimize a cost function on both of these polytopes simultaneously.

Researchers have employed two general approaches in attacking TSP or an important variant, Asymmetric TSP [Chr76, FGM82, Blä02, GLS05, KLSS05, FS07].

i) Start with a minimum cost connected subgraph, i.e., a minimum cost spanning tree, and then add edges and make it Eulerian.

ii) Start with a Eulerian subgraph, i.e., a minimum cost cycle cover, then add edges while preserving the Eulerian-ness until it becomes connected.

We say a subgraph is Eulerian if the degree of each vertex is even and we say a subgraph is connected if it includes a spanning tree. There are two general approaches that researches employed in attacking TSP [Chr76, FGM82]:

Again, these are the main two approaches people have applied to TSP. For example, the Christofides 3/2 approximation algorithm uses approach (i).

[1] If we want to be precise, a Hamiltonian cycle is in the 1-tree polytope, where a 1-tree is a union of a spanning tree and a single other edge.

1.1.1 Our Contributions

All of our new approximation algorithms for different variants of TSP are examples of approach (i), and from this point forward our discussion will be restricted to approach (i). Since the matching polytope is integral, there is an efficient algorithm that for any given connected subgraph finds the minimum cost *Eulerian augmentation* of that subgraph, i.e., the minimum cost set of edges to make it Eulerian. But the cost of the Eulerian augmentation strongly depends on the structure of the connected subgraph that we choose in the first step. The main new ingredient of our works is a new method to round a fractional solution of the Linear Programming relaxation of TSP into a random spanning tree. We show that a tree chosen from such a distribution preserves many properties of the fractional solution with high probability. Consequently, we can argue that the cost of the Eulerian augmentation corresponding to that tree is significantly smaller than the cost of a Eulerian augmentation of an arbitrary minimum spanning tree.

Next, we describe a brief outline of our method that rounds any feasible solution of the LP relaxation of TSP into a random spanning tree. We call this method, the rounding by sampling method. First, we interpret an optimal solution of the LP relaxation as a convex combination of the extreme point solutions of an integral polytope - in the case of TSP, a convex combination of the integral spanning trees of the input graph. This convex combination defines a distribution over extreme points. We then sample an extreme point from the underlying distribution and augment it into an integral solution of the problem.

Our main idea is to choose a convex combination of extreme points that has the *maximum entropy*. Intuitively, we avoid imposing any additional structure by keeping the combinatorial structure intact while maximizing the uncertainty. We use the following convex program to compute the maximum entropy distribution:

$$\begin{aligned} \max \quad & \sum_{T} -p_T \log(p_T) \\ & \sum_{T \ni e} p_T = z_e \qquad \forall e \in E, \\ & p_T \geq 0 \qquad\qquad \forall T. \end{aligned}$$

In the above program, $\mathbf{z}$ represents a fractional point inside the spanning tree polytope, and p_T represents the probability of selecting an extreme point of the spanning tree polytope, which in this case is a spanning tree T. We prove that the optimizer of this program can be interpreted as a weighted uniform distribution on spanning trees (see Section 3.1 for details), and that it can be approximated efficiently (see Subsection 3.1.2).

In summary, our machinery implies that a fractional solution of the LP relaxation of TSP can be efficiently written as a random spanning tree distribution. This allows us to use many properties of the random spanning tree distributions that have been studied for decades by mathematicians (see

Algorithm	pa561	si535	si1032
Christofides	1.21	1.24	1.07
Rounding by Sampling	1.09	1.05	1.003
Lin-Kernighan Heuristic	1.009	1.0007	1.0008

Table 1.1.1

Section 2.8 and Section 2.9 for more details). For example, we can show that, with high probability, at least 1/3 of the vertices of a tree sampled from this distribution are even.

We have applied the above machinery to various classes of problems including the Traveling Salesman Problem [AGM+10, OSS11], the Online Stochastic Matching Problem [MOS11], and the Minimum Strongly Connected Subgraph Problem [LOS12]. As a simple inspiring example, in Subsection 1.1.2 we describe a simple algorithm as an application of the rounding by sampling method to the online stochastic matching problem. Next, we explain the main results that we will provide in Part I.

Symmetric TSP: In a joint work with Saberi and Singh [OSS11], we designed a $3/2 - \epsilon$ approximation algorithm for TSP on graphic metrics, breaking the 3/2 barrier of Christofides [Chr76] where $\epsilon > 0$ is a universal constant. Graphic metrics are the cost functions corresponding to the shortest path metric of an unweighted graph (see Section 2.1 for more details).

Our algorithm is very simple to describe: it chooses a random spanning tree based on the solution of the LP relaxation using the rounding by sampling method. Then it adds the minimum cost matching on the odd degree vertices of the tree (see Algorithm 9).

Our analysis, on the other hand, is sophisticated. It builds on properties of uniform spanning tree distributions Section 2.8 and polygon representation of near minimum cuts Section 2.7. Very recently, Borcea, Branden and Liggett [BBL09] used tools from complex geometry and proved that strongly Rayleigh measures that are a generalization of uniform spanning tree measures satisfy the strongest forms of negative dependence and are closed under certain operations Section 2.9. These properties are one of the fundamental parts of our analysis. As a byproduct of our results, we show new properties of near minimum cuts of any graph in Section 3.2, and new properties of random spanning tree distributions in Section 3.3.

Although we only prove that our algorithm beats Christofides' 3/2 approximation algorithm on graphic metrics, we conjecture that its approximation factor is strictly better than 3/2 in the worst case. We also compared our algorithm with Christofides' algorithm and one of the best heuristics for TSP, namely the Lin-Kernighan Heuristic on several of the test cases of the public TSP library, TSPLIB[2]. The result of the comparison is shown in Table 1.1.1. Each column label is the name of a test case of TSPLIB. Each entry of the table represents the ratio of the cost of the solution computed by an algorithm with respect to the cost of the optimum solution of the LP relaxation.

[2]The test cases can be downloaded from http://elib.zib.de/pub/mp-testdata/tsp/tsplib/tsp/index.html.

Observe that our algorithm performs significantly better than Christofides' algorithm but it cannot beat the Lin-Kernighan Heuristic. One explanation is that the heuristics usually implement the best of many ideas.

Asymmetric Traveling Salesman Problem (ATSP): ATSP is a generalization of TSP in which the distances between the vertices need not be symmetric. In a joint work with Asadpour, Goemans, Madry and Saberi [AGM+10], we designed an $O(\log n/\log\log n)$ approximation algorithm for ATSP, breaking the $O(\log n)$ barrier developed in 1982 [FGM82]. The algorithm is very similar to our algorithm for the TSP that we described above. The main difference is in the computation of the minimum cost Eulerian augmentation. In this case, the minimum cost Eulerian augmentation of a given spanning tree can be computed efficiently by solving a minimum cost flow problem (see Algorithm 6 for details). Also, in a joint work with Saberi [OS11], we managed to design the first constant factor approximation algorithm ATSP on planar or bounded-genus graphs.

Part I of this thesis is organized as follows. In Chapter 2 we provide background on convex optimization, matroids, linear programming relaxations, structure of minimum cuts and near minimum cuts, properties of random spanning trees and strongly Rayleigh measures. Chapter 3 is specifically organized to provide new machineries developed in this part of the thesis that we expect to see in several applications in the future. We describe the rounding by sampling method in Section 3.1, new properties of near minimum cuts in Section 3.2, and new properties of random spanning trees in Section 3.3. We provide our $O(\log n/\log\log n)$ approximation algorithm for ATSP in Chapter 4 and our constant factor approximation for planar ATSP in Chapter 5. Finally, in Chapter 6 we provide our $3/2-\epsilon$ approximation algorithm for graphic TSP.

1.1.2 Rounding by Sampling and Online Stochastic Matching Problem

The goal of this section is to provide a simple and inspiring application of the rounding by sampling method in a very different context. We design a very simple algorithm for the Online Stochastic Matching Problem that beats the previous algorithm Feldman et al. [FMMM09]. The result of this section is based on a joint work with Vahideh Manshadi and Amin Saberi [MOS11].

The online stochastic matching problem proposed by Feldman et al. [FMMM09] as a model of display ad allocation. We are given a bipartite graph $G(A,B,E)$; where B represent one side of the graph corresponds to a fixed set of n bins and A represent the other side which is a set of n possible ball types. At each time step $1,2,\ldots,n$ a ball of type $a\in A$ is chosen independently at random (with replacement). The algorithm can assign this ball to at most one of the empty bins that are adjacent to it; each bin can be matched to at most one ball. The goal of the algorithm is to maximize the expected number of non-empty bins at time n.

We compare the expected size of the matching computed by an algorithm to the optimum offline solution, the expected size of the maximum matching of the revealed graph at time n. Given the

sequence of arrived balls $\omega = (a_1, a_2, \ldots, a_n)$, one can compute the optimum allocation in polynomial time by solving a maximum matching problem. Let $f_\omega : E \to \{0, 1\}$ be the indicator function of edges that are used in the optimum allocation given ω.

Let $\mathsf{OPT}(\omega) = \langle \mathbf{1}, f_\omega \rangle$ be the size of the maximum matching and $\mathsf{ALG}(\omega)$ be the size of a matching computed by an online algorithm ALG. The competitive ratio of ALG is defined as $\frac{\mathbb{E}[\mathsf{ALG}(\omega)]}{\mathbb{E}[\mathsf{OPT}(\omega)]}$. turns out that the algorithms that we study in this section $\mathsf{ALG}(\omega)$ and $\mathsf{OPT}(\omega)$ are concentrated around their expected values, therefore the above competitive ratio is fairly robust. Feldman, Mehta, Mirrokni and Muthukrishnan [FMMM09] designed an algorithm with competitive ratio of 0.67 for the online stochastic matching problem. In this section we use the rounding by sampling method to design a very simple algorithm with competitive ratio of 0.68.

Our algorithm crucially uses the optimum offline solution for making decisions. For any edge e let $f(e) := \mathbb{E}[f_\omega(e)]$. Also, let us abuse the notation and use $f(b) := \sum_{a:(a,b)\in E} f(a, b)$.

It turns out that f is in the matching polytope of G, i.e., the vector $\mathbf{z}$ where $z_e = f(e)$ is a feasible solution of the following linear program.

$$\begin{aligned} \sum_{b:(a,b)\in E} z_{a,b} &\leq 1 \qquad && \forall a \in A, \\ \sum_{a:(a,b)\in E} z_{a,b} &\leq 1 \qquad && \forall b \in B, \\ z_e &\geq 0 && \forall e \in E. \end{aligned} \tag{1.1.1}$$

Fact 1.1.1. *The optimum offline solution f is in the matching polytope of G.*

Proof. Given ω and $a \in A$, let $N_\omega(a)$ be the number of balls of type a in ω. Clearly $\sum_{b:(a,b)\in E} f_\omega(a, b) \leq N_\omega(a)$. Taking expectations from both sides show that $\sum_{b:(a,b)\in E} f(a, b) \leq 1$. where we used the fact that each ball type is sampled once in expectation. On the other hand, since in any instance of the problem any bin y can be matched to at most one ball, $\sum_{a:(a,b)\in E} f(a, b) \leq 1$. So, $f(.)$ is in the matching polytope of G. □

Since the matching polytope is integral, $f(.)$ can be written as a convex combination of bipartite matchings of G (see Section 2.2 for background on polytopes and extreme point solutions). Therefore, using standard algorithmic versions of Caratheodory's theorem (see e.g. [GLS93, Theorem 6.5.11]) we can decompose a $f(.)$ into a convex combination of polynomially many bipartite matchings in polynomial time. More specifically, we obtain the following:

Lemma 1.1.2. *It is possible to efficiently and explicitly construct (and sample from) a distribution μ on the set of matchings in G such that*

$$\mathbb{P}_{M\sim\mu}[e \in M] = f(e), \ \forall e \in E.$$

Algorithm.

Our algorithm has some similarities with the online algorithm that Feldman et al. [FMMM09] proposed. Both algorithms start by computing two matchings M_1 and M_2 offline; When the first ball of type a arrives it will be allocated to the bin matched to a in M_1, and when the second ball arrives, we will allocate it via M_2. If the corresponding bins are already full, we drop the ball. Note that the probability that there are more than two balls of each type a in the sequence of arrivals is very small.

The main difference is in the construction of M_1, M_2. Roughly speaking, one would like that M_1, M_2 are disjoint and $|M_1|, |M_2|$ is as large as possible. So, if G has two disjoint maximum matchings one can try to find them and use them in the algorithm. But the main difficulty is when G does not have two disjoint maximum matching. Feldman et al. find M_1 and M_2 by carefully decomposing the solution of a maximum 2-flow of G into two disjoint matchings. They have to go into an extensive case analysis to adjust the size of these matchings. Here, we use the rounding by sampling method to choose M_1, M_2. Recall that we have written f as a distribution of matchings μ. First note that $\mathbb{E}_{M\sim\mu}[|M|] = \sum_e f(e) = \mathbb{E}[\mathsf{OPT}]$. So any sample from μ is as large as OPT in expectation. To make sure that M_1, M_2 share the least number of edges we just sample two matchings *independently* from μ. The details of the final algorithm is described below.

Algorithm 1 The Online Stochastic Matching Algorithm

Offline Phase:

Compute the fractional matching f, and the distribution μ using [GLS93, Theorem 6.5.11].
Sample two matchings M_1 and M_2 from μ *independently*; set M_1 (M_2) to be the first (second) priority matching.

Online Phase:

When the first ball of type a arrives, allocate it through the first priority matching, M_1.
When a ball of type a arrives for the second time, allocate it through the second priority matching, M_2.

We prove the following theorem.

Theorem 1.1.3. *The competitive ratio of Algorithm 1 is at least 0.684.*

Analysis.

Throughout the analysis we drop the terms of $O(1/n)$. Let X_b be the random variable indicating the event that bin b is matched with a ball during the run of the algorithm. We analyze the competitive ratio of the algorithm by lower bounding $\mathbb{E}[X_b]/f(b)$ for all $b \in B$,

$$\frac{\mathbb{E}[\mathsf{ALG}]}{\mathbb{E}[\mathsf{OPT}]} = \frac{\sum_{b\in B}\mathbb{E}[X_b]}{\sum_{b\in B} f(b)} \geq \min_{b\in B}\frac{\mathbb{E}[X_b]}{f(b)}.$$

For $b \in B, a \in A$ we abuse notation and use $M_1(b) := \{a\}$ if $(a, b) \in M_1$, and if b is not saturated

in M_1, we let $M_1(b) := \emptyset$; similarly we define $M_2(b)$. Given M_1 and M_2, $\mathbb{E}[X_b|M_1, M_2]$ can be computed similar to [FMMM09, Section 4.2.2],

$$\mathbb{E}[X_b \mid M_1, M_2] \approx \begin{cases} 0 & \text{if } M_1(b) = M_2(b) = \emptyset \\ 1 - 1/e & \text{if } M_1(b) \neq \emptyset, M_1(b) = M_2(b) \\ 1 - 1/e & \text{if } M_1(b) \neq \emptyset, M_2(b) = \emptyset \\ 1 - 2/e & \text{if } M_1(b) = \emptyset, M_2(b) \neq \emptyset \\ 1 - 2/e^2 & \text{if } M_1(b) \neq \emptyset, M_2(b) \neq \emptyset, M_1(b) \neq M_2(b). \end{cases} \tag{1.1.2}$$

Let us describe the fourth case, say $M_1(b) = \emptyset, M_2(b) = a$ for some $a \in A$: in this case b is matched if and only if at least two balls of type a arrive. Let $N_\omega(a)$ be the number of balls of type a in the arriving sequence ω. Then,

$$\begin{aligned} \mathbb{E}[X_b \mid M_1(b) = \emptyset, M_2(b) = a] = \mathbb{P}[N_\omega(a) \geq 2] &= 1 - \mathbb{P}[N_\omega(a) = 0] - \mathbb{P}[N_\omega(a) = 1] \\ &= 1 - \left(1 - \frac{1}{n}\right)^n - \frac{1}{n}\binom{n}{1}\left(1 - \frac{1}{n}\right)^{n-1} = 1 - \frac{2}{e}. \end{aligned}$$

where we dropped a term of $O(1/n)$ in the RHS.

For a bin $b \in B$ let $\delta(b)$ be the set of edges adjacent to b. Since we choose M_1, M_2 independently at random,

$$\begin{aligned} \mathbb{E}[X_b] &= (1 - 1/e) \sum_{(a,b)\in\delta(b)} \mathbb{P}[M_1(b) = a] \cdot \mathbb{P}[M_2(b) = a \text{ or } M_2(b) = \emptyset] \\ &\quad + (1 - 2/e) \sum_{(a,b)\in\delta(b)} \mathbb{P}[M_1(b) = \emptyset] \cdot \mathbb{P}[M_2(b) = a] \\ &\quad + (1 - 2/e^2) \sum_{(a,b),(a',b)\in\delta(b), a\neq a'} \mathbb{P}[M_1(b) = a] \cdot \mathbb{P}[M_2(b) = a'] \\ &= (1 - 1/e) \sum_{e\in\delta(b)} f(e)(1 - f(b) + f(e)) + (1 - 2/e) \sum_{e\in\delta(b)} f(e)(1 - f(b)) \\ &\quad + (1 - 2/e^2) \sum_{\substack{e,e'\in\delta(b) \\ e\neq e'}} f(e)f(e') \\ &= f(b)(2 - 3/e) - f(b)^2(1 + 2/e^2 - 3/e) - (1/e - 2/e^2) \sum_{e\in\delta(b)} f(e)^2 \end{aligned}$$

where in the second inequality we used Lemma 1.1.2, in the last equality we used $\sum_{e\in\delta(b)} f(e) = f(b)$.

It remains to prove a lower bound the RHS of the above equation. We show,

$$\frac{\mathbb{E}[X_b]}{f(b)} = (2-3/e) - (1+2/e^2-3/e)f_b - (1/e-2/e^2)\frac{\sum_{e\in\delta(b)} f_e^2}{f_b} \geq 0.684 \tag{1.1.3}$$

Let us first fix f_b and find the minimum of the LHS in terms of f_b. For any f_b, the LHS is minimized when $\sum_{e\in\delta(b)} f_e^2$ is maximized. Consider any k edges $(a_1,b),\ldots,(a_k,b)$.

$$\sum_{i=1}^{k} f(a_i,b) \leq \mathbb{P}\left[\sum_{i=1}^{k} N_\omega(a_i) \geq 1\right] = 1-\mathbb{P}\left[\sum_{i=1}^{k} N_\omega(a_i) = 0\right] = 1-(1-k/n)^n = 1-1/e^k.$$

For example, it follows that $\max_{e\in\delta(b)} f(e) \leq \min\{f(b), 1-1/e\}$. With the above constraint, it is easy to see that the LHS of (1.1.3) is minimized when $f_b = 1$, otherwise we may add a dummy ball type $a \in A$, and connect it to b by an edge $e = (a,b)$ with very small probability, $f(e) = \epsilon$, and see that this only decreases the value of LHS. If $f(b) = 1$, then by the above equation $\sum_{e\in\delta(b)} f(e)^2 \leq 0.463$. Plugging this into (1.1.3) completes the proof of Theorem 1.1.3.

1.2 New Analysis of Spectral Graph Algorithms through Higher Eigenvalues

Spectral graph algorithms are simple heuristics that explore the structure of a graph using eigenvalues and eigenvectors of the adjacency matrix of the graph or any of its normalizations like the Laplacian or the normalized Laplacian matrix. These algorithms are widely used in practice because they typically run in near linear time, provide high quality solutions, and with the aid of a linear algebra library are very simple to implement.

In practical applications spectral graph algorithms, like a strong hammer, have been used to attack problems in a variety of areas including Image Segmentation [SM00, YGS02, YS03, BJ03, TM06], data clustering [NJW02, BH03], community detection [DM05, WS05, SC10] and VLSI design [CSZ94, AKY95]. Although spectral graph algorithms are very simple to implement, they are non-trivial and interesting; and although these algorithms are widely used in practice, we still do not have a rigorous justification for their performance.

1.2.1 Spectral Clustering Algorithm

Let us describe an application of spectral graph algorithms in data clustering. Clustering is one of the fundamental primitives in machine learning and data analysis with a variety of applications in information retrieval, pattern recognition, recommendation systems, etc. Suppose we have a set of data points that we want to cluster (see Figure 1.2.1). Suppose the distance between the data points represents their similarity, i.e., two points are more similar if they get closer to each other. A

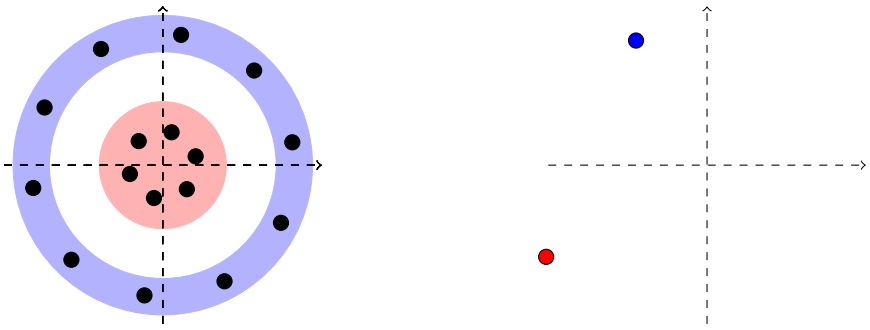

Figure 1.2.1: A natural clustering of the left points divide them into a group of points on the outer circle and a group on the inner circle. But, it is quite unlikely that a Heuristic like kmeans finds this clustering. Instead, if we map the vertices based on spectral embedding the points on the outer circle will map to the blue point on the right and the points on the inner circle map to the red point. Now we can find the natural clustering using kmeans.

natural clustering of the points in the example of Figure 1.2.1 divides the points into two parts as we have shown in this figure. However, it is quite unlikely that a heuristic like kmeans recovers this natural clustering. Recall that kmeans is a clustering heuristic where we map a set of points into k sets such that each point is mapped to the set with the closest mean.

Data clustering may be modeled as a graph partitioning problem, where one models each of the data points as a vertex of a graph $G = (V, E)$ and the weight of an edge connecting two vertices represents the similarity of the corresponding data points. There are many ways to define the graph G; for example, G can be a complete graph where for any two vertices $u, v \in V$,

$$w(u, v) \propto \exp(-\|\mathbf{x}_u - \mathbf{x}_v\|^2 / \sigma^2) \tag{1.2.1}$$

where we used $\mathbf{x}_u, \mathbf{x}_v$ to denote the coordinates of the points corresponding to u, v and $\sigma > 0$ is a free parameter. To take another example, one can let G be an unweighted graph where there is an edge between two vertices u, v if and only if $\|\mathbf{x}_u - \mathbf{x}_v\| \leq \epsilon$ for some threshold $\epsilon > 0$.

Once we construct a weighted graph, we can use the *spectral graph clustering algorithm* to partition the vertices. First we compute multiple eigenfunctions of a normalization of the adjacency matrix of the graph, called the normalized Laplacian matrix $\mathcal{L} = I - D^{-1/2}AD^{-1/2}$, where I is the identity matrix, D is the diagonal matrix of vertex degrees, and A is the adjacency matrix. Say $f_1, f_2, \ldots, f_k$ are the eigenfunctions of $\mathcal{L}$ corresponding to the first k eigenvalues. The spectral embedding of G is the function $F : V \to \mathbb{R}^k$ where for any $v \in V$ $F(v) := (f_1(v), f_2(v), \ldots, f_k(v))$ (see Section 8.1 for the properties of spectral embedding). We embed the vertices to a new space using $F(.)$ and then we run the kmeans on the spectral embedding and we return its output (the details of the algorithm is described in Algorithm 2). For the example of Figure 1.2.1, the spectral embedding maps the points of the outer circle close to each other and likewise those of the inner

circle. Therefore, using the eigenfunctions we managed to change the basis and make the task of clustering significantly easier for kmeans.

Algorithm 2 Spectral Graph Clustering Algorithm

Input: A graph $G = (V, E)$, and a weight function $w : E \to \mathbb{R}_+$, $k \geq 2$.
Output: A k-partitioning of V.
Let $\mathcal{L} = I - D^{-1/2} A D^{-1/2}$ be the normalized Laplacian where for any $u, v \in V$, $A(u, v) = w(u, v)$.
Let $f_1, f_2, \ldots, f_k : V \to \mathbb{R}$ be an orthonormal set of functions corresponding to the first k eigenvalues of $\mathcal{L}$.
Let $F : V \to \mathbb{R}^k$ where for any $v \in V$, $F(v) := (f_1(v), \ldots, f_k(v))$. be the spectral embedding.
Run k-means on the vectors $F(v)/\|F(v)\|$ for all $v \in V$ and return its output.

The spectral clustering algorithm that we just described is one of the fundamental tools in data clustering. An enormous number of articles apply this idea to cluster images, movies, music, webpages, etc. We can intuit that the reason people previously used eigenfunctions in this algorithm is that eigenfunctions can be seen as the optimizers of a continuous relaxation of the graph k-partitioning problem (see Subsection 7.7.1 for more details). We refer interested readers to a recent survey by Luxburg [Lux07] for more information.

1.2.2 Spectral Graph Algorithms in Theory

Spectral graph algorithms are one of the fundamental tools in theoretical computer science. From a very high level point of view, they relate combinatorial properties of graphs to the algebraic properties of matrices. For example, they relate properties of cuts to the eigenvalues of adjacency matrix or the Laplacian matrix of the graph (see Subsection 1.2.4 at the end of this section for an inspiring example). This relation is one of the key insights in various areas of theory including approximation algorithms [ST96, Kel04, ABS10], probability theory and the analysis of random walks [SJ89, JSV04], construction of error-correcting codes [Spi96], and complexity theory [RVW00, Rei05].

To the best of our knowledge, all of the classical analyses of spectral graph algorithms only exploit the first or last two eigenvalues of graphs and relate them to the properties of cuts [AM85, Alo86, AK97, Tre09][3]. Roughly speaking, a classical spectral algorithm works as follows: first we map the vertices of G to a line using an eigenfunction of $\mathcal{L}$, then we try several possibilities for cutting the line into two pieces and we choose the best cut that we find. Such algorithms are limited because they only exploit one dimensional mapping of a graph.

Let us be more specific and describe Cheeger's inequality, which is one of the most influential results in the field of spectral graph theory. A basic fact in algebraic graph theory is that the number of connected components in any undirected graph is equal to the multiplicity of the eigenvalue 0 in

[3]We remark that in some random/semi-random models there are results that use matrix perturbation theory and multiple eigenvectors, but here I do not have any assumption on the structure of the graphs there are examples of spectral algorithms for random or semi-random graphs that use use matrix perturbation theory and multiple eigenvectors, see e.g., [McS01], but here we do not have any prior assumption on the structure of the graphs.

the normalized Laplacian matrix of the graph. In particular, a graph is disconnected if and only if the second eigenvalue of the normalized Laplacian matrix, λ_2, is zero. Cheeger's inequality provides an "approximate" version of the latter fact.

Let us first provide a robust version of connectivity. There are several combinatorial measures for the quality of a multiway partitioning of a graph including diameter, k-center, k-median, conductance, etc. Kannan, Vempala and Vetta [KVV04] show that several of these measures fail to capture the natural clustering in simple examples. They also argue that conductance is one of the best objective functions for measuring the quality of a cluster. For a set $S \subseteq V$, the *conductance* of S, $\phi(S)$ is the following ratio:

$$\phi(S) := \frac{|E(S, \overline{S})|}{\text{vol}(S)}$$

where $\text{vol}(S)$ is the summation of the degree of vertices of S. The conductance of G,

$$\phi(G) = \min_{\text{vol}(S) \leq \text{vol}(V)/2} \phi(S)$$

is the minimum conductance among all sets that have at most half of the total volume. Observe that for any graph G, $0 \leq \phi(G) \leq 1$. Furthermore, if $\phi(G) \approx 0$, there is a cut $(S, \overline{S})$ such that $|E(S, \overline{S})| \ll \text{vol}(S), \text{vol}(\overline{S})$, so we can say G is almost disconnected (see Section 7.7 for more background on the conductance).

Cheeger's inequality for graphs [AM85, Alo86] states that a graph is almost disconnected if and only if the second smallest eigenvalue of $\mathcal{L}$, λ_2, is close to 0. Quantitatively, for any graph G,

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

Observe the close relation between an algebraic quantity of normalized Laplacian, λ_2, and a combinatorial property of G, namely the conductance of G. We will provide a detailed proof of this inequality in Section 7.8.

Cheeger's inequality has significant applications in graph partitioning [ST96, KVV04], explicit construction of expander graphs [JM85, HLW06, Lee12], approximate counting [SJ89, JSV04], and image segmentation [SM00]. The proof of Cheeger's inequality gives a simple, nearly linear time algorithm (the *spectral partitioning algorithm*) that finds cuts with nearly-minimal conductance. Given an eigenfunction f_2 of λ_2, the algorithm finds the best threshold cut. That is the cut separating the vertices where $f_2(v) \leq t$, for the best threshold t (see Algorithm 10 for details). The spectral partitioning algorithm is widely used in practice for its efficiency and the high quality of solutions that it provides [Sim91, HL92, BS93].

Let us summarize the above discussion. Spectral graph algorithms, like strong hammers, have been used in a variety of problems in practice because they are fast, simple to implement, and provide high quality solutions. On the other hand, spectral graph algorithms are one of the fundamental

tools in theory. Although more eigenvalues and eigenfunctions provide better quality solutions in practice [AKY95], theoretical analyses can only justify and analyze the algorithms that use the second or the last eigenfunction.

1.2.3 Our Contribution

Our main goal in the second part of this thesis is to close this gap. We will understand higher eigenvalues of graphs, provide tools to handle them and control them, and of course we will design new spectral algorithms with our knowledge. Here is a summary of our contribution. We will analyze several spectral graph algorithms using higher eigenvalues and eigenfunctions. As a consequence of this we can provide a rigorous justification for the spectral graph clustering algorithms and we can provide new ideas to improve this algorithm. We also provide faster clustering algorithms and several new ideas for designing graph clustering algorithms.

Our main machinery in controlling higher eigenvalues of graphs is the same spectral embedding function that has been used by practitioners for many years. In Section 8.1 we prove several important properties of this embedding like isotropy, spreading and the energy and we use these properties throughout Part II.

Now, let us provide a more detailed list of our contribution. We use $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ to denote the eigenvalues of $\mathcal{L}$. We use $\rho(k)$ to denote the k-way conductance constant,

$$\rho(k) = \min_{\text{disjoint } S_1,\ldots,S_k} \max_{1\leq i\leq k} \phi(S_i).$$

Higher Order Cheeger's Inequality. In a joint work with Lee and Trevisan [LOT12], we prove the first generalization of Cheeger's inequality to higher eigenvalues of a graph. We show that the vertices of any graph can be partitioned into k subsets each defining a sparse cut, if and only if λ_k is close to 0. Quantitatively, for any $k \geq 2$,

$$\lambda_k/2 \leq \rho(k) \leq O(k^2)\sqrt{\lambda_k}.$$

Compared to the classical analyses of spectral graph algorithm, our analysis uses high dimensional embedding of graphs, namely the spectral embedding. We also use several of recent developments in high dimensional geometry on random partitioning of metric spaces (see Section 7.9 for background).

Our result provides a rigorous justification for the spectral clustering algorithm that we described in Subsection 1.2.1. In particular, our theorem shows that the spectral graph clustering algorithm finds a "good" k-partitioning of a given graph if and only if the k-th smallest eigenvalue of the normalized Laplacian matrix is close to zero. Using our machinery we justify the folklore belief that the number of clusters, k, in the spectral clustering algorithm must be chosen based on the largest gap between eigenvalues (see Section 10.3). Our proof also justifies the application of kmeans

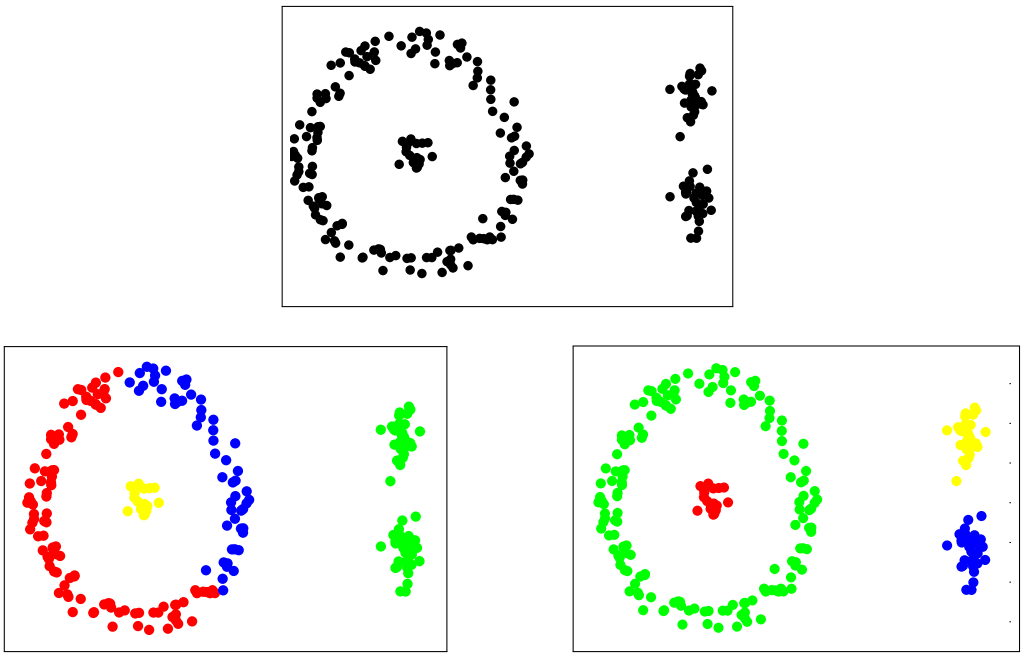

Figure 1.2.2: The set of data points in the top figure is from Jianbo Shi's website on data clustering, http://www.cis.upenn.edu/~jshi/software/demo1.html. The goal is to find a coloring of these points with 4 colors, red, blue, green and yellow such that the points in each cluster have the same color. We run the spectral graph clustering algorithm on a complete graph G where the weight of each edge is computed by (1.2.1) for $\sigma := 0.05 \max_{u,v} \|\mathbf{x}_u - \mathbf{x}_v\|$. The left figure shows the output of Algorithm 2, and the right figure shows the output of the modified spectral clustering, Algorithm 3, where we randomly project the points to a 2 dimensional space and then we use the kmeans Heuristic.

heuristic in the last step of spectral clustering for the graphs where there is a large gap between successive eigenvalues.

Our analysis provides several new theoretical tools. The first is a new way of upper bounding higher eigenvalues of graphs by defining a smooth localization of the spectral embedding function (see Subsection 10.2.1 for more details). The second is a new type of dimension reduction that bypasses the $O(\log n)$ barrier in the well-known dimension reduction of Johnson and Lindenstrauss [JL84]. Specifically, we study dimension reductions from a k dimensional space to an $O(\log k)$ dimensional space and we show that it preserves several properties of spectral embedding (see Section 8.2).

We also provide new ideas to improve the quality of the spectral graph clustering algorithm. Our analysis suggests randomly projecting the points of the spectral embedding to an $O(\log k)$ dimensional space using independently chosen Gaussian vectors and then applying the kmeans algorithm to the new space. In Figure 1.2.2 we show that in some data clustering examples this idea can help obtain better quality solutions, but we do not know if this idea always improves the quality of clustering in practice. We refer to Section 10.6 for details of our algorithm.

Algorithm 3 Spectral Clustering Algorithm with Dimension Reduction

Input: A graph $G = (V, E)$, and a weight function $w : E \to \mathbb{R}_+$, $k \geq 2$.
Output: A k-partitioning of V.

Let $\mathcal{L} = I - D^{-1/2}AD^{-1/2}$ be the normalized Laplacian where for any $u, v \in V$, $A(u, v) = w(u, v)$.
Let $f_1, f_2, \ldots, f_k : V \to \mathbb{R}$ be an orthonormal set of functions corresponding to the first k eigenvalues of $\mathcal{L}$.
Let $F : V \to \mathbb{R}^k$ where for any $v \in V$, $F(v) := (f_1(v), \ldots, f_k(v))$. be the spectral embedding.
Choose $l = \Theta(\log(k))$ random k dimensional Gaussian vectors $\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_l$ and let

$$\Gamma(v) = \frac{1}{\sqrt{l}}(\langle F(v), \boldsymbol{\zeta}_1, \rangle, \langle F(v), \boldsymbol{\zeta}_2 \rangle, \ldots, \langle F(v), \boldsymbol{\zeta}_l \rangle)$$

Run k-means on the vectors $\Gamma(v)/\left\|\Gamma(v)\right\|$ for all $v \in V$ and return its output.

Improved Cheeger Inequality through Higher Order Spectral Gap. In a joint work with Kwok, Lau, Lee, and Trevisan [KLL+13], we *strengthen* the right side of Cheeger's inequality and we show that for any $k \geq 2$,

$$\phi(G) \leq O(k)\frac{\lambda_2}{\sqrt{\lambda_k}}$$

Consequently, we can characterize graphs for which the right side of the Cheeger's inequality is tight; in all these graphs we must have approximate multiplicity of eigenvalues, i.e., $\lambda_k \approx \lambda_2$ for all constant k.

Our result describes why the spectral partitioning algorithm performs significantly better than the worst case guarantee of Cheeger's inequality in practical applications. If for a constant k, λ_k is bounded away from 0 for some graph G, then the spectral partitioning algorithm provides a constant factor approximation for the sparsest cut problem. In practical instances of image segmentation, there are usually only a few outstanding objects in the image, and thus λ_k is bounded away from 1 for a constant k [SM00].

Almost Optimal Local Graph Clustering. A local graph algorithm is one that finds a solution around a given vertex of the graph by looking only at the local neighborhood of a vertex. In a joint work with Trevisan [OT12], we design a local graph clustering algorithm with almost the same guarantee as the spectral partitioning algorithm. This is the first sublinear (in the size of the input) time algorithm with almost the same guarantee as the Cheeger's inequality.

Another advantage of our algorithm is that if there are both large and small sets with near-optimal conductance, our algorithm is more likely to find the smaller sets. Indeed, for any given target size k, our local algorithm can find sets of size approximately k with near-minimal sparsity around the starting vertex. Small communities generally contain more interesting and substantial information than large communities.

Our analysis provides new properties of simple random walks on graphs. We show that for any

set $S \subseteq V$ a simple t step random walk started at a uniformly chosen vertex of S remains in S with probability at least $(1-\phi(S))^t$ (see Section 8.3 for more details). This has been also used to provide improved lower bounds on the *mixing time* of reversible random walks. Our analysis builds on the recent works of Andersen, Morris and Peres [MP03, AP09].

Universal Bounds on Laplacian Eigenvalues. In a joint work with Lyons [LO12], we use the spectral embedding to provide a *unifying framework* for lower bounding all the eigenvalues of the normalized Laplacian matrix of graphs. For example, we show that for any graph G with n vertices $\lambda_k \leq 1 - \Omega(k^3/n^3)$, this upper bound improves to $1 - \Omega(k^2/n^2)$ if the graph is regular (note that there is no dependency to the degree). We generalize these results and we provide sharp bounds on the eigenvalues of various classes of graphs including vertex transitive graphs, and infinite graphs in terms of specific graph parameters like the volume growth.

Using these bounds we design a slightly sub-exponential time algorithm that beats the $O(\sqrt{\log n})$ approximation algorithm of [ARV09] for the sparsest cut. Our work introduces the spectral embedding as a new tool in analyzing the reversible Markov Chains. We have used our machinery to provide (improved) upper bounds on the return probabilities and mixing time of random walks with considerably shorter and more direct proofs. Furthermore, building on an earlier work of Lyons [Lyo05a], we design fast local algorithms to approximate the number of spanning trees of massive graphs.

Partitioning into Expanders There is a basic fact in algebraic graph theory that $\lambda_k > 0$ if and only if G has at most $k-1$ connected components. In a joint work with Trevisan [OT13] we prove a robust version of this fact. If $\lambda_k > 0$, then for some $1 \leq l \leq k-1$, V can be *partitioned* into l sets $P_1, \ldots, P_l$ such that each P_i is a low-conductance set in G and induces a high conductance induced subgraph. In particular, $\phi(P_i) \lesssim l^3\sqrt{\lambda_l}$ and $\phi(G[P_i]) \gtrsim \lambda_k/k^2$.

We design a simple polynomial time spectral algorithm to find such partitioning of G with a quadratic loss in the inside conductance of P_i's. Unlike the traditional spectral clustering algorithms, our algorithm does not use higher order eigenfunctions of G. Furthermore, if there is a sufficiently large gap between λ_k and λ_{k+1}, more precisely, if $\lambda_{k+1} \gtrsim \text{poly}(k)\lambda_k^{1/4}$ then our algorithm finds a k partitioning of V into sets $P_1, \ldots, P_k$ such that the induced subgraph $G[P_i]$ has a significantly larger conductance than the conductance of P_i in G. Such a partitioning may represent the best k clusterings of G. Our algorithm is a simple local search that only uses the Spectral Partitioning algorithm as a subroutine. We expect to see further applications of this simple algorithm in clustering applications.

Part II of this thesis is organized as follows. In Chapter 7 we provide background on spectral graph theory, Laplacian matrix, random walks, eigenfunction computation, conductance, Cheeger's inequality and random partitioning of metric spaces. Chapter 8 is specifically organized to provide

new machineries developed in this part of the thesis that we expect to see in several applications in the future. We describe our new machinery of spectral embedding and its properties in Section 8.1. We prove our new dimension reduction framework in Section 8.2, and in Section 8.3 we prove our improved upper bound on the escape probability of random walks. In Chapter 9 we prove universal lower bounds on eigenvalues of the normalized Laplacian matrix. We prove our higher order Cheeger's inequality in Chapter 10, and our improved Cheeger's inequality in Chapter 11. Finally, we design our local graph clustering algorithm in Chapter 12 and our new graph clustering algorithm in Chapter 13.

1.2.4 An Upper Bound on Graph Diameter based on Laplacian Eigenvalues

The goal of this section to provide a simple and inspiring proof of relating combinatorial properties of graphs to algebraic properties of matrices. We relate the dimeter of a graph to eigenvalues of the normalized Laplacian matrix. Our proof uses some of the new machineries on higher eigenvalues that we will develop in this thesis. The result of this section is based on a joint work with Luca Trevisan [GT12].

Let $G = (V, E)$ be a connected, undirected and unweighted graph, and let $d(v)$ be the degree of vertex v in G. Let D be the diagonal matrix of vertex degrees and A be the adjacency matrix of G. Let $L := D - A$ be the *Laplacian* of G, and let $\mathcal{L} := I - D^{-1/2}AD^{-1/2}$ be the *normalized Laplacian* matrix of G where I is the identity matrix (see Section 7.2 for background and properties of the Laplacian and the normalized Laplacian matrices). The matrix $\mathcal{L}$ is positive semi-definite. Let

$$0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n \leq 2$$

be the eigenvalues of $\mathcal{L}$. For any pair of vertices $u, v \in G$, we define their distance, $\mathrm{dist}(u, v)$, to be the length of the shortest path connecting u to v. The diameter of the graph G is the maximum distance between all pairs of vertices, i.e.,

$$\mathrm{diam}(G) := \max_{u,v} \mathrm{dist}(u, v).$$

Alon and Milman [AM85] show that if Δ is the maximum degree of vertices of G, and λ is the second smallest eigenvalue of the Laplacian of G, then

$$\mathrm{diam}(G) \leq 2\sqrt{2\Delta/\lambda}\log_2 n. \tag{1.2.2}$$

Chung [Chu89] improved the above result for regular graphs and show that,

$$\mathrm{diam}(G) \leq \left\lceil \frac{\log(n-1)}{\log \frac{\Delta}{\Delta-\lambda}} \right\rceil$$

To the best of our knowledge, none of the above results are generalized to higher eigenvalues of the (normalized) Laplacian matrix of G. The following question is asked by Gil Kalai in a personal communication [Kal12]. Is it true that for any connected graph G, and any $k \geq 2$, $\mathrm{diam}(G) = O(k \log(n)/\lambda_k)$. Equation (1.2.2) shows that this question already holds for $k = 2$. Therefore, Kalai's question can be seen as a generalization of the result of [AM85] to higher eigenvalues of $\mathcal{L}$.

In this section we answer his question affirmatively and we prove the following theorem

Theorem 1.2.1. *For any unweighted, connected graph G, and any $k \geq 2$,*

$$\mathrm{diam}(G) \leq \frac{48k \log n}{\lambda_k}.$$

Observe that the above theorem relates a combinatorial property of G to an algebraic property of the normalized Laplacian matrix. This is because the eigenvalues are the zeros of the characteristic polynomial of the determinant of $\mathcal{L} - \lambda I$ (see Section 7.1 for background).

Our proof uses the easy direction of our higher order Cheeger inequalities that we prove in Chapter 10. For a set $S \subseteq V$, let $E(S, \overline{S}) := \{\{u, v\} : |\{u, v\} \cap S| = 1\}$ be the set of edges with in the cut $(S, \overline{S})$. Let $\mathrm{vol}(S) := \sum_{v \in S} d(v)$ be the volume of the set S, and let

$$\phi(S) := \frac{|E(S, \overline{S})|}{\min\{\mathrm{vol}(S), \mathrm{vol}(\overline{S})\}}$$

be the *conductance* of S. Let $\rho(k)$ be the worst conductance of any k disjoint subsets of V, i.e.,

$$\rho(k) := \min_{\text{disjoint } S_1, S_2, \ldots, S_k} \max_{1 \leq i \leq k} \phi(S_i).$$

In Theorem 10.1.1 we will show that for any graph G and any $k \geq 2$,

$$\frac{\lambda_k}{2} \leq \rho(k) \leq O(k^2)\sqrt{\lambda_k}. \tag{1.2.3}$$

We will use the left side of the above inequality, a.k.a. easy direction of higher order Cheeger inequality, to prove Theorem 1.2.1.

Proof. We construct k disjoint sets $S_1, \ldots, S_k$ such that for each $1 \leq i \leq k$, $\phi(S_i) \leq O(k \log n / \mathrm{diam}(G))$, and then we use (1.2.3) to prove the theorem.

First, we find $k + 1$ vertices $v_0, ..., v_k$ such that the distance between each pair of the vertices is at least $\mathrm{diam}(G)/2k$. We can do that by taking the vertices v_0 and v_k to be at distance $\mathrm{diam}(G)$.

Then, we consider a shortest path connecting v_0 to v_k and take equally spaced vertices on that path.

For a set $S \subseteq V$, and radius $r \geq 0$ let

$$B(S, r) := \{v : \min_{u \in S} \operatorname{dist}(v, u) \leq r\}$$

be the set of vertices at distance at most r from the set S. If $S = \{v\}$ is a single vertex, we abuse notation and use $B(v, r)$ to denote the ball of radius r around v. For each $i = 0, \ldots, k$, consider the ball of radius $\operatorname{diam}(G)/6k$ centered at v_i, and note that all these balls are disjoint. Therefore, at most one of them can have a volume of at least $\operatorname{vol}(V)/2$. Remove that ball from consideration, if present. So, maybe after renaming, we have k vertices $v_1, ..., v_k$ such that the balls of radius $\operatorname{diam}(G)/6k$ around them, $B(v_1, \operatorname{diam}(G)/6k), \ldots, B(v_k, \operatorname{diam}(G)/6k)$, are all disjoint and all contain at most a mass of $\operatorname{vol}(V)/2$.

The next claim shows that for any vertex v_i there exists a radius $r_i < \operatorname{diam}(G)/6k$ such that $\phi(B(v_i, r_i)) \leq 24k \log n / \operatorname{diam}(G)$.

Claim 1.2.2. *For any vertex $v \in V$ and $r > 0$, if $\operatorname{vol}(B(v, r)) \leq \operatorname{vol}(V)/2$, then for some $0 \leq i < r$, $\phi(B(v, i)) = 4 \log n / r$.*

Proof. First observe that for any set $S \subseteq V$, with $\operatorname{vol}(S) \leq \operatorname{vol}(V)/2$,

$$\operatorname{vol}(B(S, 1)) = \operatorname{vol}(S) + \operatorname{vol}(N(S)) \geq \operatorname{vol}(S) + |E(S, \overline{S})| = \operatorname{vol}(S)(1 + \phi(S)) \tag{1.2.4}$$

where the inequality follows from the fact that each edge $\{u, v\} \in E(S, \overline{S})$ has exactly one endpoint in $N(S)$, and the last equality follows from the fact that $\operatorname{vol}(S) \leq \operatorname{vol}(V)/2$. Now, since $B(v, r) \leq \operatorname{vol}(V)/2$, by repeated application of (1.2.4) we get,

$$\begin{aligned} \operatorname{vol}(B(v, r)) \geq \operatorname{vol}(B(v, r-1))(1 + \phi(B(v, r-1))) \geq \ldots &\geq \prod_{i=0}^{r-1} (1 + \phi(B(v, i))) \\ &\geq \exp\left(\frac{1}{2} \sum_{i=0}^{r-1} \phi(B(v, i))\right). \end{aligned}$$

where the last inequality uses the fact that $\phi(S) \leq 1$ for any set $S \subseteq V$. Since G is unweighted, $\operatorname{vol}(B(v, r)) \leq \operatorname{vol}(V) \leq n^2$. Therefore, by taking logarithm from both sides of the above inequality we get,

$$\sum_{i=0}^{r-1} \phi(B(v, i)) \leq 2 \log(\operatorname{vol}(B(v, r))) \leq 4 \log n.$$

Therefore, there exists $i < r$ such that $\phi(B(v, i)) \leq 4 \log n / r$. □

Now, for each $1 \leq i \leq k$, let $S_i := B(v_i, r_i)$. Since $r_i < \operatorname{diam}(G)/6k$, $S_1, \ldots, S_k$ are disjoint. Furthermore, by the above claim $\phi(S_i) \leq 24k \log n / \operatorname{diam}(G)$. Therefore, $\rho(k) \leq 24k \log n / \operatorname{diam}(G)$.

Finally, using (1.2.3), we get

$$\lambda_k \leq 2\rho(k) \leq \frac{48k \log n}{\text{diam}(G)}.$$

This completes the proof of Theorem 1.2.1.

Part I

New Approximation Algorithms to the Traveling Salesman Problem

Chapter 2

Background

Throughout this part we assume $G = (V, E)$ with $n := |V|$ vertices. Unless otherwise specified, we allow G to have parallel edges, so we think of E as a multi-set of edges. We often assume that G is an undirected graph and we use $e = \{u, v\}$ to denote an edge of G. if G is directed we use $a = (u, v)$ to denote an arc of G. For any undirected graph H we use $V(H)$ to denote the vertex set of H and $E(H)$ to denote the edge set of H. We use bold lower letters to refer to vectors. For a vector $\mathbf{x} \in \mathbb{R}^E$ (resp. $\mathbf{x} \in \mathbb{R}^A$), we use x_e (resp. x_a) to denote the value assigned to an edge e (or an arc a) of G. For a given function $f : A \to \mathbb{R}$, the cost of f is defined as follows:

$$c(f) := \sum_{e \in E} c(e) f(e).$$

For a set $S \subseteq E$, we define

$$f(S) := \sum_{e \in S} f(e).$$

We use the same notation for a function defined on the edge set A of an directed graph.

For a set $S \subseteq V$ we use

$$\begin{aligned} E(S) &:= \{\{u, v\} \in E : u, v \in S\}, \\ \delta(S) &:= \{\{u, v\} \in E : u \in S, v \in \overline{S}\}. \end{aligned}$$

If $S = \{v\}$ for a vertex $v \in V$, we may abuse the notation and use $\delta(v)$ instead. If $G = (V, A)$ is directed, we use

$$\delta^+(S) := \{(u, v) \in A : u \in S, v \in \overline{S}\}, \delta^-(S) := \{(u, v \in S : u \in \overline{S}, v \in S\}.$$

For disjoint $S, T \subseteq V$ we use $E(S, T) := \{\{u, v\} : u \in S, v \in T\}$.

2.1 The Traveling Salesman Problem

In an instance of the traveling salesman problem (TSP) we are given a set of cities (V) with a non-negative cost function $c\colon V \times V \to \mathbb{R}_+$ that satisfies the triangle inequality, i.e., for any $u, v, w \in V$,

$$c(u, v) \leq c(u, w) + c(w, v).$$

The goal is to find the shortest *tour* that visits each vertex *at least* once. A sequence of vertices $v_1, v_2, \ldots, v_k$ is a tour when $\{v_1, \ldots, v_k\} = V$. More precisely, our goal is to find a tour $v_1, \ldots, v_n$ such that

$$c(v_1, v_2) + c(v_2, v_3) + \ldots + c(v_{n-1}, v_n) + c(v_n, v_1)$$

is as small as possible.

Often this problem is formulated such that each vertex must be visited *exactly* once. It is easy two see that the two definitions are indeed equivalent. If we find a tour that visits each vertex at least once, then we can shortcut the tour and avoid visiting a vertex more than once. By triangle inequality the cost of the new tour can only be smaller than the original one.

TSP is proved to be NP-complete since the Hamiltonian Circuit problem is NP-complete. If we do not assume $c(.,.)$ satisfies the triangle inequality and we are asked to find a tour that visits each vertex exactly once, then the problem does not admit any approximation algorithm because of a simple reduction to the Hamiltonian circuit problem [SG76].

Next we describe several important variants of TSP all of them known to be NP-hard, and we recall the best known approximation algorithms for each of them prior to our works.

Symmetric TSP. Symmetric TSP (STSP) is the most well-known variant of TSP where we assume that the cost function is *symmetric*, i.e., for all $u, v \in V$, $c(u, v) = c(v, u)$. After a long line of work [Eng99, BS00, PV06, Lam12] the best known lower bound for approximating TSP is by Lampis [Lam12] who show that it is NP-hard to approximate TSP with a factor better than $\frac{185}{184}$. The best known approximation algorithm for TSP has an approximation factor of 3/2 and is due to Christofides [Chr76]. It is conjectured that there is a $\frac{4}{3}$ approximation algorithm for TSP but this conjecture is proved only in very special cases [AGG11, BSvdSS11, MS11]. Although many researchers tried to improve the 3/2 factor during the last 30 years, no-one is ever succeeded. It remains a central open problem in the field of computing to design an approximation algorithm for TSP that beats the Christofides' 3/2 approximation factor.

Asymmetric TSP. Asymmetric TSP (ATSP) is the most general variant of TSP where any cost function that is not necessarily symmetric but satisfies the triangle inequality is allowed. The first approximation algorithm for ATSP is due to Frieze, Galbiati and Maffioli [FGM82] who obtained a $\log(n)$ approximation algorithm. Although many researchers tried, after a long line of work [Blä02, KLSS05, FS07] the approximation factor is only improved to $0.66 \log(n)$ and it

remained an important open problem to break the $\Theta(\log(n))$ barrier. On the hardness side, the best known inapproximability result is $\frac{117}{116}$ due to Papadimitriou and Vempala [BS00, PV06].

Euclidean TSP. Euclidean TSP is a special case where it is assumed that the vertices are mapped to a plane and the cost function is simply the Euclidean distance between the vertices. Papadimitriou proved that even Euclidean TSP is an NP-complete problem [Pap77]. Unlike the general version of TSP, there is no inapproximability result for Euclidean TSP. Indeed, Arora [Aro96] and Mitchell [Mit99] independently designed a polynomial time approximation scheme (PTAS) for Eulidean TSP. We recall that a PTAS is an algorithm that for any given $\epsilon > 0$ in a polynomial time in $|V|$ finds a solution that is within $(1 + \epsilon)$ factor of the optimum.

Graphic TSP. Graphic TSP is a natural special case of TSP where we are given an underlying connected graph $G_0 = (V, E_0)$, and for all $u, v \in V$, $c(u, v)$ is the length of the shortest path that connects u to v. Equivalently, we can reformulate graphic TSP as follows: we are given an unweighted graph, and we want to find an Eulerian connected subgraph with the minimum number of edges. We recall that a graph is Eulerian if every vertex has an even degree. Similarly, one can also define graphic ATSP problem where G_0 is directed graph and $c(u, v)$ is the length of the shortest directed path from u to v.

Observe that if G_0 is a allowed to be weighted then we would recover the general version of symmetric TSP. This is because for a given instance of symmetric TSP one can construct a complete graph G_0 where the weight of each edge $\{u, v\}$ is $c(u, v)$.

The importance of graphic TSP is that all of the known hard instances of TSP are essentially instances of graphic TSP (e.g., in the Lampis [Lam12] construction, although the graph G_0 is weighted, the weight of the edges are between 1 and 2). So, it seems graphic TSP capture the main difficulty of the problem. Also, similar to TSP, graphic TSP is APX-hard, meaning that under the $P \neq NP$ conjecture there is no PTAS for graphic TSP. Prior to our work the best known approximation algorithm for graphic TSP is also the 3/2 approximation algorithm of Christofides [Chr76].

Planar TSP. Planar TSP is a special case of TSP where the cost function is the shortest path completion metric of a weighted planar graph. In other words, we are given a weighted planar graph $G_0 = (V, E_0)$ and for all pair of vertices $u, v \in V$, $c(u, v)$ is the weight of the shortest path from u to v in G_0. Polynomial time approximation schemes have been found for planar TSP [GKP95, AGK+98, Kle05], and bounded genus TSP [DHM07]. Similarly, Planar ATSP is a special case of ATSP where the cost function is the shortest path metric of a weighted directed planar graph. Prior to our works nothing better than $O(\log(n))$ approximation algorithm was known for Planar ATSP.

2.2 Linear and Convex Programming

In this short section we overview useful properties of Linear and Convex programs. We refer the interested readers to [GLS93, BV06] for more information.

A polytope is an intersection of a number of half-planes. A polytope is *bounded* if there is a ball of finite radius that contains it. A polytope is *finite* if it is an intersection of a finite number of half-planes.

An extreme point solution of a linear program, or a vertex of a polytope is a feasible point $\mathbf{x}$ such that $\mathbf{x}$ cannot be written as a linear combination of two other points of the program.

Theorem 2.2.1 (Carathéodory's theorem)**.** *Any feasible point* $\mathbf{x}$ *of a linear program over* $\mathbb{R}^n$ *can be written as a convex combination of at most* $n+1$ *extreme point solutions.*

Algorithmic versions of the above theorem can be found in [GLS93, Thm 6.5.11], that is there is a polynomial time algorithm that, for any given feasible solution $\mathbf{x}$, writes $\mathbf{x}$ as a convex combination of at most $n+1$ extreme point solutions.

Consider the following generic convex program over points $\mathbf{x} \in \mathbb{R}^n$.

$$\begin{aligned} &\text{minimize } f_0(\mathbf{x}) \\ &\text{subject to } f_i(\mathbf{x}) \leq 0 \quad \forall 1 \leq i \leq m \\ &A\mathbf{x} = b \end{aligned}$$

where $f_0, \ldots, f_m$ are convex functions, $A \in \mathbb{R}^{m' \times n}$. We say a feasible solution $\mathbf{x}$ is in the *relative interior* of the above program if for all $1 \leq i \leq m$, $f_i(\mathbf{x}) < 0$. We say a convex program satisfies the *Slater's condition* if there is a feasible solution in the relative interior of the program. If a convex program satisfies the Slater's condition then it satisfies the strong duality, i.e., the primal optimum is equal to the Lagrangian dual (see [BV06, Section 5.2.3] for more information).

Next, we show that if $\mathbf{x}$ is a vector in the relative interior of a finite bounded polytope P, then $\mathbf{x}$ can be written as a convex combination of *all* vertices of P such that each vertex has a positive coefficient. Let $\mathbf{y}$ be the summation of all vertices of P. Since P is finite and bounded y is well defined. Since $\mathbf{x}$ is in the relative interior of P, for a sufficiently small $\epsilon > 0$, $\mathbf{x}' = \mathbf{x} - \epsilon\mathbf{y} \in P$. Now we can write $\mathbf{x}' = \sum_{i=1}^{k} \alpha_i \mathbf{x}_i$ as a convex combination of vertices of P, then write $\mathbf{x} = \epsilon\mathbf{y} + \sum_{i=1}^{k} \alpha_i \mathbf{x}_i$ as a convex combination of all vertices of P.

Many of the linear or convex programs that we study in this thesis have an exponential or even infinite number of constraints. To efficiently find a feasible or an extreme point solution of these programs we need to provide a *separating hyperplane oracle* and use the *ellipsoid algorithm*. Let $P \subset \mathbb{R}^n$ be an arbitrary bounded polytope. Let $R > 0$ such that for a point $\mathbf{y}_0 \in \mathbb{R}^n$,

$$P \subseteq \{\mathbf{y} : \|\mathbf{y} - \mathbf{y}_0\| \leq R\}.$$

Also, let $r > 0$ such that for a point $\mathbf{x}_0 \in P$, $\{\mathbf{x} \in P : \|\mathbf{x} - \mathbf{x}_0\| \leq r\} \subseteq P$. A separating hyperplane oracle is a deterministic algorithm that for any given point $\mathbf{y} \in \mathbb{R}^n$ either decides $\mathbf{y} \in P$, or finds a vector $\mathbf{a} \in \mathbb{R}^n$ such that for all $\mathbf{x} \in P$,

$$\langle \mathbf{a}, \mathbf{y} \rangle < \langle \mathbf{a}, \mathbf{x} \rangle.$$

The following theorem follows from Khachiyan's ellipsoid algorithm.

Theorem 2.2.2. *If the separating hyperplane oracle runs in in time polynomial in n and $\log(R/r)$, then the ellipsoid algorithm finds a feasible solution of P in time polynomial in n and $\log(R/r)$.*

Note that the running time is independent of the number of constraints, or the number of faces of P.

2.3 Matroids and Spanning trees

For a ground set of elements E and $\emptyset \neq \mathcal{I} \subseteq 2^E$, $\mathcal{M} = (E, \mathcal{I})$ is called a *matroid* if

1. For any $A \in \mathcal{I}$ and $B \subseteq A$ we have $B \in \mathcal{I}$. In other words, $\mathcal{I}$ is a downward closed family of subsets of E.

2. For any $A, B \in \mathcal{I}$ such that $|A| < |B|$ there exists an element $e \in B - A$ such that $A \cup \{e\} \in \mathcal{I}$. This property is called the *extension* property of matroids.

A well-known example of matroids is the *graphical* matroid defined on a graph G where E is the set of edges of G, and $\mathcal{I}$ is the subsets of E that does not include a cycle.

The *rank function* of a matroid $\mathcal{M}$ assigns to every $S \subseteq E$ a number

$$\text{rank}(S) := \max\{|A| : A \in \mathcal{I}, A \subseteq S\}.$$

If $\mathcal{M}$ is a graphic matroid of a graph $G = (V, E)$ then for any $S \subseteq E$, $\text{rank}(S) = |V| - \#\,\text{components}(S)$. In other words, $\text{rank}(S)$ is the number of edges in the maximum spanning forest of S.

A set $A \in \mathcal{I}$ is called a base of a matroid $\mathcal{M}$ if S has the largest number of elements among all sets in $\mathcal{I}$, i.e., $|A| = \max\{|B| : B \in \mathcal{I}\}$. For example, if G is a connected graph, and $\mathcal{M}$ is the graphic matroid on G, then the bases of $\mathcal{M}$ are exactly the spanning trees of G.

The following lemma will be useful later.

Lemma 2.3.1. *Suppose we have assigned weights $w : E \to \mathbb{R}$ such that $w(e_1) \leq \ldots \leq w(e_m)$. If $w(e_k) > w(e_{k+1})$ for some $1 \leq k < m$, then any maximum weight base of $\mathcal{M}$ has exactly $\text{rank}\{e_1 \ldots, e_k\}$ elements from $\{e_1, \ldots, e_k\}$.*

Proof. Let $S = \{e_1, \ldots, e_k\}$. Let B be a maximum weight base of $\mathcal{M}$ such that $|B \cap S| < \text{rank}(S)$. Let $A = B \cap S$. Since $A \in \mathcal{I}$, by the extension property, there is an element $e \in S$ such that $A \cup \{e\} \in \mathcal{I}$ (note that $e \notin B$). Now, again by the extension property we can extend $A \cup \{e\}$ by adding elements from B and obtain a base B'. By construction, $\{e\} = B' - B$, and for some $e' \in E$, $\{e'\} = B - B'$. Since $B \cap S \subseteq B' \cap S$, we must have $e' \notin S$. Therefore, by the definition of S, $w_e > w_{e'}$, and we get $w(B') > w(B)$ which is a contradiction. □

2.4 Linear Programming Relaxation

As we elaborated in Section 1.1 our approach for solving the Traveling salesman problem is by rounding an optimal solution to the linear programming (LP) relaxation of the problem. In this section we overview this relaxation and several other LP relaxation of related problems.

The following linear program first formulated by Dantzig, Fulkerson and Johnson [DFJ54] is known as subtour elimination polytope or Held-Karp LP relaxation (see also [HK70]).

$$\begin{aligned} \text{minimize} \quad & \sum_{\{u,v\}} c(u,v) x_{\{u,v\}} & & \\ \text{subject to} \quad & \sum_{u \in S, v \in \overline{S}} x_{\{u,v\}} \geq 2 & & \forall\, S \subsetneq V \\ & \sum_{v \in V} x_{\{u,v\}} = 2 & & \forall\, u \in V \\ & x_{\{u,v\}} \geq 0 & & \forall\, u, v \in V. \end{aligned} \tag{2.4.1}$$

Observe that an optimal *integral* tour that visits each vertex exactly once (i.e., a Hamiltonian circuit) is a feasible solution to the above LP. In particular, each vertex is adjacent to exactly two edges of the tour and each cut separates the endpoints of at least two edges of the tour. Therefore, the solution of above LP provides a *lower bound* on the cost of the optimum tour.

We use $G = (V, E, \mathbf{x})$ to denote the fractional support graph corresponding to a feasible solution $\mathbf{x}$, i.e., $E = \{e : x_e > 0\}$. For an edge $e \in E$ we use x_e to denote the fraction of e in G. In this sense, the degree of a vertex in G is the sum of the fractions of edges incident to that vertex. Therefore, G is fractionally 2-regular and 2-edge connected. We use $c(\mathbf{x}(E')) = \sum_{e \in E'} c(e) \cdot x_e$. In particular, we use $c(\mathbf{x}) := c(\mathbf{x}(E))$.

Next we describe the LP relaxation of asymmetric TSP.

$$\begin{aligned} \text{minimize} \quad & \sum_{u,v \in V} c(u,v) x_{(u,v)} & \\ \text{subject to} \quad & \sum_{u \in S, v \in \overline{S}} x_{(u,v)} \geq 1 & \forall\, S \subsetneq V \\ & \sum_{v \in V} x_{(u,v)} = \sum_{v \in V} x_{(v,u)} = 1 & \forall\, u \in V \\ & x_{(u,v)} \geq 0 & \forall\, u, v \in V. \end{aligned} \tag{2.4.2}$$

Note that in (2.4.1) we have a variable $x_{\{u,v\}}$ for each un-ordered pair of vertices, but in the above program we have $x_{(u,v)}$ for each ordered pair. Again a directed Hamiltonian circuit is a feasible solution to the above program, so the program provides a lower bound on the cost of the optimum tour of ATSP. Similarly, we can define $G = (V, A, \mathbf{x})$ as the support graph, where in this case G is a directed graph.

Both of LP (2.4.1) and (2.4.2) have an exponential number of constraints. An optimum solution can be computed in polynomial-time either by the ellipsoid algorithm or by reformulating the above programs as LPs with polynomially-bounded size. Note that in both cases a separating hyperplane oracle needs to find a global minimum cut in the fractional graph $G = (V, E, \mathbf{x})$ or $G = (V, A, \mathbf{x})$. The global minimum cut can be found by an almost linear time algorithm of Karger [Kar00].

2.4.1 Integrality Gap

Since STSP and ATSP are NP-complete we do not expect that the optimum solution of (2.4.1) and (2.4.2) provide an exact estimate of the cost of the optimum tour. (as it turns out in many of the practical applications the value of these LP are very close to the integral optimum [ABCC07]). Consequently, an optimum solution $\mathbf{x}$ is a fractional vector.

The integrality gap of a family of LP is the supremum of the ratio of the cost of the optimum integral solution to the cost of the optimum fractional solution of LP. Next, we describe a folklore example that shows the integrality gap of (2.4.1) is at least 4/3. Consider an infinite family of graphs illustrated in Figure 2.4.1 where each is an instance of Graphic TSP. It is easy to see that the cost of the optimum integral tour is at least $4n/3 - 2$ and the cost of the optimum fractional solution is exactly n (the latter is because the cost of each edge is exactly 1). Therefore,

$$\text{Integrality Gap (2.4.1)} \geq \lim_{n \to \infty} \frac{4n/3 - 2}{n} = \frac{4}{3}.$$

Wolsey [Wol80] proved a new analysis of Christofides' 3/2 approximation algorithm [Chr76] and he show that the integrality gap of (2.4.1) is at most 3/2. It is conjectured that the true value of the integrality gap is 4/3 but this remained open for several decades. Schalekamp, Williamson and

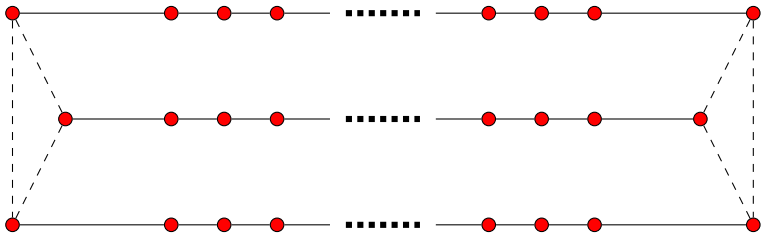

Figure 2.4.1: An illustration of the integrality gap example of LP (2.4.1). Each long path contains exactly $n/3$ vertices. For each edge $e = \{u, v\}$, $c(u, v) = 1$, and the $c(u, v)$ is the length of the shortest path connecting u to v for the remaining pair of vertices. Here, in an optimal solution of (2.4.1), $x_e = 1$ for solid edges and $x_e = 1/2$ for dashed edges.

van Zuylen [SWvZ12] conjectured the worst-case ratio of (2.4.1) occurs for half-integral fractional solution, i.e., vectors $\mathbf{x}$ where x_e is $1/2$ or 1 for all of the edges in the support of $\mathbf{x}$, but still we don't know the worst case integrality gap of even half-integral solutions. In efforts of better understanding the Held-Karp relaxation, extreme point solutions of this linear program are intensely studied in the literature [BP91, CV00, Goe06].

Integrality gap of (2.4.2) is also studied. Prior to our works, the best upper bound is $O(\log(n))$ by the analysis of Frieze et al. [FGM82]. But, the best lower bound is only 2 by a recent work of Charikar, Goemans and Karloff [CGK06]. Note that compared to TSP, we have a significantly larger gap between the upper bound and lower bound. It is conjectured that the integrality gap is a constant. As we describe in Chapter 5 under Conjecture 5.3.2 the integrality gap of (2.4.2) is a constant.

2.4.2 Integral Polytopes

One can also write an LP relaxation for problems in P. These LPs are typically integral (i.e., their integrality gap is 1). In other words, any non-integral feasible solution can be written as a convex combination of two other feasible solutions, so any extreme point solution is an integral vector.

We start by describing the spanning tree polytope. For $G = (V, E)$, Edmonds [Edm70] gave the following LP relaxation of spanning trees of G.

$$\begin{aligned} &\mathbf{z}(E) = n - 1 && \\ &\mathbf{z}(E(S)) \leq |S| - 1 && \forall S \subseteq V \\ &z_e \geq 0 && \forall e \in E. \end{aligned} \tag{2.4.3}$$

Edmonds [Edm70] proved that above linear program is exactly the convex-hull of all spanning trees of graph G, i.e., extreme point solutions of above linear program are exactly the spanning trees of G. Therefore, above program is known as the *spanning tree polytope*.

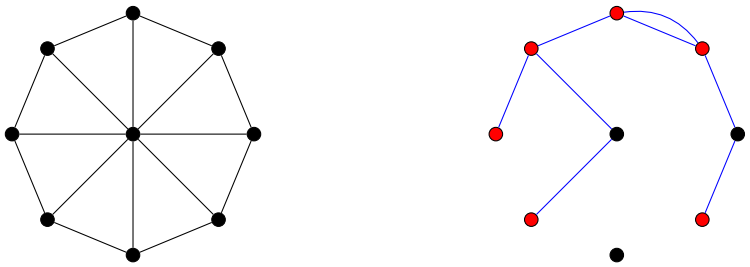

Figure 2.4.2: Consider the wheel graph shown at left. The right diagram shows an O-join, the red vertices are the vertices of O and the edges of the O-join are shown in blue.

We made the following simple observation in [AGM+10]

Fact 2.4.1. *For any feasible solution* $\mathbf{x}$ *of* (2.4.1), $\mathbf{z} = (1-1/n)\mathbf{x}$ *is a feasible solution in the relative interior of* (2.4.3)

Proof. First, observe that

$$\mathbf{z}(E) = (1-1/n)\mathbf{x}(E) = \frac{1-1/n}{2}\sum_{v\in V}\mathbf{x}(\delta(v)) = n-1.$$

On the other hand, for any $S \subseteq V$,

$$\mathbf{z}(E(S)) = (1-1/n)\mathbf{x}(E(S)) = \frac{1-1/n}{2}\Big(\sum_{v\in S}\mathbf{x}(\delta(v)) - \mathbf{x}(\delta(S))\Big) = \frac{1-1/n}{2}(2|S|-2)$$
$$< |S|-1.$$

So, $\mathbf{z}$ is in the relative interior of (2.4.3). □

For a graph $G = (V, E)$ and a set $O \subseteq V$ with even number of vertices, an O-join is a *multiset* F of edges of E such that in the subgraph (V, F) every vertex of O has an odd degree and every vertex of $V - O$ has an even degree. Note that F can have multiple copies of an edge in E (see Figure 2.4.2 for an example). We remark that conventionally the term T-join is used, here we would rather reserve the notation T for spanning trees.

Edmonds and Johnson [EJ73] proved the following characterization of the O-join polytope.

Proposition 2.4.2. *For any graph* $G = (V, E)$ *and cost function* $c : E \to \mathbb{R}_+$, *and* $O \subseteq V$ *with even number of vertices, the minimum weight of an* O*-join equals the optimum value of the following linear program.*

$$\begin{array}{lll} \text{minimize} & c(\mathbf{y}) & \\ \text{subject to} & \mathbf{y}(\delta(S)) \geq 1 \quad \forall S \subseteq V, |S \cap O| \text{ odd} & \\ & y_e \geq 0 \qquad\qquad\qquad\qquad \forall e \in E & \end{array} \tag{2.4.4}$$

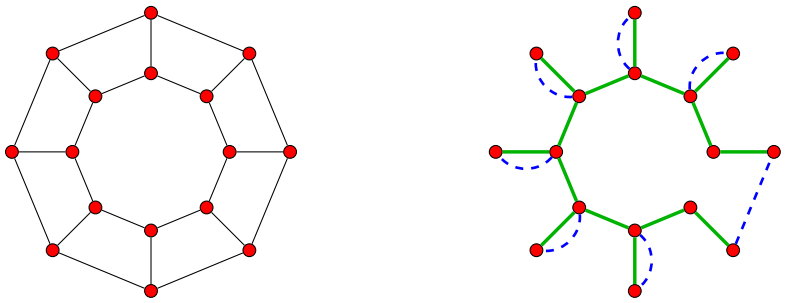

Figure 2.5.3: The left graph shows an instance of Graphic TSP. For each pair of vertices $u, v \in V$, $c(u, v)$ is the length of the shortest path from u to v. In the right graph we show a possible output of Algorithm 4. The green edges are the edges of a minimum spanning tree, and the blue dashed edges are the edges of a minimum cost perfect matching on the odd degree vertices of the tree. Observe that the cost of the computed solution is 23 while the optimum 16. If we increase the number of vertices to infinity, then the approximation factor approach 3/2

Note that (2.4.4) is the up-hull of all O-joins of G. We recall that the up-hull of a polytope is the set of points that are at least as large as some point in the polytope.

2.5 The Christofides' Algorithm

In this section we describe Christofides' 3/2 approximation algorithms for symmetric TSP. This algorithm is one of the first approximation algorithms designed in the field of computing. Because of the simplicity and elegance of this algorithm, it is taught in many of the introduction to algorithm courses.

The details of the algorithm is described in Algorithm 4. In Figure 2.5.3 we illustrated an output of this algorithm in an instance of Graphic TSP.

Algorithm 4 Christofides' 3/2 approximation algorithm for TSP

Compute a minimum cost spanning tree of the complete graph where cost of each edge $\{u, v\}$ is $c(u, v)$.
Compute a minimum cost perfect matching on the odd degree vertices of the spanning tree.
Return the union of the tree and the matching.

Observe that the union of a spanning tree and the matching is a feasible TSP tour, because it is connected and Eulerian. It is also easy to see that the cost of the solution is at most 3/2 of the optimum. First, observe that the cost of the spanning tree is always at most the cost of the optimum tour, because by removing an edge of any Hamiltonian circuit we obtain a spanning tree. On the other hand, we show for any $S \subseteq V$, where $|S|$ is even, the cost of the minimum matching on the vertices of S is at most half of the optimum. First observe that by the triangle inequality we can

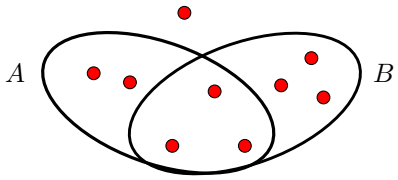

Figure 2.6.4: An example of two crossing sets.

shortcut the vertices in $V - S$ and obtain a tour of cost at most the optimum that visits each vertex of S exactly once. This tour defines two disjoint perfect matching on S one of which have cost at most half of the optimum tour.

Figure 2.5.3 shows that the above analysis is tight and the cost of the solution of the Christofides' algorithm may be 3/2 of the optimum solution even in instances of Graphic TSP. The above analysis shows that the cost of the tour obtained in Algorithm 4 is at most 3/2 of the optimum integral tour, but it does not bound the integrality gap of LP (2.4.1). Next, we describe an argument of Wolsey [Wol80] that shows the cost of tour in the output is at most 3/2 of the optimum value of LP (2.4.1).

Let $\mathbf{x}$ be an optimum solution of (2.4.1). Let $\mathbf{z} = (1 - 1/n)\mathbf{x}$. Then, by Fact 2.4.1, $\mathbf{z}$ is inside the LP (2.4.3). So, the cost of the minimum spanning tree is at least $c(\mathbf{z}) \leq c(\mathbf{x})$. It remains to upper bound the cost of the matching by $c(\mathbf{x})/2$. Let $\mathbf{y} = \mathbf{x}/2$. By (2.4.1) for each cut $(S, \overline{S})$, $\mathbf{y}(\delta(S)) \geq 1$. So $\mathbf{y}$ is inside the LP (2.4.4), for any set $O \subseteq V$, and in particular, for the set O being the odd degree vertices of the chosen minimum spanning tree. By Proposition 2.4.2 the cost of the minimum O-join on G is at most $c(\mathbf{y}) = c(\mathbf{x})/2$, So, by the triangle inequality the cost of the minimum cost perfect matching is at most $c(\mathbf{x})/2$.

2.6 Structure of Minimum Cuts

In this section we overview the structure of minimum cuts of any Δ-edge connected undirected graph G. All of the statements hold for fractional graphs as well but for simplicity of notation we assume G is integral, but we allow G to have parallel edges. Before describing the structure of minimum cuts we discuss general properties of any family of cuts in G.

Two subsets $A, B \subseteq V$ are crossing if $A \cap B, A - B, B - A, \overline{A \cup B}$ are all non-empty (see Figure 2.6.4). Two cuts $(A, \overline{A})$ and $(B, \overline{B})$ are crossing if A, B are crossing. A cut $(A, \overline{A})$ is a *trivial* cut if $|A| = 1$ or $|\overline{A}| = 1$.

Definition 2.6.1 (Atom)**.** *For a collection $\mathcal{F}$ of cuts, the atoms of $\mathcal{F}$ are the members of a partition $\mathcal{P}$ of the vertex set V such that*

- *no cut of $\mathcal{F}$ divides any of the atoms of $\mathcal{F}$, and*
- *$\mathcal{P}$ is the coarsest partition with this property.*

We say an atom is singleton *if it is a set of a single vertex of V.*

See Figure 2.6.5 for example of atoms of a family of cuts. We say a cut class is *trivial* if it has two atoms and one of them is a singleton.

Definition 2.6.2 (Cross Graph)**.** *For a collection $\mathcal{F}$ of cuts of a graph G,* cross graph *$\mathcal{G}$ is a graph on vertex set $\mathcal{F}$ that has an edge between two cuts in $\mathcal{F}$ if they cross. Each connected component of $\mathcal{G}$ is called a* cut class*, we use $\mathcal{C}$ to denote a single cut class.*

For example, if $\mathcal{F}$ is the set of three cuts that are shown by dashed lines in Figure 2.6.5, then $\mathcal{G}$ has just one connected component.

For a cut class $\mathcal{C}$ we use $\psi(\mathcal{C})$ to denote the set of atoms of $\mathcal{C}$. We say a cut $(A, \overline{A})$ is *non-proper*, if it separates an atom of its cut class from the rest of the atoms, and it is *proper* otherwise. Observe that a cut class $\mathcal{C}$ has a non-proper cut if and only if it has exactly one cut, or equivalently two atoms. This is because no cut crosses a non-proper cut. Since a cut class with 3 atoms can only have non-proper cuts, and no two non-proper cuts cross, we cannot have any cut class with 3 atoms.

In the rest of this section we prove several properties of cut classes and their atoms. Our first lemma relates atoms in different cut classes. The proof is based on [Ben97, Lemma 4.1.7].

Lemma 2.6.3 ([Ben97, Lemma 4.1.7])**.** *For any two distinct cut classes $\mathcal{C}_1, \mathcal{C}_2$ of a collection of cuts $\mathcal{F}$, there are unique atoms $A^* \in \psi(\mathcal{C}_1)$ and $B^* \in \psi(\mathcal{C}_2)$ such that $A^* \cup B^* = V$. So,*

i) For any $B \in \psi(\mathcal{C}_2)$ such that $B \neq B^*$, we have $B \subseteq A^*$.

ii) For any $A \in \psi(\mathcal{C}_1)$ such that $A \neq A^*$, we have $A \subseteq B^*$.

iii) For any $A \in \psi(\mathcal{C}_1), B \in \psi(\mathcal{C}_2)$, A does not cross B.

iv) If there are $A \in \psi(\mathcal{C}_1)$, $B \in \psi(\mathcal{C}_2)$ such that $A = B$, then exactly one of $\mathcal{C}_1, \mathcal{C}_2$ is non-proper. So, we cannot have three atoms A_1, A_2, A_3 in three distinct cut classes such that $A_1 = A_2 = A_3$.

Proof. Consider a cut $(C, \overline{C}) \in \mathcal{C}_1$, and form two subsets of the cuts $(D, \overline{D}) \in \mathcal{C}_2$ depending on whether D or $\overline{D} \subseteq C$, or D or $\overline{D} \subseteq \overline{C}$. Since no cut of $\mathcal{C}_2$ crosses a cut of $\mathcal{C}_1$, all cuts of $\mathcal{C}_2$ are put into one of these two sets. Furthermore, since no cut from one of the subsets of above type may cross a cut from the other set and since the cuts in $\mathcal{C}_2$ form a connected component, one of the two sets must be empty. So, (perhaps after renaming) assume that for all $(D, \overline{D}) \in \mathcal{C}_2$, we have $D \subseteq C$.

The above argument holds for any $(C, \overline{C}) \in \mathcal{C}_1$. So,

$$\bigcup_{(D,\overline{D})\in\mathcal{C}_2} D \subseteq \bigcap_{(C,\overline{C})\in\mathcal{C}_1} C.$$

But by the definitions of atoms the LHS is the complement of an atom B^* in $\psi(\mathcal{C}_2)$, and the RHS is an atom A^* in $\psi(\mathcal{C}_1)$. These two atoms satisfy the lemma's conclusion. We prove the uniqueness below.

Conclusions (i), (ii) simply follow from the fact that the atoms of any cut class form a partition of the vertex set V. Conclusion (iii) follows from (i) and (ii).

Now, let us prove (iv). Since $A = B$ we must either have $B \neq B^*$ or $A \neq A^*$. Without loss of generality assume $B \neq B^*$. Since by (i) $B \subseteq A^*$, $A = A^*$. So, $B = A^*$ and by (i) $\mathcal{C}_2$ must have exactly two atoms.

Finally, let us prove the uniqueness of A^*, B^*. Suppose there are another pair of atoms $A \in \psi(\mathcal{C}_1)$ and $B \in \psi(\mathcal{C}_2)$ such that $A \cup B = V$ and $A \neq A^*$. Since $A \subseteq B^*$, we must also have $B \neq B^*$. Since by (ii) $\mathcal{C}_1 \neq \mathcal{C}_2$, at least one of $\mathcal{C}_1, \mathcal{C}_2$, say $\mathcal{C}_1$ is non-proper and has an atom $A' \neq A, A^*$. But then by (ii) $A' \subseteq B^*$ and $A' \subseteq B$ which is a contradiction. □

In the following lemma we upper bound the number of atoms of any family of cut classes.

Corollary 2.6.4. *For any k cut classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$ of any collection $\mathcal{F}$ of cuts, we can find a family P of atoms such that atoms in P are pairwise disjoint and*

$$|P| \geq -2(k-1) + \sum_{i=1}^{k} |\psi(\mathcal{C}_i)|.$$

Proof. Without loss of generality, we assume $|\psi(\mathcal{C}_i)| > 2$ for all $1 \leq i \leq k$. So, by Lemma 2.6.3 for any $i \neq j$ and any $A \in \psi(\mathcal{C}_i), B \in \psi(\mathcal{C}_j)$ we must have $A \neq B$. Also, let $l = \sum_{i=1}^{k} |\psi(\mathcal{C}_i)|$.

By Lemma 2.6.3 for any $i \geq 2$ there is an atom $B_i^* \in \psi(\mathcal{C}_i)$ that is a complement of an atom $A_i^* \in \psi(\mathcal{C}_1)$. Note that A_i^* and A_j^* are not necessarily different atoms of $\mathcal{C}_1$. Let

$$Q = \psi(\mathcal{C}_1) \cup (\psi(\mathcal{C}_2) - B_2^*) \cup \ldots \cup (\psi(\mathcal{C}_k) - B_k^*).$$

By part (i) of Lemma 2.6.3 for any $i \geq 2$, and any $B \in \psi(\mathcal{C}_i) \cap Q$, $B \subseteq A_i^*$. So, by part (iii) of Lemma 2.6.3 for any $A, B \in Q$, we have $A \cap B \neq \emptyset$ if and only if $A \subset B$ or $B \subset A$. Furthermore, for any atom $A \in Q$ that is a superset of an atom $B \in \psi(\mathcal{C}_i) \cap Q$, A is also a superset of any other atoms in $\psi(\mathcal{C}_i) \cap Q$. To obtain a set of pairwise disjoint atoms we only need to remove supersets of any atom of Q. But, we have to do this removal carefully such that we can count the number of remaining atoms.

By above argument, for any cut class $\mathcal{C}_i$ ($i \geq 2$) there is a chain of atoms $A_{i,1} \subset A_{i,2} \subset A_{i,3} \ldots \subset A_i^*$ such that they are supersets of all atoms of $\psi(\mathcal{C}_i) \cap Q$, and these are the only supersets of atoms of $\psi(\mathcal{C}_i) \cap Q$. Now, we define

$$P = Q - A_{2,1} - A_{3,1} - \ldots - A_{k,1}.$$

First, observe that

$$|P| \geq |Q| - (k-1) = -2(k-1) + \sum_{i=1}^{k} |\psi(\mathcal{C}_i)|$$

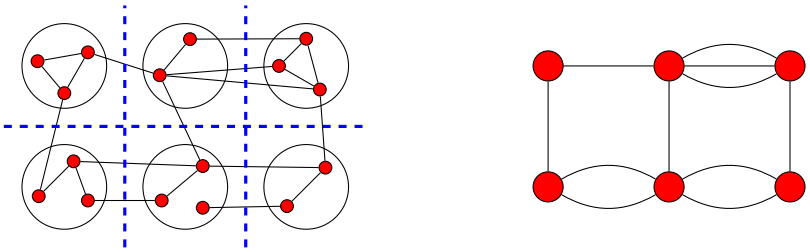

Figure 2.6.5: Left graph shows a connected family of three crossing cuts. This particular family has 6 atoms that are represented by 6 circles. In the right graph we identify the set of the vertices in each of the 6 atoms and we remove the loops.

On the other hand, for any $i \geq 2$, any atom $A_{i,j}$ in the chain $A_{i_1}, A_{i,2}, \ldots$ is removed by $A_{i,j-1}$. Therefore, for any A, B in P we must have $A \not\subset B$ and $B \not\subset A$. Since $P \subseteq Q$ atoms of P are pairwise disjoint. □

2.6.1 Properties of Minimum Cuts

Next, we discuss specific properties of structure of minimum cuts. For $A \subseteq V$, $(A, \overline{A})$ is a minimum cut of G, if $|\delta(A)| = \Delta$. It turns out that if $(A, \overline{A})$ and $(B, \overline{B})$ are crossing minimum cuts, then the cuts defined by any of the four sets $A \cap B, A - B, B - A, A \cup B$ are also minimum cuts.

Lemma 2.6.5. *For any two crossing minimum cuts $(A, \overline{A})$, $(B, \overline{B})$,*

$$|\delta(A \cap B)|, |\delta(A - B)|, |\delta(B - A)|, |\delta(A \cup B)| = \Delta$$

Proof. Observe that,

$$2\Delta \leq |\delta(A \cap B)| + |\delta(A \cup B)| \leq |\delta(A)| + |\delta(B)| = 2\Delta.$$

So, $|\delta(A \cap B)| = |\delta(A \cup B)| = \Delta$. The other cases can be proved similarly. □

For a partitioning $P = \{A_1, A_2, \ldots, A_k\}$ of vertices in G, let $G(P)$ be the graph obtained by identifying the vertex set of each part A_i, and removing the self-loops afterwards. In particular, for a cut class $\mathcal{C}$, each vertex of $G(\psi(\mathcal{C}))$ is an atom of $\mathcal{C}$ (see Figure 2.6.5 for an example).

The following lemma characterizes the structure of minimum cuts in a single cut class. The lemma follows from the work of Dinits, Karzanov and Lomonosov [DKL76] (the proof below is based on a technical report by Fleiner and Frank [FF09]).

Lemma 2.6.6 ([DKL76])**.** *Let $\mathcal{C}$ denote a cut class of minimum cuts of G. Then $G(\psi(\mathcal{C}))$ is a cycle where between each adjacent pair of vertices there are $\Delta/2$ parallel edges, and every pair of edges of the cycle corresponds to a minimum cut of G.*

Proof. The proof essentially follows by repeated applications of Lemma 2.6.5. Let $H = G(\psi(\mathcal{C}))$ (note that H is also Δ-edge connected). We say $A \subseteq V(H)$ is *tight* if $|\delta(V(A))| = \Delta$. We say a tight set A is *non-trivial* if $|A| \geq 2$. The next fact follows from the fact that the atoms of $\mathcal{C}$ are the coarsest partition with respect to cuts in $\mathcal{C}$.

Fact 2.6.7. *Any non-trivial set A is crossed by a tight set.*

First, we show that the degree of each vertex of H is exactly Δ. For $v \in V(H)$, let A be the smallest tight set containing v. If $|A| > 1$, there must be a tight set B crossing A such that $v \in B$. But then $A \cap B$ is a smaller set containing v, and by Lemma 2.6.5 it is a tight set. So, we must have $|A| = 1$, and we get $|\delta(v)| = \Delta$.

Since each vertex has a degree of Δ, all we need to show is that each vertex $u \in V(H)$ is contained in at least two tight sets $\{u, v\}$ and $\{u, w\}$ for $v \neq w$. In the next claim we show that every vertex is contained in at least one tight set of size 2.

Claim 2.6.8. *Let A be a non-trivial tight set containing u. Then, A includes a two element tight set containing u.*

Proof. We prove by induction. If $|A| = 2$ we are already done. So assume $|A| \geq 3$. Let B be a tight set containing u that crosses A. If $|A \cap B| \geq 2$, then we are done by induction. If $|A \cap B| = 1$, then since $A - B$ is a non-trivial set there is also a non-trivial tight set B' that contains u and crosses $A - B$. Now, either $B' \subset A$, or B' crosses A. In either case B', or $B' \cap A$ gives a smaller non-trivial tight set containing u, and we are done by induction. □

It remains to show that a vertex u is contained in two tight sets of size 2. By the above claim let $\{u, v\}$ be a tight set. Then, there is a non-trivial tight set, A containing u that crosses $\{u, v\}$. So $v \notin A$. By above lemma A includes a two element tight set, $\{u, w\}$. Since $v \notin A$, $v \neq w$. □

The above lemma nicely characterizes any proper cut class of minimum cuts: we can place the atoms around a cycle such that each two non-consecutive edges of the cycle is a minimum cut in that cut class. Consequently, if $\mathcal{C}$ is a proper cut class of minimum cuts with k atoms, then it contains exactly $k(k-3)/2$ minimum cuts.

2.6.2 The Cactus Representation

Next, we discuss the representation of different cut classes of minimum cuts as represented by a cactus graph [DKL76] and then the representation of different cut classes of an arbitrary collection of cuts in a tree hierarchy [Ben95].

Definition 2.6.9 (Cactus Graph)**.** *A* cactus graph *is a graph with no cut edges in which no two simple cycles share an edge.*

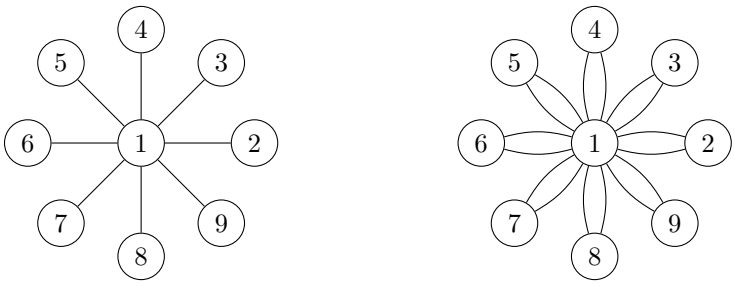

Figure 2.6.6: The cactus representation of a star graph

Observe that one can assign a tree to any cactus K; the vertices of the tree are the cycles of K and there is an edge between two vertices if corresponding cycles share a vertex. Also, the minimum cut of any cactus is exactly 2.

The cactus representation of G consists of a partition P of vertices of G and a cactus graph K whose vertices may have either an element of P or the empty set, each element of P appearing exactly once. Each minimum cut of G corresponds to a minimum cut of the cactus and each minimum cut of the cactus correspond to a minimum cut of G. There are many ways to construct a cactus representation of the minimum cuts of G. The following construction is based on [Ben97, BG08].

Lemma 2.6.10 ([BG08, Prop 24])**.** *Let $\mathcal{F}$ be the collection of minimum cuts of an unweighted Δ-edge connected graph G. For each cut class $\mathcal{C}$, let P_i be the partition of V corresponding to $\psi(\mathcal{C})$. Then there exists a cactus $K = (N, E')$ and a mapping $f : V \to N$ such that*

1. *K has no cycle of length 3,*
2. *there is a 1-to-1 correspondence between the connected components $\mathcal{C}$ of the cross graph and the cycles C_i of K,*
3. *the removal of the edges of $C_i = u_1 - u_2 \ldots - u_l - u_1$ break K into k (depending on i) connected components, $A_1, \ldots, A_l \subset N$ where $u_j \in A_j$ such that $P_i = \{f^{-1}(A_j) : 1 \leq j \leq l\}$,*
4. *for each cut $(B, \overline{B}) \in \mathcal{F}$,there is a unique cycle C_i in K and two edges of C_i which are non-consecutive if the cycle is not of length 2, whose removal partitions N into U and $N - U$ where $B = f^{-1}(U)$.*

Proof. The proof is a simple application of Lemma 2.6.3 and Lemma 2.6.6. First, if all of the minimum cuts of G are degree cuts, then the cactus is simply a star with two parallel edges between the center and any any other vertex. Each vertex of degree Δ in G is mapped to a distinguished leaf and all other vertices are mapped to the center (see Figure 2.6.6).

Otherwise, there is a cut class, $\mathcal{C}$ with a non-degree cut. Let $A_1, \ldots, A_l$ be the atoms of $\psi(\mathcal{C})$ in the order of vertices in the cycle graph $G(\psi(\mathcal{C}))$. For each atom A_i, Let G_i be the graph obtained

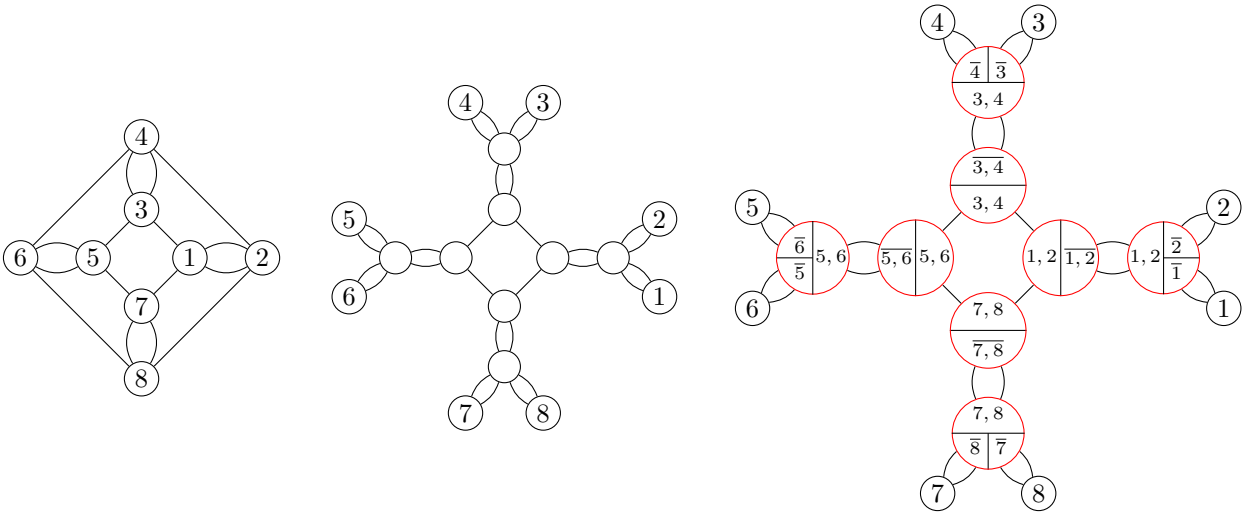

Figure 2.6.7: The left graph is a 4-edge connected graph. The middle graph shows the cactus representation of the minimum cuts of the left graph. Observe that each vertex is mapped to exactly one vertex of the cactus, and 4 vertices of the cactus are empty. Also, each minimum cut of the cactus, i.e., each cut of of size 2, corresponds to a minimum cut of the left graph. The right graph shows the tree hierarchy of the minimum cuts of the left graph. We used $\overline{1,2}$ to denote the set $\{3,4,\ldots,8\}$, similarly, $\overline{1} = \overline{\{1\}}$ Observe that each vertex of left graph is mapped to one vertex of each cycle of the cactus. Three atoms of three different cut classes are mapped to each vertex of the middle cycle.

by identifying the vertices in $\overline{A_i}$. Find a cactus representation K_i of G_i recursively. We also assume that $K_1, \ldots, K_l$ have disjoint vertex sets. Now, we are ready to construct K. First, we add a cycle, C, of length $|\psi(\mathcal{C})|$. Then, we identify the i-th vertex of the cycle with a vertex of $K(G_i)$ where $\overline{A_i}$ is mapped. This vertex will be an empty vertex. So, the vertices of A_i are only mapped to the K_i subgraph of K. This completes the construction

Conclusion 1 holds since no cut-class has 3 atoms. By Lemma 2.6.3 for every other cut class $\mathcal{C}' \neq \mathcal{C}$, there is an atom A_i of $\mathcal{C}$ such that all but one of the atoms of $\mathcal{C}'$ are subsets of A_i. So, the cuts in $\mathcal{C}'$ are mapped to the subgraph K_i of K. So contracting the vertices of atoms in the recursive construction preserve all of the cuts in $\mathcal{F}-\mathcal{C}$. Conclusions 2 and 3 hold by the construction. Conclusion 4 holds by Lemma 2.6.6. □

See Figure 2.6.7 for an example of the cactus representation.

Benczur, in his thesis [Ben97], studied representations of *any* collection of cuts. He slightly changed the cactus representation and extended the above construction to any collection of cuts [Ben97, Theorem 4.1.6]. The following representation is called *tree hierarchy*.

Theorem 2.6.11 ([Ben97])**.** *Let $\mathcal{F}$ be a collection of cuts of G, and let P_i be the partition of V corresponding to $\psi(\mathcal{C})$. There exists a cactus $K = (N, E')$ such that*

1. *K has no cycle of length 3.*

2. *there is a 1-to-1 correspondence between the cut classes of $\mathcal{F}$ and the cycles of K such that the atoms of each cut class $\mathcal{C}$ are mapped to the vertices of a unique cycle C_i,*

3. *all pairs of atoms $A \in \psi(\mathcal{C})$ and $B \in \psi(\mathcal{C}')$ with $\mathcal{C} \neq \mathcal{C}'$ that are mapped to a coinciding vertex of the cactus satisfy $A \cup B = V$.*

See Figure 2.6.7 for an example of the tree hierarchy. Observe that the only difference with the construction of cactus representation in the proof of Lemma 2.6.10 is that here the vertices of G are mapped to exactly one vertex of *each* cycle of K.

There is a simple way to construct a tree hierarchy from Lemma 2.6.10. All we need to do is to give a partitioning of $V(G)$ for each cycle of K. For a cycle of length l, this partitioning is exactly the pre-images of $A_1, \ldots, A_l$ as defined in conclusion 3 of Lemma 2.6.10. The proof of above theorem follows from Lemma 2.6.3. Recall that Lemma 2.6.3 hold for *any* collection of cuts. The construction is very similar to Lemma 2.6.10 except in this case the ordering of atoms in each cycle is arbitrary.

In this thesis we only work with tree hierarchy representation of cuts. Because of the generality of above theorem we can use it when we discuss the structure of near minimum cuts in Section 2.7.

2.6.3 Properties of Tree Hierarchy and Cactus Representation

Before concluding this section we prove some general properties of and tree hierarchy of any collection of cuts.

The following lemma provides an upper bound on the number of cut classes with τ number of atoms.

Corollary 2.6.12. *The cut classes of any collection $\mathcal{F}$ of cuts satisfy the following properties.*

i) For any $\tau \geq 4$, the number of cut classes with at least τ atoms is at most $n/(\tau - 2)$.

ii) The number of all cut classes is at most $2n - 3$.

iii) The number of atoms of all cut classes is at most $5n$.

Proof. Let $\mathcal{C}_1, \mathcal{C}_2 \ldots, \mathcal{C}_k$ be cut classes with at least τ atoms, By definition,

$$\sum_{i=1}^{k} |\psi(\mathcal{C}_i)| \geq k \cdot \tau.$$

Applying Corollary 2.6.4 to these cut classes, there is a P of mutually disjoint atoms of them such that

$$|P| \geq -2(k-1) + \sum_{i=1}^{k} |\psi(\mathcal{C}_i)|.$$

So,

$$k \cdot \tau \leq |P| + 2k \leq n + 2k,$$

where we used the fact that the atoms in P are pairwise disjoint. So, $k \leq n/(\tau - 2)$. This proves part (i).

Next we prove part (ii). We prove it by induction. If $n = 2$, then the claim obviously holds. First, assume that $\mathcal{F}$ has no crossing cuts, i.e., every cut class is proper. Fix a vertex $v \in V$. Let $\mathcal{S}$ be a family of subsets of V such that for any cut $(A, \overline{A}) \in \mathcal{F}$, if $v \notin A$, the $A \in \mathcal{S}$ and otherwise $\overline{A} \in \mathcal{S}$. It follows that $\mathcal{S}$ is a *laminar* family of subsets of V, i.e., no two sets in $\mathcal{S}$ cross. So, $|\mathcal{F}| = |\mathcal{S}| \leq 2n - 3$ and we are done.

Now, assume that $\mathcal{F}$ has a proper cut class $\mathcal{C}$ with atoms $A_1, \ldots, A_k$. For $1 \leq i \leq k$, let G_i be the graph where $\overline{A_i}$ is contracted. By Lemma 2.6.3 every cut in $\mathcal{F} - \mathcal{C}$ is in exactly one G_i. By the induction hypothesis the number of cut classes of each G_i is at most $2(|A_i| + 1) - 3$. So, the number of cut classes of G is at most

$$1 + \sum_{i=1}^{k} (2|A_i| - 1) \leq \sum_{i=1}^{k} 2|A_i| - 3 \leq 2n - 3,$$

where the first inequality follows by the assumption that $\mathcal{C}$ is a proper cut class and $k \geq 4$, and the second inequality follows by the fact that atoms of $\mathcal{C}$ form a partition of V.

It remains to prove part (iii). Suppose there are k cut classes with at least 4 atoms. By Corollary 2.6.4,

$$\sum_{\mathcal{C}} |\psi(\mathcal{C})| = \sum_{\mathcal{C}:|\psi(\mathcal{C})| \geq 4} |\psi(\mathcal{C})| + \sum_{\mathcal{C}:|\psi(\mathcal{C})|=2} 2 \leq n + 2k + 2(2n - k) = 5n.$$

where the last inequality follows by part (ii) that the number of all cut classes is at most 2n. □

Lemma 2.6.13. *For any collection $\mathcal{F}$ of cuts of a graph G, in all (except possibly one) of the cut classes there is an atom which is a subset of at least $n/2$ of the vertices of G.*

Proof. Let $\mathcal{C}$ be a cut class such that for any $A \in \psi(\mathcal{C})$, $|A| < n/2$. If such a cut class does not exist, we are already done. By Lemma 2.6.3 for any cut class $\mathcal{C}' \neq \mathcal{C}$ there exists an atom $A^* \in \psi(\mathcal{C})$ and $B^* \in \psi(\mathcal{C}')$ such that $A^* \cup B^* = |V|$. Since $|A^*| < n/2$, $|B^*| \geq n/2$ and we are done. □

Lemma 2.6.14. *Let $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_{2l+2}$ be cut classes of a collection $\mathcal{F}$ of cuts of G for $l \geq 1$. Then there exists a cut class $\mathcal{C}_i$ that has an atom with at least $\min\{n/2 + l, 2n/3\}$ vertices.*

Proof. For a cut class $\mathcal{C}$ let

$$f(\mathcal{C}) := \max_{A \in \psi(\mathcal{C})} |A|.$$

Let $\mathcal{C}_1$ be a cut class that maximizes $f(.)$ and if there are multiple maximizers, let $\mathcal{C}_1$ be one with maximum number of atoms. Also, let $A_1 \in \psi(\mathcal{C}_1)$ be an atom of $\mathcal{C}_1$ with maximum number of vertices.

By Lemma 2.6.3, for any $2 \leq i \leq 2l+2$ there is an atom $A_i \in \psi(\mathcal{C}_i)$ such that for some $B_i \in \psi(\mathcal{C}_1)$, $A_i \cup B_i = V$. First, we show that $B_i = A_1$ for all $2 \leq i \leq 2l+2$. By definition of A_i,

$$|A_i| \geq n - |B| \geq |A_1|.$$

By definition of $\mathcal{C}_1$, both of the above inequalities must be equality, so $|A_i| = |A_1|$ and $\mathcal{C}_1$ must be a non-proper cut class. But then $f(\mathcal{C}_i) \geq f(\mathcal{C}_1)$ and $|\psi(\mathcal{C}_i)| \geq |\psi(\mathcal{C}_1)|$ which is in contradiction with the definition of $\mathcal{C}_1$. So, $B_i = A_1$.

Next, we show that if $f(\mathcal{C}_1) < 2n/3$, then for any $i, j \geq 2$ such that $i \neq j$ we either have $A_i \subseteq A_j$ or $A_j \subseteq A_i$. Since $A_i, A_j \subseteq A_1$, by part (iii) of Lemma 2.6.3, A_i, A_j do not cross. So, if $A_i \not\subseteq A_j$ and $A_i \not\subseteq A_i$ we must have $A_i \cup A_j = V$. Therefore,

$$|A_1| = f(\mathcal{C}_1) \geq \max\{|A_i|, |A_j|\} \geq |\overline{A_1}| + |A_1|/2 = n - |A_1|/2.$$

Therefore, $|A_1| \geq 2n/3$ and we are done.

So, perhaps after renaming assume that

$$A_{2l+2} \subseteq A_{2l} \subseteq \ldots \subseteq A_2 \subseteq A_1.$$

Next, we show that $|A_{2l+2}| \geq |A_1 - 2l$. By part (iv) of Lemma 2.6.3 for any $1 < i < 2l+2$ either $A_i \neq A_{i+1}$ or $A_i \neq A_{i-1}$. Furthermore, if for $1 \leq i < 2l+1$, $A_i = A_{i+1}$, then one of $\mathcal{C}_i, \mathcal{C}_{i+1}$ has two atoms and the other one hast at least 4 atoms. But this means that $|A_{i+2}| \leq |A_i| - 2$. Putting them together, for any $i \geq 1$, either $|A_{i+1}| \leq |A_i| - 1$ or $|A_{i+2}| \leq |A_i| - 2$. So,

$$|A_{2l+2}| \leq |A_1| - 2l$$

But by definition of A_{2l+2}, $A_1 \cup A_{2l+2} = V$. Therefore,

$$|A_1| \geq n - |A_{2l+2}| \geq n - |A_1| + 2l.$$

So $|A_1| \geq n/2 + l$ which completes the proof. □

2.7 Structure of Near Minimum Cuts

In this section we discuss properties of near minimum cuts of a Δ-edge connected graph. Similar to the previous section all of our statements hold for fractional graphs as well but for simplicity of

notations we work with integral graphs.

For $\eta > 0$, a cut $(A, \overline{A})$ is a $(1+\eta)$-near minimum cut if $|\delta(A)| < (1+\eta)\Delta$. In the first part of this section we discuss structure of near minimum cuts for values of $\eta \leq 1/5$. In the second part we characterize larger values of η.

2.7.1 Structure of $(1+\eta)$ near min cuts for small η

We start by providing some basic lemmas. The following lemma proves a generalization of Lemma 2.6.5 to crossing near minimum cuts.

Lemma 2.7.1. *Let $(A, \overline{A})$ and $(B, \overline{B})$ be two crossing cuts of G and let $(A, \overline{A})$ be a $(1+\eta)$ near minimum cut. Then,*

$$\max\{|\delta(A \cap B)|, |\delta(A \cup B)|, |\delta(A - B)|, |\delta(B - A)|\} \leq |\delta(B)| + \eta\Delta.$$

Proof. We prove the lemma only for $A \cap B$; the rest of the cases can be proved similarly.

$$|\delta(A \cap B)| + |\delta(A \cup B)| \leq |\delta(A)| + |\delta(B)| \leq (1+\eta) \cdot \Delta + |\delta(B)|.$$

Since $|\delta(A \cup B)| \geq \Delta$, we have $|\delta(A \cap B)| \leq |\delta(B)| + \eta \cdot \Delta$. This completes the proof of the lemma. □

The following lemma is proved in [Ben97]

Lemma 2.7.2 ([Ben97, Lem 5.3.5])**.** *Let $(A, \overline{A})$ and $(B, \overline{B})$ be two crossing $(1+\eta)$ near minimum cuts of G. Then $|E(A \cap B, A - B)| \geq (1-\eta)\frac{\Delta}{2}$.*

Note that by symmetry it can be derived from the above lemma that

$$|E(A \cap B, B - A)| \geq (1-\eta)\frac{\Delta}{2}$$
$$|E(A \cup B, A - B)| \geq (1-\eta)\frac{\Delta}{2}$$
$$|E(A \cup B, B - A)| \geq (1-\eta)\frac{\Delta}{2}$$

It turns out the representation of near minimum cuts is more complicated than the cactus representation. Unfortunately, Lemma 2.6.6 does not naturally extend to near minimum cuts. In fact, one of our significant contributions is a generalization of Lemma 2.6.6 to the system $(1+\eta)$ near minimum cuts for some absolute constant value of η (see Section 3.2). Next, we describe the polygon representation defined in [Ben95] and then generalized in [BG08] to represent the set of near minimum cuts of each of the cut classes of G.

Definition 2.7.3. *The* polygon representation *of a cut class $\mathcal{C}$ possesses the following properties:*

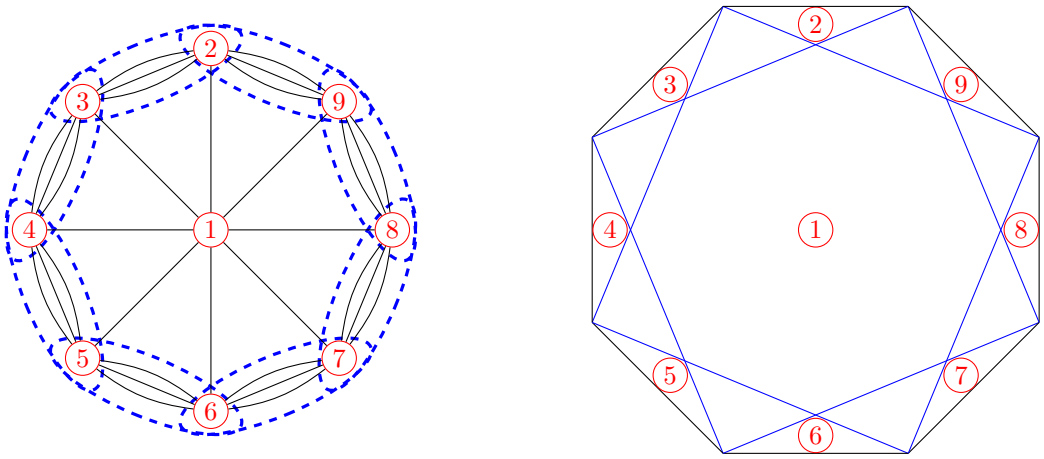

Figure 2.7.8: The left graph shows A 7-edge connected graph. The dashed blue lines show a cut class $\mathcal{C}$ of cuts of value at most 8, i.e., (1+1/7)-near minimum cuts. The right image shows the polygon representation of $\mathcal{C}$. The blue lines in the right image are the representing diagonals. This representation has 8 outside atoms and $\{1\}$ is the only inside atom. The system of near minimum cuts corresponding to sets $(\{2,3\},\overline{\{2,3\}}),\ldots,(\{8,9\},\overline{\{8,9\}}),(\{9,2\},\overline{\{9,2\}})$ shows an 8-cycle for the inside atom $\{1\}$.

- *A* representing polygon *is a regular polygon with a collection of distinguished* representing diagonals*, with all polygon-edges and diagonals drawn by straight lines in the plane. These diagonals divide the polygon into* cells*.*
- *Each atom of $\psi(\mathcal{C})$ is mapped to a (different) cell of this polygon; some cells may contain no atoms.*
- *No cell has more than one incident polygon edge; each cell incident to the polygon boundary contains an atom which we call an* outside *atom. The rest of the atoms are called* inside *atoms.*
- *Each representing diagonal defines a cut, with sides being the union of the atoms contained by cells on each side of the diagonal; The collection of cuts $\mathcal{C}$ is equal to the collection of cuts defined by representing diagonals.*

See Figure 2.7.8 for an example of a polygon representation of a cut class of near minimum cuts.

Benczur [Ben97] show that any cut class of $(1+\eta)$ near minimum cuts for small value of η possesses a polygon representation.

Theorem 2.7.4 ([Ben97, Theorem 5.2.2])**.** *If $\mathcal{C}$ represents a cut class of $(1+\eta)$ near minimum cuts and $\eta \leq 1/5$, then it possesses a polygon representation.*

The proof of above theorem is rather complicated. We refer the interested reader to [Ben97, Ch. 5]. Note that once we have the representation of a single cut class, by Theorem 2.6.11 we can plug it into the tree hierarchy to represent all near minimum cuts. Observe that if $\mathcal{C}$ is a cut class of minimum cuts then every diagonal of the representing polygon is a representing diagonal. Furthermore, we do not have any inside atoms. From above theorem for any cut class $\mathcal{C}$ of $(1+\eta)$ near minimum cuts for $\eta \leq 1/5$, the number of cuts in $\mathcal{C}$ may be significantly smaller than $|\psi(\mathcal{C})| \cdot (|\psi(\mathcal{C})| - 3)$.

Corollary 2.7.5. *For any $\eta \leq 1/5$ the number of cuts in any cut class $\mathcal{C}$ of $(1+\eta)$ near minimum cuts is at most $|\psi(\mathcal{C})| \cdot (|\psi(\mathcal{C})| - 3)$.*

Stronger forms of above theorem are proved in [NNI94].

Before concluding this part we describe more properties of inside atoms. Benczur [Ben97] suggested a simple combinatorial argument to distinguish between inside atoms and outside atoms. To describe this argument we first need to define a k-cycle of cuts.

Definition 2.7.6 (Benczur [Ben97, Definition 5.3.1])**.** *We say that $k \geq 4$ sets $S_i \subset V$, for $1 \leq i \leq k$, form a k-cycle if*

- *B_i crosses both B_{i-1} and B_{i+1};*
- *$B_i \cap B_j = \emptyset$ for $j \neq i-1, i$ or $i+1$; and*
- *$\bigcup_{1 \leq i \leq k} B_i \neq V$.*

See Figure 2.7.8 for an example of a 8-cycle. One of the main differences between a collection of minimum and a collection of near minimum cuts is that minimum cuts do not admit any k-cycle. Benczur in [Ben97] showed that the collection $(1+\eta)$ near minimum cuts does not have a k-cycle for $k \leq 1/\eta$-cycle. In particular, $6/5$ near minimum cuts do not have any k-cycle for $k \leq 5$.

Lemma 2.7.7 (Benczur [Ben97])**.** *For any $\eta < 1$, the set of $(1+\eta)$ near minimum cuts of G does not have any k-cycle for $k \leq 1/\eta$.*

Proof. The proof is based on [BG08, Lem 22]. Suppose there is a cycle of $B_1, \ldots, B_k$ of $(1+\eta)$ near minimum cuts, and let $A = V - \bigcup_{1 \leq i \leq k} B_i$. The following inequality follows from the fact that $B_1, \ldots, B_k$ is a k-cycle.

$$|\delta(A)| + \sum_{i=1}^{k} |\delta(B_i \cap B_{i+1})| \leq \sum_{i=1}^{k} |\delta(B_i)|.$$

In any edge that contributes to the LHS also contributes to the RHS. Furthermore, only edges in $E(B_i \cap B_{i+1}, A)$ contribute more than once to the LHS, but these edge contribute exactly twice to the LHS and the RHS.

Now, we get

$$\Delta \leq |\delta(A)| \leq \sum_{i=1}^{k} (|\delta(B_i)| - |\delta(B_i \cap B_{i+1})|) < \sum_{i=1}^{k} ((1+\eta)\Delta - \Delta) = k\eta.$$

So, $k > 1/\eta$. □

Benczur and Goemans [BG08] show that the set of inside atoms of the polygon representation of a cut class $\mathcal{C}$ are atoms A such that there is a k-cycle that does not intersect

Theorem 2.7.8 ([BG08, Def 4, Prop 5]). *Let $\mathcal{C}$ be a collection of $(1+\eta)$-near minimum cuts for $\eta \leq 1/5$, and $A \in \psi(\mathcal{C})$. Then any $A \in \psi(\mathcal{C})$ is an inside atom iff there exists a k-cycle $B_1, \ldots, B_k$ such that $A \subseteq V - \bigcup_{i=1}^{k} B_i$.*

In the example in Figure 2.7.8, $\{1\}$ is an inside atom. We say k-cycle $B_1, \ldots, B_k$ is for an inside atom A if $A \subseteq V - \bigcup_{i=1}^{k} B_i$. The following technical lemma also proved in [BG08] will be useful later.

Lemma 2.7.9 ([BG08, Lem 12]). *Let $\mathcal{C}$ be a cut class of $(1+\eta)$ near minimum cuts of G for $\eta \leq 1/5$, and let $(S, \overline{S}) \in \mathcal{C}$ contain an inside atom $A \in \psi(\mathcal{C})$. For any k-cycle $B_1, \ldots, B_k$ for A, there exists i such that $B_i \subset S$.*

2.7.2 Structure of α near minimum cuts for large α

In this part we discuss the structure of cuts of value at most α times the minimum cut for values of α that can depend on n. To the best of our knowledge there is no unified representation for values of $\alpha > 6/5$. So, in this part we mainly study the number of these cuts.

We prove the following result due to Karger [Kar93].

Theorem 2.7.10 (Karger [Kar93]). *For any Δ-edge connected graph $G = (V, E)$ and any $\alpha \geq 1$ the number of cuts of value at most $\alpha \cdot \Delta$ is at most $4n^{2\alpha}$.*

Proof. Nash and Williams [NW61] show that any Δ-edge connected graph contains $\Delta/2$ edge disjoint spanning trees. Let $T_1, \ldots, T_{\Delta/2}$ be $\Delta/2$ edge disjoint spanning trees of G. Consider any cut $(S, \overline{S})$ of size at most $\alpha \cdot \Delta$. Therefore, the expected number of edges of $(S, \overline{S})$ in a uniformly random tree among $T_1, \ldots, T_{\Delta/2}$ is at most 2α. Now, by Markov inequality with probability $1/(4\alpha)$ a uniformly random tree has at most 2α edges in $(S, \overline{S})$.

Now, fix a tree T_i, and suppose a set F of edges of T_i are in a cut. We claim that for any $F \subseteq T_i$ this cut is uniquely defined based on edges in F. First, we fix a vertex u on one side of the cut. Then, for every other vertex v we count the number of edges of F in the path from u to v. If this number is even, then v is in the same side of the cut as u, otherwise u is on the other side.

Now we can give a simple procedure to count the number of cuts of size $\Delta\alpha$. First we choose a random spanning tree T_i and then we count all cuts that cut at most 2α edges of T_i. The number of these cuts is at most $2\binom{n-1}{2\alpha}$. In this way we have counted at least $1/(4\alpha)$ fraction of all $\alpha \cdot \Delta$ near minimum cuts. So, the number of $\alpha \cdot \Delta$ near minimum cuts is at most

$$8\alpha \binom{n-1}{2\alpha} \leq 4n^{2\alpha}.$$

□

The above proof is based on [Kar00, Lemma 3.2].

2.8 Random Spanning Trees

Let $G = (V, E)$ be an unweighted graph, and let $\mathcal{T}$ be the set of all spanning trees of G. For non-negative weighted $\lambda : E \to \mathbb{R}_+$ assigned to the edges of G, a λ-*random* spanning tree distribution is a distribution μ where the probability of each tree T is

$$\mathbb{P}_\mu [T] \propto \prod_{e \in E} \lambda_e.$$

Observe that if $\lambda_e = 1$ for all $e \in E$, then μ is exactly the uniform distribution on spanning trees of G. For rational λ_e's, if we replace each edge e with $C\lambda_e$ parallel edges (for an appropriate choice of C), then a uniform spanning tree in the corresponding multigraph gives a λ-random spanning tree of G. Because of this almost all properties of uniform spanning trees naturally generalize to (λ)-random spanning tree distributions. We refer the interested readers to [LP13].

We start this section by stating the The Kirchoeff's matrix tree theorem (see [Bol02]). The weighted Laplacian L of G is defined as follows (see Section 7.2 for properties of this matrix).

$$L(u, v) = \begin{cases} -\lambda_e & e = (u, v) \in E \\ \sum_{e \in \delta(\{u\})} \lambda_e & u = v \\ 0 & \text{otherwise.} \end{cases} \tag{2.8.1}$$

Matrix tree theorem states that for any graph G, $\sum_{T \in \mathcal{T}} \prod_{e \in T} \lambda_e$ is equal to the absolute value of any cofactor of the weighted Laplacian.

Theorem 2.8.1 (Matrix Tree Theorem)**.** *For any graph G any $\lambda : E \to \mathbb{R}_+$ and any $u, v \in V$.*

$$\sum_{T \in \mathcal{T}} \prod_{e \in T} \lambda_e = \det(L_{u,v}) = \det'(L).$$

where by $L_{u,v}$ we mean the minor of L where the u-th row and v-th column are removed, and by $\det'(L)$ we mean the product of the non-zero eigenvalues of L.

One of the important consequence of above theorem is that it gives an efficient algorithm to compute the probability that an edge e is in a λ-random tree T. For this purpose, we can evaluate $\sum_{T \in \mathcal{T}} \prod_{e \in T} \lambda_e$ for both G and $G/\{e\}$ (in which edge e is contracted) using the matrix tree theorem,

$$\mathbb{P}[e \in T] = \frac{\sum_{T \ni e} \prod_{e' \in T} \lambda_{e'}}{\sum_T \prod_{e' \in T} \lambda_{e'}}. \tag{2.8.2}$$

Similarly, we can compute the probability that $e \notin T$,

$$\mathbb{P}[e \notin T] = \frac{\sum_{T \not\ni e} \prod_{e' \in T} \lambda_{e'}}{\sum_T \prod_{e' \in T} \lambda_{e'}} \tag{2.8.3}$$

We remark that an alternative approach for computing $\mathbb{P}[e \in T]$ is to use the fact that $\mathbb{P}[e \in T]$ is equal to λ_e times the effective resistance of e in G treated as an electrical circuit with conductances of edges given by λ (see Subsection 2.8.2 below).

2.8.1 Sampling a λ-Random Tree

There is a host of results (see [Gué83, Kul90, CMN96, Ald90, Bro89, Wil96, KM09] and references therein) on obtaining polynomial-time algorithms for generating a uniform spanning tree, or a λ-random spanning tree. Almost all of them can be easily modified to allow arbitrary λ; however, not all of them still guarantee a polynomial running time for general λ_e's. The problem is that many of these algorithms are based on running a random walk on G, and if there is an exponential gap (in n) between λ_e's then the mixing time of the random walk can also be exponential in n (see Section 7.4 for background on random walks).

So, in this section we describe an iterative approach similar to [Kul90]. The idea is to order the edges $e_1, \ldots, e_m$ of G arbitrarily and process them one by one, deciding probabilistically whether to add a given edge to the final tree or to discard it. More precisely, when we process the j-th edge e_j, we decide to add it to a final spanning tree T with probability p_j being the probability that e_j is in a λ-random tree *conditioned* on the decisions that were made for edges $e_1, \ldots, e_{j-1}$ in earlier iterations. Clearly, this procedure generates a λ-random tree, and its running time is polynomial as long as the computation of the probabilities p_j can be done in polynomial time.

To compute these probabilities efficiently we note that, by (2.8.2) p_1 is easy to compute. Now, if we choose to include e_1 in the tree then:

$$\begin{aligned} p_2 = \mathbb{P}[e_2 \in T | e_1 \in T] &= \frac{\sum_{T' \ni e_1, e_2} \prod_{e \in T'} \lambda_e}{\sum_{T' \ni e_1} \prod_{e \in T'} \lambda_e} \\ &= \frac{\sum_{T' \ni e_1, e_2} \prod_{e \in T' - e_1} \lambda_e}{\sum_{T' \ni e_1} \prod_{e \in T' - e_1} \lambda_e}. \end{aligned}$$

As one can see, the probability that $e_2 \in T$ conditioned on the event that $e_1 \in T$ is equal to the probability that e_2 is in a λ-random tree of a graph obtained from G by contracting the edge e_1. (Note that one can sum up the λ's of the multiple edges formed during the contractions and replace them with one single edge.) Similarly, if we choose to discard e_1, the probability p_2 is equal to the probability that e_2 is in a λ-random tree of a graph obtained from G by removing e_1. In general, p_j is equal to the probability that e_j is included in a λ-random tree of a graph obtained from G by contracting all edges that we have already decided to add to the tree, and deleting all edges that we

have already decided to discard. But, we can compute all p_j's simply using the matrix tree theorem and (2.8.2).

2.8.2 Electrical Networks and λ-Random Trees

We may view the graph $G = (V, E)$ as an electrical network. For any edge $e = (u, v) \in E$ there is a wire with *resistance* $1/\lambda_e$, or equivalently with conductance λ_e between u and v.

We adopt a similar notation as in [SS11] to describe the electrical networks. Consider an arbitrary orientation of the edges of G. Define the matrix $B \in \mathbb{R}^{|E|\times|V|}$ as follows,

$$B(e,v) = \begin{cases} 1 & \text{if } v \text{ is } e\text{'s head} \\ -1 & \text{if } v \text{ is } e\text{'s tail} \\ 0 & \text{otherwise.} \end{cases}$$

Also, let $\Lambda = \mathbb{R}^{|E|\times|E|}$ be the diagonal matrix where $\Lambda(e,e) = \lambda_e$. It follows from (2.8.1) that $L = B^T \Lambda B$.

Let $g : V \to \mathbb{R}_+$ be a function of currents that we injected at the vertices of G, and let $i(e)$ be the current induced in an edge e in the direction of the current and $\phi(v)$ be the potential at a vertex v. We say $i(.)$ and $\phi(.)$ define a valid electrical flow if they satisfy the Kirchhoff's law and the Ohm's law. By Kirchhoff's law the sum of the currents entering a a vertex u is equal to the current that injected at u, i.e., for any vertex u,

$$\sum_{e=(u,v)} i(e) = g(u),$$

or in a matrix form,

$$B^T i = g.$$

By Ohm's Law, the current flow in an edge is equal to the potential difference across its endpoints times the conductance of that edge, for any edge $e = (v, u)$

$$i(e) = \lambda_e (B(e,v)\phi(v) + B(e,u)\phi(u)),$$

or in a matrix form,

$$i = \Lambda B \phi$$

Combining these facts,

$$g = B^T i = B^T \Lambda B \phi = L\phi.$$

Note that the Laplacian matrix L is singular, the first eigenvalue is always zero and the first eigenfunction is the constant function (see Section 7.1 for background on eigenvalues and eigenfunctions). If G is connected, then all other eigenvalues of L are positive. If $g \perp \ker(L)$ then we

have

$$L^\dagger g = \phi, \tag{2.8.4}$$

where $L^\dagger$ is the pseudo-inverse of L (see Section 7.1 for definition).

Suppose we inject one unit of flow at u and extract one from v, i.e., $g = \mathbf{1}_{u\to v} = \mathbf{1}_u - \mathbf{1}_v$. Since $g \in \ker(L)$, the potential at each vertex v' is $L^\dagger \mathbf{1}_{u\to v}(v')$. We can now define the *effective resistance*, $\mathrm{Reff}(u, v)$ between two vertex $u, v \in V$. If $g = \mathbf{1}_{u\to v} = \mathbf{1}_u - \mathbf{1}_v$, that is we inject one unit of flow at u and extract one from v, then $\mathrm{Reff}(u, v) = \phi(u) - \phi(v)$. In other words, by (2.8.4),

$$\mathrm{Reff}(u, v) = \langle \mathbf{1}_{u\to v}, L^\dagger \mathbf{1}_{u\to v} \rangle. \tag{2.8.5}$$

Observe that effective resistance is always non-negative (see [GBS08] for the properties of the effective). The name effective resistance follows from the fact that if we replace the whole network with a wire of conductance $1/\mathrm{Reff}(u, v)$ between u and v, then $\phi(u)$ and $\phi(v)$ remains unchanged.

Say e is oriented from u to v. For any edge $f = \{u', v'\}$ oriented from u' to v' let $i^e(f)$ be the current that flows from u' to v' when a unit current is imposed between the endpoints of e, i.e.,

$$i^e(f) := \lambda_f \cdot \langle \mathbf{1}_{u'\to v'}, L^\dagger \mathbf{1}_{u\to v} \rangle. \tag{2.8.6}$$

For example, $\mathrm{Reff}(u, v) = i^e(e)$.

Now we are ready to describe the connection between electrical networks and λ-random spanning trees.

Proposition 2.8.2. *For any graph G and any $\lambda : E \to \mathbb{R}_+$, and any $e = (u, v) \in E$, a λ-random spanning tree T satisfies the following*

$$\mathbb{P}[e \in T] = i^e(e) = \mathrm{Reff}(u, v) \cdot \lambda_e.$$

Proof. For a function $g : V \to \mathbb{R}$, let gg^T be the matrix where $gg^T(u, v) = g(u) \cdot g(v)$ for all $u, v \in V$. We use the following lemma which is based on Sherman-Morrison Formula (see [GL96, Section 2.1.3]).

Lemma 2.8.3. *For any symmetric matrix $M \in \mathbb{R}^{V\times V}$, and any function $g \perp \ker(M)$,*

$$\det(M + gg^T) = \det(M)(1 + \langle g, M^\dagger g \rangle).$$

Now, let $g = \sqrt{\lambda_e}\mathbf{1}_{u\to v}$. Observe that $L - gg^T$ is the same as the Laplacian matrix of the graph

$G - \{e\}$. Therefore, by (2.8.3),

$$\mathbb{P}[e \in T] = 1 - \mathbb{P}[e \notin T] = 1 - \frac{\det'(L - gg^T)}{\det'(L)} = 1 - \frac{\det'(L)(1 + \langle g, L^\dagger g\rangle)}{\det'(L)}$$
$$= \lambda_e \langle \mathbf{1}_{u\to v}, L^\dagger \mathbf{1}_{u\to v}\rangle = \lambda_e \mathrm{Reff}(u, v),$$

where the last equation follows by (2.8.5). □

The following properties of the function $i^e(.)$ are useful.

Fact 2.8.4. *For any two edges e, f, $i^e(f)\lambda_e = i^f(e)\lambda_f$.*

Proof. Say $e = (u, v), f = (u', v')$. Since $L^\dagger$ is symmetric,

$$i^e(f)\lambda_e = \lambda_e \cdot \lambda_f \langle \mathbf{1}_{u'\to v'}, L^\dagger \mathbf{1}_{u\to v}\rangle = \lambda_e \cdot \lambda_f \langle \mathbf{1}_{u\to v}, L^\dagger \mathbf{1}_{u'\to v'}\rangle = i^f(e)\lambda_f.$$

□

Fact 2.8.5 (Lyons, Peres [LP13, Exercise 4.29])**.** *Let $e, f \in E$ not sharing the same endpoints. Let $i^e_c(.)$ be the function $i^e(.)$ in the graph $G/\{f\}$. Then,*

$$i^e = i^e_c + \frac{i^e(f)}{i^f(f)} i^f.$$

2.8.3 Negative Correlation and Concentration Inequalities.

One of the important consequence of Proposition 2.8.2 is that we can show a negative correlation between the edges of a λ-random spanning tree distributions (see [LP13, Chapter 4] for a proof using random walks).

Lemma 2.8.6. *For any graph G, $\lambda : E \to \mathbb{R}_+$, and any two edges $e, f \in E$, the event that e is in a λ-random spanning tree T is negatively correlated with the event f is in T,*

$$\mathbb{P}[e, f \in T] \le \mathbb{P}[e \in T] \cdot \mathbb{P}[f \in T]. \tag{2.8.7}$$

More generally, for any $F \subseteq E$,

$$\mathbb{P}_T[\forall_{e\in F}, e \in T] \le \prod_{e\in F} \mathbb{P}[e \in T]. \tag{2.8.8}$$

Proof. Here, we prove (2.8.7); equation (2.8.8) can be proved similarly. First, by the Bayes rule, it is sufficient to show that

$$\mathbb{P}[e \in T | f \in T] \le \mathbb{P}[e \in T].$$

Also, $\mathbb{P}[e \in T | f \in T]$ is the same as the probability of e being in the random spanning tree of the graph $G/\{f\}$ (in which edge f is contracted). Contracting an edge f is the same as letting $\lambda_f = \infty$, or equivalently, shortcutting f in the electrical network.

By the Rayleigh monotonicity property, (see [LP13, Section 2]), if we increase the conductance of any edge of G the effective resistance between any pair of vertices only decreases. So, by Proposition 2.8.2,

$$\text{Reff}_{G/\{f\}}(e) \leq \text{Reff}_G(e) \quad \Rightarrow \quad \mathbb{P}[e \in T | f \in T] = \lambda_e \cdot \text{Reff}_{G/\{f\}}(e) \leq \lambda_e \cdot \text{Reff}_G(e) = \mathbb{P}[e \in T].$$

□

One of the direct consequences of above lemma is that the set of edges of G satisfy Chernoff types of bounds, namely, for any subset $F \subseteq E$ the number of edges of F in a random tree T is concentrated around its expectation (we will provide a much stronger characterization in Subsection 2.9.3).

Theorem 2.8.7. *For each edge e, let X_e be an indicator random variable associated with the event $[e \in T]$, where T is a λ-random tree. Also, for any subset F of the edges of G, define $X(F) = \sum_{e \in F} X_e$. Then, we have*

$$\mathbb{P}[X(F) \geq (1+\delta)\mathbb{E}[X(F)]] \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^{\mathbb{E}[X(F)]}.$$

Usually, when we want to obtain such concentration bounds, we prove that the variables $\{X_e\}_e$ are independent and we use the Chernoff bound. Although in our case the variables $\{X_e\}_{e \in E}$ are not independent, they are negatively correlated, and it follows directly from the result of Panconesi and Srinivasan [PS97] that the upper tail part of the Chernoff bound requires only negative correlation and not the full independence of the random variables. So, the above theorem follows directly from [PS97].

2.8.4 λ-Random Trees and Determinantal Measures

In this last part we describe a beautiful result of Burton and Pemantle that derives the exact probability that a set of edges appear in a random spanning tree as a determinant. Recall that $i^e(f)$ is the current that flows across f when a unit current is imposed between the endpoints of e (see (2.8.6)). Burton and Pemantle [BP93] proved the following, known as Transfer-Current Theorem which also gives the exact value of correlation between any two edges in the graph.

Theorem 2.8.8 (Burton, Pemantle [BP93])**.** *For any distinct edges $e_1, \ldots, e_k \in G$, let $M \in \mathbb{R}^{k \times k}$ where $M(i,j) = i^{e_i}(e_j)$. Then,*

$$\mathbb{P}[e_1, \ldots, e_k \in T] = \det(M).$$

In particular, for any two edges e and f, we get a quantitative version of the negative correlation between e and f,

$$\mathbb{P}[e, f \in T] - \mathbb{P}[e \in T] \cdot \mathbb{P}[f \in T] = -i^e(f) \cdot i^f(e) = -\frac{\lambda_e}{\lambda_f} i^e(f)^2.$$

where the last equality follows by Fact 2.8.4. Observe that if $i^e(f) \cdot i^f(e) = 0$, then we can say e is independent of f.

One of the important consequence of the above theorem is that λ-random spanning tree distributions are *determinantal measures* (see [Lyo03] for the definition of determinantal measures). Therefore, they are special cases of *strongly Rayleigh measures*.

2.9 Strongly Rayleigh Measures

In this section we continue discussing properties of λ-random spanning trees. But instead of directly working with these probability distributions we discuss a more general class, called *strongly Rayleigh probability measures*. It turns out that, strongly Rayleigh measures satisfy several properties that will be crucially used in analyzing our algorithm for the traveling salesman problem. While discussing properties of the strongly Rayleigh measures, we provide several examples and we discuss the consequences to the λ-random spanning tree distributions.

Most of the materials of this section are based on a recent work of Borcea, Brändén and Liggett [BBL09] on strongly Rayleigh probability measures. Strongly Rayleigh measures include determinantal measures (in particular uniform and λ-random spanning tree measures), product measures, etc. They also enjoy all the virtues of negative dependence and negative association.

Let E be the ground set of elements with $m = |E|$ (note that we intentionally use E for the ground set of elements, this is because later we use strongly Rayleigh measures to analyze distribution of edges in λ-random spanning trees sampled). A non-negative function $\mu : 2^E \to \mathbb{R}_+$ is a probability measure on subsets of E, if

$$\sum_{S \subseteq E} \mu(S) = 1.$$

Let $\mathcal{B}_E$ be the set of all probability measures on the Boolean algebra 2^E. For an element $e \in E$, let X_e be the indicator random variable for e, and for $S \subseteq E$, let $X_S = \sum_{e \in S} X_e$.

Let $\mathcal{P}_m$ be the set of all multi-affine polynomials in m variables $g(y_e : e \in E)$ with non-negative coefficients such that $g(1, 1, \ldots, 1) = 1$. There is a one-one correspondence between $\mathcal{B}_E$, and $\mathcal{P}_m$: For $\mu \in \mathcal{B}_E$ we may form its generating polynomial, namely $g(y) = \sum_{S \subseteq E} \mu(S) y^S$, where $y^S = \prod_{e \in S} y_e$.

A polynomial $g \in \mathcal{P}_m$ is called *real stable* if $g(y_e : e \in E) \neq 0$ whenever $\mathrm{Im}(y_e) > 0$ for all $e \in E$. For example, this simply implies that a polynomial in one variable is real stable if and only if all its roots are real (this is because if $c \in \mathbb{C}$ is a root of $g \in \mathcal{P}_m$, then so is the conjugate of c). A measure

$\mu \in \mathcal{B}_E$ is called *strongly Rayleigh* if its generating function is real stable. Equivalently, Brändén [Brä07] proved that a multi-affine polynomial $g \in \mathcal{P}_m$ is real stable if and only if

$$\frac{\partial g}{\partial y_e}(x)\frac{\partial g}{y_{e'}}(x) \geq \frac{\partial^2 g}{\partial y_e \partial y_{e'}}(x)g(x),$$

for all $x \in \mathbb{R}^m$, and $e, e' \in E$. As an example, if $x_e = 1$ for all $e \in E$, then $\frac{\partial g}{\partial y_e} = \mathbb{E}[X_e]$. So, above inequality implies $\mathbb{E}[X_e] \cdot \mathbb{E}[X_{e'}] \geq \mathbb{E}[X_e \cdot X'_e]$, i.e., elements are negatively correlated in any strongly Rayleigh measure.

2.9.1 Operations defined on Strongly Rayleigh Measures

First, we describe several operations (Projection, Conditioning and Truncation) that maintain the strongly Rayleigh property.

Definition 2.9.1 (Projection)**.** *For any $\mu \in \mathcal{B}_E$ and $F \subset E$ the projection of μ onto 2^F is the measure μ' obtained from μ by restricting the samples to the subsets of E', i.e.:*

$$\forall S' \subseteq F : \mu'(A) := \sum_{S \subseteq E : S \cap F = S'} \mu(B)$$

Borcea et al. [BBL09] show that any projection of a strongly Rayleigh measures is still a strongly Rayleigh measure. For example, if μ is a uniform measure on the spanning trees of $G = (V, E)$, and $F = \delta(A) \subset E$ is the set of edges in the cut $(A, \overline{A})$, the projection of μ on $2^{\delta(A)}$ is a strongly Rayleigh measure.

Definition 2.9.2 (Conditioning)**.** *For any element $e \in E$, the measure obtained from μ by conditioning on $X_e = 0$ is defined as follows:*

$$\forall S' \subseteq E - \{e\} : \mu'(S') := \frac{\mu(S')}{\sum_{S \subseteq E - \{e\}} \mu(S)}$$

Similarly, we can define the measure obtained by conditioning on $X_e = 1$.

For example, if μ is a uniform measure on the spanning trees of G, the measure obtained by conditioning on $X_e = 0$ for some $e \in E$ is a uniform measure on spanning trees of $G - \{e\}$, so is a strongly Rayleigh measure. Similarly, we can condition on the set of spanning trees that contain all of the edges of a set S, and none of the edges of S'. More generally, for $A \subseteq V$, the measure obtained by conditioning on $\sum_{e \in E(A)} X_e = |A| - 1$ (i.e. having an spanning tree inside A), can be seen as a product of a uniform spanning tree measure on $G[A]$ and a uniform spanning tree measure on G/A (G where all vertices of A are contracted). Such a measure is not a uniform spanning tree measure on E, but is a product of two spanning tree measures, and so is a strongly Rayleigh measure on E.

Borcea et al. proved something stronger, they showed for any integer k, if μ is strongly Rayleigh, then so is μ conditioned on $\sum_e X_e = k$, First we define the truncation of a measure.

Definition 2.9.3 (Truncation)**.** *For any $1 \leq k \leq l \leq |E|$, the* truncation *of μ to $[k,l]$ is the conditional measure $\mu_{k,l}$, where*

$$\forall k \leq |S'| \leq l \colon \mu_{k,l}(S') = \frac{\mu(S')}{\sum_{S \subseteq E : k \leq |S| \leq l} \mu(S)}$$

Borcea et al. proved that if $l - k \leq 1$, the truncation of any strongly Rayleigh measure is still a strongly Rayleigh measure.

Theorem 2.9.4 ([BBL09, Corollary 4.18])**.** *If $\mu \in \mathcal{B}_E$ is a strongly Rayleigh probability measure and $0 \leq p \leq q \leq |E|$ such that $l - k \leq 1$, then $\mu_{k,l}$ is strongly Rayleigh.*

For example, let μ be the uniform measure on the spanning trees of $G = (V, E)$, and $S \subset E$. Let μ' be the projection of μ on S, and let μ'' be the projection of μ on $\overline{S}$. For any $1 \leq k \leq |S|$ such that $k \leq 1$, μ'_k and μ''_{n-1-k} are strongly Rayleigh measures. Moreover, since any spanning tree sampled from μ has exactly $|V| - 1 = n - 1$ edges, $X_S = k$ if and only if $X_{\overline{S}} = n - 1 - k$. So, μ'_k and μ''_{n-1-k} are projections of the same set spanning trees into the complementary sets $S, \overline{S}$.

It is worth noting that random spanning tree distributions are not closed under the projection or truncation operations. So, once we generalize a random spanning tree distribution to a strongly Rayleigh measure, we are allowed to use several properties that we could not use if we restrict our analysis to the random spanning tree distributions.

2.9.2 Properties of Strongly Rayleigh Measures

Next we describe several properties of the strongly Rayleigh measures that are essential in our proofs. We start with the negative association property.

Definition 2.9.5 (Increasing Events and Functions)**.** *An* increasing event*, $\mathcal{A}$, on 2^E is a collection of subsets of E that is closed under upward containment, i.e. if $A \in \mathcal{A}$ and $A \subseteq B \subseteq E$, then $B \in \mathcal{A}$. Similarly, a decreasing event is closed under downward containment. An* increasing function *$f : 2^E \to \mathbb{R}$, is a function where for any $A \subseteq B \subseteq E$, we have $f(A) \leq f(B)$.*

For example, an indicator function of an increasing event is an increasing function. If E is the set of edges of a graph G, then the existence of a Hamiltonian cycle is an increasing event, and the 3-colorability of G is a decreasing event.

A measure $\mu \in \mathcal{B}_E$ is *positively associated* if for any increasing functions $f, g : 2^E \to \mathbb{R}$,

$$\mathbb{E}_\mu [f \cdot g] \geq \mathbb{E}_\mu [f] \cdot \mathbb{E}_\mu [g] . \tag{2.9.1}$$

One of the fundamental inequalities in probability theory is the FKG inequality, named after FortuinKasteleynGinibre. FKG inequality states that for $\mu \in \mathcal{B}_E$, if μ satisfies the *positive lattice condition*, i.e., for any $S, S' \subseteq E$,

$$\mu(S \cap S') \cdot \mu(S \cup S') \geq \mu(S) \cdot \mu(S'),$$

then μ is positively associated. For example, consider an Erdös-Réyni random graph distribution on G, i.e., let μ be the product distribution where each edge $e \in E$ is included with a fixed probability p independent of other edges, and let f be the indicator function that a sample has a Hamiltonian cycle, and g be the indicator function that a sample is *not* 3-colorable. It is an easy exercise that μ satisfies the positive lattice condition, so we can deduce that if we know a sample has a Hamiltonian cycle, it is less likely that it is 3-colorable.

In this thesis we are mainly interested in measures with negative association. The definition of negative association is not simply the inverse of (2.9.1). The reason is that no measure would satisfy such a property: say f and g are indicator functions of a fixed $e \in E$. Then, for all measures $\mu \in \mathcal{B}_E$, $\mathbb{E}[f] \cdot \mathbb{E}[g] \geq \mathbb{E}[f \cdot g]$. Instead, we require that f and g are functions of disjoint subsets of E,

Definition 2.9.6 (Negative Association)**.** *A measure $\mu \in \mathcal{B}_E$ is* negatively associated *or NA if for any increasing functions $f, g : 2^E \to \mathbb{R}$, that depend on* disjoint *sets of edges,*

$$\mathbb{E}_\mu[f] \cdot \mathbb{E}_\mu[g] \geq \mathbb{E}_\mu[f \cdot g]$$

Feder and Mihail [FM92] proved that uniform measures on balanced matroids (and in particular λ-random spanning tree distributions) are negative associated. Borcea et al. in [BBL09] proved that, more generally, strongly Rayleigh measures are negatively associated.

Theorem 2.9.7 ([BBL09])**.** *Strongly Rayleigh measures are negatively associated.*

Observe that negative association is a stronger property compared to negative correlation, since we can prove $e, e' \in E$ are negatively correlated by letting f, g be the indicator functions of edges e, e' in Definition 2.9.6. The following fact is a simple application of negative association.

Fact 2.9.8. *If μ is a λ-random spanning tree distribution on $G = (V, E)$, then for any $S \subset E$, and $p \in \mathbb{R}$ we have*

1. $\forall e \in E - S : \mathbb{E}_\mu\left[X_e \middle| X_S \geq p\right] \leq \mathbb{E}_\mu[X_e]$
2. $\forall e \in E - S : \mathbb{E}_\mu\left[X_e \middle| X_S \leq p\right] \geq \mathbb{E}_\mu[X_e]$

The following corollary is a simple consequence of this:

Corollary 2.9.9. *Let μ be a uniform measure on spanning trees of a graph $G = (V, E)$, $S \subset E$, $s = \mathbb{E}_\mu[X_S]$. Recall that* $\operatorname{rank}(S)$ *be the rank of S in the graphical matroid on G (see Section 2.3). For*

any set $S' \subseteq \overline{S}$*, we have* $\mathbb{E}_\mu[X_{S'}|X_S = 0] \leq \mathbb{E}_\mu[X_{S'}] + s$*, and* $\mathbb{E}_\mu[X_{S'}|X_S = \text{rank}(S)] \geq \mathbb{E}_\mu[X_{S'}] - \text{rank}(S) + s$*.*

Proof. First, since μ is a measure on spanning trees, and each spanning tree has $n-1 = |V|-1$ vertices, we have $\mathbb{E}[X_{\overline{S}}|X_S = 0] = \mathbb{E}[X_{\overline{S}}] + s$, and $\mathbb{E}[X_{\overline{S}}|X_S = \text{rank}(S)] = \mathbb{E}[X_{\overline{S}}] - \text{rank}(S) + s$. Second, since any spanning tree selects at least zero, and at most rank(S) edges from S, the $X_S = 0$ is a decreasing event and $X_S = \text{rank}(S)$ is an increasing event. So, the statement of corollary follows from the negative association property. □

The next important property that we describe in this part is the stochastically dominance property on truncations of strongly Rayleigh measures.

Definition 2.9.10 ([BBL09, Definition 2.14])**.** *For* $\mu, \nu \in \mathcal{B}_E$*, we say* μ stochastically dominates ν *(*$\nu \preceq \mu$*) if for any increasing event* $\mathcal{A}$ *on* 2^E*, we have* $\mu(\mathcal{A}) \geq \nu(\mathcal{A})$*.*

Borcea et al. showed that a truncation of strongly Rayleigh measures is stochastically dominated by a truncation of a larger value:

Theorem 2.9.11 ([BBL09, Theorem 4.19])**.** *For any a strongly Rayleigh probability measure* $\mu \in \mathcal{B}_E$ *and* $1 \leq k \leq |E|$*, if* $\mathbb{P}[X_E = k], \mathbb{P}[X_E = k-1] > 0$*, then* $\mu_{k-1} \preceq \mu_k$*.*

As an example, let μ be the uniform measure on spanning trees of $G = (V, E)$, and $S \subseteq F \subseteq E$. Let μ' be the projection of μ on 2^F. Since μ' is strongly Rayleigh, we have $\mu'_k \preceq \mu'_{k+1}$, for any integer $k \geq 0$, where μ'_k, μ'_{k+1} are well defined. Therefore, for any $l \in \mathbb{R}$

$$\mathbb{P}_{\mu'_{k+1}}[X_S \geq l] \geq \mathbb{P}_{\mu'_k}[X_S \geq l].$$

2.9.3 Properties of Rank Function of Strongly Rayleigh Measures

In this part we describe the ultra log-concavity (ULC) property of the rank function of strongly Rayleigh measures. Recall that in Subsection 2.8.3 we proved the upper-tail of Chernoff-bound for λ-random spanning tree distributions on any subset of edges using the negative correlation property. In this section we show that there is a direct way to prove all Chernoff types of bounds for λ-random spanning tree distributions.

Let $\mu \in \mathcal{B}_E$ be a strongly Rayleigh measure. The *rank sequence* of μ is the sequence

$$\mathbb{P}[X_E = 0], \mathbb{P}[X_E = 1], \ldots, \mathbb{P}[X_E = m].$$

Let $g(y)$ be the generating polynomial of μ. The *diagonal specialization* of μ, $\bar{g}(.)$ is a univariate polynomial obtained by pretending $g(.)$ as a univariate polynomial (i.e., considering $g(y, y, \ldots, y)$). Observe that $\bar{g}(.)$ is the generating polynomial of the rank sequence of μ. If $g(c) = 0$ for $c \in \mathbb{C}$, then $g(c, c, \ldots, c) = 0$. So, if $g(.)$ is a real stable polynomial then so is $\bar{g}$. Since a univariate polynomial

with real coefficients is stable if and only if all of its roots are real, $\bar{g}(.)$ is a polynomial with real roots.

Generating polynomials of probability distributions with real root are very well studied in the literature. If $\bar{g}(.)$ is a univariate polynomial of degree m with real roots, then coefficients of $\bar{g}(.)$ are corresponds to the probability density function of the convolution of a set of m independent Bernoulli random variables. In other words, there are m independent Bernoulli random variables $B_1, \ldots, B_m$ with success probabilities $p_1, \ldots, p_m \in [0, 1]$ such that the probability that exactly k variables succeed is the coefficient of y^k in $\bar{g}(.)$.

Fact 2.9.12. *[BBL09, Pit97] The rank sequence of a strongly Rayleigh measure is the probability distribution of the number of successes in m independent trials for some sequence of success probabilities $p_1, \ldots, p_m \in [0, 1]$.*

Now, if μ is a λ-random spanning tree measure, and μ' is the projection of μ on $S \subseteq E$, then the rank sequence of μ', equivalently the distribution of X_S, is the same as the distribution of $|S|$ independent bernoulli random variables. So, all Chernoff bounds that hold for independent bernoulli random variables hold for X_S as well.

The distribution of the number of successes of m independent trials is well studied in the literature [Dar64, Hoe56, Gle75, Wan93, Pit97]. Dorrach [Dar64] proved that such a distribution is unimodal (i.e., it has a single mode), and the mode differs from the mean by less than 1. Recall that the mode is the value at which the probability mass function takes its maximum value. Moreover, by Newton's inequality [HLP52], a sum of independent bernoulli random variables is an Ultra Log-concave distribution.

Definition 2.9.13 (Ultra Log Concavity [BBL09, Definition 2.8])**.** *A real sequence $\{a_k\}_{k=0}^m$ is log-concave if $a_k^2 \geq a_{k-1} \cdot a_{k+1}$ for all $1 \leq k \leq m-1$, and it is said to have no internal zeros if the indices of its non-zero terms form an interval (of non-negative integers). We say that a non-negative sequence $\{a_k\}_{k=0}^m$ is* ULC *(ultra log-concave) if the sequence $\{a_k / \binom{m}{k}\}_{k=0}^m$ is log-concave and have no internal zeros.*

The following proposition follows,

Proposition 2.9.14 ([HLP52, Dar64, BBL09])**.** *The rank sequence of any strongly Rayleigh measure is ULC, unimodal, and its mode differs from the mean by less than 1.*

Suppose $\mu \in \mathcal{B}_E$ is a strongly Rayleigh measure, and let $\mathbb{E}[X_E] = p$ such that $k < p < k+1$. In the rest of this section we would like to lower bound $\mathbb{P}[X_e = k]$ and $\mathbb{P}[X_e = k+1]$ with a function of p, k that is independent of $|E|$.

For $p \in [0, m]$, let $\mathcal{B}_m(p)$ be the set of all probability distribution of the number of successes in m independent trials such that the expected number of successes is p. For any integer $k \geq 0$ and

any $p \in [0, m]$, let

$$\mathrm{Ber}(p, k) := \min_{m, \mu \in \mathcal{B}_m(p)} \mathbb{P}_\mu \left[\text{exactly } k \text{ trials succeed}\right]. \tag{2.9.2}$$

First, observe that if $|k - p| \geq 1$, then the distribution with $m = \lceil p \rceil$ trials such that $\lfloor p \rfloor$ of Bernoullis succeed with probability 1 (and the only possible last one succeed with probability $p - \lfloor p \rfloor$) implies $\mathrm{Ber}(p, r) = 0$. Therefore, we assume $|k - p| < 1$. Let $\mu^* \in \mathcal{B}_m(p)$ be a minimizer of $\mathrm{Ber}(p, k)$. Hoeffding [Hoe56, Corollary 2.1] showed that Bernoullis take only one of three different success probabilities, $0, x, 1$, in μ^*, where x is any number in $[0, 1]$. Since μ^* is allowed to have any arbitrary number of trials $m \geq 0$, we can further assume that none of the success probabilities in μ^* is 0. Let

$$\mathrm{Bin}(p, k) := \min_{\substack{m \geq k \\ m \geq p}} \binom{m}{k} (p/m)^k (1 - p/m)^{m-k}.$$

Let $0 \leq l^* \leq \min\{p, k\}$ be the number of Bernoullis in μ^* that have success probability 1. Then, $\mathrm{Ber}(p, k) = \mathrm{Bin}(p - l^*, k - l^*)$. So, for $k - 1 < p < k + 1$,

$$\mathrm{Ber}(p, k) = \min\{\mathrm{Bin}(p, k), \mathrm{Bin}(p - 1, k - 1), \ldots, \mathrm{Bin}(p - k + \mathbb{I}\left[p < k\right], \mathbb{I}\left[p < k\right])\}. \tag{2.9.3}$$

So, to lower bound $\mathrm{Ber}(p, k)$, it is sufficient find a lower bound $\mathrm{Bin}(p - l, k - l)$ for any integer $0 \leq l \leq k - \mathbb{I}\left[p < k\right]$.

Lemma 2.9.15. *For any integer $k \geq 1$ and $k - 1 < p \leq k$,*

$$\mathrm{Bin}(p, k) \geq \left(\frac{p}{k}\right)^k \cdot e^{-p},$$

and for any $k < p < k + 1$,

$$\mathrm{Bin}(p, k) \geq \left(\frac{p}{k}\right)^k \cdot \min\left\{\left(1 - \frac{p}{m^*}\right)^{m^* - k}, \left(1 - \frac{p}{m^* + 1}\right)^{m^* + 1}\right\}.$$

where $m^* = \lfloor p + 1 \rfloor$.

Proof. Since, $k \geq 1$, for any $m \geq \max\{p, k\}$,

$$\binom{m}{k} (p/m)^k (1 - p/m)^{m-k} \geq \frac{m^k}{k^k} \cdot \frac{p^k}{m^k} (1 - p/m)^{m-k}.$$

Since m must be an integer, if $m = p$, then $p = k$. So, $\mathrm{Bin}(p, k) = 1$ and we are done. Otherwise, if $m = m^*$, we are done by above equation. So, assume $m \geq m^* + 1$.

The lemma follows from the fact that $(1 - p/x)^{x-k}$ is an increasing function of x for $x \geq k$ when $p \leq k$ and $(1 - p/x)^x$ is an increasing function of x for $x \geq p$.

First, assume that $p \leq k$. For all $x \geq k$,

$$\frac{\partial(1-p/x)^{x-k}}{\partial x} = (1-p/x)^{x-k} \cdot \Big(\log(1-p/x) + \frac{p(x-k)}{x^2(1-p/x)} \Big)$$

We show that the RHS is non-positive. Computing the Maclaurin series of

$$\log(1-p/x) + \frac{p(x-k)}{x^2(1-p/x)} = -\sum_{i=1}^{\infty} \frac{1}{i}(p/x)^i + \sum_{i=1}^{\infty}(p/x)^i - \frac{k}{x}\sum_{i=1}^{\infty}(p/x)^i \leq \sum_{i=2}^{\infty}(p/x)^i - \frac{p}{x}\sum_{i=1}^{\infty}(p/x)^i = 0.$$

where the last inequality uses $p \leq k$. This completes the first conclusion of the lemma. To prove the second conclusion we just need to let $k = 0$ in the above argument and we get

$$\log(1-p/x) + \frac{px}{x^2(1-p/x)} = -\sum i = 1^{\infty} \frac{1}{i}(p/x)^i + \sum_{i=1}^{\infty}(p/x)^i \geq 0.$$

This completes the proof of lemma. □

Now, we are ready to lower bound $\mathrm{Ber}(p,k)$.

Proposition 2.9.16. *For any integer k, and $k-1 < p \leq k$,*

$$\mathrm{Ber}(p,k) \geq (p-k+1) \cdot e^{-p}.$$

Otherwise, if $k \leq p < k+1$,

$$\mathrm{Ber}(p,k) \geq \min\big\{1 - \frac{p}{k+1}, \big(1 - \frac{p}{k+2}\big)^{k+2}\big\}$$

Proof. The proof is a simple algebraic manipulation, and follows from the following inequalities, First, for $p \geq 1$ and $m^* = \lfloor p+1 \rfloor$,

$$1 - \frac{p}{m^*} \leq 1 - \frac{p-1}{m^*-1} \quad \text{and} \quad 1 - \frac{p}{m^*+1} \leq 1 - \frac{p-1}{m^*}.$$

Second, if $1 \leq p < k$, then

$$\Big(\frac{p-1}{k-1}\Big)^{k-1} \leq (p/k)^k.$$

Otherwise if $p \geq k \geq 1$, then $(p/k)^k \geq 1$. □

Next we describe a simple example to show an application of the above proposition.

Example 2.9.17. *Let μ be a λ-random spanning tree distributions, and let $v \in V$ such that $\mathbb{E}[X(\delta(v)) = 2]$. Then,*

$$\mathbb{P}[X(\delta(v)) = 2] \geq \mathrm{Ber}(2,2) \geq \min\{(1-2/3), (1-2/4)^4\} = 1/16.$$

In other words, v will have an even degree in a random tree T with probability at least $1/16$.

If we do a more careful analysis we can improve $1/16$ to $1/e$. Roughly speaking, we should first use the fact that in any spanning tree T the degree of v is at least 1. So, indeed $\mathbb{P}[X(\delta(v)) = 2] \geq \mathrm{Ber}(1,1)$. Furthermore, the minimizer of $\mathrm{Ber}(1,1)$ is the poisson distribution of rate 1.

In the final example we show that the bound $1/e$ in the above example (that we did not prove rigorously) is tight.

Example 2.9.18. *Let G be a complete graph and let μ be a uniform spanning tree on G. By symmetry, the expected degree of each vertex is $2(1-1/n)$. Next, we show that the degree distribution of each vertex is essentially one plus a poisson distribution of rate 1 when n goes to infinity.*

Using Prüfer code there is a one-to-one correspondence between the spanning trees of G and all sequences of length $n-2$ of vertices of V, where the degree of a vertex v in a tree T is the number of occurrences of v in the corresponding sequence plus one. But the distribution of the number of occurrences of v in a uniformly random sequence of length $n-2$ of V is essentially a poisson distribution. More precisely, the probability that v appears exactly once is

$$\binom{n-2}{1} \cdot \frac{1}{n} \cdot \left(1 - \frac{1}{n}\right)^{n-3} \approx (1 - 1/n)^n \approx 1/e.$$

where the approximations become equality when $n \to \infty$. So the degree of each vertex is 2 with probability at most $1/e$.

Chapter 3

New Machineries

3.1 Rounding by Sampling Method

In this section we describe the rounding by sampling method. We describe the approach by rounding feasible solutions of an LP relaxation of TSP (see LP (2.4.1)). Let $G = (V, E, \mathbf{x})$ be the underlying graph of a feasible solution $\mathbf{x}$ of LP (2.4.1).

Let us first describe the result of applying the classical randomized rounding method of Raghavan and Thompson [RT87]. In this method we construct a new H by independently rounding each edge of the LP solution. More precisely, for each edge e include e in H with probability x_e independent of other edges. Let X_e be a random variable indicating that $e \in H$. The two main properties of the independent randomized rounding are the following.

i) Any linear function of variables X_e's will remain the same in expectation, i.e., for any $f\colon E \to \mathbb{R}$,

$$\mathbb{E}\left[\sum_{e\in E} f(e)X_e\right] = \sum_{e\in E} f(e)x_e.$$

For example, $\mathbb{E}[c(H)] = c(x)$.

ii) Since the variables X_e for $e \in E$ are independent, we can use strong concentration bounds such as Chernoff bounds to argue that any Lipschitz function of these indicator variables is concentrated around its expected value.

However, this method does not preserve combinatorial properties of the LP solution. Although G is fractionally 2-edge connected, H may be disconnected with high probability.

Let us provide a concrete example. Let G be a complete graph where $x_{u,v} = 2/(n-1)$ and $c(u, v) = 1$ for all $u, v \in V$. It is easy to see that, for each $v \in V$, the degree of v in H is zero with

probability

$$(1-2/(n-1)^{n-1} \approx \exp(-2).$$

So, with high probability, H has $n \cdot \exp(-2)$ degree zero vertices. So, not only is H disconnected, it has $\Theta(n)$ many connected components.

Now, let us describe the rounding by sampling method. Our goal is to provide a rounding method that preserves the underlying combinatorial structure of $\mathbf{x}$ while satisfying the two main properties of the independent randomized rounding method as much as possible. The first observation is that we can write $\mathbf{x}$ as a point in an integral polytope, in this case the spanning tree polytope. In Fact 2.4.1 we showed that for any feasible solution $\mathbf{x}$ of LP (2.4.1), the vector

$$\mathbf{z} = (1-1/n)\mathbf{x} \tag{3.1.1}$$

is in LP (2.4.3). Any feasible point of a polytope can be written as a convex combination of vertices of that polytope (see Theorem 2.2.1). So, we can write $\mathbf{z}$ as a convex combination of vertices of the spanning tree polytope. However, we know that the vertices of the spanning tree polytope are integral spanning trees of our graph G. So, we can write,

$$\mathbf{z} = \alpha_1 T_1 + \ldots + \alpha_k T_k.$$

where $T_1, \ldots, T_k$ are integral spanning trees of G. Any convex-combination defines a distribution. Therefore, we can define a distribution μ, where $\mathbb{P}_{T\sim\mu}[T = T_i] = \alpha_i$. Now, we can round the solution $\mathbf{x}$ simply by choosing a random spanning tree from μ. Observe that by definition μ preserves the marginal probabilities imposed by $\mathbf{z}$, $\mathbb{P}_{T\sim\mu}[e \in T] = z_e$. Therefore, the quantitative properties are preserved in expectation: for any function $f : V \to \mathbb{R}$,

$$\mathbb{P}_\mu \left[\sum_{e\in E} f(e)X_e\right] = \sum_{e\in E} f(e)z_e. \tag{3.1.2}$$

For example,

$$\mathbb{E}[c(T)] = c(\mathbf{z}) = (1-1/n)c(\mathbf{x}).$$

Furthermore, unlike the independent randomized rounding method, the rounded solution is always *connected*.

An, Kleinberg and Shmoys [AKS12] used the above simple idea to design an improved approximation algorithm for the TSP path problem. The Algorithm 1 for the online stochastic matching problem applies this idea to the matching polytope. Also, in a joint work with Laekhunakit and Singh [LOS12], we use this idea to design an approximation algorithm for the minimum strongly connected subgraph problem. Unfortunately, it turns out that this idea is not enough to break the 3/2 approximation algorithm of Christofides [Chr76] even on a Graphic metric. If we write the

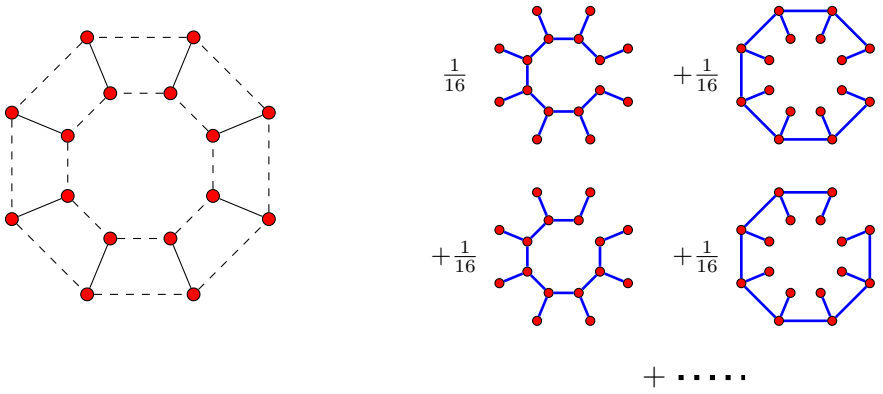

Figure 3.1.1: The left graph shows a feasible solution of LP (2.4.1) for an instance of Graphic TSP. The solid edges have fraction 1 and the dashed edges have fraction 1/2, and the cost of each edge is 1. In the right we write the vector $\mathbf{z} = (1 - 1/n)\mathbf{x}$ as a convex combination of spanning trees. Note that the cost of the minimum cost matching on the odd degree vertices of any of the spanning trees in the support of the distribution is 7, approximately half of $c(\mathbf{x})$.

fractional solution for the family graphs in the left graph of Figure 3.1.1 as a convex combination of spanning trees as shown in the right, then as n goes to infinity, the cost of the minimum cost perfect matching on the odd degree vertices of any spanning tree in the support of the distribution converges to $c(\mathbf{x})/2$.

Let us give some intuitions for the above failure. Roughly speaking, the reason is that the simple rounding by sampling method does not satisfy property (ii) of the independent randomized rounding method. Using the Chernoff bound, one can show that, in a sample from the independent randomized rounding method, the degree of every vertex is 2 with constant probability. On the other hand, since we chose μ in the above argument arbitrarily, we don't have any bound on the correlation between different edges. Although the expected degree of every vertex in a sample tree is $2(1 - 1/n)$, in the above example, almost all vertices have odd degree with high probability. Therefore, the cost of the minimum matching on the odd degree vertices of the tree is about $n/2$.

In summary, although the rounding by sampling method promises that the rounded solution is connected, it may lose the parity of the degree of vertices or other structural properties of LP solution. In the rest of this section we show that if we carefully choose the distribution μ, we are guaranteed to preserve (ii) and indeed almost all advantages of the independent randomized rounding method.

There are many ways to write a feasible point of a polytope as a convex combination of its vertices. Our idea is to use a distribution that maximizes the randomness while preserving the marginal probability of the edges. Roughly speaking, we don't want to enforce any additional artificial structure when writing $\mathbf{z}$ as a convex combination of spanning trees. More formally, we write $\mathbf{z}$ as a distribution of spanning trees that has the maximum possible *entropy* among all distributions that

preserve marginal probability of edges. Asadpour and Saberi first studied and used the *maximum entropy rounding* scheme for sampling a random matching in a bipartite graph with given marginal probabilities [AS07, AS09].

3.1.1 Maximum Entropy Rounding by Sampling Method

Let $\mathcal{T}$ be the collection of all the spanning trees of $G = (V, E)$. Recall that the entropy of a probability distribution $p : \mathcal{T} \to \mathbb{R}_+$, is simply $\sum_{T\in\mathcal{T}} -p(T)\log(p(T))$. The maximum entropy distribution $p^*(\cdot)$ with respect to given marginal probabilities z is the optimum solution of the following convex program (CP):

$$\begin{aligned} \inf \quad & \sum_{T\in\mathcal{T}} p(T)\log p(T) & & \\ \text{subject to} \quad & \sum_{T\ni e} p(T) = z_e & \forall e \in E, & \\ & p(T) \geq 0 & \forall T \in \mathcal{T}. & \end{aligned} \tag{3.1.3}$$

The above convex program is feasible whenever $\mathbf{z}$ belongs to the spanning tree polytope P defined on $G = (V, E)$. As the objective function is bounded and the feasible region is compact (closed and bounded), the infimum is attained and there exists an optimum solution $p^*(\cdot)$. Furthermore, since the objective function is strictly convex, this maximum entropy distribution $p^*(\cdot)$ is unique.

The value $p^*(T)$ determines the probability of sampling any tree T in the maximum entropy rounding scheme. Note that it is implicit in the constraints of this convex program that, for any feasible solution $p(.)$, we have $\sum_T p(T) = 1$ since

$$n - 1 = \sum_{e\in E} z_e = \sum_{e\in E}\sum_{T\ni e} p(T) = (n-1)\sum_T p(T).$$

Let $\mathsf{OPT}_{\text{Ent}}$ denote the optimum value of convex program (3.1.3). Observe that if we remove the equality constraint the optimum distribution of the above convex program is just the uniform spanning tree distribution. Since any graph has at most n^{n-2} spanning trees [Cay89], $\mathsf{OPT}_{\text{Ent}}$ always satisfies

$$\mathsf{OPT}_{\text{Ent}} \geq \log(1/|\mathcal{T}|) \geq -\log(n^{n-2}) \geq -n\log n. \tag{3.1.4}$$

We now want to show that, if we assume that the vector $\mathbf{z}$ is in the *relative interior* of the spanning tree polytope of G then $p^*(T) > 0$ for every $T \in \mathcal{T}$ and $p^*(T)$ admits a simple exponential formula (see Theorem 3.1.1 below). Note that the vector $\mathbf{z}^*$ obtained from the LP relaxation of the ATSP indeed satisfies this assumption (see Fact 2.4.1).

For this purpose, we write the Lagrange dual to CP (3.1.3), see for example [Nem05]. For every $e \in E$, we associate a Lagrange multiplier δ_e to the constraint corresponding to the marginal

probability z_e, and define the Lagrange function by

$$L(p,\delta) = \sum_{T\in\mathcal{T}} p(T)\log p(T) - \sum_{e\in E}\delta_e\left(\sum_{T\ni e} p(T) - z_e\right).$$

This can also be written as:

$$L(p,\delta) = \sum_{e\in E}\delta_e z_e + \sum_{T\in\mathcal{T}}\left(p(T)\log p(T) - p(T)\sum_{e\in T}\delta_e\right).$$

The Lagrange dual to CP (3.1.3) is now

$$\sup_{\delta}\inf_{p\geq 0} L(p,\delta). \tag{3.1.5}$$

The inner infimum in this dual is easy to solve. As the contributions of the $p(T)$'s are separable, we have that, for every $T\in\mathcal{T}$, $p(T)$ must minimize the convex function

$$p(T)\log p(T) - p(T)\delta(T),$$

where, as usual, $\delta(T) = \sum_{e\in T}\delta_e$. Taking derivatives, we derive that

$$0 = 1 + \log p(T) - \delta(T),$$

or

$$p(T) = e^{\delta(T)-1}. \tag{3.1.6}$$

Thus,

$$\inf_{p\geq 0} L(p,\delta) = \sum_{e\in E}\delta_e z_e - \sum_{T\in\mathcal{T}} e^{\delta(T)-1}.$$

Using the change of variables $\gamma_e = \delta_e - \frac{1}{n-1}$ for $e\in E$, the Lagrange dual (3.1.5) can therefore be rewritten as

$$\sup_{\gamma}\left[1 + \sum_{e\in E} z_e\gamma_e - \sum_{T\in\mathcal{T}} e^{\gamma(T)}\right]. \tag{3.1.7}$$

Our assumption that the vector $\mathbf{z}$ is in the relative interior of the spanning tree polytope implies that z can be expressed as a convex combination of all spanning trees in $\mathcal{T}$ such that the coefficient corresponding to any spanning tree is positive. But this means that there is a point $p(.)$ in the relative interior of program (3.1.3). So, the convex program (3.1.3) satisfies the Slater's condition. This implies that the sup in (3.1.7) is attained by some vector γ^*, and the Lagrange dual value equals the optimum value $\mathsf{OPT}_{\text{Ent}}$ of our convex program (see Section 2.2 for background). Furthermore, we have that the (unique) primal optimum solution p^* and any dual optimum solution γ^* must

satisfy

$$L(p, \gamma^*) \geq L(p^*, \gamma^*) \geq L(p^*, \gamma), \tag{3.1.8}$$

for any $p \geq 0$ and any γ, where we have implicitly redefined L due to our change of variables from δ to γ. Therefore, p^* is the unique minimizer of $L(p, \gamma^*)$ and from (3.1.6), we have that

$$p^*(T) = e^{\gamma^*(T)}. \tag{3.1.9}$$

Observe that the above distribution is indeed a λ-random spanning tree distribution for $\lambda_e = e^{\gamma_e^*}$ (see Section 2.8 for background). Summarizing, the following theorem holds.

Theorem 3.1.1. *Given a vector $\mathbf{z}$ in the relative interior of the spanning tree polytope on $G = (V, E)$, there exist λ_e for all $e \in E$ such that if we sample a spanning tree T of G according to $p^*(T) := \prod_{e \in T} \lambda_e$ then $\mathbb{P}[e \in T] = z_e$ for every $e \in E$.*

It is worth noting that the requirement that z is in the relative interior of the spanning tree polytope (as opposed to being just in this polytope) is necessary (the fact that being in the spanning tree polytope was not sufficient had been observed before, see [LP13, Exercise 4.19]). Let G be a triangle and $\mathbf{z}$ be the vector $(\frac{1}{2}, \frac{1}{2}, 1)$. In this case, $\mathbf{z}$ is in the polytope (but not in its relative interior) and there are no λ_e^*'s that would satisfy the statement of the theorem (however, one can get arbitrarily close to z_e for all $e \in E$).

One we have a random spanning tree distribution we can use all properties of them that we discussed in Section 2.8 and Section 2.9. For example, suppose $\mathbf{z}$ is a fractional spanning tree obtained from a feasible solution of Held-Karp relaxation for TSP (2.4.1), and let $\lambda : E \to \mathbb{R}_+$ be the corresponding λ-random spanning tree distribution that preserves z_e as the marginal probability of any edge e. Since for any $v \in V$,

$$\mathbb{E}[|T \cap \delta(v)|] = 2(1 - 1/n),$$

by Example 2.9.17 a constant fraction of the vertices of T have an even degree.

In the next example we describe a maximum entropy distribution of spanning trees for a given marginal vector $\mathbf{z}$.

Example 3.1.2. *Consider the family of fractional spanning trees illustrated at the left of Figure 3.1.2 for a large n. The solid edges have fraction $1 - 1/n$ and the dashed edges have fraction $(1 - 1/n)/2$. By symmetry, the weights λ_e of the solid edges are the same, and λ_e of the dashed edges are the same. Let us normalize the weights such that $\lambda_e = 1$ for all dashed edges. Then, the weights of the solid edges will be $n - 1.5$ (we verified numerically).*

Note that each spanning tree has a non-zero probability. But, it turns out that with high probability a random sample has a very nice shape. First, since the solid edges have a large weight, with high

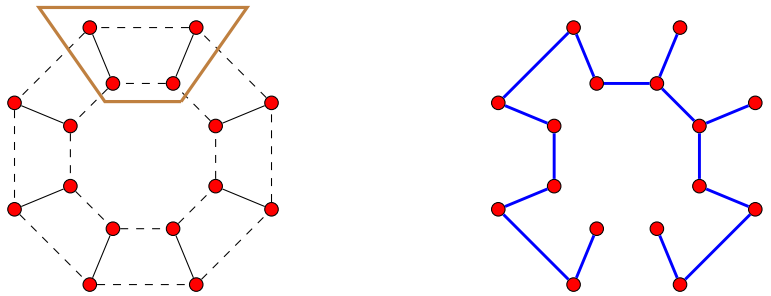

Figure 3.1.2: The left graph shows a fractional spanning tree obtained from a feasible solution of LP (2.4.1), where the solid edges have fraction $1-1/n$, and the dashed edges have fraction $(1-1/n)/2$. In the maximum entropy distribution the λ value of the solid edges is about n times the dashed edges. In the right we show a sample from the corresponding λ-random spanning tree distribution.

probability all of them (except possibly a constant number) will be in a random spanning tree. We say two dashed edges are paired, if their endpoints are matched by two solid edges (a paired dashed edges are shown in the left of Figure 3.1.2 inside a trapezoid). Let e, e' be a paired dashed edges; we show a random tree T has exactly one of them with high probability. Suppose the endpoints of e, e' are matched by the edges f, f'. Since by the union bound $f, f' \in T$ with probability $1 - O(1/n)$ the probability that both of e, e' are in T is $O(1/n)$. On the other hand, the probability that none of e, e' is in T is $O(1/n)$, because any spanning tree contains at least one edge of each (except possibly one) of the paired edges. So, with high probability T contains exactly one of e, e'. Furthermore, since $\lambda_e = \lambda'_e$, T will choose one of the uniformly at random. We shown a random spanning tree with these properties at the right of Figure 3.1.2

Now suppose we have an instance of Graphic-TSP where $c(e) = 1$ for all edges of our graph. Recall that this graph is a tight example for Christofides' algorithm, see Figure 2.5.3, and the rounding by sampling method without exploiting the maximum entropy distribution also gives a $3/2$ approximation, see Figure 3.1.1. It is an instructive exercise to show that the cost of the minimum perfect matching on odd degree vertices of the a λ-random spanning tree is $n/4 + O(1)$ with high probability. So, maximum entropy rounding by sampling method has a $5/4$-approximation ratio on this family of instances.

As a final remark observe that one can study maximum entropy distributions of the convex hull P of any set of discrete objects on a ground set of elements. Above analysis shows that the maximum entropy distribution is always a production distribution, i.e., one can assign non-negative weights to the elements of the ground set such that the probability of each of the objects in the maximum entropy distribution preserving a given marginal vector in the interior of P is proportional to the product of the weight of its elements. In order to use these distributions in computation first we need to compute the weight of the ground elements and second we need to be able to sample efficiently from the corresponding product distribution. Interestingly, very recently, Singh and Vishnoi [SV13] show

that these two problems are equivalent (see more details in the next section). In Subsection 2.8.1 we showed how to efficiently sample from a λ-random distribution of spanning trees. In the next section we design an efficient algorithm to approximately find the weight of the edges in the maximum entropy distribution.

3.1.2 Computation of Maximum Entropy Distribution

In this part we design an algorithm to find a λ-random spanning tree distribution that preserves the marginal probability of all the edges within multiplicative error of $1+\epsilon$. Our algorithm runs in time polynomial in $n, 1/\epsilon, -\log z_{\min}$ where $z_{\min} = \min_{e \in E} z_e$ is the smallest non-zero value assigned to the edges. Note that if $\mathbf{z} = (1-1/n)\mathbf{x}$ for $\mathbf{x}$ being an extreme point solution of (2.4.1) then $z_{\min} \geq 2^{-n\log(n)}$.

Theorem 3.1.3. *Given $\mathbf{z}$ in the relative interior of the spanning tree polytope of $G=(V,E)$. For any $e^{-n^2/2} < \epsilon \leq 1/4$, values $\tilde{\gamma}_e$ for all $e \in E$ can be found, so that if we let $\tilde{\lambda}_e = \exp(\tilde{\gamma}_e)$ for all $e \in E$, then the corresponding $\tilde{\lambda}$-random spanning tree distribution, $\tilde{\mu}$, satisfies*

$$\sum_{T \in \mathcal{T}: T \ni e} \mathbb{P}_{\tilde{\mu}}[T] \leq (1+\varepsilon) z_e, \quad \forall e \in E,$$

i.e., the marginals are approximately preserved. Furthermore, the running time is polynomial in $n = |V|$, $-\log z_{\min}$ and $\log(1/\epsilon)$.

Very recently, Singh and Vishnoi [SV13] generalized and improved the above theorem; they show that for any family of discrete objects, $\mathcal{M}$, and any given marginal probability vector in the interior of the convex hull of $\mathcal{M}$, one can efficiently compute the approximate weight of the ground elements in the maximum entropy distribution if and only if there is an efficient algorithm that approximates the weighted sum of all the objects for any given weights, i.e., an efficient algorithm that approximates $\sum_{M \in \mathcal{M}} \exp(\gamma(M))$ for any vector γ. For example, since there is an efficient algorithm that approximates the weighted sum of all perfect matchings a bipartite graph with respect to given weights γ, [JSV04], one can approximately compute the maximum entropy distribution of the perfect matchings of any bipartite graph with respect to any given marginals in the interior of the perfect matching polytope (see [SV13] for more details).

In the rest of this section we prove the above theorem. We will use the Ellipsoid method, Theorem 2.2.2, so we just need to provide a separating hyperplane oracle, a polynomial in n bound on the radius of a ball that contains our polytope, and an inversely polynomial in n bound on the radius of a ball in the interior of our polytope. In [AGM+10] we also provide a combinatorial algorithm to approximate the maximum entropy distribution that we do not include in this thesis; we refer the interested reader to [AGM+10].

First, we show that the optimum value of the following convex program is the same as the

optimum value of the original dual program (3.1.7).

$$\begin{aligned} \sup_{\gamma} \quad & \sum_{e} z_e \gamma_e, \\ \text{subject to} \quad & \sum_{T \ni e} e^{\gamma(T)} \leq z_e \quad \forall e \in E \end{aligned} \tag{3.1.10}$$

This is because on one hand for any vector γ that is a feasible solution of above program,

$$1+\sum_{e\in E} z_e\gamma_e - \sum_{T\in\mathcal{T}} e^{\gamma(T)} = 1+\sum_{e\in E} z_e\gamma_e - \frac{1}{n-1}\sum_{e\in E}\sum_{T\ni e} e^{\gamma(T)} \geq 1+\sum_{e\in E} z_e\gamma_e - \frac{1}{n-1}\sum_{e\in E} z_e = \sum_{e\in E} z_e\gamma_e,$$

where the last equation holds since $\mathbf{z}$ is a fractional spanning tree. So the optimum of CP (3.1.10) is at most the optimum of CP (3.1.7). On the other hand, since $\mathbf{z}$ is in the interior of spanning tree polytope, there is a unique optimum γ^* to CP (3.1.7) that satisfies (3.1.8), so for all $e \in E$, $\sum_{T\ni e} \exp(\gamma^*(T)) = \sum_{T\ni e} p^*(T) = z_e$, and γ^* is a feasible solution of (3.1.10). Furthermore,

$$1 + \sum_{e\in E} z_e\gamma_e^* - \sum_{T\in\mathcal{T}} \exp(\gamma^*(T)) = 1 + \sum_{e\in E} z_e\gamma_e^* - \sum_{T\in\mathcal{T}} p^*(T) = \sum_{e\in E} \gamma_e^* z_e.$$

Therefore, the optimum of (3.1.7) is at most the optimum of (3.1.10). Therefore, they are equal, and the optimum of (3.1.10) is $\mathsf{OPT}_{\text{Ent}}$.

Next, we use the ellipsoid method, Theorem 2.2.2, to find a near optimal solution of CP (3.1.10). The main difficulty is that the coordinates of the optimizers of CP (3.1.10) are not necessarily bounded by a function of n. First, we simply turn the optimization problem into a feasibility problem by doing a binary search on the value of the optimum, so suppose we guess the optimum is t. Now, instead of proving that every feasible solution of CP (3.1.10) that satisfies $\sum_e z_e\gamma_e \geq t$ falls in a ball of radius that is a polynomial function of n, we *restrict* the set of feasible solutions of (3.1.10) to the vectors whose coordinates are bounded by a polynomial function of n. Furthermore, to ensure that the new polytope has a non-empty interior, we relax the RHS of the constraint $\sum_{T\ni e} \exp(\gamma(T)) \leq z_e$. More precisely, for any $\alpha > 0, M > 0$ and $t \in \mathbb{R}$, let $\mathcal{F}(\alpha, t, M)$ be the following feasibility convex program

$$\begin{aligned} & \sum_{e} z_e\gamma_e \geq t, \\ & \sum_{T\ni e} e^{\gamma(T)} \leq (1+\alpha) z_e \qquad \forall e \in E, \\ & -M \leq \gamma_e \leq M \qquad \forall e \in E. \end{aligned} \tag{3.1.11}$$

The following lemma relates the above CP to CP (3.1.10).

Lemma 3.1.4. *For any $t \leq \mathsf{OPT}_{Ent}$, $\mathcal{F}(e^{-n^2/2}, t, n^4 - n^2 \log z_{min})$ is non-empty.*

Proof. We say that a vector $\gamma : E \to \mathbb{R}$ has a gap at an edge $f \in E$ if for any $e \in E$, either $\gamma_e \le \gamma_f$ or $\gamma_e > \gamma_f + \mathsf{gap}$ where $\mathsf{gap} := n^2 - \log z_{\min}$. Observe that for any $\gamma : E \to \mathbb{R}$, the number of gaps of γ is at most $|E| \le \binom{n}{2}$.

In the following claim we show that if γ has a gap at an edge e then we can construct another vector $\tilde{\gamma}$ with fewer number of gaps while losing a small amount in the value of the objective function.

Claim 3.1.5. *Let $\gamma : E \to \mathbb{R}$ that has at least one gap. Let $T_{\max}$ be a maximum spanning tree of G with respect to weights γ, i.e., $T_{\max} = \operatorname{argmax}_T \gamma(T)$. There exists $\tilde{\gamma} : E \to \mathbb{R}$ with at least one fewer gap such that for any $e \in E$,*

$$\sum_{T \ni e} e^{\tilde{\gamma}(T)} \le \sum_{T \ni e} e^{\gamma(T)} + n^{n-2} e^{-\gamma(T_{\max}) - \mathsf{gap}}. \tag{3.1.12}$$

and

$$\sum_e z_e \tilde{\gamma}_e \ge \sum_e z_e \gamma_e. \tag{3.1.13}$$

Proof. Suppose that γ has a gap at an edge $e^* \in E$. Let $F := \{e \in E : \gamma_e > \gamma_{e^*}\}$. Let $k = \operatorname{rank}(F)$ be the size of the maximum spanning forest of F (see Section 2.3 for background). Recall that by definition any spanning tree of G has at most k edges from F, so $\mathbf{z}(F) \le k$. We reduce the γ_e for all $e \in F$ and increase it for the rest of the edges. In particular,

$$\tilde{\gamma}_e = \begin{cases} \gamma_e + \frac{k\Delta}{n-1} & \text{if } e \notin F, \\ \gamma_e - \Delta + \frac{k\Delta}{n-1} & \text{if } e \in F, \end{cases}$$

where $\Delta = \min_{e \in F} \gamma_e - \gamma_{e^*} - \mathsf{gap}$. Note that by the assumption of the claim $\Delta > 0$. By above definition, $\tilde{\gamma}$ does not have a gap at e^*, and for any edge $e \neq e^*$, $\tilde{\gamma}$ has a gap at e if γ has a gap at e.

First, observe that,

$$\sum_e z_e \tilde{\gamma}_e = \sum_e z_e \gamma_e + \frac{k\Delta}{n-1} \sum_e z_e - \mathbf{z}(F)\Delta \ge \sum_e z_e \gamma_e + k\Delta - k\Delta = \sum_e z_e \gamma_e.$$

where we used $\mathbf{z}(F) \le k$. This proves (3.1.13).

It remains to prove (3.1.12). If a spanning tree T has exactly k edges from F, then $\tilde{\gamma}(T) = \gamma(T)$, and $\exp(\tilde{\gamma}(T)) = \exp(\gamma(T))$. By Lemma 2.3.1 any maximum weight spanning tree of (V, E, γ) or $(V, E, \tilde{\gamma})$ has exactly k edges of F. Since $\tilde{\gamma}(T) = \gamma(T)$ for any tree where $|T \cap F| = k$, the maximum spanning trees of (V, E, γ) are the same as the maximum spanning trees of $(V, E, \tilde{\gamma})$. So, $T_{\max}$ is also a maximum weight spanning tree of $(V, E, \tilde{\gamma})$.

Now, suppose a spanning tree T has less than k edges in F. Since $|T \cap F| < k$, there exists an edge $f \in (T_{\max} \cap F) - T$ such that $(T \cap F) \cup \{f\}$ is a forest of G. Therefore, the unique circuit in

$T \cup \{f\}$ contains an edge $e \notin F$. Thus $T' = T \cup \{f\} - \{e\}$ is a spanning tree. By the definition of γ',

$$\tilde{\gamma}(T_{\max}) \geq \tilde{\gamma}(T') = \tilde{\gamma}(T) - \tilde{\gamma}_e + \tilde{\gamma}_f > \tilde{\gamma}(T) + \mathsf{gap}, \tag{3.1.14}$$

which yields the desired inequality. Therefore, for any tree T,

$$e^{\tilde{\gamma}(T)} \leq e^{\gamma(T)} + e^{\tilde{\gamma}(T_{\max})-\mathsf{gap}} = e^{\gamma(T)} + e^{\gamma(T_{\max})-\mathsf{gap}},$$

where the last equality follows by the fact that $|T_{\max} \cap F| = k$. Now, (3.1.12) follows by the fact that any graph at most n^{n-2} spanning trees [Cay89]. □

Let γ^* be an optimum solution of CP (3.1.10). If γ^* does not have any gap we let $\tilde{\gamma} = \gamma^*$. Otherwise, we repeatedly apply the above claim and remove all of the gaps and find a vector $\tilde{\gamma}$ such that $\sum_e z_e\tilde{\gamma}_e \geq \sum_e z_e\gamma_e$, and for any edge $e \in E$,

$$\sum_{T \ni e} e^{\tilde{\gamma}(T)} \leq \sum_{T \ni e} e^{\gamma^*(T)} + |E|n^{n-2}e^{\gamma^*(T_{\max})-\mathsf{gap}} \leq z_e + n^n e^{-n^2} z_{\min} \leq (1 + n^{-n^2/2})z_e. \tag{3.1.15}$$

where the first inequality follows by the fact that γ^* has at most $|E|$ gaps, the second inequality follows by the feasibility of γ^* in CP (3.1.10) and that $e^{\gamma^*(T_{\max})} \leq \max_e z_e \leq 1$.

Since $\tilde{\gamma}$ does not have any gap

$$\max_e \tilde{\gamma}_e - \min_e \tilde{\gamma}_e \leq |E| \cdot \mathsf{gap}.$$

So, it is sufficient to lower bound $\max_e \tilde{\gamma}_e$ and upper bound $\min_e \tilde{\gamma}_e$. Let $f = \operatorname{argmax}_e \tilde{\gamma}_e$. By (3.1.4),

$$-n \log n \leq \mathsf{OPT}_{\mathrm{Ent}} = \sum_e z_e\gamma^*_e \leq \sum_e z_e\tilde{\gamma}_e \leq n \cdot \max_e \tilde{\gamma}_e.$$

On the other hand, by (3.1.15), $e^{\tilde{\gamma}(T)} \leq 2$ for any tree T, so $\min_e \tilde{\gamma}_e \leq 1$. Therefore,

$$\begin{aligned} \max_e \tilde{\gamma}_e &\leq \min_e \tilde{\gamma}_e + |E| \cdot \mathsf{gap} \leq 1 + |E| \cdot \mathsf{gap} \leq n^4 - n^2 \log z_{\min}, \\ \min_e \tilde{\gamma}_e &\geq \max_e \tilde{\gamma}_e - |E| \cdot \mathsf{gap} \geq -\log(n) - |E| \cdot \mathsf{gap} \geq -n^4 + n^2 \log z_{\min}. \end{aligned}$$

This completes the proof of Lemma 3.1.4. □

In Algorithm 5 we provide a separating hyperplane oracle for CP (3.1.11). Note that all the steps of the algorithm can be done in polynomial time. The only one which may need some explanation is computing $q_e(\gamma)$ for some edge e and its gradient.

$$q_e(\gamma) = e^{\gamma_e} \sum_{T \ni e} e^{\gamma(T-\{e\})} \quad \text{and} \quad \frac{\partial q_e(\gamma)}{\partial \gamma_{e'}} = e^{\gamma_e + \gamma_{e'}} \sum_{T \ni e,e'} e^{\gamma(T-\{e,e'\})}.$$

Algorithm 5 Separating hyperplane oracle for CP (3.1.11)

Input: $\gamma \in \mathbb{R}^{|E|}$
  if γ violates any of the linear constraints **then**
    Return the violated inequality as a separating hyperplane.
  else
    Compute $q_e(\gamma) = \sum_{T \ni e} e^{\gamma(T)}$ for every e.
    if $q_e(\gamma) \leq (1+\alpha)z_e$ for all edges e **then**
      report $\gamma \in \mathcal{F}(\alpha, t, M)$.
    else
      Let $\hat{e}$ be an edge for which the constraint is violated. Compute the gradient of $q_{\hat{e}}(\gamma)$.
      Return the hyperplane $\{(\gamma' - \gamma).\nabla q_{\hat{e}}(\gamma) > 0, \gamma' \in \mathbb{R}^{|E|}\}$ as a violated constraint.
    end if
  end if

Both of the above expressions can be computed efficiently by the matrix tree theorem (see Theorem 2.8.1).

Now, we are ready to prove Theorem 3.1.3.

Proof of Theorem 3.1.3. Let $\alpha = \epsilon/6$. By Lemma 3.1.4, $\mathcal{F}(\alpha, \mathsf{OPT}_{\text{Ent}}, M)$ where $M = n^4 - n^2 \log z_{\min}$ is non-empty. Let γ^* be a point in $\mathcal{F}(\alpha, \mathsf{OPT}_{\text{Ent}}, M)$ and let $B = \{\gamma : \|\gamma - \gamma^*\|_\infty \leq \beta\}$, where $\beta = \epsilon/4n$. For any $\gamma \in B$,

$$\sum_e z_e \gamma_e \geq \sum_e z_e(\gamma_e^* - \beta) \geq \mathsf{OPT}_{\text{Ent}} - n\beta = \mathsf{OPT}_{\text{Ent}} - \epsilon/4.$$

Also, for any edge $e \in E$,

$$\sum_{T \ni e} e^{\gamma(T)} \leq \sum_{T \ni e} e^{\gamma^*(T) + n\beta} \leq e^{n\beta}(1+\alpha)z_e \leq (1+\epsilon/2)z_e.$$

where the last inequality follows by the assumption $\epsilon < 1/4$.

So $B \subseteq \mathcal{F}(\epsilon/2, \mathsf{OPT}_{\text{Ent}} - \epsilon/4, M + \beta)$. Therefore, $\mathcal{F}(\epsilon/2, \mathsf{OPT}_{\text{Ent}} - \epsilon/4, M+1)$ is non-empty and contains a ball of radius $\beta = \epsilon/4n$ and is contained in a ball of radius $|E| \cdot n$ which is a polynomial function of n, $-\log z_{min}$ and $1/\epsilon$. Using binary search on t and the ellipsoid method, we can find a point γ in $\mathcal{F}(\epsilon/2, \mathsf{OPT}_{\text{Ent}} - \epsilon/4, M+1)$ in time polynomial in n, $-\log z_{\min}$, and $\log 1/\epsilon$.

Since γ is a feasible solution of of the CP, (3.1.7). $1 + \sum_e z_e \gamma_e - \sum_T e^{\gamma(T)} \leq \mathsf{OPT}_{\text{Ent}}$. On the other hand, since $\gamma \in \mathcal{F}(\epsilon/2, \mathsf{OPT}_{\text{Ent}} - \epsilon/4, M+1)$,

$$\sum_e z_e \gamma_e \geq OPT_{CP} - \epsilon/4.$$

These two imply that

$$\sum_T e^{\gamma(T)} \geq 1 - \epsilon/3.$$

The last step is to normalize γ_e's. Define $\tilde{\gamma}(e) = \gamma(e) - \frac{\log \sum_T e^{\gamma(T)}}{n-1}$. By definition, $\sum_T e^{\tilde{\gamma}(T)} = 1$. So, for any edge $e \in E$,

$$\sum_{T \ni e} e^{\tilde{\gamma}(T)} = \frac{\sum_{T \ni e} e^{\gamma(T)}}{\sum_T e^{\gamma(T)}} \leq \frac{(1+\epsilon/2)z_e}{1 - \mathsf{OPT}_{\text{Ent}} + \sum_e z_e \gamma_e} \leq \frac{(1+\epsilon/2)z_e}{1-\epsilon/4} \leq (1+\epsilon)z_e.$$

where the first inequality follows by the fact that the optimum of (3.1.10) is $\mathsf{OPT}_{\text{Ent}}$ and γ is a feasible point of that program. □

3.1.3 Other Methods for Rounding with Negative Correlation Property

Since the appearance of our works, other ways of producing *negatively correlated* probability distributions on trees (or, more generally, matroid bases) satisfying some given marginals $\mathbf{z}$ have been proposed [CVZ10]. Chekuri, Vondrak and Zenklusen [CVZ10] proposed a randomized variant of *pipage rounding method* [CCPV07], and a new randomized selection method, called swap rounding, which is based on the basis exchange property of matroid bases. Since these approaches preserve negative correlation, they satisfy Chernoff types of bound that we discussed in Subsection 2.8.3.

The distributions produced by pipage rounding or swap rounding methods are not necessarily a λ-random spanning tree distribution. In fact, we are not aware of any closed form description of these distributions. So, these distributions do not fit in the class of strongly Rayleigh probability measures, and they do not necessarily satisfy their properties. In general, these distributions satisfy weaker form of negative dependence compared to the maximum entropy rounding by sampling method, but the corresponding sampling algorithm is faster. Furthermore, both of the pipage rounding and swap rounding methods can be applied to any matroid polytope, but, as we discussed in Subsection 3.1.2, the maximum entropy rounding by sampling method can only be applied to polytopes that has an efficient exact or approximate counting algorithm [SV13].

It turns out that the analysis of our new approximation algorithm for ATSP (see Chapter 4) works with any distribution of spanning trees that satisfies Chernoff types of bounds on cuts and (approximately) preserves the vector $\mathbf{z}$ as the marginal probability of the edges. But, the analysis of our approximation algorithm for Symmetric TSP (see Chapter 6) uses several properties of strongly Rayleigh measures and λ-random spanning tree distributions that are stronger than the negative correlation property (e.g., the negative association property defined in Definition 2.9.6). Therefore, we can use either of pipage rounding method, or randomized swap rounding method in place of maximum entropy rounding by sampling method in Algorithm 6 and prove the same approximation factor of $O(\log n / \log \log n)$ for ATSP. But, we are not aware of any analysis of an algorithm for STSP that uses pipage rounding instead of maximum entropy distributions.

3.2 The Cactus-like Structure of Near Minimum Cuts

Let $G = (V, E)$ be an unweighted Δ-edge connected graph. In this section we prove new properties of $(1+\eta)$ near minimum cuts for small but constant values of η. Then, in Subsection 3.2.2 we discuss applications to the instances of traveling salesman problem.

Our main result in this section is a generalization of Lemma 2.6.6 to the structure of near minimum cuts in an approximate sense (see Theorem 3.2.5 below). Roughly speaking, we show that for a sufficiently small η, for any cut class $\mathcal{C}$ of $(1+\eta)$ near minimum cuts, $G(\psi(\mathcal{C}))$ is very close to a cycle. In particular, we can find $|V|(1-\epsilon)$ pair of vertices such that there are $\Delta(1-\epsilon')$ parallel edges between each pair where ϵ, ϵ' are functions of η and *do not depend on* $|V|$. In some sense, our theorem adds a characterization of the placement of the edges between the atoms to the polygon representation, when η is sufficiently small.

Before getting into our main theorem we provide a simple proposition that shows a direct generalization of the argument in the proof of Lemma 2.6.6 only works for cut classes with "small" number of atoms. As it will be clear later, the argument for "large" cut classes is much more complicated and requires many properties of the polygon representation discussed in Section 2.7.

Proposition 3.2.1. *Let $\mathcal{F}$ be the collection of $(1+\eta)$ near minimum cuts of graph G. For any proper cut class $\mathcal{C}$ the contracted graph $H = G(\psi(\mathcal{C}))$ satisfy the following properties.*

1. *For any vertex $u \in V(H)$, we have $|\delta(u)| \leq \Delta(1+\eta(|V(H)|-2))$,*
2. *For any vertex $u \in V(H)$ there are two distinct vertices $v, w \in V(H)$ such that*

$$|E(u,v)|, |E(u,w)| \geq \frac{\Delta}{2}(1-\eta(|V(H)|-3)).$$

where by $E(u,v)$ we mean the set of edges between the atoms corresponding to u and v in G.

Proof. A set $A \subseteq V(H)$ is *tight* if $(A, \overline{A})$ is a near minimum cut. Also, A is a non-trivial subset if $|A| \geq 2$.

Claim 3.2.2. *Any non-trivial set A containing a vertex u includes a two element set A^* containing u such that*

$$|\delta(A^*)| \leq |\delta(A)| + \eta(|A|-2)\Delta.$$

Proof. We prove by induction. If $|A| = 2$ then we are already done. So assume $|A| \geq 3$, and let $A' = A - \{u\}$. Since A' is non-trivial, by Fact 2.6.7 there is a tight set B containing u that crosses A'. Now, either $B \subset A$, or B crosses A. In either case B or $B \cap A$ gives a smaller non-trivial set that contains u. Furthermore, since B is a tight set, and $|\delta(A)| \geq \Delta$,

$$|\delta(B)| \leq (1+\eta)\Delta \leq |\delta(A)| + \eta\Delta.$$

Also, by Lemma 2.7.1 $|\delta(B \cap A)| \leq |\delta(A)| + \eta\Delta$. So, by the induction hypthesis there is a two element set A^* containing u and

$$\begin{aligned}\delta(A^*) \leq \max\{|\delta(A \cap B)|, |\delta(B)|\} + \eta(\max\{|A \cap B|, |B|\} - 2)\Delta &\leq \delta(A) + \eta\Delta + \eta(|A| - 3)\Delta \\ &= |\delta(A)| + \eta(|A| - 2)\Delta,\end{aligned}$$

and we are done. □

For any vertex $u \in V(H)$, let $(A, \overline{A})$ be a near minimum cut and let $u \in A$. Then, by the above claim there is A^* of size 2 such that $u \in A^*$ and

$$|\delta(A^*)| \leq |\delta(A)| + \eta(|A| - 2)\Delta \leq (1 + \eta)\Delta + \eta(|V(H)| - 4)\Delta \leq \Delta(1 + \eta(|V(H)| - 3)).$$

where we used the fact that $\mathcal{C}$ is a proper cut class in showing $|A| \leq |V(H)| - 2$. Since A^* is a non-trivial set there is a tight set B crossing A. By Lemma 2.7.1,

$$|\delta(u)| = |\delta(A^* \cap B)| \leq |\delta(A^*)| + \eta\Delta \leq \Delta(1 + \eta(|V(H)| - 2)).$$

This proves the first conclusion.

Let $A^* = \{u, v\}$. Applying above claim to B gives a two element set $B^* = \{u, w\}$ such that similar to A^*, $|\delta(B^*)| \leq \Delta(1 + \eta(|V| - 3))$. But since $B^* \subseteq B$, $v \neq w$. So, $|E(u, v)|, |E(u, w)| \geq (1 - \eta(|V(H)| - 3))\frac{\Delta}{2}$. □

Observe that if $|\psi(\mathcal{C})| \ll 1/\eta$, then we can arrange the vertices of H around a cycle such that *almost* all edges adjacent to each vertex go to its neighbors in the cycle. This is made rigorous in the following corollary.

Corollary 3.2.3. *Let $\mathcal{F}$ be the collection of $1 + \eta$-near minimum cuts of G for $\eta < 1/16$. For any proper cut class $\mathcal{C}$ such that $2|\psi(\mathcal{C})| \leq 1/\eta$,*

1. *The polygon representation of $\mathcal{C}$ does not have any inside atoms.*
2. *For any two consecutive outside atoms in polygon representation, A, B, $|E(A, B)| \geq \Delta/2(1 - \eta(|\psi(\mathcal{C})| - 3))$.*

Proof. The proof of first conclusion uses Lemma 3.2.15 that we prove later in this section. It follows from Lemma 3.2.15 that if $\mathcal{C}$ has an inside atom, then there is an inside atom A that has at most $4\eta\Delta$ edges to all other inside atoms. On the other hand, by Claim 3.2.17 for any outside atom B, $|E(A, B)| \leq 2\eta\Delta$. Since $2|\psi(\mathcal{C})| \leq 1/\eta$,

$$\frac{\Delta}{2}(1 - \eta|\psi(\mathcal{C})|) \geq \Delta/4 > 4\eta\Delta$$

So, A does not satisfy the second conclusion of Proposition 3.2.1 and it cannot be an atom of $\mathcal{C}$.

It remains to prove the second conclusion. It is easy to see that the same inductive argument of Claim 3.2.2 shows that each atom of $\psi(\mathcal{C})$ has $\Delta/2(1-\eta(|\psi(\mathcal{C})|-3))$ to its neighboring atoms in the polygon representation. It is sufficient to strengthen the inductive hypothesis and assume that the non-trivial set A corresponds to a diagonal (not necessarily a representing diagonal) of the polygon representation. □

If $|\psi(\mathcal{C})| > 1/\eta$, then the above proposition does not say anything about the structure of edges in $G(\psi(\mathcal{C}))$. So, in the rest of this section our main concern are "large" cut classes.

Before describing our main theorem we first need to define cactaceous structures as a cycle-like structure for "large" cut classes.

Definition 3.2.4 ((α, α', β)-cactaceous)**.** *A Δ-edge connected graph $G = (V, E)$ is (α, α', β)-cactaceous if for some $\eta \geq 0$:*

i) There exists at least $m := (1-\alpha\sqrt{\eta})|V|$ pairs of vertices of G, $\{\{v_1, u_1\}, \{v_2, u_2\}, \ldots, \{v_m, u_m\}\}$ such that for each $1 \leq i \leq m$, $|E(v_i, u_i)| \geq \frac{\Delta}{2}(1-\alpha'\sqrt{\eta})$, and each vertex $v \in V$ is contained in at most two such pairs.

ii)
$$\frac{\Delta}{2}|V| \leq |E| \leq (1+\beta\eta)\frac{\Delta}{2}|V|.$$

Our main theorem in this section shows that any cut class of a collection of near minimum cuts is cactaceous.

Theorem 3.2.5. *For any $\eta < 1/100$ and any cut class $\mathcal{C}$ of $(1+\eta)$ near minimum cuts of G, $G(\psi(\mathcal{C}))$ is $(14, 3, 3)$-cactaceous.*

Observe that if we let $\eta = 0$ in the above theorem, then we obtain Lemma 2.6.6. So, above theorem can be seen as a generalization of Lemma 2.6.6 to the system near minimum cuts. We just need to prove the above theorem for a fixed cut class $\mathcal{C}$ of G. So, for the sake of brevity, we work with $G(\psi(\mathcal{C}))$, and we prove the following equivalent statement.

Theorem 3.2.6. *Let $G = (V, E)$ be a Δ-edge connected graph, and $\eta < 1/100$ such that any $S \subseteq V$ of size $|S| \geq 2$ is crossed by a $(1+\eta)$ near minimum cut, and the cross graph of the $(1+\eta)$ non-trivial near minimum cuts of G is connected. Then G is $(14, 3, 3)$-cactaceous.*

Observe that the two properties of graph G in above theorem follows from Fact 2.6.7 and Lemma 2.6.3. Unless otherwise specified, throughout this section by a near minimum cut we mean a $(1+\eta)$-near minimum cut.

We start by proving that G satisfies (ii) property of cactaceous structures for $\beta = 3$. Since G is Δ-edge connected, the degree each vertex is at least Δ. Therefore, it is sufficient to show that the *average degree* of vertices of G is significantly larger than Δ.

Unfortunately, there are examples where some vertices of G have very large degrees. For example, consider the wheel graph in Figure 2.7.8, and observe that this graph satisfies the assumption of Theorem 3.2.6. Now, we can increase the number of vertices in the surrounding cycle, this increases the degree of vertex 1 linearly with n. So, although the average degree of the vertices is only 8, G has a vertex of degree n.

The above example shows that maximum degree of G is unbounded, so we prove (ii) property by upper bounding the average degree of vertices.

Lemma 3.2.7. *The average degree of the vertices of G is at most $(1+3\eta)\cdot\Delta$.*

Proof. We run the following procedure, and we incrementally construct a partitioning of V.

Let $(A, \overline{A})$ be a near minimum cut of G.
We maintain a partition $\mathcal{P}$ of vertices of V.
Initialize $\mathcal{P} = \{A, \overline{A}\}$.
while there is $S \in \mathcal{P}$ s.t. $|S| \geq 2$ **do**
  Find a minimum cut $(T, S-T)$ of the induced subgraph $G[S]$.
  Substitute T with B and $S-T$.
end while

Observe that each edge $\{u, v\} \in E$ is inside exactly one of the cuts $(T, S-T)$ considered in above procedure. Therefore, if we show that the size of each cut $|E(T, S-T)| \leq (1+3\eta)\frac{\Delta}{2}$, then the average degree of the vertices of G is at most:

$$\frac{|E|}{|V|} \leq \frac{|E(A,\overline{A})|}{|V|} + \frac{|V|-2}{|V|}(1+3\eta)\frac{\Delta}{2} \leq (1+3\eta)\frac{\Delta}{2},$$

and we done.

We say a near minimum cut $(B, \overline{B})$ is *outside* of a set $S \subseteq V$, if either $S \subseteq B$, or $S \subseteq \overline{B}$ or B crosses S. In Claim 3.2.8 we show that that if a set $S \subset V$ satisfies

$$\min_{T \subset S} |E(T, S-T)| > (1+3\eta)\frac{\Delta}{2},$$

then any near minimum cut $(B, \overline{B})$ of G that is outside of S, crosses S. But, all S considered throughout the procedure are either a subset of A, or $\overline{A}$, so $(A, \overline{A})$ is outside of any such set S. Therefore, the size of the minimum cut of of induced subgraph $G[S]$ is at most $(1+3\eta)\frac{\Delta}{2}$, and we are done.

Claim 3.2.8. *Let $(S, \overline{S})$ is a non-trivial cut in G. If for any set $T \subset S$,*

$$|E(T, S-T)| > (1+3\eta)\frac{\Delta}{2},$$

then any near minimum cut of G that is outside of S, crosses S.

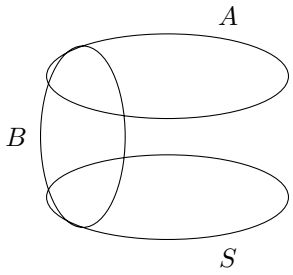

Figure 3.2.3: Setting in the proof of Lemma 3.2.7.

Proof. We prove by contradiction. Suppose there exists a near minimum cut outside of S, and not crossing S. On the other hand, since $(S, \overline{S})$ is a non-trivial cut of G, there exists a near minimum cut crossing $(S, \overline{S})$. But, since the set of near minimum cuts of G are connected, there is a path of crossing cuts which connects these two cuts. So, without loss of generality, we can assume there are two crossing near minimum cuts $(A, \overline{A})$ and $(B, \overline{B})$ such that A is outside of S and B crosses S (see Figure 3.2.3).

Since B crosses S, by Claim's assumption we have $|E(B_l \cap S, \overline{B_l} \cap S)| > (1+3\eta)\frac{\Delta}{2}$. On the other hand, since $(B, \overline{B})$ crosses $(A, \overline{A})$ by Lemma 2.7.2 we have $|E(B_l \cap B_{l-1}, \overline{B_l} \cap B_{l-1})| \geq (1-\eta)\frac{\Delta}{2}$. Therefore:

$$\begin{aligned} |E(B, \overline{B})| &\geq |E(B \cap S, \overline{B} \cap S)| + |E(B \cap A, \overline{B} \cap A)| \\ &> (1+3\eta)\frac{\Delta}{2} + (1-\eta)\frac{\Delta}{2} = (1+\eta)\Delta, \end{aligned}$$

where the first inequality holds by the fact that $A \cap S = \emptyset$. Therefore, $(B, \overline{B})$ is not a near minimum cut of G which is a contradiction. □

This completes the proof of Lemma 3.2.7. □

It remains to prove G satisfies the (i) property of the cactaceous structures for $\alpha = 20, \alpha' = 4$. This is the most technical part of the proof of Theorem 3.2.6. We start by proving the (i) property holds in the special case where the polygon representation does not have any inside atoms. Then, in Subsection 3.2.1 we extend the proof to the general case where inside atoms are allowed. So for now, we may assume that we have a (regular) convex $|V|$-gon, such that each vertex of G is mapped to a distinguished edge of the polygon, and the diagonals represent the near minimum cuts (see Section 2.7 for background on the polygon representation).

Lemma 3.2.9. *Suppose that the polygon representation of G does not have any inside atoms. Then, for any integer $s \geq 12$, there exists at least $m = (1 - \frac{8}{s-10})|V|$ pairs of vertices of G, $\{\{v_1, u_1\}, \{v_2, u_2\}, \ldots, \{v_m, u_m\}\}$ such that for each $1 \leq i \leq m$, $|E(v_i, u_i)| \geq (1 - s\eta)\frac{\Delta}{2}$, and each vertex $v \in V$ is contained in at most two such pairs.*

Proof. Let $n := |V|$, and let us label the points of the polygon representation cyclically (in a clockwise order) as $p_1, p_2, \ldots, p_n$. Moreover, assume that the n vertices $v_1, v_2, \ldots, v_n$ of G are placed on the edges of the polygon such that v_i is placed on the edge (p_i, p_{i+1}) (note that by p_{i+1} we mean the point next to p_i in the cyclic ordering). We show that there is a set S^* of vertices of G such that $|S^*| \geq (1 - \frac{8}{s-10})n$ such that for any $v_i \in S^*$, we have $|E(v_i, v_{i+1})| \geq (1 - s\eta)\frac{\Delta}{2}$. Observe that the lemma follows from the existence of S^*. Since G is Δ-edge connected, if $|\delta(\{v_i, v_{i+1}\})| \leq (1 + s\eta)\Delta$, then $|E(v_i, v_{i+1})| \geq (1 - s\eta)\frac{\Delta}{2}$. So, equivalently, we can show there is a set $S^* \subseteq V$ such that $|S^*| \geq (1 - \frac{8}{s-10})n$ and for any $v_i \in S^*$ $|\delta(\{v_i, v_{i+1}\})| \leq (1 + s\eta)\Delta$.

Before defining the set S^*, we need to define some notations. Since each near minimum cut of G is corresponding to a representing diagonal (and a consecutive sequence of vertices), we will use intervals to represent (near minimum) cuts. For any two points p_i and p_j let

$$[p_i, p_j] := \{p_i, p_{i+1}, \ldots, p_j\},$$
$$[p_i, p_j) := \{p_i, p_{i+2}, \ldots, p_{j-1}\}.$$

Note that $[p_i, p_j] \cup [p_j, p_i]$ is all of the points of the polygon. Also, let

$$C(p_i, p_j) := \delta(\{v_i, v_{i+1}, \ldots, v_{j-1}\}).$$

be the cut corresponding to an interval of the points of the polygon. We say two intervals cross, if their corresponding cuts cross. For example, the intervals $[p_i, p_{i+2}]$ and $[p_{i+2}, p_{i+4}]$ do not cross, while $[p_i, p_{i+2}]$ and $[p_{i+1}, p_{i+3}]$ cross each other (assuming $n \geq 4$). We say an interval $[p_{i'}, p_{j'}]$ is a *subinterval* of $[p_i, p_j]$ if the set of polygon vertices contained in $[p_{i'}, p_{j'}]$ is a subset of $[p_i, p_j]$. For example, $[p_2, p_4]$ is a subinterval of $[p_1, p_5]$, but $[p_4, p_2]$ is not a subinterval of $[p_1, p_5]$.

For each point p_i, let q_i be the nearest point to p_i (in terms of the cyclic distance), such that $C(p_i, q_i)$ is a near minimum cut (note that since there is a near minimum cut crossing $C(p_{i-1}, p_{i+1})$, each vertex p_i is adjacent to at least one representing diagonal). Since we only consider the non-trivial near minimum cuts of G, $q_i \neq p_{i+1}$ and $q_i \neq p_{i-1}$.

For any point p_i of the polygon, we define a *chain* as a sequence of points $q_{i_0}, q_{i_1}, q_{i_2}, \ldots, q_{i_l}$ satisfying the following properties:

1. $q_{i_0} = q_i$,
2. $q_{i_l} = p_{i+2}$,
3. for all $j \geq 1$, we have $q_{i_j} \in [p_{i+2}, q_{i_{j-1}})$; and
4. for all $j \geq 1$, there exists a point r_j such that the cut $C(q_{i_j}, r_j)$ is a near-minimum cut, and it crosses the cut $C(p_i, q_{i_{j-1}})$.

See Figure 3.2.4 for an example of a chain. The length of a chain is the number of its points. In the next claim we show that every point p_i has at least one chain.

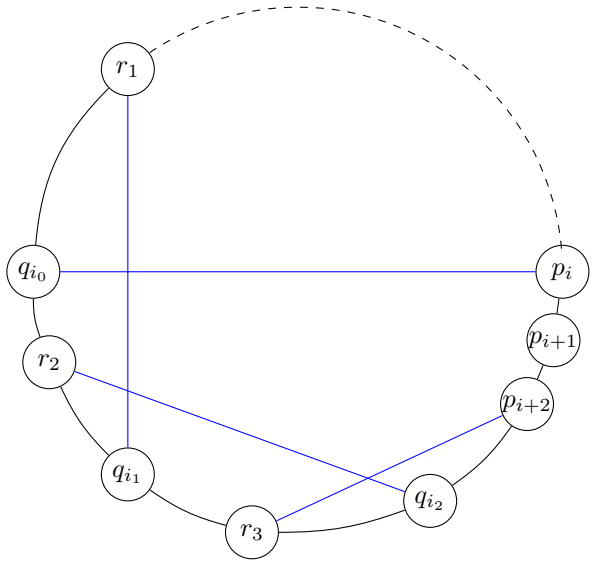

Figure 3.2.4: An example of a chain of length 4 with vertices $q_{i_0}, q_{i_1}, q_{i_2}, p_{i+2}$ for a vertex p_i. The blue edges are the representing diagonals in the polygon representation

Claim 3.2.10. *Any point p_i satisfy the following properties.*

1) There exists at least one chain for p_i.

2) If there is a chain of length l for p_i, then $|\delta(\{v_i, v_{i+1}\})| \leq (1 + \eta \cdot l)\Delta$.

3) Let $P_i := \{q_{i_0}, q_{i_1}, \ldots, q_{i_l}\}$ be a chain for vertex p_i, and let $C(p_a, p_b)$ be a near minimum cut such that $p_a \in [p_{i+2}, q_i)$, and more than two vertices of P_i are contained in the interval $[p_a, p_b)$. If $C(p_a, p_b)$ crosses $C(p_i, q_i)$, or $[p_a, p_b]$ is a subinterval of $[p_i, q_i]$, then there is a shorter chain for p_i.

Proof. We start by proving the first property. We construct the chain inductively starting with $q_{i_0} := q_i$. Each time we add a new vertex q_{i_j} that is closer to p_i compared to $q_{i_{j-1}}$ such that there is a near minimum cut $C(q_{i_j}, r_j)$ that crosses $C(p_i, q_{i_{j-1}})$. Since $q_{i_{j-1}} \in [p_{i+3}, q_i]$ (i.e., the chain is not completed yet), $C(p_{i+1}, q_{i_{j-1}})$ is a non-trivial cut, and by the assumptions of Theorem 3.2.6, there exists a near minimum $C(p_a, p_b)$ crossing it. Assume (perhaps after renaming) that $p_a \in [p_{i+2}, q_{i_{j-1}})$. On the other hand, since q_i is the closest vertex to p_i, $b \neq i$. Therefore, $C(p_a, p_b)$ crosses $C(p_i, q_{i_{j-1}})$. and we can simply set $q_{i_j} := p_a$. This completes the proof of conclusion (1). It is instructive to compare this proof with the simple proof of Proposition 3.2.1.

For the second conclusion we use a simple induction to show that $|C(p_i, q_{i_j})| \leq \Delta + (j+1)\eta\Delta$. Then for $j = l - 1$ we get $|C(p_i, p_{i+2})| = |C(p_i, q_{i_{l-1}})| \leq \Delta + l\eta\Delta$ which implies conclusion (2). For $j = 0$ we have $|C(p_i, q_{i_0})| = |C(p_i, q_i)| \leq (1+\eta)\Delta$. On the other hand, by Lemma 2.7.1,

$$|C(p_i, q_{i_{j+1}})| \leq |C(p_i, q_{i_j})| + \eta\Delta \leq \Delta + (j+1)\eta + \eta\Delta = \Delta + (j+2)\eta\Delta,$$

where the last inequality holds by the induction hypothesis. This completes the proof of conclusion (2).

To prove the last property, it suffices to construct a shorter chain for p_i using the near minimum cut $C(p_a, p_b)$. Let q_{i_c}, q_{i_d} be the first, and the last points of the chain that are contained in $[p_a, p_b)$. Since either $C(p_a, p_b)$ crosses $C(p_i, q_i)$ or $[p_a, p_b]$ is a subinterval of $[p_i, q_i]$, q_{i_c}, q_{i_d} are well-defined. By claim's assumptions we have $d \geq c + 2$.

Let $P_i' := \{q_{i_0}, q_{i_1}, \ldots, q_{i_c}, p_a, q_{i_d+1}, \ldots, q_{i_l}\}$. Since $d \geq c+2$, the length of P_i' is smaller than P_i. So, it remains to show P_i' is indeed a chain for p_i. Suppose $C(q_{i_{d+1}}, p)$ is a near minimum cut that crosses $C(p_i, q_{i_d})$ where p is a point of the polygon (note that this cut exists since P_i is a chain). We need to show that $C(p_a, p_b)$ crosses $C(p_i, q_{i_c})$, and $C(q_{i_{d+1}}, p)$ crosses $C(p_i, p_a)$. The latter follows from $p_a \in (q_{i_{d+1}}, q_{i_d}]$.

It remains to prove $C(p_a, p_b)$ crosses $C(p_i, q_{i_c})$. If $C(p_a, p_b)$ crosses $C(p_i, q_i)$, then we have $q_{i_c} = q_0$. Therefore, $C(p_a, p_b]$ crosses $[p_i, q_{i_c})$, and we are done. Otherwise, by claim's assumption, $[p_a, p_b]$ is a subinterval of $[p_i, q_i]$. Then, since $q_{i_c} \in [p_a, p_b)$, $C(p_a, p_b)$ crosses $C(p_i, q_{i_c})$, and we are done. This completes the proof of conclusion (3). □

For any vertex p_i, let P_i be its shortest length chain. Now we are ready to define the set S^*. We define

$$S^* := \{v_i : \ \text{len}(P_i) \leq s\}$$

to be the set of vertices v_i such that the length P_i is at most s.

By property 2 of the above claim, if length of P_i is at most s, then we have $|\delta(\{v_i, v_{i+1}\})| \leq (1 + s\eta)\Delta$. It remains to prove that $|S^*| \geq (1 - \frac{8}{s-10})n$. We say a point p_i is *bad* if $\text{len}(P_i) > s$ (i.e., $v_i \notin S^*$). In the next lemma we show that the number of bad vertices is at most $\frac{8}{s-10}n$. This completes the proof of Lemma 3.2.9. □

Lemma 3.2.11. *The number of bad vertices is at most* $\frac{8}{s-10}n$.

Proof. We prove this claim by a double counting argument. Consider a graph D, with n vertices one for each point of the polygon and β additional "bad" vertices one for each interval $[p_i, q_i]$ corresponding to a bad point p_i. We construct a directed acyclic graph (DAG) by adding directed edges from the bad vertices to the rest (we allow directed edges between the bad vertices too). We will show that the in-degree of each vertex is at most 4, while the out-degree of each bad vertex is at least $\frac{s-2}{2}$. The lemma follows by noting that the sum of the in-degrees is equal to the sum of the

out-degrees, thus:

$$4(n+\beta) \geq \beta\frac{s-2}{2} \Rightarrow \beta \leq \frac{8}{s-10}n.$$

The construction of the graph D is quite intricate and we will do it in several steps. We say a point p_j is a *potential child* of $[p_i, q_i]$ iff $p_j \in [p_{i+2}, q_i)$. We say an interval $[p_j, q_j]$ is a *potential child* of $[p_i, q_i]$ iff it is a subinterval of $[p_i, q_i]$ and both of its endpoints are potential children of $[p_i, q_i]$. The directed edges of D are from a bad vertex to a *subset* of the potential children of its interval. Since the edges are directed only from an interval to the intervals/vertices inside it, D will be a directed acyclic graph.

We add the directed edges inductively in a bottom up manner. Consider an interval $[p_i, q_i]$ and suppose we have added the edges between the potential children of $[p_i, q_i]$, such that the outdegree of all intervals is at least $\frac{s-2}{2}$.

Let D_i be the induced subgraph of the set of potential children of $[p_i, q_i]$. We show that D_i contains at least $(s-2)/2$ *sources* (vertices of in-degree 0). Then we connect $[p_i, q_i]$ to some specific set of the sources of D_i.

Let $P_i = q_{i_0}, q_{i_1}, \ldots, q_{i_l}$ be the shortest chain assigned to a bad point p_i. Since p_i is a bad point we have $l \geq s$. Let s' be the largest odd integer smaller than s (i.e., $s' := s - \mathbb{I}[s \text{ is even}]$). Define the set of vertices $S_i := \{q_{i_3}, q_{i_5}, \ldots, q_{i_{s'}}\}$ in D_i. Note that S_i contains all vertices of P_i with odd index except q_{i_1}; this is a technical requirement and will be useful later in the proof of Claim 3.2.13.

In the next claim we show each source vertex in D_i has a directed path to at most one of the vertices of S_i. This implies that D_i contains at least $|S_i| = (s'-1)/2 \geq \frac{s-2}{2}$ sources since either the vertex in S_i is a source or there is a unique source connected to it.

Claim 3.2.12. *Any source of D_i is connected by directed paths to at most one of the vertices of S_i.*

Proof. Let $[p_j, q_j]$ be a potential child of $[p_i, q_i]$, connected by directed paths to two vertices $q_{i_a}, q_{i_b} \in S_i$, where $a + 2 \leq b$. We show that using $[p_j, q_j]$, we can obtain a shorter chain for p_i, which is a contradiction.

First note the transitivity: if x is a potential child of y, and y is a potential child of z, then x is also a potential child of z. Since each interval is only adjacent to its potential children, and this property is transitive, q_{i_a} and q_{i_b} are potential children of $[p_j, q_j]$. Therefore, $q_{i_a}, q_{i_b} \in [p_{j+2}, q_j) \subset (p_j, q_j)$. Hence, all the vertices between them in P_i, and in particular $q_{i_{a+1}}$, are also contained in (p_j, q_j).

Since $[p_j, q_j]$ is a potential child of $[p_i, q_i]$, $[p_j, q_j]$ is a subinterval of $[p_i, q_i]$, and $p_j \in [p_{i+2}, q_i)$. Therefore, since $C(p_j, q_j)$ is a near minimum cut, and at least three vertices of P_i are included in (p_j, q_j), by conclusion (3) Claim 3.2.10 we may obtain a shorter chain for p_i which is a contradiction. □

Now we are ready to define the $\frac{s'-1}{2}$ directed edges from (p_i, q_i) to its potential children: for each vertex $q_{i_j} \in S_i$, we add an edge from $[p_i, q_i]$ to one of the sources (i.e., vertices with in-degree 0) in D_i that has a directed path to q_{i_j}.

It remains to show that after the construction of D the in-degree of each vertex is at most 4. It is worth noting that, indeed some of the vertices may have in-degree more than 1. Let us give an example. Suppose $[p_i, q_i]$, and $[p_{i+1}, q_{i+1}]$ are two bad intervals. Since $[p_{i+1}, q_{i+1}]$ is not a potential child of $[p_i, q_i]$, both of $[p_i, q_i]$ and $[p_{i+1}, q_{i+1}]$ can have an edge to a source vertex in $V(D_i) \cap V(D_{i+1})$.

In the next claim we show that if two intervals are both adjacent to a vertex of D_i, then their corresponding cuts do not cross.

Claim 3.2.13. *Assume that two bad intervals $[p_i, q_i]$ and $[p_j, q_j]$ are both directly adjacent to a vertex $x \in V(D)$. Then we either have $p_j = p_{i+1}$, or $p_i = p_{j+1}$, or $p_j, q_j \in [p_i, q_i)$, or $p_i, q_i \in [p_j, q_j)$.*

Proof. Assume that $p_i \neq p_{j+1}$ and $p_j \neq p_{i+1}$. Since each bad vertex has exactly one associated interval in D, we also have $p_i \neq p_j$. First of all, note that by claim's assumption, x is a potential child of both of the intervals. Therefore, $[p_{i+2}, q_i)$ and $[p_{j+2}, q_j)$ have a non-empty intersection. Without loss of generality, assume that $p_j \in [p_i, q_i)$ (the other cases are equivalent). We need to show that $q_j \in [p_i, q_i)$. Suppose not; let $P_i = q_{i_0}, \ldots, q_{i_l}$ be the shortest chain assigned to p_i. We show that we may modify P_i and obtain a shorter chain for p_i.

Since $[p_i, q_i]$ is adjacent to x, by definition x is a source in D_i that has a directed path to one of the vertices of S_i (say q_{i_a}) in D_i (note that x may be equal to q_{i_a}). Since $[p_j, q_j]$ is also adjacent to x, by transitivity, q_{i_a} is a potential child of $[p_j, q_j]$. Moreover, since $q_j \notin [p_i, q_i)$, all of the vertices $q_{i_1}, q_{i_2}, \ldots, q_{i_a}$ are also potential children of $[p_j, q_j]$ (note that since we may have $q_j = q_{i_0} = q_i$, q_{i_0} is not necessarily a potential child of $[p_j, q_j]$). Therefore, we have $q_{i_1}, q_{i_2}, \ldots, q_{i_a} \in (p_j, q_j)$. Since by construction of S_i, $a \geq 3$, at least 3 vertices of the chain P_i is contained in (p_j, q_j).

Since $p_j \in [p_i, q_i)$, but $p_j \notin [p_i, p_i + 1]$, we have $p_j \in [p_{i+2}, q_i)$. Moreover, since $q_j \notin [p_i, q_i)$, $C(p_j, q_j)$ crosses $C(p_i, q_i)$. Therefore, since $C(p_j, q_j)$ is a near minimum cut that contains three consecutive vertices of P_i, by conclusion 3 of Claim 3.2.10 we may obtain a shorter chain for p_i, which is a contradiction. □

Now we can show that the in-degree of each vertex in D is at most 4:

Claim 3.2.14. *The in-degree of each vertex of D is at most 4.*

Proof. We prove the claim by contradiction. Let $x \in V(D)$ be a vertex with in-degree at least 5. We show that one of the 5 intervals adjacent to x is indeed a potential child of another one, and thus x was not a source vertex for at least one of the induced DAGs associated to one of these intervals, which is a contradiction.

First of all, since each bad vertex has exactly one associated interval in D, the 5 intervals must start at distinct vertices of the polygon. Therefore, among these 5 intervals we can find 3 intervals $[p_i, q_i], [p_j, q_j], [p_k, q_k]$ such that $p_i \neq p_{j+1}, p_{j-1}$, $p_j \neq p_{l+1}, p_{l-1}$, and $p_k \neq p_{i+1}, p_{i-1}$.

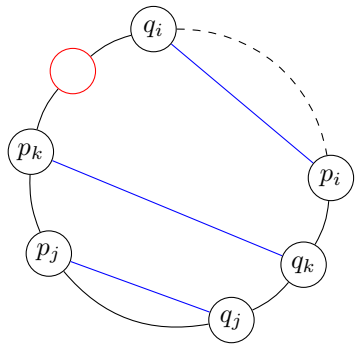

Figure 3.2.5: Setting in the proof of Claim 3.2.14. The red point is the point corresponding to vertex x.

By Claim 3.2.13, we can assume (perhaps after renaming) that $p_j, q_j, p_k, q_k \in [p_i, q_i)$. So, we must have $p_j, p_k \in [p_{i+2}, q_i)$. Therefore, $[p_j, q_j], [p_k, q_k]$ are potential children of $[p_i, q_i]$ unless they are not a subinterval of $[p_i, q_i]$ (i.e. $q_j \in [p_i, p_j)$, and $q_k \in [p_i, p_k)$).

Suppose they are not a subinterval of $[p_i, q_i]$; by Claim 3.2.13 we can assume $p_k, q_k \in [p_j, q_j)$, and $p_k \in [p_{j+2}, q_j)$. But since $[p_j, q_j], [p_k, q_k]$ are not a subinterval of $[p_i, q_i]$, $[p_k, q_k]$ must be a subinterval of $[p_j, q_j]$ (see Figure 3.2.5. So, $[p_k, q_k]$ is a potential child of $[p_j, q_j]$. This means that $[p_k, q_k] \in V(D_j)$, so x was not a source vertex for $[p_j, q_j]$ in the construction of D, and there is no directed edge from $[p_j, q_j]$ to x which is a contradiction. □

This completes the proof of Lemma 3.2.11. □

3.2.1 Cut Classes with Inside Atoms

In this subsection we show that inside atoms (vertices) of the polygon representation do not have a significant impact on the number of pairs of vertices that are highly connected. First, in Lemma 3.2.15 we show that the inside atoms are loosely connected to each other. We use this lemma in Corollary 3.2.16 to extend the proof of Lemma 3.2.9 to the case where the inside atoms are allowed. Finally, in Corollary 3.2.18 we show that only $O(\sqrt{\eta})$ fraction of atoms of any cut class can be inside atoms.

Lemma 3.2.15. *For any $S \subseteq V_{in}$, $|E(S)| \leq 2\eta \cdot (|S| - 1) \cdot \Delta$.*

Proof. We show that for any $S \subseteq V_{\text{in}}$ the minimum cut of the induced subgraph $G[S]$ is at most $2\eta \cdot \Delta$. Then, the lemma follows by a simple induction. If $(A, S - A)$ is a minimum cut of $G[S]$, then,

$$\begin{aligned} |E(G[S])| = |E(G[A])| + |E(G[S - A])| + |(A, S - A)| &\leq 2\eta \cdot (|A| - 1) \cdot \Delta + 2\eta \cdot (|S - A| - 1) \cdot \Delta + 2\eta \cdot \Delta \\ &\leq 2\eta \cdot |S| \cdot \Delta. \end{aligned}$$

Now, fix $S \subseteq V_{\text{in}}$. In the rest of the proof we show that minimum cut of $G[S]$ is at most $6\eta \cdot |S| \cdot \Delta$. We consider two cases.

There is a k-cycle $B_1, \ldots, B_k$ such that B_i crosses S for some $1 \leq i \leq k$. Let $T = B_i \cap S$. Then,

$$\begin{aligned} |E(T, S-T)| &\leq |E(\delta(B_i))| - |E(B_i \to B_{i+1})| - |E(B_i \to B_{i-1})| \\ &\leq (1+\eta)\Delta - (1-\eta)\frac{\Delta}{2} - (1-\eta)\frac{\Delta}{2} = 2\eta \cdot \Delta. \end{aligned}$$

where we used $E(B_i \to B_{i+1})$ to denote $E(B_i \cap B_{i+1}, B_{i+1} - B_i)$. The last equation follows by Lemma 2.7.2,

For any $u, v \in S$ a k-cycle for u is also a k-cycle for v. Without loss of generality assume $|S| > 1$. Let $(A, \overline{A})$ be a non-trivial near minimum cut crossing S, and let $B_1, \ldots, B_k$ be a k-cycle for S, i.e.,

$$S \subseteq V - \bigcup_{i=1}^{k} B_i.$$

Since $|S \cap A| \neq \emptyset$ by Lemma 2.7.9 for some $1 \leq i \leq k$ $B_i \subset A$. Similarly, since $|S \cap \overline{A}| \neq \emptyset$ by Lemma 2.7.9 for some $1 \leq j \leq k$ $B_j \subset \overline{A}$. Note that by definition $j - i > 1$. Following the path of crossing sets $B_i, B_{i+1}, \ldots, B_j$ we find $B_{i'}$ that crosses $(A, \overline{A})$. Similarly, following the path of crossing sets $B_j, B_{j+1}, \ldots, B_i$ we find $B_{j'}$ that crosses $(A, \overline{A})$. By definition $B_{i'} \cap B_{j'} = \emptyset$. So,

$$\begin{aligned} (1+\eta)\Delta &\geq |E(\delta(A))| \geq |E(A \to B_{i'})| + |E(A \to B_{j'})| + |E(A \cap S, \overline{A} \cap S)| \\ &\geq (1-\eta)\Delta + |E(A \cap S, \overline{A} \cap S)|. \end{aligned}$$

where the last inequality follows by Lemma 2.7.2. So, $|E(A \cap S, \overline{A} \cap S)| \leq 2\eta \cdot \Delta$. □

In the next corollary we extend the proof o Lemma 3.2.9 to the polygon representation of the cut classes that may contain inside atoms:

Corollary 3.2.16. *For any integer $s \geq 12$ and $\eta \leq 1/8$, there exists at least $m = (1 - \frac{8}{s-10})|V_{\text{out}}|$ pairs of vertices of G, $\{\{v_1, u_1\}, \{v_2, u_2\}, \ldots, \{v_m, u_m\}\}$ such that for each $1 \leq i \leq m$, $|E(v_i, u_i)| \geq (1 - s\eta)\frac{\Delta}{2}$, and each vertex $v \in V$ is contained in at most two such pairs.*

Proof. We essentially use the same proof strategy of Lemma 3.2.9. The main difference is that because of the existence of inside vertices the cuts $C(p_i, p_j)$ considered throughout the proof are not well defined (recall that for arbitrary i, j $C(p_i, p_j)$ is not necessarily a near minimum cut in G, so it may not correspond to a representing diagonal of the polygon representation).

For an interval $[p_i, p_j]$, let

$$I(p_i, p_j) := \text{argmin}_{I \subseteq V_{\text{in}}} |\delta(\{v_i, \ldots, v_{j-1}\} \cup I)|$$

be a subset of inside vertices which together with the outside vertices in the interval $[p_i, p_j]$ makes the minimum possible cut, and let

$$C(p_i, p_j) := \delta(\{v_i, v_{i+1}, \ldots, v_{j-1}\} \cup I(p_i, p_j)),$$

be that cut. Observe that if there is a representing diagonal between p_i and p_j then $C(p_i, p_j)$ is the near minimum cut corresponding to that diagonal.

The only part of the proof of Lemma 3.2.9 that we used the number of edges in $C(p_i, q_i)$ is in the proof of conclusion 2 of Claim 3.2.10. Following above definition, the same proof shows that if there is a path of length l for a point p_i, then $C(p_i, p_{i+2}) \leq (1 + \eta \cdot l)\Delta$. Unfortunately, $|C(p_i, p_{i+2})|$ being small does not immediately imply that $|E(v_i, v_{i+1})|$ is large, since $C(p_i, p_{i+2})$ may contain inside vertices. It remains to show that if $|C(p_i, p_{i+2})| \leq \Delta(1 + s\eta)$ for a point p_i, then $|E(v_i, v_{i+1})| \geq (1 - s\eta)\frac{\Delta}{2}$.

Let $I := I(p_i, p_{i+2})$. Then,

$$\begin{aligned} \Delta(1 + s\eta) \geq |C(p_i, p_i + 2)| &\geq \Delta(|I| + 2) - 2(|E(I)| - |E(I, \{v_i, v_{i+1}\})| - |E(v_i, v_{i+1})|) \\ &\geq \Delta(|I| + 2) - 4\eta(|I| - 1)\Delta - 4\eta|I| \cdot \Delta - 2|E(v_i, v_{i+1})| \end{aligned}$$

where the last inequality follows by Lemma 3.2.15, and the next claim. So,

$$|E(v_i, v_{i+1})| \geq (1 - s\eta)\Delta + \Delta(|I| - 8\eta) \geq (1 - s\eta)\Delta.$$

where the last inequality follows by the assumption $\eta \leq 1/8$.

Claim 3.2.17. *For any consecutive outside vertices v_i, v_{i+1} and any inside vertex $u \in V_{in}$,*

$$|E(u, \{v_i, v_{i+1}\})| \leq 2\eta\Delta.$$

Proof. Let $B_1, B_2, \ldots, B_k$ be a k-cycle for u. By Theorem 2.7.8 $v_i, v_{i+1} \in \cup_{i=1}^{k} B_i$. Since v_{i+1} is next to v_i, we must have $v_i, v_{i+1} \in B_i$ for some $1 \leq i \leq k$. Therefore,

$$|E(u, \{v_i, v_{i+1}\})| \leq |\delta(B_i)| - |E(B_i \to B_{i+1})| - |E(B_i \to B_{i_1})| \leq 2\eta\Delta.$$

where we used $E(B_i \to B_{i+1})$ to denote $E(B_i \cap B_{i+1}, B_{i+1} - B_i)$. The last inequality follows by Lemma 2.7.2. □

This completes the proof of Corollary 3.2.16. □

Corollary 3.2.18. *The number of inside vertices is no more than* $|V_{in}| \leq \frac{1}{1-11\eta}|V_{\text{out}}|\left[(s+3)\eta + \frac{8}{s-10}\right]$.

Proof. We use a double counting argument. Let E_{in} be the set of edges between the inside vertices, E_{out} be the set of edges between the outside vertices and $E_{\text{in-out}}$ be the set edges from the inside to outside vertices. By Lemma 3.2.15 we have $|E_{\text{in}}| \leq 2\eta|V_{\text{in}}|\Delta$. Since the degree of each inside vertex is at least Δ, we have

$$|E_{\text{in-out}}| \geq \Delta(1-4\eta)|V_{\text{in}}|. \tag{3.2.1}$$

Let $s \geq 12$ be an integer (it will be fixed later), by Corollary 3.2.16, there are $m := (1-\frac{8}{\alpha-10})|V_{\text{out}}|$ pairs of outside vertices such that the vertices in each pair are connected by at least $(1-s\eta)\frac{\Delta}{2}$ edges. We have

$$|E_{\text{out}}| \geq \frac{\Delta}{2}(1-s\eta)(1-\frac{8}{s-10})|V_{\text{out}}|. \tag{3.2.2}$$

Finally, by Lemma 3.2.7 we have

$$|V|(1+3\eta)\frac{\Delta}{2} \geq |E_{\text{in}}| + |E_{\text{out}}| + |E_{\text{in-out}}|. \tag{3.2.3}$$

By combining equations (3.2.1),(3.2.2), and (3.2.3) we obtain:

$$(|V_{\text{in}}| + |V_{\text{out}}|)(1+3\eta)\frac{\Delta}{2} \geq \Delta(1-4\eta)|V_{\text{in}}| + \frac{\Delta}{2}(1-s\eta)(1-\frac{8}{s-10})|V_{\text{out}}|$$

Therefore,

$$|V_{\text{in}}| \leq \frac{1}{1-11\eta}|V_{\text{out}}|\left((s+3)\eta + \frac{8}{s-10}\right)$$

□

Now we may complete the proof of Theorem 3.2.6:

Proof of Theorem 3.2.6. By Corollary 3.2.16 there are $m := (1-\frac{8}{s-10})|V_{\text{out}}|$ pairs of vertices of G that are connected by at least $(1-s\eta)\frac{\Delta}{2}$ edges (for any integer $s \geq 12$). By Corollary 3.2.18,

$$(1-\frac{8}{s-10})|V_{\text{out}}| \geq \left(1 - \frac{8}{s-10} - \frac{1}{1-11\eta}\left((s+3)\eta + \frac{8}{s-10}\right)\right)|V|$$

pairs of vertices of H such that each pair is connected by at least $(1-s\eta)\frac{\Delta}{2}$ edges. Let $s = \lfloor\sqrt{9/\eta}\rfloor$, Then, we can let $\alpha' = 3$ since $(1-s\eta)\frac{\Delta}{2} \geq (1-3\sqrt{\eta})\frac{\Delta}{2}$. On the other hand, since $\eta < 1/100$,

$$\begin{aligned}\frac{8}{s-10} + \frac{1}{1-11\eta}((s+3)\eta + 8/(s-10)) &\leq \frac{2-11\eta}{1-11\eta}\cdot\frac{8}{\sqrt{9/\eta}-11} + \frac{(\sqrt{9/\eta}+3)\eta}{1-11\eta} \\ &\leq 2.13\cdot 4.22\sqrt{\eta} + 1.13(3\sqrt{\eta}+3\eta) \leq 14\sqrt{\eta}\end{aligned}$$

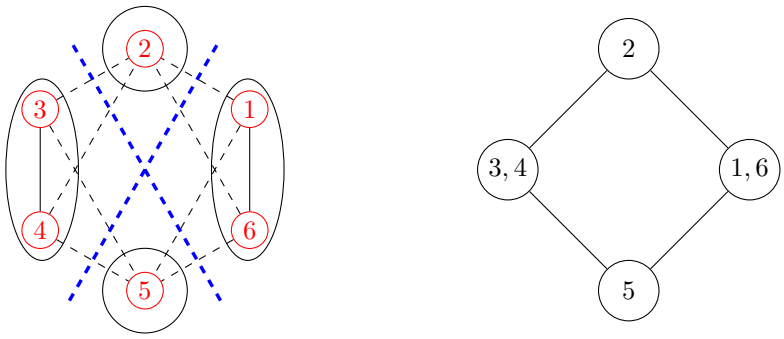

Figure 3.2.6: The left graph shows a fractional solution of LP (2.4.1). The dashed edges have fraction 1/2 and the solid edges have fraction 1. Consider the cut class $\mathcal{C}$ corresponding to the two minimum cuts shown with dashed blue lines. This cut class has 4 atoms. In the right graph we show $G(\psi(\mathcal{C}))$ where the fraction of edges between any two adjacent vertices is 1, e.g., the fraction of edge between $\{2\}$ and $\{1,6\}$ is $x_{1,2} + x_{1,6} = 1$.

so we can let $\alpha = 14$. □

3.2.2 Applications to TSP

In this section we describe an application of the machinery that we just developed to the feasible solutions of the LP relaxation of TSP. Assume that $G = (V, E, \mathbf{x})$ is a fractionally 2-edge connected 2-regular grap, i.e., $\mathbf{x}$ is a feasible solution to LP (2.4.1), and let $n := |V|$. The following corollary follows from Theorem 3.2.5.

Corollary 3.2.19. *For any $\eta < 1/100$ and any cut class $\mathcal{C}$ of $(1+\eta)$-near minimum cuts of graph $G = (V, E, \mathbf{x})$, $H = G(\psi(\mathcal{C}))$ satisfies the following:*

i) There exists at least $m := (1-14\sqrt{\eta})|\psi(\mathcal{C})|$ pairs of vertices of H, $\{(v_1, u_1), (v_2, u_2), \ldots, (v_m, u_m)\}$ such that for each $1 \le i \le m$, $x_{\{v_i,u_i\}} \ge 1 - 3\sqrt{\eta}$, and each vertex $v \in V(H)$ is contained in at most two such pairs.

ii) $|\psi(\mathcal{C})| \le \mathbf{x}(E(H)) \le (1+3\eta)|\psi(\mathcal{C})|$.

We remark that if G is a fractional graph, then $G(\psi(\mathcal{C}))$ is also a fractional graph where the fraction of each edge is the sum of the fraction of edges between the contracted vertices (see Figure 3.2.6).

Using above corollary we show that for η small enough, either a constant fraction of edges appear in a constant number of $(1+\eta)$-near minimum cuts, or x is nearly integral. If x is integral, that is if G is a Hamiltonian cycle, then every edge belongs to $n-1$ minimum cuts (see Figure 3.2.7). Our characterization proves an approximate converse of this statement: for some large constant τ, if almost all the edges are in more than τ near minimum cuts, then the graph is close to a Hamiltonian cycle in the sense that almost all of its edges are nearly integral.

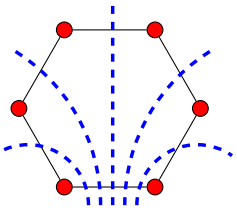

Figure 3.2.7: Each edge of a cycle of length n is contained in $n-1$ minimum cuts.

Definition 3.2.20. *For $\tau \geq 4$, we say a cut class $\mathcal{C}$ is* large *if $|\psi(\mathcal{C})| \geq \tau$, and* small *otherwise.*

Let $\text{Large}(\tau)$ be the multi-set of all atoms of the large cut classes, i.e.

$$\text{Large}(\tau) = \bigcup_{\mathcal{C}:|\psi(\mathcal{C})|\geq\tau} \psi(\mathcal{C}).$$

Note that since $\tau \geq 4$, by Lemma 2.6.3 for any two atoms A, B in $\text{Large}(\tau)$, we have $A \neq B$.

The size of $\text{Large}(\tau)$ plays an important role. It can be shown that we always have $|\text{Large}(\tau)| \leq n(1+\frac{2}{\tau-2})$ (see the proof of Lemma 3.2.21). We show that if $|\text{Large}(\tau)| < (1-\epsilon)n$, then a constant fraction of edges are in constant number of near minimum cuts. Otherwise x is nearly integral.

Lemma 3.2.21. *For any $\tau \geq 4$, $\epsilon \geq \frac{1}{\tau-2}$ and $\eta < 1/100$, if $|\text{Large}(\tau)| \geq (1-\epsilon)n$ then there is a set $E' \subseteq E$ such that $|E'| \geq (1-14\sqrt{\eta}-17\epsilon)n$ and for any $e \in E'$, $x_e \geq 1-3\sqrt{\eta}$.*

Proof. Recall that an atom $A \in \psi(\mathcal{C})$ is a singleton if $|A| = 1$. The main idea of the proof is that, for any cut class $\mathcal{C}$, an edge of fraction close to 1 in $G(\psi(\mathcal{C}))$ between two *singleton* atoms correspond to an edge of fraction close to 1 in G. Note that even an edge of fraction 1 between two non-singleton atoms many not correspond to any edge of fraction 1 in G (see Figure 3.2.6).

Let L be the number of large cut classes. By Corollary 2.6.12,

$$L \leq \frac{n}{\tau-2}.$$

Applying Corollary 2.6.4 to the set of large cut classes, there is a set P of mutually disjoint atoms of $\text{Large}(\tau)$ such that

$$|P| \geq -2(L-1) + |\text{Large}(\tau)|.$$

So, by lemma's assumption

$$|P| \geq -2(L-1) + |\text{Large}(\tau)| \geq \frac{2n}{\tau-2} + (1-\epsilon)n \geq n\Big(1-\epsilon-\frac{2}{\tau-2}\Big)$$

Since the atoms in P are mutually disjoint, there are at least $n(1-2\epsilon-\frac{4}{\tau-2})$ singletons in P. Therefore, the number of non-singleton atoms of $\text{Large}(\tau)$ is at most $n(2\epsilon+\frac{6}{\tau-2})$.

Let $H = (N, F, x)$ be the disjoint union of graphs $G(\psi(\mathcal{C}))$ for all large cut classes $\mathcal{C}$. Observe that each vertex of H corresponds to an atom of a large cut class, so H has $|\operatorname{Large}(\tau)|$ vertices. By Corollary 3.2.19, there is a subset F' of edges of H such that for each $e \in F'$, $x_e \geq 1 - 3\sqrt{\eta}$, and

$$|F'| \geq \sum_{\mathcal{C}:|\psi(\mathcal{C})|\geq\tau} |\psi(\mathcal{C})|(1 - 14\sqrt{\eta}) = |\operatorname{Large}(\tau)|(1 - 14\sqrt{\eta}) \geq n(1 - \epsilon - 14\sqrt{\eta})$$

Furthermore, each vertex of H is adjacent to at most two edges of F'. Now, let $F'' \subseteq F'$ be the set of edges such that both of their endpoints are singletons. Observe that $|F'' - F'|$ is at most twice the number of non-singleton atoms of $\operatorname{Large}(\tau)$. So,

$$|F''| \geq n(1 - \epsilon - 14\sqrt{\eta}) - 2n\left(2\epsilon + \frac{6}{\tau - 2}\right) \geq n(1 - 14\sqrt{\eta} - 15\epsilon).$$

In the last inequality we used the assumption $\epsilon > \frac{1}{\tau-2}$. The lemma follows from the fact that each edge in F'' correspond to an actual edge of G. □

Next, we show that if $|\operatorname{Large}(\tau)| < (1 - \epsilon)n$, then there is a set E'_{Small} of edges of G such that $\mathbf{x}(E'_{\text{Small}}) \geq \Omega(n)$ and each edge $e \in E'_{\text{Small}}$ is contained in only a constant number of near minimum cuts.

Definition 3.2.22. *We say an edge e is* contained *in a cut class $\mathcal{C}$ if the endpoints of e belong to two distinct atoms of $\mathcal{C}$.*

Let E_{Small} be the set of edges that are not contained in any of the large cut classes, i.e., they are only contained in small cut classes. In the next lemma we show that if $|\operatorname{Large}(\tau)| < (1 - \epsilon)n$, then $\mathbf{x}(E_{\text{Small}})$ is large:

Lemma 3.2.23. *For any $\eta < 1/100$ and $\epsilon > 6\eta$, If $|\operatorname{Large}(\tau)| < (1-\epsilon)n$ then $\mathbf{x}(E_{\text{Small}}) \geq n(\epsilon-3\eta)$. Furthermore, there is a set $E'_{\text{Small}} \subseteq E_{\text{Small}}$ such that $\mathbf{x}(E'_{\text{Small}}) \geq \epsilon \cdot n/4$ and each $e \in E'_{\text{Small}}$ is contained in at most $21\tau \cdot (\tau - 3)/\epsilon$ near minimum cuts.*

Proof. By Corollary 3.2.19, for any cut class $\mathcal{C}$,

$$\mathbf{x}(E(G(\psi(\mathcal{C})))) \leq |\psi(\mathcal{C})|(1 + 3\eta).$$

So,

$$\sum_{\mathcal{C}:|\psi(\mathcal{C})|\geq\tau} \mathbf{x}(E(G(\psi(\mathcal{C})))) \leq |\operatorname{Large}(\tau)|(1 + 3\eta) < n(1 - \epsilon)(1 + 3\eta) \leq n(1 - \epsilon + 3\eta).$$

Since $\mathbf{x}$ is a feasible solution of LP (2.4.1), $\mathbf{x}(E) = n$. For $E_{\text{Small}} = E - \cup_{\mathcal{C}:|\psi(\mathcal{C})|\geq\tau} E(G(\psi(\mathcal{C})))$, we have $\mathbf{x}(E_{\text{Small}}) \geq n(\epsilon - 3\eta) \geq n\epsilon/2$.

By part (iii) Corollary 2.6.12,

$$\sum_{\mathcal{C}:|\psi(\mathcal{C})|\leq\tau} x(E(G(\psi(\mathcal{C})))) \leq \sum_{\mathcal{C}:|\psi(\mathcal{C})|\leq\tau} |\psi(\mathcal{C})|(1+3\eta) \leq 5n(1+3\eta).$$

So, by Markov inequality there is a set $E'_{\text{Small}} \subseteq E_{\text{Small}}$ such that $\mathbf{x}(E'_{\text{Small}}) \geq \mathbf{x}(E_{\text{Small}})/2$ and any edge $e \in E'_{\text{Small}}$ is contained in at most

$$2 \cdot \frac{5n(1+3\eta)}{\mathbf{x}(E_{\text{Small}})} \leq 21/\epsilon$$

small cut classes. But by Corollary 2.7.5, each small cut class has at most $\tau \cdot (\tau - 3)$ near minimum cuts. So, each edge in E'_{Small} is contained in at most $21\tau \cdot (\tau - 3)/\epsilon$ near minimum cuts. □

3.3 Locally Hamiltonian Property of Random Spanning Trees

Throughout this section, we assume μ is a λ-random spanning tree distribution with marginals $\mathbf{z} = (1 - 1/n)\mathbf{x}$, and we use $T \sim \mu$ to denote a random tree sampled from μ. Also, for the sake of brevity of arguments we ignore $1/n$ factors by letting n sufficiently larger such than all of the constants that we consider throughout. So, we assume T preserves the marginals in $\mathbf{x}$. Unless otherwise specified, we let $\mathbf{x}$ be a feasible solution of Held-Karp relaxation (2.4.1), and $\mathbb{P}[.]$ is the probability under samples from μ.

If $\mathbf{x}$ is a feasible solution of (2.4.1), then the expected degree of all vertices under μ is 2. So, $T \sim \mu$ in expectation looks like a Hamiltonian path. But, we do not expect T to be a Hamiltonian cycle; as an example if G is a complete graph, and $x_e = x_{e'}$ for any two edges e, e', then μ is just the uniform spanning tree distribution, so the probability that T is a Hamiltonian path is exponentially small in n.

So, instead of looking for a Hamiltonian path globally, in this section we prove T *locally* looks like a Hamiltonian path. Recall that in Example 2.9.17 we show that if the expected degree of a vertex is 2 under a λ-random spanning tree distribution, then degree of that vertex is 2 with a constant probability. This shows that if we only consider a single vertex of G we see an induced Hamiltonian path with a constant probability. Now, let us consider two vertices $u, v \in V$. Can we show *both* u and v have degree 2 in T with a constant probability? In this section we answer this question. We remark that it is a fascinating open problem to extend this to more than 2 vertices, i.e., find a necessary and sufficient condition for the joint event that 3 or any constant number of vertices of G have degree 2. We have some partial results, e.g., in Chapter 6 we show any small cut class of near minimum cuts satisfy a locally Hamiltonian property.

Let us fix two vertices $u, v \in V$. By Example 2.9.18, the probability that each of them have a degree 2 can be as small as $1/e$. So, we cannot use the union bound to show both u, v have degree

2 in T with a non-zero probability. We can neither argue that the degree of u in T is independent of the degree of v, e.g., if $z_{\{u,v\}} > 0$, then conditioning on $|\delta(u) \cap T| = 2$ changes the distribution of edges adjacent to v. Therefore, we use several properties of strongly Rayleigh measures to bound this probability.

Before getting into the details of the argument we include an example that exhibit a graph where the expected degree of u, v is 2 but the probability that u, v have degree 2 simultaneously is inversely polynomial in n.

Example 3.3.1. *In this example we show in the graph illustrated in left diagram of Figure 3.3.8 the expected degree of u, v is $2 \pm O(1/\sqrt{n})$, but the probability that u, v have degree 2 simultaneously is $O(1/\sqrt{n})$.*

Consider the λ-random spanning tree distribution corresponding to the graph in the left diagram of Figure 3.3.8 where the λ values are shown next to each edge. (note that the corresponding marginal vector is not a feasible solution of (2.4.1)). Since $\lambda_{\{u',u\}}, \lambda_{\{v',v\}} = \infty$ both of these edges are in T with high probability. We show that with probability $1 - O(1/\sqrt{n})$ exactly one of the edges $\{u, v\}, \{u', v'\}$ is in T and exactly one of the edges $\{u, w\}, \{v, w\}$ is in T. This shows that the expected degree of each of u, v is about 2, but $|T \cap \delta(u)| + |T \cap \delta(v)|$ is odd with probability $1 - O(1/\sqrt{n})$ which proves our claim.

First, since $\{u, u'\}, \{v, v'\} \in T$ with high probability, at most one of $\{u, v\}, \{u', v'\}$ is in T. Also, if both $\{u, w\}, \{v, w\} \in T$, then since $\lambda_{\{u,v\}} = \sqrt{n}\lambda_{\{u,w\}}$, we can remove one of them and add $\{u, v\}$ and obtain a new tree T' such that $\mathbb{P}[T'] = \sqrt{n} \cdot \mathbb{P}[T]$. So, with probability $1 - 1/\sqrt{n}$ T has at most one of $\{u, v\}, \{u', v'\}$ and at most one of $\{u, w\}, \{v, w\}$.

It remains to show that with high probability none of $\{u, v\}, \{u', v'\}$ is in T, and none of $\{u, w\}, \{v, w\}$ is in T. Here, we show that latter, the former can be proved similarly. Suppose $\{u, w\}, \{v, w\} \notin T$. Then, T must have all of the edges of one of the two paths of length $n/2 - 1$ to connect w to u' or v'. But, in this case we can add $\{u, w\}$ to T and remove one of the $n/2 - 1$ edges of that path to obtain $\Theta(n)$ distinct trees $T_1, \ldots, T_{n/2-1}$. By definition of vector λ,

$$\sum_{i=1}^{n/2-1} \mathbb{P}[T_i] \geq \Omega(\sqrt{n})\mathbb{P}[T].$$

So, with probability $1 - O(1/\sqrt{n})$ a random tree T has at least one of $\{u, w\}, \{v, w\}$.

By the above example, we cannot show that any arbitrary pair of vertices u, v with expected degree 2 have even degree in a λ-random spanning tree with a constant probability. As we prove next, the reason is that in the above example $\mathbb{P}[\{u, v\} \in T] \approx 1/2$. In the next proposition we show that for any λ-random spanning tree distribution such that the expected degree of u, v is 2, if the probability of the edge connecting u to v is bounded away from 2 (if there is no edge between u, v then this probability is zero), then both of u, v have even degree with constant probability.

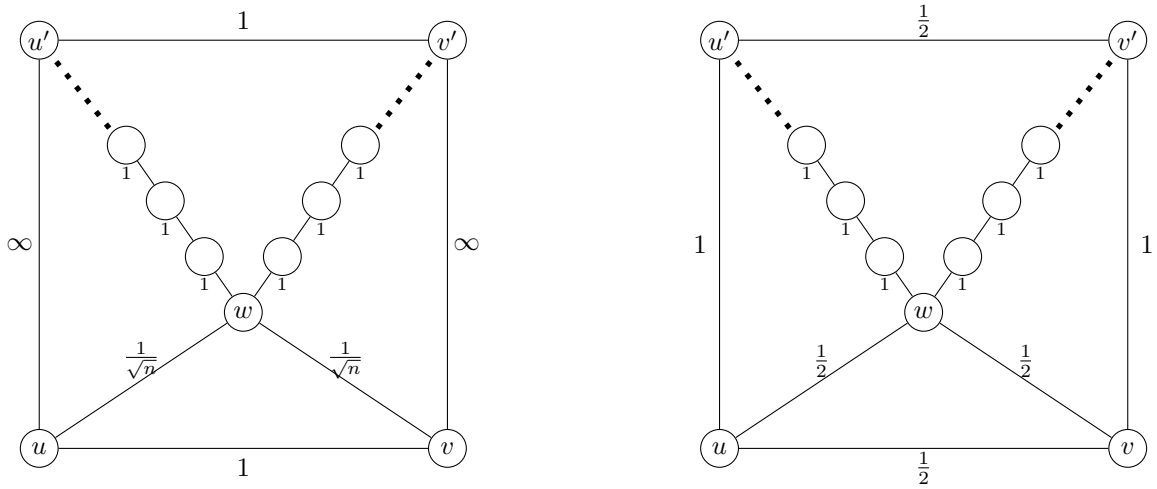

Figure 3.3.8: The left diagram represents the λ values assigned to the edges of a graph. There are two paths from w to v', u', where each path has a length of $n/2 - 1$. In the right we show the approximate probability of each edge (we dismissed $O(1/n)$ factors). Although the expected degree of u, v are two, the probability that both of them have two edges in λ-random spanning tree is $O(1/\sqrt{n})$.

Throughout this section, for vertices $u, w \in V$ we use the notation $\delta_w(u) := \{\{u, v\} : v \neq w\}$.

Proposition 3.3.2. *Let μ be a λ-random spanning tree distribution, $u, v \in V$ and $e = \{u, v\}$. If $\mathbb{E}[|\delta(u) \cap T|] = \mathbb{E}[|\delta(v) \cap T|] = 2$ and for $\epsilon < 1/10$ we either have $\mathbb{P}[e \in T] < \frac{1}{2} - \epsilon$ or $\mathbb{P}[e \in T] > \frac{1}{2} + \epsilon$, then*

$$\mathbb{P}_\mu[|T \cap \delta(u)| = 2, |T \cap \delta(v)| = 2] \geq \frac{\epsilon}{10000}.$$

We emphasize that the above proposition holds for any random spanning tree distribution as long as the expected degree of u, v is 2 and the probability of edge e is bounded away from $1/2$. We show even if $\mathbb{P}[\{u, v\} \in T] \approx 1/2$, but $\mathbb{E}[|(\delta_v(u) \cup \delta_u(v)) \cap T|]$ is bounded away from 3 then the conclusion of the above proposition holds (up to some constant loss in the right hand side.

We prove the above proposition through a sequence of lemmas. The proofs use several properties of the strongly Rayleigh measures that we discussed in Section 2.9.

Lemma 3.3.3. *Let $E_1, E_2 \subseteq E$ be two disjoint subset of edges, and let $X := |T \cap E_1|$, $Y := |T \cap E_2|$. If for $\epsilon, \alpha \geq 0$ and $0 \leq \beta \leq 1$ the following conditions are satisfied,*

$$\epsilon \leq \mathbb{P}[X + Y = 2] \tag{3.3.1}$$

$$\alpha \leq \mathbb{P}[X \leq 1], \mathbb{P}[Y \leq 1] \tag{3.3.2}$$

$$\beta \leq \mathbb{E}[X], \mathbb{E}[Y] \tag{3.3.3}$$

then

$$\mathbb{P}[X=1,Y=1] \geq \frac{\epsilon\alpha}{3}\beta(1-\beta/2)^2.$$

Proof. By (3.3.1), we have:

$$\mathbb{P}[X=1,Y=1] = \mathbb{P}[X=1,Y=1|X+Y=2]\,\mathbb{P}[X+Y=2] \geq \epsilon\mathbb{P}[X=1,Y=1|X+Y=2]$$

Let $\gamma := \alpha\beta(1-\beta/2)^2/3$. It is sufficient to show that $\mathbb{P}[X=1,Y=1|X+Y=2] \geq \gamma$. Observe that by Theorem 2.9.4 $\{\mu|X+Y=2\}$ is strongly Rayleigh, let μ' be this measure. So, by Proposition 2.9.14

$$\mathbb{P}[X=1,Y=1|X+Y=2]^2 \geq \mathbb{P}[X=2,Y=0|X+Y=2]\,\mathbb{P}[X=0,Y=2|X+Y=2]$$

Since either of the terms in the RHS can be zero, to prove the lemma it is enough to lower bound $\mathbb{P}[X\geq 1|X+Y=2]$ and $\mathbb{P}[Y\geq 1|X+Y=2]$ by $3\gamma/2$. This is because by definition $\gamma \geq 1/6$, and $\mathbb{P}_{\mu'}[X=0]+\mathbb{P}_{\mu'}[X=1]+\mathbb{P}_{\mu'}[X=2]=1$.

By symmetry, it is sufficient to show

$$\mathbb{P}[X\geq 1|X+Y=2] = \mathbb{P}[Y\leq 1|X+Y=2] \geq 3\gamma/2.$$

Since $X\geq 1$ and $Y\leq 1$ are an upward (resp. downward) event, by Theorem 2.9.11

$$\begin{aligned}\mathbb{P}[Y\leq 1|X+Y=2] &\geq \mathbb{P}[Y\leq 1|X+Y\geq 2]\\ \mathbb{P}[X\geq 1|X+Y=2] &\geq \mathbb{P}[X\geq 1|X+Y=1]\end{aligned}$$

In the second equation, if $X+Y=1$ has a zero probability, we let $\mathbb{P}[.|X+Y=1]=0$. Putting above equations together, it is sufficient to show the following,

$$\mathbb{P}[Y\leq 1|X+Y\geq 2]+\mathbb{P}[X\geq 1|X+Y=1] \geq 3\gamma. \tag{3.3.4}$$

By, the Bayes rule,

$$\begin{aligned}\mathbb{P}[Y\leq 1] &= \mathbb{P}[Y\leq 1|X+Y\geq 2]\,\mathbb{P}[X+Y\geq 2]+\mathbb{P}[Y\leq 1|X+Y\leq 1]\,\mathbb{P}[X+Y\leq 1]\\ &\leq \mathbb{P}[Y\leq 1|X+Y\geq 2]+\mathbb{P}[X+Y\leq 1]\\ &= \mathbb{P}[Y\leq 1|X+Y\geq 2]+\mathbb{P}[X=1,Y=0]+\mathbb{P}[X=0,Y\leq 1]\\ &= \mathbb{P}[Y\leq 1|X+Y\geq 2]+\mathbb{P}[X\geq 1|X+Y=1]+\mathbb{P}[X=0,Y\leq 1]. \end{aligned}\tag{3.3.5}$$

It remains to upper bound $\mathbb{P}[X=0,Y\leq 1]$. By the negative association property and (3.3.3)

$$\mathbb{E}[X|Y\leq 1] \geq \mathbb{E}[X] \geq \beta.$$

Now, if $\beta \geq 1$, by Proposition 2.9.14 the mode of X under the distribution $\{.|Y \leq 1\}$ is at least 1, so $\mathbb{P}[X = 0|Y \leq 1] \leq \frac{1}{2}$. On the other hand, if $\beta \leq 1$, by Proposition 2.9.16

$$\mathbb{P}[X = 1|Y \leq 1] \geq \mathrm{Ber}(\beta, 1) = \beta(1 - \beta/2)^2.$$

and $\mathbb{P}[X = 0|Y \leq 1] \leq 1 - \beta \cdot (1 - \beta/2)^2$. Since $\beta \leq 1$, we can put them together and obtain

$$\begin{aligned} \mathbb{P}[X = 0, Y \leq 1] = \mathbb{P}[X = 0|Y \leq 1]\,\mathbb{P}[Y \leq 1] &\leq \max\{1/2, 1 - \beta(1 - \beta/2)^2\} \cdot \mathbb{P}[Y \leq 1] \\ &\leq (1 - \beta(1 - \beta/2)^2) \cdot \mathbb{P}[Y \leq 1] \end{aligned}$$

the last inequality follows since $\beta \leq 1$. By (3.3.5) we get,

$$\mathbb{P}[Y \leq 1|X + Y \geq 2] + \mathbb{P}[X \geq 1|X + Y = 1] \geq \beta(1 - \beta/2)^2 \cdot \mathbb{P}[Y \leq 1]$$

Finally, using equation (3.3.2) we obtain equation (3.3.4) which completes the proof. □

For any edge $e \in E$ we use $I_e := |T \cap \{e\}|$ to denote the bernoulli random variable indicating that e is in T.

Corollary 3.3.4. *Let $u, v \in V$, and $e = \{u, v\}$, $X := |T \cap d_v(u)|$, $Y := |T \cap d_u(v)|$. If $\mathbb{E}[X + I_e] = 2, \mathbb{E}[Y + I_e] = 2$, $\mathbb{P}[I_e] \geq \frac{1}{2} - \frac{1}{10}$ and $\mathbb{P}[X + Y = 2|I_e = 1] \geq \epsilon$, then*

$$\mathbb{P}[X = 1, Y = 1, I_e = 1] \geq \frac{\epsilon}{200}.$$

Proof. Let $\mu' = \{\mu|I_e = 1\}$ be the corresponding λ-random spanning tree distribution on $G/\{e\}$. Observe that for any edge f,

$$\mathbb{P}_{\mu'}[f \in T] = \mathbb{P}_{\mu}[f \in T|I_e = 1].$$

We show that μ' satisfies all of the conditions of Lemma 3.3.3. Since $\mathbb{P}_{\mu'}[X + Y = 2] \geq \epsilon$, the first condition is satisfied. Since $\mathbb{E}_{\mu}[X] = \mathbb{E}_{\mu}[Y] = 2 - \mathbb{P}[e]$, by the negative association property, Corollary 2.9.9,

$$\begin{aligned} 1 &\leq \mathbb{E}_{\mu'}[X] \leq 1.5 + \frac{1}{10}, \text{ and} \\ 1 &\leq \mathbb{E}_{\mu'}[Y] \leq 1.5 + \frac{1}{10}. \end{aligned}$$

Now it is straightforward to see that $\alpha \geq \frac{3}{20}$, and $\beta = 1$ in the assumptions of Lemma 3.3.3. Therefore,

$$\mathbb{P}_{\mu}[X = 1, Y = 1, I_e = 1] = \mathbb{P}_{\mu'}[X = 1, Y = 1] \cdot \mathbb{P}_{\mu}[I_e = 1] \geq \frac{\epsilon\alpha}{3}\beta(1 - \beta/2)^2(1/2 - 1/10) \geq \frac{\epsilon}{200}$$

□

Lemma 3.3.5. *Let μ be a λ-random spanning tree distribution, $u, v \in V$, such that $\delta(u) \cap \delta(v) = \emptyset$ and $X := |T \cap \delta(u)|$ and $Y := |T \cap \delta(v)|$. If for $\epsilon, \alpha \geq 0$ and $1 \leq \beta \leq 2$ the following conditions are satisfied,*

$$\epsilon \leq \mathbb{P}[X + Y = 4] \tag{3.3.6}$$

$$\alpha \leq \mathbb{P}[X \leq 2], \mathbb{P}[Y \leq 2] \tag{3.3.7}$$

$$\beta \leq \mathbb{E}[X], \mathbb{E}[Y] \tag{3.3.8}$$

then

$$\mathbb{P}[X = 2, Y = 2] \geq \frac{\epsilon\alpha}{3}(\beta - 1)(1 - (\beta - 1)/2)^2.$$

Proof. The proof is very similar to Lemma 3.3.3. The only difference is that in this case the random variables X, Y are at least 1 with probability 1. So, if we define $X' = X - 1, Y' = Y - 1$, we can rewrite the same equations for X', Y', e.g., observe that $\{\mu | X' + Y' = 2\}$ is the same as $\{\mu | X + Y = 4\}$ and is strongly Rayleigh. So, we obtain the same conclusion as of Lemma 3.3.3 except that β is replaced with $\beta - 1$. □

Corollary 3.3.6. *Let μ be a λ-random spanning tree distribution, $u, v \in G$ and $e = \{u, v\}$, $X := |T \cap d_v(u)|$, $Y := |T \cap d_u(v)|$, and $I_e := |T \cap \{(u, v)\}|$. If $\mathbb{E}[X + I_e] = \mathbb{E}[Y + I_e] = 2$, $\mathbb{P}_\mu[I_e] \leq \frac{1}{2} + \frac{1}{10}$ and $\mathbb{P}_{T \sim \mu}[X + Y = 4 | I_e = 0] \geq \epsilon$, then*

$$\mathbb{P}[X = 2,\ Y = 2,\ I_e = 0] \geq \frac{\epsilon}{100}.$$

Proof. Let $\mu' = \{\mu | I_e = 0\}$ be the measure obtained from μ conditioned on $I_e = 0$. We show that μ' satisfies all of the conditions of Lemma 3.3.5. Since $\mathbb{P}_{\mu'}[X + Y = 4] \geq \epsilon$, the first condition is satisfied. Since $\mathbb{E}_\mu[X] = \mathbb{E}_\mu[Y] = 2 - \mathbb{P}_\mu[I_e]$, Fact 2.9.8 implies that

$$1.5 - \frac{1}{10} \leq \mathbb{E}_{\mu'}[X] \leq 2$$

$$1.5 - \frac{1}{10} \leq \mathbb{E}_{\mu'}[Y] \leq 2$$

Now it is straightforward to see that $\alpha \geq \frac{1}{3}$ using Markov's inequality, and $\beta \geq 1.5 - \frac{1}{10}$ in the assumptions of Lemma 3.3.5. Therefore,

$$\mathbb{P}_\mu[X = 2, Y = 2, I_e = 0] = \mathbb{P}_{\mu'}[X = 2, Y = 2] \cdot \mathbb{P}[I_e = 0] \geq \frac{\epsilon\alpha}{3}(\beta - 1)(1 - (\beta - 1)/2)^2(1/2 - 1/10) \geq \frac{\epsilon}{100}.$$

□

Now we are ready to prove Proposition 3.3.2.

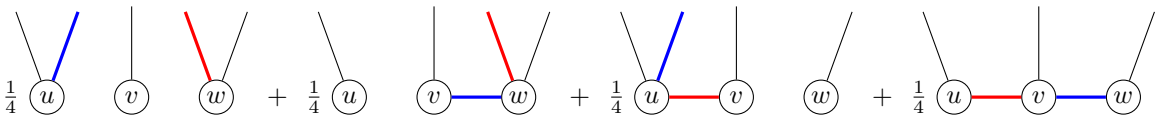

Figure 3.3.9: A strongly Rayleigh distribution of subgraphs of a graph G where the marginal probability of each of $\{u, v\}$ and $\{v, w\}$ is $1/2$, the expected degree of each of u, v, w is 2, but the degree of v is even if and only if the degree of each of u and w is odd.

Proof of Proposition 3.3.2. Let $X := |T \cap \delta_v(u)|$ and $Y := |T \cap \delta_u(v)|$ be the random variables indicating the number of edges of $\delta_v(u)$ and $\delta_u(v)$ that are sampled in $T \sim \mu$. If $\mathbb{E}[I_e] < 1/2 - \epsilon$, then by negative association property, Fact 2.9.8 and Corollary 2.9.9, we have

$$3 + 2\epsilon \leq \mathbb{E}[X + Y | I_e = 0] \leq 4.$$

So, by Proposition 2.9.16

$$\mathbb{P}[X + Y = 4 | I_e = 0] \geq \mathrm{Ber}(3 + 2\epsilon, 4) \geq 2\epsilon e^{-3.5} \geq \epsilon/20.$$

and therefore the conclusion follows by Corollary 3.3.6.

If $x_e > 1/2 + \epsilon$ the statement can be proved similarly by conditioning on $I_e = 1$. By Corollary 2.9.9,

$$2 \leq \mathbb{E}[X + Y | I_e = 1] \leq \mathbb{E}[X + Y] \leq 3 - 2\epsilon.$$

Therefore, by Proposition 2.9.16

$$\mathbb{P}[X + Y = 2 | I_e = 1] \geq \mathrm{Ber}(3 - 2\epsilon, 2) \geq \min\{2\epsilon/3, (1 - 3/4)^4\} \geq \epsilon/30.$$

The proposition follows by Corollary 3.3.4. □

Now we discuss the case where there is an edge of probability close to $1/2$ between u and v. Recall that in Example 2.9.17 we show that if $x_{\{u,v\}} \approx 1/2$, then it could be that $\mathbb{P}[|T \cap \delta(u)| = 2, |T \cap \delta(v)| = 2] \approx 0$. In the following lemma we show that if $x_{\{u,v\}} \approx 1/2$ and $x_{\{v,w\}} \approx 1/2$ then with a constant probability either u and v will have an even degree, or v and w will have an even degree. The proof of the lemma is different from the above proofs, and we use properties of λ-random spanning trees that do not necessarily extend to generalize to strongly Rayleigh measures. In particular, we use the connection between random spanning trees and electrical flows in the proof.

Let us show that this is necessary, i.e., there is a strongly Rayleigh distribution of subgraphs of G such that $x_{\{u,v\}}, x_{\{v,w\}} = 1/2$ and the expected degree of each of u, v, w is 2, but the probability that both u and v have even degree is zero, and the probability that both v and w have even degree is zero. Consider the graph in Figure 3.3.9. This is a product distribution on blue/red edges, so is

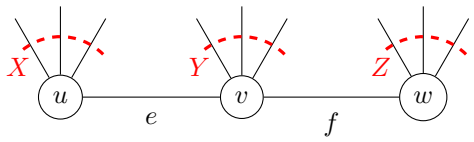

Figure 3.3.10: Setting of Lemma 3.3.7. Random variables X, Y, Z count the number of edges adjacent to u, v, w (except e, f that are in a random tree T.

strongly Rayleigh. It is also easy to verify that the marginal probability of each of $\{u, v\}, \{v, w\}$ is $1/2$ and the expected degree of the vertices are 2.

For any two random variables X, Y, let $\mathrm{Cov}(X, Y) := \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$. For example, by Theorem 2.8.8, for any two edges $e, f \in E$,

$$\mathrm{Cov}(I_e, I_f) = \mathbb{P}[I_e](\mathbb{E}[I_f | I_e = 1] - \mathbb{P}[I_f]) = -i^e(f)i^f(e). \tag{3.3.9}$$

As another example, for any random variable X and any edge e,

$$\mathbb{E}[X | I_e = 1] = \mathbb{E}[X] + \frac{\mathrm{Cov}(I_e, X)}{\mathbb{P}[I_e = 1]}, \tag{3.3.10}$$

$$\mathbb{E}[X | I_e = 0] = \mathbb{E}[X] - \frac{\mathrm{Cov}(I_e, X)}{\mathbb{P}[I_e = 0]}. \tag{3.3.11}$$

Lemma 3.3.7. *Let μ be a λ-random spanning tree distribution, $u, v, w \in V$, $e = \{u, v\}, f = \{v, w\}$. If $\mathbb{E}[|\delta(u) \cap T| = 2]$, $\mathbb{E}[|\delta(v) \cap T|] = 2$ and for $\epsilon = 1/1000$,*

$$\frac{1}{2} - \epsilon \leq \mathbb{P}[e \in T], \mathbb{P}[f \in T] \leq \frac{1}{2} + \epsilon,$$

then at least one of the following two equations hold,

$$\mathbb{P}[|T \cap \delta(u)| = 2, |T \cap \delta(v)| = 2] \geq \frac{\epsilon}{2500} \quad , \quad \mathbb{P}[|T \cap \delta(v)| = 2, |T \cap \delta(w)| = 2] \geq \frac{\epsilon}{2500}.$$

Proof. Let $X := |T \cap \delta_v(u)|, Y := |T \cap \delta_u(v) \cap \delta_w(v)|, Z := |T \cap \delta_v(w)|$ be the random variables indicating the number of edges of $\delta_v(u), \delta_u(v) \cap \delta_w(v), \delta_v(w)$ sampled in T, see Figure 3.3.10. Note that by definition $\mathbb{E}[X] \approx \mathbb{E}[Z] \approx 1.5$, and $\mathbb{E}[Y] \approx 1$).

Let $c := -\mathrm{Cov}(I_e, I_f)$ for an orientation of the edges that we fix later in the proof. First, assume that $c \geq \epsilon + \epsilon^2$ is large. We show the lemma follows by Corollary 3.3.4. Let $\mu' = \{\mu | I_e = 1\}$. By (3.3.10)

$$\mathbb{P}_{\mu'}[I_f] = \frac{\mathrm{Cov}(I_e, I_f)}{\mathbb{P}[I_e = 1]} + \mathbb{P}[I_f] \leq \mathbb{P}[I_f] - 2\epsilon.$$

where we used the assumption that $\mathbb{P}[I_e] \leq 1/2 + \epsilon$. Therefore, $\mathbb{E}_{\mu'}[X + Y + I_f] \leq 3 - \epsilon$. By

Proposition 2.9.16

$$\mathbb{P}_{\mu'}\left[X+Y+I_f=2\right] \geq \mathrm{Ber}(3-\epsilon,2) \geq \min\{\epsilon/3,(1-3/4)^4\}.$$

So, by Corollary 3.3.4,

$$\mathbb{P}\left[X=1,Y+I_f=1,I_e=1\right] \geq \frac{\min\{\epsilon/3,4^{-4}\}}{200},$$

and we are done.

In the rest of the proof assume $c<\epsilon(1+\epsilon)$. First we show $\mathrm{Cov}(I_f,X)$ or $\mathrm{Cov}(I_e,Z)$ is very close to 0, i.e., we show I_e is essentially independent of Z or I_f is essentially independent of X. Then we use Corollary 3.3.4 to prove the lemma.

By (3.3.9),

$$\min\{|i^e(f)|,|i^r(e)|\} \leq \sqrt{c} \leq \sqrt{\epsilon(1+\epsilon)}.$$

Without loss of generality, assume $|i^f(e)| \leq i^e(f)$ satisfies the above equation. We fix the following orientation of the edges: any edge $\{u,u'\}$ adjacent to u is oriented outwards from u to u', f is oriented from w to v and the rest of the edges have arbitrary orientations. Let $i^f_c(.)$ be the function $i^f(.)$ in the graph $G/\{e\}$. By Fact 2.8.5, for any edge e' adjacent to u

$$i^f(e') = i^f_c(e') + \frac{i^f(e)}{i^e(e)} i^e(e') \leq \frac{\sqrt{c}}{i^e(e)} i^e(e'). \tag{3.3.12}$$

In the above inequality we used an important observation that $i^f_c(e') \leq 0$, this is because e' is oriented outwards from u, and $i^f_c(.)$ is the current function when we are sending one unit of current from w to v in $G/\{e\}$. By Theorem 2.8.8,

$$\begin{aligned}
-\mathrm{Cov}(I_f,X) = \sum_{e'\in\delta_v(u)} -\mathrm{Cov}(I_f,I_{e'}) = \sum_{e'\in\delta_v(u)} i^f(e')i^{e'}(f) \leq \sum_{e'\in\delta(u)} |i^f(e')| &\leq 2 \sum_{e'\in\delta(u):i^f(e')\geq 0} i^f(e') \\
&\leq \frac{2\sqrt{c}}{i^e(e)} \sum_{e'\in\delta(u):i^f(e')\geq 0} i^e(e') \\
&\leq \frac{2\sqrt{c}}{i^e(e)} \sum_{e'\in\delta(u)} i^e(e') \leq \frac{2\sqrt{c}}{i^e(e)}.
\end{aligned}$$

where the first inequality follows by Fact 2.8.4, the second inequality follows by Kirchhoff, that is $\sum_{e'\in\delta(u)} = i^e(e') = 0$, the third inequality follows by (3.3.12), and the fourth inequality follows by the fact that $i^e(e') \geq 0$ for any edge $e' \in \delta(u)$. Since we assumed $c \approx \sqrt{\epsilon}$ is close to zero, the above inequality shows I_f is essentially independent of X. Building on this observation we can use Lemma 3.3.3 to prove the lemma.

Let $\mu' = \{\mu | I_f = 0\}$. By (3.3.11) we can upper bound $\mathbb{E}_{\mu'}[X]$ as follows,

$$\mathbb{E}_{\mu'}[X] = \mathbb{E}[X|I_f = 0] = \mathbb{E}[X] - \frac{\mathrm{Cov}(I_f, X)}{\mathbb{P}[I_f = 0]} \leq \mathbb{E}[X] + \frac{2\sqrt{c}}{\mathbb{P}[I_e = 1] \cdot \mathbb{P}[I_f = 0]} \leq \mathbb{E}[X] + 9\sqrt{\epsilon}. \tag{3.3.13}$$

If $\mathbb{E}_{\mu'}[Y + I_e + Z] \geq 3 + \epsilon$, then by Proposition 2.9.16

$$\mathbb{P}_{\mu'}[Y + I_e + Z = 4] \geq \mathrm{Ber}(3 + \epsilon, 4) \geq \epsilon e^{-3-\epsilon} \geq \epsilon/25,$$

and by Corollary 3.3.6

$$\mathbb{P}[Y + I_e = 2, I_f = 0, Z = 2] \geq \frac{\epsilon}{2500}$$

and we are done. Therefore, we assume $\mathbb{E}_{\mu'}[Y + I_f + Z] \leq 3 + \epsilon$.

Now, let $\mu'' = \{\mu | I_e = 1\}$. We want to use Lemma 3.3.3 to show that $\mathbb{P}_{\mu''}[X = 1, Y = 1]$ is a constant. By negative association, Fact 2.9.8,

$$\begin{aligned} 1 \quad &\leq \mathbb{E}_{\mu''}[X] \leq \quad \mathbb{E}[X] + 9\sqrt{\epsilon} \leq 1.5 + \epsilon + 9\sqrt{\epsilon} \\ 0.5 - \epsilon \quad &\leq \mathbb{E}_{\mu''}[Y] \leq \quad 1 + 3\epsilon. \end{aligned}$$

where the first equations hold by equation (3.3.13). Letting $\epsilon = 0.001$, by Proposition 2.9.16

$$\mathbb{P}[X + Y = 2] \geq \min\{\mathrm{Ber}(2.8, 2), \mathrm{Ber}(1.5 - 0.001, 2)\} \geq \frac{1}{150}$$

and we let $\beta = 0.5 - 0.001$ and $\alpha = 1/10$, by Lemma 3.3.3

$$\mathbb{P}_{\mu''}[X = 1Y = 1] \geq \frac{\alpha}{150 \cdot 3} \cdot \beta(1 - \beta/2)^2 \geq \frac{1}{20000}.$$

So,

$$\begin{aligned} \mathbb{P}[X = 1, Y = 1, I_e = 1, I_f = 0] \quad &\geq \quad \mathbb{P}_{\mu'}[X = 1, Y = 1, I_e = 1]\,\mathbb{P}[I_f = 0] \\ &\geq \quad \mathbb{P}_{\mu''}[X = 1, Y = 1] \cdot \mathbb{P}[I_e = 1 | I_f = 0]\,\mathbb{P}[I_f = 0] \geq 10^{-5}. \end{aligned}$$

where we used $\mathrm{Cov}(I_e, I_f) \leq \epsilon + \epsilon^2$. This completes the proof of Lemma 3.3.7. □

Chapter 4

Asymmetric TSP

In this section we provide an $O(\log n/\log\log n)$ randomized approximation algorithm for the Asymmetric TSP. Our algorithm with high probability returns solution of cost $O(\log n/\log\log n)$ of the optimum solution of the LP (2.4.2). Our proof also shows that the integrality gap of the Held-Karp relaxation for the Asymmetric TSP (2.4.2) is bounded above by $O(\log n/\log\log n)$. We refer the reader to Section 2.1 for background. The results in this section are based on a joint work with Arash Asadpour, Michel Goemans, Aleksander Madry and Amin Saberi [AGM+10].

4.1 Introduction

Our approach for ATSP (and for STSP) is very similar to the Christofides algorithm (see Algorithm 4). First we choose a spanning tree T and then we add the *minimum cost Eulerian augmentation*. An Eulerian augmentation is a set of edges to make T an Eulerian graph (i.e., for the case of ATSP an Eulerian augmentation makes the in-degree of each vertex equal to its out-degree, and in the case of STSP it makes the degree of each vertex even). In the case of STSP the minimum cost Eulerian augmentation can be computed efficiently by finding a minimum cost perfect matching on odd degree vertices of the tree, and for ATSP it can be computed efficiently by solving a minimum cost flow problem that makes the in-degree of vertices equal to their out-degree (see equation (4.2.1) for details).

The main difference between our algorithm and Algorithm 4 is in the selection of the tree. Instead of choosing an arbitrary minimum spanning tree, we select our tree randomly by rounding an optimal solution of the LP relaxation of TSP using the rounding by sampling method that we discussed in Section 3.1. By (3.1.2), for any feasible solution $\mathbf{x}$, $\mathbb{E}[c(T)] \leq c(\mathbf{x})$. So, the main difficulty in our proofs is in upper bounding the cost of the Eulerian augmentation. For the Asymmetric TSP we show this cost is at most $O(\log n/\log\log n)\cdot c(\mathbf{x})$, and for Symmetric TSP we upper bound this cost by $(1/2-\epsilon)\cdot c(\mathbf{x})$ for some constant $\epsilon > 0$. In this chapter we define a combinatorial object known

as "thin tree" and we use it to bound the cost of the Eulerian augment. For Symmetric TSP we use the O-join polytope (2.4.4) to upper bound the cost of Eulerian augmentation.

A simple flow argument using Hoffman's circulation theorem [Sch03] shows that if the tree chosen in the first step is "thin" with respect to a feasible solution $\mathbf{x}$ of LP (2.4.2), then the cost of the Eulerian augmentation is within a factor of the "thinness" of $c(\mathbf{x})$. This flow argument works irrespectively of the actual directions of the (directed) arcs corresponding to the (undirected) edges of the tree. Roughly speaking, a *thin* tree with respect to a feasible solution $\mathbf{x}$ of LP (2.4.2) is a spanning tree that, for every cut, contains a small multiple (the *thinness*) of the corresponding value of $\mathbf{x}$ in this cut when the direction of the arcs are disregarded.

The high level description of our algorithm can be found in Algorithm 6. The proof of our main Theorem 4.3.3 also gives a more formal overview of the algorithm.

Algorithm 6 An $O(\log n/\log\log n)$ Approximation Algorithm for the Asymmetric TSP

1: Let $\mathbf{x}^*$ be an optimal solution of LP (2.4.2). Let $\mathbf{z}^*$ be a fractional spanning tree where for all $u, v \in V$, $z^*_{\{u,v\}} = (1-1/n)(x^*_{u,v} + x^*v, u)$, and let $G = (V, E, \mathbf{z}^*)$ be underlying graph of the fractional spanning tree z.

2: Find weights $\lambda : E \to \mathbb{R}_+$ such that the λ-random spanning tree distribution, μ, approximates the marginal probabilities imposed by $\mathbf{z}^*$, i.e., for all $e \in E$,

$$\mathbb{P}_\mu [e \in T] \leq (1 + 1/n) z^*_e$$

3: Sample a tree T^* from μ.

4: Orient each edge of T so as to minimize its cost. Find a minimum cost integral circulation that contains the oriented tree $\vec{T}^*$. Shortcut this multigraph and output the resulting Hamiltonian cycle.

4.1.1 Notations

Throughout this chapter we assume $\mathbf{x}$ is a feasible solution Held-Karp relaxation for Asymmetric TSP, i.e., LP (2.4.2). Let A denote the support of $\mathbf{x}$, i.e. $A = \{(u, v) : x_{u,v} > 0\}$.

Similar to (3.1.1) we can construct a fractional spanning tree from $\mathbf{x}$. We first make this solution symmetric and then we scale it down by $(1 - 1/n)$,

$$z_{\{u,v\}} := (1 - 1/n) \cdot (x_{u,v} + x_{v,u}). \tag{4.1.1}$$

We let E be the support of $\mathbf{z}$. For every edge $e = \{u, v\}$ of E, we can define its cost as $\min\{c(a) : a \in \{(u, v), (v, u)\} \cap A\}$; with the risk of overloading the notation, we denote this new cost of this edge e by $c(e)$. This implies that $c(\mathbf{z}^*) < c(\mathbf{x}^*)$.

4.2 Thin Trees and Asymmetric TSP

In this section we reduce the problem of approximating Asymmetric TSP to the problem of finding a thin tree. Then, in the next section we show that the tree sampled from the maximum entropy distribution is almost surely "thin".

First, we define the thinness property.

Definition 4.2.1. *We say that a tree T of $G = (V, E)$ is α-thin with respect to a vector $\mathbf{z} : E \to \mathbb{R}_+$ if for each set $S \subset V$,*

$$|T \cap \delta(S)| \leq \alpha \mathbf{z}(\delta(S)).$$

Also we say that T is (α, s)-thin if it is α-thin and moreover,

$$c(T) \leq sc(\mathbf{z}).$$

Note that thinness property can be defined for any subgraph of a graph G. Observe that if a subgraph H is (α, s)-thin with respect to a vector $\mathbf{z}$, then any subgraph of H is also (α, s)-thin with respect to $\mathbf{z}$.

We give a few remarks about the above definition. In the next theorem we show that if we have a (α, s)-thin tree with respect to a feasible solution $\mathbf{x}$ of LP (2.4.2), then we can obtain an ATSP tour of cost at most $(2\alpha + s)c(\mathbf{x})$. So, this reduces the problem of approximating ATSP to the problem of finding a (α, s) thin tree for a small values of α and s. The advantage of this new problem is that we do not need to worry about the direction of the arcs in the graph, and we can simply treat them as undirected edges. In Chapter 5 we show that an algorithm for finding an α-thin tree is also sufficient to find an $O(\alpha)$-approximation to ATSP, in other words we can also drop the cost of the edges and only look for α-thin trees for small values of α. Apart from these advantage, the problem of finding a thin tree has several inherent hardness. For example, for a given tree there is no efficient algorithm to find its thinness (see Chapter 5 for a more detailed discusson).

Theorem 4.2.2. *Let $\mathbf{x}$ be a feasible solution $\mathbf{x}$ of LP (2.4.2) and $\mathbf{z}$ be a fractional spanning tree as defined in (4.1.1), and let A be the support of $\mathbf{x}$ and E be the support of $\mathbf{z}$. Given a (α, s)-thin spanning tree $T \subseteq E$ with respect to $\mathbf{z}$, we can find a Hamiltonian cycle of cost no more than $(2\alpha + s)c(\mathbf{x})$.*

Before proceeding to the proof of Theorem 4.2.2, we recall some basic network flow results related to circulations. A function $f : A \to \mathbb{R}$ is called a *circulation* if $f(\delta^+(v)) = f(\delta^-(v))$ for each vertex $v \in V$. Hoffman's circulation theorem [Sch03, Theorem 11.2] gives a necessary and sufficient condition for the existence of a circulation subject to lower and upper capacities on arcs.

Theorem 4.2.3 (Hoffman's circulation theorem)**.** *For any directed graph $G = (V, A)$, and any two functions $f_l, f_u : A \to \mathbb{R}$, there exists a circulation f satisfying $f_l(a) \leq f(a) \leq f_u(a)$ for all $a \in A$* if and only if

1. $f_l(a) \le f_u(a)$ *for all* $a \in A$ *and*

2. *for all subsets* $S \subset V$, *we have* $f_l(\delta^-(S)) \le f_u(\delta^+(S))$.

Furthermore, if f_l *and* f_u *are integer-valued,* f *can be chosen to be integer-valued.*

Proof of Theorem 4.2.2. We first orient each edge $\{u, v\}$ of T to $\arg\min\{c(a) : a \in \{(u, v), (v, u)\} \cap A\}$, and denote the resulting directed tree by $\vec{T}$. Observe that by definition of our undirected cost function, we have $c(\vec{T}) = c(T)$. We then find a minimum cost augmentation of $\vec{T}$ into an Eulerian directed graph; this can be formulated as a minimum cost circulation problem with integral lower capacities (and no or infinite upper capacities). Indeed, set

$$f_l(a) = \begin{cases} 1 & a \in \vec{T} \\ 0 & a \notin \vec{T}, \end{cases}$$

and consider the minimum cost circulation problem

$$\min\{c(f) : f \text{ is a circulation and } f(a) \ge f_l(a)\ \forall a \in A\}. \tag{4.2.1}$$

An optimum circulation f^* can be computed in polynomial time and can be assumed to be integral, see e.g. [Sch03, Corollary 12.2a]. This integral circulation f^* corresponds to a directed (multi)graph H which contains $\vec{T}$. Every vertex in H has an in-degree equal to its out-degree. Therefore, every cut has the same number of arcs in both directions. On the other hand, since H contains $\vec{T}$, it is weakly connected, i.e., it has at least one edge in every cut in G, Therefore H is strongly connected, so H is an Eulerian directed multigraph. We can extract an Eulerian walk of H and shortcut it to obtain our Hamiltonian cycle of cost at most $c(f^*)$ since the costs satisfy the triangle inequality. Note that we only used the tree T to make sure that H is weakly connected, so any other connected subgraph of $G = (V, E)$ would give us an Eulerian directed multigraph.

To complete the proof of Theorem 4.2.2, it remains to show that $c(f^*) \le (2\alpha + s)c(\mathbf{x})$. For this purpose, we define

$$f_u(a) = \begin{cases} 1 + 2\alpha x_a & a \in \vec{T} \\ 2\alpha x_a & a \notin \vec{T}. \end{cases}$$

We claim that there exists a circulation g satisfying $f_l(a) \le g(a) \le f_u(a)$ for every $a \in A$. To prove this claim, we use Hoffman's circulation theorem 4.2.3. Indeed, by construction, $l(a) \le u(a)$ for every $a \in A$; furthermore, Lemma 4.2.4 below shows that, for every $S \subset V$, we have $f_l(\delta^-(S)) \le f_u(\delta^+(S))$. Thus the existence of the circulation g is established. Furthermore,

$$c(f^*) \le c(g) \le c(f_u) = c(\vec{T}) + 2\alpha c(\mathbf{x}) \le (2\alpha + s)c(\mathbf{x}),$$

establishing the bound on the cost of f^*. This completes the proof of Theorem 4.2.2. □

Lemma 4.2.4. *For the capacities f_l and f_u as constructed in the proof of Theorem 4.2.2, the following holds for any subset $S \subset V$:*

$$f_l(\delta^-(S)) \leq f_u(\delta^+(S)).$$

Proof. Since $\mathbf{x}(\delta^-(v)) = \mathbf{x}(\delta^+(v))$ for all $v \in V$,

$$\mathbf{x}(\delta^-(S)) = \mathbf{x}(\delta^+(S)) \tag{4.2.2}$$

Irrespective of the orientation of T into $\vec{T}$, the number of arcs of $\vec{T}$ in $\delta^-(S)$ is at most $\alpha \cdot \mathbf{z}(\delta(S))$ by definition of α-thinness. Thus,

$$f_l(\delta^-(S)) \leq |T \cap \delta(S)| \leq \alpha \cdot \mathbf{z}(\delta(S)) < \alpha \cdot (\mathbf{x}(\delta^-(S)) + \mathbf{x}(\delta^+(S))) = 2\alpha \cdot \mathbf{x}(\delta^-(S)),$$

where the last equality follows by (4.2.2). On the other hand, we have

$$f_u(\delta^+(S)) \geq 2\alpha \cdot \mathbf{x}(\delta^+(S)) = 2\alpha \cdot \mathbf{x}(\delta^-(S)) \geq f_l(\delta^-(S)),$$

where we have used (4.2.2). The lemma follows. □

4.3 Construction of a Thin Tree

In this section prove our main theorem. First, we show that a λ-random spanning tree T sampled from the (approximately) maximum entropy distribution that preserve z_e as the marginal probability of each edge $e \in E$ is $\alpha = O(\log n / \log\log n)$-thin with high probability. Combining this with Theorem 4.2.2 proves our main theorem.

Proposition 4.3.1. *Let $\mathbf{z}$ be a fractional spanning tree and let E be the support of $\mathbf{z}$. Let $\lambda : E \to \mathbb{R}_+$ such that the λ-random spanning tree distribution μ satisfies,*

$$\forall e \in E, \quad \mathbb{P}_{T \sim \mu}[e \in T] \leq (1+\epsilon) z_e.$$

If $\epsilon < 0.2$, then $T \sim \mu$ is α-thin with probability at least $1 - 4/\sqrt{n}$ for $\alpha = 4 \log n / \log\log n$.

Proof. In Lemma 4.3.2 we show that the probability that a particular cut $\delta(S)$ violates the α-thinness of T is at most $n^{-2.5\mathbf{z}^*(\delta(S))}$. Now, by Theorem 2.7.10 there are at most $4n^{2\beta}$ cuts of size at most β times the minimum cut of $G = (V, E, \mathbf{z})$ for any $\beta \geq 1$. Since, $\mathbf{z}$ is a convex combination of all spanning trees of G, $\mathbf{z}(\delta(S)) \geq 1$ for any $S \subseteq V$. Therefore there are at most $4n^{2\beta}$ cuts $\delta(S)$ with $\mathbf{z}(\delta(S)) \leq \beta$ for any $\beta \geq 1$. Therefore, by applying the union bound, we derive that the probability

that there exists some cut $\delta(S)$ with $|T \cap \delta(S)| > \alpha \mathbf{z}^*(\delta(S))$ is at most

$$\int_{x=1}^{\infty} 4n^{2x} n^{-2.5x} dx \leq \int_{x=1}^{\infty} \frac{4}{n^{x/2}} dx \leq 4/\sqrt{n},$$

where each term is an upper bound on the probability that there exists a violating cut of size within $[i-1, i]$. Thus, indeed, $T \sim \mu$ is a α-thin spanning tree with high probability.

Lemma 4.3.2. *If $\epsilon < 0.2$, then for any set $S \subset V$,*

$$\mathbb{P}\left[|T \cap \delta(S)| > \alpha \cdot \mathbf{z}(\delta(S))\right] \leq n^{-2.5\mathbf{z}(\delta(S))},$$

where $\alpha = 4 \log n / \log \log n$.

Proof. By linearity of expectation,

$$\mathbb{E}\left[|T \cap \delta(S)|\right] = (1+\epsilon)\mathbf{z}(\delta(S)).$$

Applying Theorem 2.8.7 with

$$1 + \delta = \alpha \frac{\mathbf{z}(\delta(S))}{\mathbb{E}\left[|T \cap \delta(S)|\right]} \geq \frac{\alpha}{1+\epsilon},$$

we derive that $\mathbb{P}\left[|T \cap \delta(S)| > \alpha \cdot \mathbf{z}(\delta(S))\right]$ can be bounded from above by

$$\begin{aligned}
\mathbb{P}\left[|T \cap \delta(S)| > (1+\delta)\mathbb{E}\left[|T \cap \delta(S)|\right]\right] \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^{\mathbb{E}[|T\cap\delta(S)|]} &\leq \left(\frac{e}{1+\delta}\right)^{(1+\delta)\mathbb{E}[|T\cap\delta(S)|]} \\
&= \left(\frac{e}{1+\delta}\right)^{\alpha\mathbf{z}(\delta(S))} \\
&\leq \left[\left(\frac{e(1+\varepsilon)}{\alpha}\right)^{\alpha}\right]^{\mathbf{z}(\delta(S))} \\
&\leq n^{-4(1-1/e)\mathbf{z}(\delta(S))},
\end{aligned}$$

where, in the last inequality, we have used that

$$\begin{aligned}
\log\left[\left(\frac{e(1+\varepsilon)}{\alpha}\right)^{\alpha}\right] &= 4\frac{\log n}{\log\log n} \cdot (1 + \log(1+\varepsilon) - \log(4) - \log\log n + \log\log\log n) \\
&\leq -4\log n\left(1 - \frac{\log\log\log n}{\log\log n}\right) \leq -4\left(1 - \frac{1}{e}\right)\log n \leq -2.5\log n,
\end{aligned}$$

since $e(1+\varepsilon) < 4$ and $\frac{\log\log\log n}{\log\log n} \leq \frac{1}{e}$ for all $n \geq 5$ (even for $n \geq 3$). □

This completes the proof of Proposition 4.3.1. □

Now, we are ready to prove the main theorem of this chapter.

Theorem 4.3.3. *For a suitable choice of parameters, Algorithm 6 finds a $(2 + 8 \log n / \log \log n)$-approximate solution to the Asymmetric Traveling Salesman Problem with high probability and in time polynomial in the size of the input.*

Proof. The algorithm start by finding an optimal extreme-point solution $\mathbf{x}^*$ to the Held-Karp LP relaxation of ATSP (2.4.2). Then, we compute a fractional spanning tree $\mathbf{z}^*$ using (4.1.1).

Then, using the algorithm Algorithm 5 with $\epsilon = 0.2$, we obtain $\lambda_e : E \to \mathbb{R}_+$'s such that the λ-random spanning tree distribution μ satisfies

$$\mathbb{P}[e \in T] \leq (1 + \epsilon) z_e^* \qquad \forall e \in E.$$

Since $\mathbf{x}^*$ was an extreme point solution of (2.4.2), it can be expressed as the unique solution of an invertible system with only $0 - 1$ coefficients, and therefore, every entry x_a^* is rational with integral numerator and denominator bounded by $2^{O(n \log n)}$. In particular, $z^*_{\min} = \min_{e \in E} z_e^* > 2^{-O(n \log n)}$. So, the algorithm Algorithm 5 indeed runs in polynomial time.

Then, we use the polynomial time sampling procedure described in Subsection 2.8.1 to sample $T^* \sim \mu$. By Proposition 4.3.1, we know that T^* is $4 \log n / \log \log n$-thin with high probability. Furthermore, $\mathbb{E}[c(T^*)] \leq c(\mathbf{x}^*)$. By Markov inequality, $c(T^*) \leq 10 c(\mathbf{x}^*)$ with probability $9/10$. So, T^* is $(4 \log n / \log \log n, 10)$ thin with high probability (we can also sample several trees and choose the best one to increase the probability of success to $1 - O(1/\sqrt{n})$.

Finally, we use Theorem 4.2.2 to obtain, in polynomial time, a $O(\log n / \log \log n)$-approximation of our ATSP instance. □

The proof also shows that the integrality gap of the Held-Karp relaxation for the Asymmetric TSP is bounded above by $O(\log n / \log \log n)$. The best known lower bound on the integrality gap is only 2, as shown in [CGK06]. Closing this gap is a challenging open question, and this possibly could be answered using thinner spanning trees. It is worth noting that even a proof of existence of a thin tree provides an upper bound on the integrality gap of the Held-Karp relaxation.

Corollary 4.3.4. *If there always* exists *a (α, s)-thin spanning tree where α and s are absolute constants, then the integrality gap of the ATSP Held-Karp LP relaxation (2.4.2) is a constant.*

4.4 Tight Examples

We conclude this chapter by providing several tight example for our analysis. First, we construct a graph where the tree sampled from the maximum entropy distribution is $\Omega(\log n / \log \log n)$-thin with high probability.

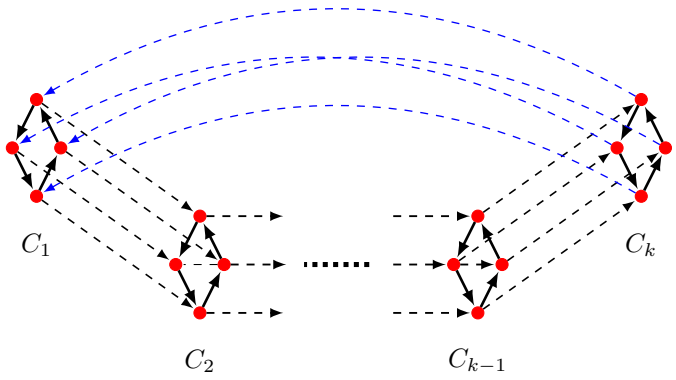

Figure 4.4.1: A tight example for Theorem 4.3.3. Suppose we have k cycles $C_1, \ldots, C_k$ each with k vertices. The solid edges in the figure have fraction $x_a = 1 - 1/k$, and the dashed edges have fraction $x_a = 1/k$. All black edges have cost $c(a) = 1$, the blue edges have cost $c(a) = k^2$. The optimum is about $2k^2$. But, if we use the independent randomized rounding method, then the cost of the Eulerian augmentation is $\Omega(\log n / \log \log n)$ with high probability.

Suppose that we have a complete graph $G = (V, E)$, and $z_e = 2/(n-1)$ for all $e \in E$. It is straightforward that $\mathbf{z}$ is a fractional spanning tree. Since all of the edges of G has the same probability, by symmetry, the maximum entropy distribution that preserve the marginal probabilities of the edges is exactly the uniform distribution on all spanning trees of G.

Next, we show that a uniformly random spanning tree of a complete graph has a vertex of degree $\Omega(\log n / \log \log n)$ with high probability. Since for each $v \in V$, $\mathbf{z}(\delta(v)) \leq 2$, this implies that the maximum entropy distribution gives a tree that is $\Omega(\log n / \log \log n)$-thin with high probability. We use the Prüfer code to obtain a one to one correspondence between all spanning trees of the complete graph and sequences of length $n-2$ of vertices of V (recall that this correspondence implies that a complete graph has exactly n^{n-2} spanning trees). Furthermore, for any vertex $v \in V$, the number of appearances of v in a Prüfer sequence is equal to the degree of v minus 1 in the corresponding spanning tree. But, we know that in a sequence of length $\Omega(n)$ of n items, there is an item that appears at least $\Omega(\log n / \log \log n)$ times (see [MR95] for background on balls and bins processes). So, a random spanning tree has a vertex v of degree $\Omega(\log n / \log \log n)$ with high probability.

Above example shows that the tree sampled from a maximum entropy distribution may indeed be $\Theta(\log n / \log \log n)$ thin, but note that the thinness of a spanning tree is just a sufficient condition in upper bounding the size of a ATSP tour. Our second example shows that if instead of sampling a random spanning tree, we use the independent randomized rounding method (see Section 3.1), then the cost of the minimum Eulerian augmentation is $\Omega(\log n / \log \log n)$ with high probability.

Let G be a union of k disjoint cycles $C_1, \ldots, C_k$ each of length k. For each $1 \leq i \leq k$, the j-th vertex of the i-th cycle has an arc to the j-th vertex of the $i+1$-st cycle (see Figure 4.4.1). Consider the following extreme point solution of LP (2.4.2). Let $x_a = 1 - 1/k$ for all arcs inside the cycles

and $x_a = 1/k$ for all arcs between the cycles (see [CV00] for a proof that this is indeed an extreme point solution of Held-Karp relaxation). Furthermore, let $c(a) = 1$ for all arcs except the arcs from C_k to C_1. For arcs from C_k to C_1, let $c(a) = k^2$. Observe that $c(\mathbf{x}) = \Theta(k^2)$.

Now if we choose the edges independently at random, by the properties of the balls and bins processes, we choose $\Omega(\log k/\log\log k)$ arcs between C_i, C_{i+1} for some $1 \leq i \leq k$ with high probability. But, this implies that the cost of the Eulerian augmentation is $\Omega(\log k/\log\log k)$ of the optimum, because the Eulerian augmentation must have $\Omega(\log k/\log\log k)$ of the arcs between C_k and C_1 and each of these arcs has a cost of k^2.

Recall that by Lemma 2.8.6 if μ is a λ-random spanning tree distribution, then for any two edges e, e',

$$\mathbb{P}\left[e, e' \in T\right] \leq \mathbb{P}\left[e \in T\right] \cdot \mathbb{P}\left[e' \in T\right].$$

If we have equality in the above equation, then e and e' were independent. The above example shows that unless we use the strict inequality in the above equation in some part of the analysis, we cannot prove any approximation factor better than the $\log n/\log\log n$ ratio.

Chapter 5

Planar Asymmetric TSP

In this chapter we give a deterministic 22.5 approximation algorithm for the Asymmetric TSP when the support graph of the solution of the Held-Karp LP relaxation (2.4.2) is a planar graph. Note that this class of problems include Planar ATSP as a special case. Using similar techniques we provide a $O(\sqrt{\gamma}\log\gamma)$ approximation when the support graph is embeddable on an *orientable surface* of *genus* γ. But in this chapter we only include the planar case and we refer an interested reader to [OS11]. Very recently, Erickson and Sidiropoulos [ES13] improved the later result and designed an $O(\log\gamma/\log\log\gamma)$ approximation for instances where the support graph is embeddable on an *any* surface of genus γ.

The results of this chapter are based on a joint work with Amin Saberi [OS11].

5.1 Introduction

In this chapter we design a constant factor approximation for Asymmetric TSP on solutions $\mathbf{x}$ of Held-Karp relaxation that can be embedded on a plane. By Theorem 4.2.2 we only need to find a (α, s) thin tree for constants α, s.

In this chapter, for the sake of clarity, we do not work with fractional graphs. Instead, we multiply all fractions x_e of a feasible solution $\mathbf{x}$ of Held-Karp relaxation by a large integer to make them integral So, we study integral multi-graphs with loops and parallel edges. Unless otherwise specified, by a graph we mean a multi-graph. We use the following adopted definition of a thin tree with respect to a graph G.

Definition 5.1.1. *A subset $F \subseteq E$ is α-thin with respect to G, if for any set $U \subset V$,*

$$|F \cap (U, \overline{U})| \leq \alpha |(U, \overline{U})|,$$

One of the main difficulties in finding a thin spanning tree is that we do not know any polynomial

size certificate, nor do we know a polynomial time verifier for measuring the thinness of a given tree T with respect to an arbitrary graph G. The best approximation algorithm for measuring the thinness of a tree T, that we are aware of, can be achieved by a reduction to the non-uniform sparsest cut problem (see Section 7.7 for background on the non-uniform sparsest cut problem). If we request a demand 1 between the endpoints of each edge of T, then the thinness of a tree T with respect to G is exactly the value of the optimum of the non-uniform sparsest cut problem on G with demand graph T. Since the current best approximation algorithm for the non-uniform sparsest cut problem [ALN08] only gives an $\tilde{O}(\sqrt{\log n})$ approximation, this reduction only gives an $O(\sqrt{\log n})$ approximation of the thinness of T.

It turns out that if G is a planar graph then we can measure the thinness of any tree T in polynomial time. Because of the duality of cuts and cycles [Whi32] it is sufficient to measure the thinness of dual of T, T^*, with respect to cycles of the dual graph G^*, i.e., find a cycle C in G^* such that $|T^* \cap C|/|C|$ is minimized. The latter problem can be solved efficiently by finding the minimum mean weight cycle in G^* where each edge of T^* has weight 1 and the rest of edges have weight 0.

Because of above observation, finding a thin tree in a planar graph is significantly easier. Our idea is to choose T based on a collection of edges such that the shortest path between their endpoints in the dual graph is large. In Lemma 5.2.2 we show that this is a sufficient condition for measuring the thinness of T. We can find this collection of edges by choosing middle edges of disjoint *threads* in the dual graph G^*.

5.2 Constructing a thin-tree

Let $G = (V, E)$ be a connected planar graph, and let G^* be its geometric dual. The *dual-girth* of G, denoted by $g^*(G)$ is the length of the shortest cycle in G^*. The main result of this section is the following lemma.

Lemma 5.2.1. *A connected planar graph with dual-girth g^* has a spanning tree with thinness $\frac{10}{g^*}$. Furthermore, such a tree can be found in polynomial time.*

We will prove this lemma in the rest of this section. First note that if $g^* = 1$, the lemma holds for trivial reasons. Therefore, without loss of generality assume that $g^* > 1$. That implies that no face of G can have two copies of an edge. In particular, G does not have any cut edge. (recall that an edge $e \in E$ is a cut edge of $G - \{e\}$ is disconnected).

Define the distance of two edges in a graph to be the closest shortest path distance between any pair of their endpoints. Our most basic tool for establishing the thinness of a tree T in G is to relate it to the pairwise distance of its corresponding edges T^* in G^*. If G^* does not have any short cycle and all the edges of T^* are far apart in G^*, then T is a thin tree in G.

Lemma 5.2.2. *Let F be a set of edges in G and F^* be the corresponding edges in the dual. If for some $m \leq g^*(G)$, the distance between each pair of edges in F^* is at least m, then F is $\frac{1}{m}$-thin in G.*

Proof. Consider a cut $S = (U, \overline{U})$ in G. Let us start by showing that S^* is a collection of edge-disjoint cycles $C_1, C_2, \ldots, C_l$ in G^*. This is because the number of edges from S^* incident to a vertex v^* in G^* is equal to the intersection of S with corresponding face of v^* in G and that is an even number. Otherwise, either that face contains two copies of an edge of S, or one could find a path $P \subseteq E$ in that face such that $P \cap S = \emptyset$, while the endpoints of P are in different sides of the cut, which are both impossible.

Because the distance of each pair of edges in F^* is at least m, F^* cannot have more than $\max(1, \lfloor \text{len}(C_i)/m \rfloor)$ edges in C_i, for $1 \leq i \leq l$ where by $\text{len}(C_i)$ we mean the number of edges in C_i). Therefore,

$$|F^*| \leq \sum_{i=1}^{l} \max\left\{1, \left\lfloor \frac{\text{len}(C_i)}{m} \right\rfloor\right\} = \sum_{i=1}^{l} \left\lfloor \frac{\text{len}(C_i)}{m} \right\rfloor \leq \frac{|S^*|}{m}.$$

Note that the equality holds by the assumption $\text{len}(C_i) \geq g^* \geq m$. Thus the number of edges of F in the cut $(U, \overline{U})$ is no more than $\lfloor |(U, \overline{U})|/m \rfloor$ and F is $1/m$-thin. □

Considering the above Lemma, our goal will be to find a set of edges in G^* that are sufficiently far apart. We will do this by finding long threads iteratively and selecting one edge from each thread.

A *thread* in a graph G is a maximal subgraph of G which is either

- a path whose internal vertices all have degree 2 in G and its endpoints have degree at least 2, or
- a cycle in which all vertices except possibly one have degree 2.

Let us start by showing the existence of long threads.

Lemma 5.2.3. *A planar graph with minimum degree 2 and girth g has a thread of length at least $g/5$.*

Proof. Let H be a graph satisfying the conditions of the theorem and H' be the graph obtained by iteratively replacing the vertices of degree 2 in H with an edge. In other words, let H' be the graph obtained by replacing every thread in H by an edge. We will show that H' has a cycle of length at most 5. Then, it follows that that at least one of the edges of that cycle is obtained from a thread of length at least $g/5$ in H.

It remains to show H' has a cycle of length at most 5. Let n', m', f' be the number of vertices, edges and faces of H' respectively. Then by Euler formula,

$$n' - m' + f' = 2.$$

Since by definition each vertex of H' has degree at least 3 we have $m' \geq 3n'/2$. So,

$$0 < n' - m' + f' \leq -m'/3 + f',$$

and we get $f' > m'/3$. But since each edge of H' is included in exactly two faces, if all faces of H' have at least 6 edges, then $6f' \leq 2m'$. So, there must be a face with at most 5 edges, which means the girth of H' is at most 5. □

Now we are ready to describe our algorithm for finding a thin tree in a planar graph. Iteratively, we find a thread P in the dual graph G^* add its middle edge to a set F^* and remove all edges of P from G^*. We return a spanning subgraph of the edges corresponding to F^* in G as a thin tree (see Algorithm 7 for details). Since $g^*(G)$ never decreases in the entire run of the algorithm, by the above lemma, we can always find a thread of length at least $g^*/10$ in G^*, so the algorithm terminates in $O(|E(G)|)$ steps.

Algorithm 7 An Algorithm for Finding a Thin Tree in a Planar Graph

Input: A connected planar graph G and its dual G^* with girth g^*.
Output: A spanning tree T with thinness at most $g^*/10$.

1: $F^* \leftarrow \emptyset$
2: **while** there exists an edge in G^* **do**
3: Find a thread P of length at least $g^*/5$ in G^*.
4: Add the middle edge of P to F^*. If P is a cycle, define its middle edge to be the one with the maximum distance from the large degree vertex.
5: Remove all edges of P from G^*.
6: **end while**

return A spanning tree $T \subseteq F$, where F is the set of edges corresponding to F^* in G.

The algorithm has an equivalent description in terms of the original graph G. Roughly speaking, in each iteration, we find a collection of at least $g^*/5$ consecutive parallel edges in G, we add the middle edge of that collection to F then we contract the end points. Observe that the planar embedding is crucial in the execution of this procedure because it provides a notion of a middle edge.

It is also worth noting that $|F|$ may end up being bigger than $|V(G)| - 1$ in an execution of Algorithm 7. This is because a thread in G^* may be equivalent to a collection of parallel *loops*. The next lemma immediately proves Lemma 5.2.1.

Lemma 5.2.4. *The set F computed in Algorithm 7 is connected and spanning in G. Furthermore, the pairwise distance of the edges of F^* in G^* is at least $g^*/10$.*

Proof. We start by proving F is connected in G. First, we show that Algorithm 7 selects at least one edge from each cycle of G^*. For any cycle C in G^*, observe that the first thread P selected in step 3 of Algorithm 7 that has a non-empty intersection with C must lie completely in C (i.e.

$P \subset C$). Therefore the middle edge of P, which is an edge of C, is added to F^*. Now, as we argued in the proof of Lemma 5.2.2, the dual of any cut $(U, \overline{U})$ in G, is a union of edge disjoint cycles, so F^* has at least one edge from one of these cycles, and F has at least one edge from $(U, \overline{U})$.

It remains to show that the pairwise distance of all edges of F^* in G^* is at least $g^*/10$. First, observe that after adding a middle edge e of a thread P to F^*, the algorithm immediately removes all the edges that are of distance less than $g^*/10$ from e because they are all in P. Although each iteration of the while loop may increase the distance of some pairs of edges, it never increases the distance of two edges that are closer than $g^*/10$. Therefore, the distance of any pairs of edges that are closer than $g^*/10$ remains the same until one of them is deleted. □

5.3 Thin trees, Goddyn's conjecture and ATSP

The algorithm presented in Section 5.2 imply the following result:

Theorem 5.3.1. *Any k-edge connected planar graph $G = (V, E)$ has a $\frac{10}{k}$-thin spanning tree. Such a spanning tree can be found in polynomial time.*

Proof. A set of edges $S \subseteq E$ is a minimal cut if there is no proper subset of S that also defines a cut in G. Recall that in the proof of Lemma 5.2.2 we show that the dual of any cut of G is a union of edge disjoint cycle. Whitney [Whi32] the dual of any minimal cut is exactly one cycle, and vice versa, the dual of a cycle is a minimal cut. By Whitney's theorem, g^*, the girth of G^* is at least k. So, by Lemma 5.2.1, Algorithm 7 finds a $10/k$-thin spanning tree. □

An equivalent way to state the above theorem is that there exists a function f such that, for any $\epsilon > 0$, any $f(\epsilon)$-edge connected planar graph G has an ϵ-thin spanning tree. This can proves the following conjecture of Goddyn [God04] for planar graphs.

Conjecture 5.3.2 (Goddyn [God04])**.** *There exists a function $f(\epsilon)$ such that, for any $0 < \epsilon < 1$, every $f(\epsilon)$-edge connected graph has an ϵ-thin spanning tree.*

Goddyn's conjecture is intimately related to ATSP and the integrality gap of Held-Karp relaxation. To make this precise, we generalize Theorem 4.2.2 and we show that if we can find an α/k-thin tree in any k-edge connected graph, then we can design an $O(\alpha)$ approximation algorithm for ATSP.

Proposition 5.3.3. *Suppose there is a non-decreasing function $f(k)$ such that any k-edge connected graph contains a $\frac{f(k)}{k}$-thin spanning tree. Given, a k-edge connected graph $G = (V, E)$, for any cost function $c : E \to \mathbb{R}_+$, G has a $\frac{2f(k)}{k}$-spanning tree T such that $c(T) \leq \frac{2f(k)}{k} \cdot c(E)$.*

Furthermore, if there is a polynomial time algorithm for finding a $f(k)/k$-thin tree in any k-edge connected graph, then we can find a $\frac{2f(k)}{k}$ thin tree of cost $c(T) \leq \frac{2f(k)}{k} \cdot c(E)$ in polynomial time.

Proof. Let $G_0 := G$ and select a $\frac{f(k)}{k}$-thin spanning tree T_0 in G_0, and remove its edges. Call this new graph G_1. Note that each cut $(U, \overline{U})$ of G_0 loses at most $\frac{f(k)}{k}|G_0(U, \overline{U})|$ of its edges. As the size of the minimum cut in G_0 is k, G_1 will be $(k - f(k))$-edge connected.

Similarly, find a

$$\frac{f(k - f(k))}{k - f(k)} \leq \frac{f(k)}{k - f(k)}$$

thin spanning tree T_1 in G_1. The inequality holds by the monotonicity assumption on $f(k)$. Remove the edges of T_1 to obtain a $(k - 2f(k))$-edge connected graph G_2. Repeat on G_2 to obtain $l = \lceil k/(2f(k))\rceil$ spanning trees $T_0, \ldots, T_{l-1}$, where for each i, T_i is a $\frac{f(k)}{k-if(k)}$-thin spanning tree of the $(k - if(k))$-edge connected graph G_i.

Because G_i is a spanning subgraph of G_0, any spanning and thin tree of G_i will be spanning and thin in G_0. Moreover, since $0 \leq i < l$ and

$$\frac{f(k)}{k - if(k)} \leq \frac{2f(k)}{k},$$

each T_i is a $\frac{2f(k)}{k}$-thin spanning tree in G_0. Among the selected trees find the one with the smallest cost. Then, by an averaging argument $c(T_j) \leq 2f(k) \cdot c(E)/k$. □

Now, we are ready to prove the main theorem of this chapter.

Theorem 5.3.4. *Given a feasible point* $\mathbf{x}$ *of the Held-Karp LP (2.4.2), embedded the plane, there is a polynomial-time algorithm that finds a hamiltonian cycle with a cost* $30(1 + O(1/n))$ *of* $c(x)$.

Proof. Let $\mathbf{z}$ be a fractional spanning tree as defined in (4.1.1). Construct a graph H by placing $n^3 z_e$ parallel edges between the endpoints of each edge e where $z_e > 0$ (if z_e is not an integer we just round $n^3 z_e$ down to the nearest integer). Observe that for any $U \subseteq V$,

$$n^3\mathbf{z}(\delta(U)) - n^2 \leq |H(U, \overline{U})|n^3\mathbf{z}(\delta(U)). \tag{5.3.1}$$

Since $\mathbf{z}(\delta(U)) \geq 2(1-1/n)$ for any $U \subseteq V$, H is $k = 2n^3 - 3n^2$ edge connected. By Theorem 5.3.1 and Proposition 5.3.3 we can find a tree T in H that is $20/k$-thin such that $c(T) \leq 20c(E(H))/k$. So, by (5.3.1), for any $U \subseteq V$,

$$|T \cap H(U, \overline{U})| \leq \frac{20}{k}|H(U, \overline{U})| \leq \frac{20n^3}{k}\mathbf{z}(\delta(U) \leq \frac{10}{1 - 2/n)}\mathbf{z}(\delta(U)).$$

So, T is $(20(1 + O(1/n)), 10(1 + O(1/n)))$-thin with respect to $\mathbf{z}$. Therefore, by Theorem 4.2.2 we can find a tour of cost at most $30(1 + O(1/n))$ of $c(\mathbf{x})$. □

A slightly better optimization of parameters lead to a 22.5 approximation algorithm for ATSP. Instead of the best of $k/20$ disjoint thin trees we can consider the cost when choosing a thin tree in

Algorithm 6. In particular, instead of including a middle edge of a thread P in a thine tree T (see step 3 of Algorithm 6) we choose the edge of smallest cost among the $|P|/3$ middle edges of P, i.e., the edge of smallest cost among edges of P of distance $|P|/3$ from the endpoints of P.

It is worth noting that the genus of an extreme point solution of Held-Karp relaxation instance with n vertices can be as large as $\Omega(n)$. In fact, for any odd integer k, it is possible to construct an extreme point on k^2 vertices that has the complete graph with k vertices as a minor. Such an extreme point can be obtained by the same construction as Carr and Vempala [CV00, Theorem 3.5] applied to a complete graph with k vertices.

An argument similar to the proof of above theorem shows that Goddyn's conjecture implies constant integrality gap of the Held-Karp relaxation of ATSP. Furthermore, an algorithmic proof of Goddyn's conjecture implies a constant factor approximation algorithm for ATSP.

Corollary 5.3.5. *If Goddyn's conjecture holds for some function $f(\epsilon) = O(1/\epsilon)$, then the integrality gap of Held-Karp relaxation is bounded from above by a constant.*

5.3.1 Nowhere-zero flows and Jaeger's conjecture

Goddyn's conjecture was inspired by the study of nowhere-zero flows and in particular in attempting Jaeger's conjecture [Jae84]. Here, we just state the Jaeger's conjecture and refer the reader to Seymour [Sey95] for more information.

Conjecture 5.3.6 (Jaeger [Jae84])**.** *For any $4k$-edge connected G there is an orientation of G such that for any $S \subset V(G)$,*

$$(k-1)|\delta^-(S)| \leq k|\delta^+(S)| \leq (k+1)|\delta^-(S)|.$$

Jaeger's conjecture has not been proved for any positive integer k yet. For $k = 1$, this is Tutte's 3-flow conjecture and is proved only for planar graphs. Goddyn's conjecture implies a weaker version of Jaeger's conjecture in which $4k$ is replaced by an arbitrary function of k. Even this version is still open [GHGG99].

Previously, Jaeger's conjecture was proved on planar or bounded genus graphs (see e.g. [Zha02]) Since Goddyn's conjecture implies Jaeger's conjecture with the same parameters, our result can be seen as a strengthening of these results for surfaces with orientable genus.

Chapter 6

Symmetric TSP

In this chapter, for some positive constant ϵ_0, we design a polynomial time $(\frac{3}{2}-\epsilon_0)$-approximation algorithm for graphic TSP. Our result improves on the $\frac{3}{2}$-approximation algorithm due to Christofides [Chr76] for this special case. A corollary of our analysis is that the integrality gap of the Held-Karp LP relaxation (2.4.1) is also strictly below $\frac{3}{2}$ on graph metrics

Since the first appearance of our work, several groups of people studied this problem and improved our approximation ratio. Mömke and Svensson [MS11] came up with a beautiful combinatorial algorithm for graphic TSP with an approximation ratio of 1.461. This approximation ratio was later improved by Mucha [Muc12] to $\frac{13}{9} \approx 1.444$. The latest work of Sebö and Vygen [SV12] improved the approximation factor to 1.4.

The results of this section are based on a joint work with Amin Saberi and Mohit Singh [OSS11].

6.1 Introduction

Recall that in an instance of graphic TSP we are given a graph $G_0 = (V, E_0)$, and a non-negative symmetric cost function $c : V \times V \to \mathbb{R}_+$ satisfying triangle inequality such that $c(e) = 1$ for all $e \in E_0$, and we want to find the shortest tour that visits every vertex at least once. In this chapter we design a polynomial time algorithm that beats the Christofides' 3/2-approximation algorithm on graph metrics. We indeed solve a slightly more general case in which the cost of the edges of G_0 are in an interval $[1, \Upsilon]$ for any constant $\Upsilon > 1$. We remark that none of the follow up results [MS11, SV12, Muc12] claim to beat Christofides' 3/2 approximation in this slightly more general case.

Theorem 6.1.1. *For a universal constant $\epsilon_0 > 0$ and any $\Upsilon \geq 1$, there exists a randomized polynomial time algorithm for any given graph $G_0 = (V, E_0)$ and a metric $c : V \times V \to \mathbb{R}_+$ such that $c(e) \in [1, \Upsilon]$ for any edge $e \in E_0$, finds a tour of expected cost at most $1.5 - \epsilon_0/\Upsilon^3$ of the optimum.*

Note that one can obtain a high probability result from the above theorem by running $\log(n)$ copies of the algorithm independently. A corollary of the analysis of the above theorem is that the integrality gap of the natural linear programming relaxation (due to Held and Karp [HK70]) is also strictly below $\frac{3}{2}$ on graph metrics.

As we mentioned in Section 1.1, our algorithm is designed for the general weighted version of TSP, but our analysis works for the unweighted version, or graphic TSP. In this section we describe our algorithm in its full generality, and later we discuss how we take advantage of the additional assumption that $c(.,.)$ is a graphic metric in our analysis.

Our algorithm for TSP is very similar to Algorithm 6. First, we find an optimal extreme point solution $\mathbf{x}^*$ of Held-Karp LP relaxation (2.4.1), and we let $\mathbf{z}^* = (1 - 1/n)\mathbf{x}^*$. Then, we find a λ-random spanning tree distribution that (approximately) preserves marginals of $\mathbf{z}^*$, we sample T^* from μ, and finally we add the minimum cost Eulerian augmentation to T^* (see Algorithm 8 for details). Note that similar to Algorithm 6 we require $\mathbf{x}^*$ to be an extreme point solution to obtain a lower bound of $2^{-O(n \log n)}$ on $\min_{e \in E} z^*_e$. We need this lower bound to compute an approximate maximum entropy distribution that preserves the marginals of $\mathbf{z}^*$ as we described in Theorem 3.1.3.

Algorithm 8 An Algorithm for Symmetric TSP in General Metrics

Find an optimal extreme point solution $\mathbf{x}^*$ of LP (2.4.1) and let $\mathbf{z}^* = (1 - 1/n)\mathbf{x}^*$ be a fractional spanning tree and let $G = (V, E)$ be the support graph of $\mathbf{x}^*$.
Use Theorem 3.1.3 to find weights $\lambda : E \to \mathbb{R}_+$ such that the λ-random spanning tree distribution, μ, approximates the marginal probabilities imposed by $\mathbf{z}^*$, i.e., for all $e \in E$,

$$\mathbb{P}_\mu [e \in T] \leq (1 + 1/n^3) z^*_e$$

Sample a tree T^* from μ.
Let O denote the set of odd-degree nodes in T^*. Find the minimum cost O-join J^*.
return Shortcut multigraph $J^* \cup T^*$ and output the resulting Hamiltonian cycle.

We can start analyzing above algorithm similar to Section 4.1. The expected cost of a tree sampled based on the maximum entropy rounding by sampling method is at most $c(\mathbf{x}^*)$. So, we just need to upper bound the cost of the minimum Eulerian augmentation - in this case, the minimum cost matching on the odd degree vertices of the tree, by $c(\mathbf{x}^*) \cdot (1/2 - \epsilon)$ for some universal constant ϵ. Currently, we are not aware of any such a proof, but for all instances of TSP that we are aware of, the approximation factor of the above algorithm is significantly better than 3/2. We conjecture that the above algorithm beats Christofides' Algorithm 4 in the worst case for general version of TSP.

Conjecture 6.1.2. *For a universal constant $\epsilon > 0$, Algorithm 8 is a $\frac{3}{2} - \epsilon$ approximation algorithm for the traveling salesman problem.*

We remark that we are not aware of any instance of TSP where the approximation ratio of

Algorithm 8 is worse than 4/3. In Example 3.1.2 we showed a family of examples where the approximation ratio of this algorithm is 5/4. Recall that in Subsection 2.4.1 we show that the integrality gap of the Held-Karp LP relaxation (2.4.1) is at least 4/3, so any analysis that uses $c(\mathbf{x}^*)$ as a lower bound on the value of the optimum tour cannot prove a factor better than 4/3.

In this chapter, we analyze Algorithm 8 only for graphical metrics and after a slight modification: If $\mathbf{x}^*$ is nearly integral, then we simply solve the problem by a greedy deterministic algorithm that has an approximation ratio close to $\frac{4}{3}$. More specifically, we say an edge $e \in E$ is *nearly integral* if $x_e^* \geq 1 - \gamma$, where $\gamma > 0$ is a constant. We say $\mathbf{x}$ is a nearly integral solution of LP (2.4.1) if i.e., $|\{e : x_e \geq 1 - \gamma\}| \geq (1 - \epsilon_2)n$ for a constant $\epsilon_2 > 0$. If $\mathbf{x}^*$ is a nearly integral solution of LP (2.4.1), we let T^* be the minimum cost spanning tree that contains as many nearly integral edges as possible. Then, we simply add minimum O-join on odd-degree vertices of T^*. When $\mathbf{x}^*$ is not nearly integral, we follow Algorithm 8. The details of the final algorithm are described in Algorithm 9.

Algorithm 9 Improved Approximation Algorithm for Graphic TSP

1: Find an optimal solution $\mathbf{x}^*$ of LP (2.4.1) and let $\mathbf{z}^* = (1 - 1/n)\mathbf{x}^*$ be a fractional spanning tree and let $G = (V, E)$ be the support graph of $\mathbf{x}^*$.
2: **if** $\mathbf{x}^*$ contains $(1 - \epsilon_2)n$ edges of fraction greater than $1 - \gamma$ **then**
3: Let $S = \{e : x_e^* \geq 1 - \gamma\}$, and let T^* be a minimum cost spanning tree with respect to the cost function $c(.)$ among all trees T such that $|T \cap S| = \text{rank}(S)$.
4: **else**
5: Use Theorem 3.1.3 to find weights $\lambda : E \to \mathbb{R}_+$ such that the λ-random spanning tree distribution, μ, approximates the marginal probabilities imposed by $\mathbf{z}^*$, i.e., for all $e \in E$,

$$\mathbb{P}_\mu [e \in T] \leq (1 + 1/n^3) z_e^*$$

6: Sample a tree T^* from μ.
7: **end if**
8: Let O denotes the set of odd-degree vertices of T^*. Find the minimum cost O-join J^*.
return Shortcut multigraph $J^* \cup T^*$ and output the resulting Hamiltonian cycle.

We remark that although μ only approximately preserves the marginal probability of $\mathbf{z}^*$ in the above algorithm, for any set $S \subseteq E$, $\mathbb{E}_\mu [|T \cap S|] - \mathbf{z}(S) = O(1/n^2)$. Letting n be sufficiently larger than the inverse of the constants that we consider throughout this chapter, for the sake of brevity, we can ignore this difference, and we assume μ preserves the marginal vector $\mathbf{z}^*$ *exactly*.

Our analysis builds on polyhedral structure of the O-join polytope, the cactus-like structure of near minimum cuts that we discussed in Section 3.2, and locally Hamiltonian properties of λ-random spanning trees that we discussed in Section 3.3.

6.2 The Structure Theorem

We consider two cases. Case 1 is the simpler case where we assume $\mathbf{x}^*$ is nearly integral.

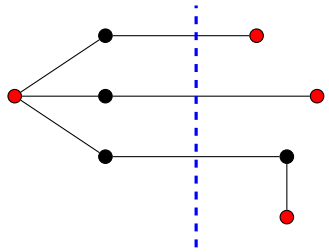

Figure 6.2.1: A set $U \subseteq V$ has an odd number of odd degree vertices of a tree T if and only if $|T \cap \delta(U)|$ is odd. Odd degree vertices are colored with red.

Case 1: $\mathbf{x}^*$ is nearly integral. We start with the simple case that $\mathbf{x}^*$ is nearly integral, and we show a simple polyhedral argument bounds the cost of the tree T^* and the O-join J^*. The following lemma bounds $c(T^*) + c(J^*)$.

Lemma 6.2.1. *Let $\mathbf{x}$ be a fractional solution of LP (2.4.1), and $S := \{e : x_e \geq 1 - \gamma\}$ for $\gamma < 1/3$. Let T^* be a minimum cost spanning tree with respect to $c(.)$ among all trees T such that $|T \cap S| = \text{rank}(S)$. Also, let J^* be a minimum cost O-join on odd degree vertices of T^*. If $(1 - \epsilon_2)n \leq |S|$, then*

$$c(T^*) + c(J^*) \leq c(\mathbf{x})\Big(\frac{4}{3} + 2\gamma + 2\Upsilon(\gamma + \epsilon_2)\Big)$$

Proof. First, since $\gamma < 1/3$, and $\mathbf{x}(\delta(v)) = 2$ for all vertices, each vertex is adjacent to at most two edges of S. So, S is a collection of disjoint paths and cycles. Furthermore, since $\mathbf{x}(\delta(U)) \geq 2$ for any set $U \subset V$, the length of each cycle of S is at least $1/\gamma$. Since $\text{rank}(P) = |P|$ for any path P and $\text{rank}(C) = |C| - 1$ for any cycle C, $\text{rank}(S) \geq |S|(1 - \gamma)$. So,

$$|T^* - S| \leq n - 1 - \text{rank}(S) \leq n - |S|(1 - \gamma) \leq n - (1 - \epsilon_2)(1 - \gamma)n \leq n(\epsilon_2 + \gamma). \tag{6.2.1}$$

Second, we bound the cost of J^*-join by constructing a fractional solution to O-join polytope (2.4.4) and using Proposition 2.4.2. For any edge $e \in E$, let

$$y_e = \begin{cases} \frac{x_e}{3(1-\gamma)} & \text{if } e \in S \cap T^*, \\ 1 & \text{if } e \in T^* - S, \\ x_e & \text{otherwise.} \end{cases} \tag{6.2.2}$$

First, we show that $\mathbf{y}$ is a feasible solution of LP (2.4.4). Let $U \subseteq V$ such that $|U \cap O|$ is odd. We want to show that $\mathbf{y}(\delta(U)) \geq 1$. Observe that, $|U \cap O|$ is odd if and only if $|\delta(U) \cap T^*|$ is odd (see Figure 6.2.1 for an example). If there is an edge $e \in (T^* - S) \cap \delta(U)$, then $\mathbf{y}(\delta(U)) \geq y_e \geq 1$ and the constraint is satisfied. Otherwise, we have $\delta(U) \cap T^* \subseteq S$. Since $|\delta(U) \cap T^*||$ is odd, $\delta(U)$

must have an odd number of edges of $S \cap T^*$. If $|S \cap T^* \cap \delta(U)| = 1$, then

$$\mathbf{y}(\delta(U)) \geq \mathbf{y}(\delta(U) - (S \cap T^*)) \geq \mathbf{x}(\delta(U) - (S \cap T^*)) \geq 1$$

Otherwise, $|S \cap T^* \cap \delta(U)| \geq 3$, and

$$\mathbf{y}(\delta(U)) \geq \mathbf{y}(\delta(U) \cap (S \cap T^*)) \geq \frac{\mathbf{x}(\delta(U) \cap S \cap T^*)}{3(1-\gamma)} \geq 3\frac{1-\gamma}{3(1-\gamma)} = 1$$

Therefore, $\mathbf{y}$ is a feasible solution of LP (2.4.4).

By Proposition 2.4.2, $c(\mathbf{y}) \geq c(J^*)$. So, it is sufficient to upper bound $c(T^*) + c(\mathbf{y})$,

$$\begin{aligned} c(T^*) + c(\mathbf{y}) &\leq \frac{1}{1-\gamma} \cdot \sum_{e \in T^* \cap S} c(e) x_e \Big(\frac{1}{1-\gamma} + \frac{1}{3(1-\gamma)} \Big) + 2c(T^* - S) + c(\mathbf{x}(E - T^*) \\ &\leq \frac{4c(\mathbf{x}(S))}{3(1-\gamma)} + 2|T^* - S| \cdot \Upsilon + c(\mathbf{x}(E - S)) \\ &\leq \frac{4c(\mathbf{x})}{3(1-\gamma)} + 2n(\epsilon_2 + \gamma) \cdot \Upsilon \leq c(\mathbf{x}) \Big(\frac{4}{3} + 2\gamma + 2\Upsilon(\gamma + \epsilon_2) \Big). \end{aligned}$$

The first inequality follows by (6.2.2) and the definition of T^*. The second inequality follows since the cost of each edge of $T^* - S$ is at most Υ. This is because $c(e) \leq \Upsilon$ for each edge $e \in E_0$ and G_0 is connected. The third inequality follows by (6.2.1), and the last inequality follows since $\gamma < 1/3$ and $c(\mathbf{x}) \geq n$. □

We remark that the above construction of the fractional O-join is similar to a construction by Monma, Munson and Pulleyblank [MMP90].

Case 2: $\mathbf{x}^*$ is not nearly integral This is the more interesting case. Let $\mathbf{x}^*$ be an optimal solution of Held-Karp relaxation that is not nearly integral. Let μ be a λ-random spanning tree distribution that approximately preserves marginals in $\mathbf{z}^*$, and let T be a sample from μ. As we proved in Section 3.1, $\mathbb{E}[c(T)] \leq c(\mathbf{x}^*)$. So, Let J be a minimum cost O-join for T. it is sufficient to show that

$$\mathbb{E}[c(J)] \leq (1/2 - \epsilon_0/\Upsilon^3) c(\mathbf{x}^*).$$

First, we recall that by Wolsy's argument [Wol80], the vector $\frac{\mathbf{x}^*}{2}$ is always a feasible fractional solution to LP (2.4.4) for *any* set $O \subseteq V$ (see Section 2.5 for more details). This is because by feasibility of $\mathbf{x}^*$, $\mathbf{x}^*(\delta(U)) \geq 1$ for any set U. To bound $c(J)$, it is sufficient to construct a feasible solution of smaller cost for the O-join polytope, when O is the set of odd degree vertices of the sampled spanning tree T. In other words, we want to construct a vector $\mathbf{y}$ such that $\mathbf{y}(\delta(U)) \geq 1$ for any set U where $|\delta(U) \cap T|$ is odd (see Figure 6.2.1). Note that the vector $\mathbf{y}$ can be a function of $\mathbf{x}^*$ and T. So, we will choose $\mathbf{y}$ with a promise that $\mathbb{E}[c(\mathbf{y})] \leq (1/2 - \epsilon_0/\Upsilon^3) c(\mathbf{x}^*)$, where the

expectation is over the randomness of T.

We say a cut $(U, \overline{U})$ is *even* with respect to a tree T if $|T \cap \delta(U)|$ is even and it is *odd* otherwise. For a sufficiently small constant $\eta > 0$ that we fix later, consider the $(1+\eta)$-near minimum cuts of $G = (V, E, \mathbf{x}^*)$. We say an edge e is *even* with respect to a tree T if any $(1+\eta)$-near minimum cut that contains e is even with respect to T, i.e., for all $U \subset V$ such that $e \in \delta(U)$ and $\mathbf{x}^*(\delta(U)) \leq 2(1+\delta)$, $|T \cap \delta(U)|$ is even. Given a tree T, we let

$$y_e = \begin{cases} \frac{x^*_e}{2(1+\eta)} & \text{if } e \text{ is even with respect to } T \\ \frac{x^*_e}{2} & \text{otherwise.} \end{cases}$$

Observe that such a vector $\mathbf{y}$ is a feasible solution of LP (2.4.4) when O is the odd degree vertices of T.

So, it is enough to find a tree T for which the cost of the even edges is large. Let $\mathcal{E}(e)$ be the event that an edge $e \in E$ is even with respect to T when T is sampled from the distribution μ. We say e is *good* if the probability of this event is bounded away from zero by some absolute constant. More precisely, e is good if for a fixed constant $\rho > 0$,

$$\mathbb{P}\left[\mathcal{E}(e)\right] = \mathbb{P}\left[\exists (U, \overline{U}) : e \in \delta(U) \text{ and } \mathbf{x}^*(\delta(U)) \leq 2(1+\eta) \text{ and } |T \cap \delta(U)| \text{ is odd}\right] \leq 1 - \rho.$$

Let $E^* \subseteq E$ be the set of good edges. It follows that

$$\mathbb{E}\left[c(\mathbf{y})\right] \leq c(\mathbf{x}^*)/2 - \rho \cdot c(\mathbf{x}^*(E^*)). \tag{6.2.3}$$

To beat the Christofides' 1.5 approximation algorithm it is sufficient to show $c(\mathbf{x}^*(E^*))$ is a constant fraction of $c(\mathbf{x}^*(E))$. Let us give some examples.

Complete graph. Let G be a complete graph and $x_e = x_{e'}$ for any two edges $e, e' \in E$. Then, the only near minimum cuts of G are the degree cuts. So an edge e is even if both of its endpoints have an even degree in a tree $T \sim \mu$. By Proposition 3.3.2 this happens with constant probability. So, in this case every edge of G is good.

Hamiltonian cycle. Let G be a cycle of length n, i.e., $x_e = 1$ for each edge of the cycle. Then, all spanning trees of G are Hamiltonian paths and each of them has a probability $1/n$ in μ. Now for any tree T in the support of μ each edge e is contained in at least one odd minimum cut (see Figure 6.2.2). So, G does not have any good edges.

Because of the above example we do not expect that all edges of G are good. In particular, G may not have any good edges at all. Note that by Proposition 2.9.16, T contains exactly two edges in any near minimum cut of G with constant probability. Let us for a moment assume that the events that a number of near minimum cuts $(U_1, \overline{U_1}), \ldots, (U_k, \overline{U_k})$ are even with respect to $T \sim \mu$

Figure 6.2.2: Consider the Hamiltonian cycle shown at the left. In any spanning tree of this graph all edges are contained in at least one odd minimum cut. The dashed blue arcs in the right shows the odd near minimum cuts for one spanning tree.

are independent. Then, Proposition 2.9.16 implies that any edge that is contained in a constant number of near minimum cuts is good. Although the independence assumption does not hold in general, it gives us a lead for finding good edges.

We find the good edge among edges of G that are contained only in a constant number of near minimum cuts. Now, one may ask why should G have any edges in a constant number of near minimum cuts? After all, if G is a Hamiltonian cycle, then every edge is contained in $n-1$ minimum cuts. To prove the existence of edges in a constant number of near minimum cuts we use the machinery that we developed in Section 3.2. Let $\mathcal{F}$ be the collection of all $(1+\eta)$-near minimum cuts, and let $\text{Large}(\tau) = \cup_{\mathcal{C}:|\psi(\mathcal{C})|\geq\tau}\psi(\mathcal{C})$ be the multi-set of all atoms of large cut classes where $\tau^2 = 1/(20\sqrt{\eta})$ (see Definition 3.2.20 for the definition of a large cut class). By Lemma 3.2.21, if $|\text{Large}(\tau)| \geq (1-\epsilon)n$ for a constant ϵ that we will fix later, then there is a set $E' \subseteq E$ such that $|E'| \geq (1-14\sqrt{\eta}-17\epsilon)n$ and for any edge $e \in E'$, $x_e \geq 1-3\sqrt{\eta}$. Now, let $\gamma = 3\sqrt{\eta}$ and $\epsilon_2 := 14\sqrt{\eta}+17\epsilon$. Since $\mathbf{x}^*$ is not nearly integral, we must have $|\text{Large}(\tau)| < (1-\epsilon)n$. Therefore, by Lemma 3.2.23 there is a set $E'_{\text{Small}} \subseteq E$ such that $\mathbf{x}^*(E'_{\text{Small}}) \geq \epsilon n/4$, and each edge $e \in E'_{\text{Small}}$ is contained in at most $O(\tau^2/\epsilon)$ near minimum cuts.

So, a constant fraction of edges are contained in a constant number of near minimum cuts. Now, if the independence assumption holds, i.e., if for any number of near minimum cuts $(U_1, \overline{U_1}), \ldots, (U_k, \overline{U_k})$ the events that these cuts are even with respect to random tree T are independent, then all edges of E'_{Small} are good, and we get

$$c(\mathbf{x}^*(E'_{\text{Small}})) \geq \epsilon n/4 \geq \epsilon \cdot c(\mathbf{x}^*)/8\Upsilon,$$

where the last inequality holds since $c(\mathbf{x}^*) \leq 2\Upsilon \cdot n$ and we would be done with the proof of Theorem 6.1.1 using (6.2.3).

Unfortunately, the above independence assumption does not hold in general (recall that in Example 3.3.1 we show that even an edge that is contained in only two near minimum cuts is not necessarily good). So, instead of showing that all edges of E'_{Small} are good, we show a constant fraction of edges in E'_{Small} are good. This is our main structure theorem.

Theorem 6.2.2 (Structure Theorem). *Let* $\mathbf{x}$ *be a feasible solution of LP (2.4.1),* $\mathbf{z} = (1-1/n)\mathbf{x}$*, and let* μ *be the* λ*-random spanning tree distribution preserving marginals of* $\mathbf{z}$*. For* $\tau^2 = 1/(20\eta)$ *and* $\epsilon \geq 1200/\tau$ *if* $\text{Large}(\tau) < (1-\epsilon)n$*, then for* $\epsilon_1 = \epsilon/10^8$ *and* $\rho = \epsilon^2/10^{12}$ *there exists a set* $E^* \subset E$ *such that* $\mathbf{x}(E^*) \geq \epsilon_1 n$*, and for all* $e \in E^*$*,* $\mathbb{P}[\mathcal{E}(e)] \geq \rho$*.*

We note that the Structure Theorem is valid for *all* feasible solutions to the Held-Karp relaxation. We also remark that we have not tried to optimize the constants.

Now, we are ready to prove Theorem 6.1.1 using the above theorem.

Proof of Theorem 6.1.1. The proof is just a summary of the above discussion. Let $\eta = 10^{-13}/\Upsilon^2$. Let $\tau^2 = 1/(20\eta)$, $\epsilon = 1200/\tau$ as defined in the Structure Theorem. Let $\epsilon_0 = \frac{\epsilon_1\eta\rho\Upsilon^2}{4(1+\eta)}$, $\gamma = 3\sqrt{\eta}, \epsilon_2 = 14\sqrt{\eta} + 17\epsilon$. Note that ϵ_0 is a universal constant. Let $\mathbf{x}^*$ be an optimal solution of Held-Karp LP relaxation (2.4.1), and let μ be a λ-random spanning tree distribution that (approximately) preserves the marginals in $(1-1/n)\mathbf{x}^*$.

Let $G = (V, E, \mathbf{x}^*)$ be the support graph of $\mathbf{x}^*$ and If $|\text{Large}(\tau)| \geq (1-\epsilon)n$, then by Lemma 3.2.21 there is a set $E' \subseteq E$ such that $|E'| \geq (1-\epsilon_2)n$ and $x^*_e \geq 1-\gamma$ for any $e \in E'$. By Lemma 6.2.1, the cost of the tour computed in Algorithm 9 is at most

$$c(\mathbf{x}^*)(4/3+2\gamma+2\Upsilon(\gamma+\epsilon_2)) \leq \frac{4}{3}c(\mathbf{x}^*)+\Upsilon\cdot c(\mathbf{x}^*)(40\sqrt{\eta}+34\epsilon) \leq \frac{4}{3}c(\mathbf{x}^*)+\Upsilon c(\mathbf{x}^*)(200000\sqrt{\eta}) \leq 1.49c(\mathbf{x}^*).$$

Now assume $\text{Large}(\tau) < (1-\epsilon)n$. In this case we use the Structure theorem. We provide a fractional solution to LP (2.4.4). For any edge $e \in E$ if e is contained in at least one odd $(1+\eta)$ near minimum cut we let $y_e = x^*_e/2$ and otherwise we let $y_e = x^*_e/2(1+\eta)$. It is easy to see that $\mathbf{y}$ is indeed a fractional solution of (2.4.4). By Theorem 6.2.2,

$$\begin{aligned}
\mathbb{E}[c(\mathbf{y})] &\leq \frac{c(\mathbf{x}^*)}{2} - \sum_{e\in E} x^*_e c(e)\mathbb{P}[\mathcal{E}(e)]\left(\frac{1}{2} - \frac{1}{2(1+\eta)}\right) \\
&\leq \frac{c(\mathbf{x}^*)}{2} - \frac{\eta}{2(1+\eta)}\sum_{e\in E^*} x^*_e\rho \\
&\leq c(\mathbf{x}^*)(\frac{1}{2} - \frac{\epsilon_1\eta\rho}{4\Upsilon(1+\eta)}) \leq c(\mathbf{x}^*)(1/2 - \epsilon_0/\Upsilon^3).
\end{aligned}$$

The second inequality holds because $c(e) \geq 1$ for all $e \in E$, and the last one because $c(\mathbf{x}^*) \leq 2\Upsilon n$. By Proposition 2.4.2, $\mathbb{E}[c(\mathbf{y})] \leq \mathbb{E}[c(J^*)]$. Therefore, the expected cost of the Eulerian tour in the output of Algorithm 9 is at most $c(\mathbf{x}^*)(1.5 - \epsilon_0/\Upsilon^3)$. □

In the rest of this chapter we prove the structure theorem. As a final remark we would like to mention that even the strong independence assumption, i.e., any constant number of near minimum cuts are even simultaneously with a constant probability, is not sufficient to prove Conjecture 6.1.2. This is because Lemma 3.2.23 only lower bounds $\mathbf{x}^*(E'_{\text{Small}})$ and not $c(\mathbf{x}^*(E'_{\text{Small}}))$. In other words, it can be the case that although $\mathbf{x}^*(E'_{\text{Small}}) = \Omega(n)$, $c(\mathbf{x}^*(E'_{\text{Small}})) = o(c(\mathbf{x}^*))$, so our gain in defining

$y_e = \frac{x^*_e}{2(1+\eta)}$ instead of $y_e = x^*_e/2$ for $e \in E'_{\text{Small}}$ is not comparable with $c(\mathbf{x}^*)$. So, we think proving Conjecture 6.1.2 requires a different approach for upper bounding $\mathbb{E}[c(\mathbf{y})]$.

6.3 In Search of Good Edges

In the rest of this chapter we let $\mathbf{x}$ be any feasible solution of Held-Karp relaxation (2.4.1), $\mathbf{z} = (1-1/n)\mathbf{x}$ and μ be a λ-random spanning tree distribution preserving the marginals of $\mathbf{z}$. We let n to be sufficiently larger than the inverse of the constants that we consider. So, for the simplicity of notation we do not differentiate between the vectors $\mathbf{x}, \mathbf{z}$ when we are dealing with $o(n)$ of edges of G. We also let $G = (V, E, \mathbf{x})$ be the support graph of $\mathbf{x}$.

In Subsection 3.2.2 we proved that if $\text{Large}(\tau) < (1-\epsilon)n$ for a constant $\epsilon > 0$ then there is a set E_{Small} of edges that are not contained in any of the large cut classes, and $\mathbf{x}(E_{\text{Small}}) = \Omega(n)$. In Example 3.3.1 we show that even an edge that is contained in only two near minimum cuts is not necessarily good, so we do not expect that all of the edges of E_{Small} are good. In this section we identify three types of good edges inside E_{Small}, and we show that they contribute to a constant fraction of all edges of G.

The first of types of good edges that we consider are "trivial edges".

Definition 6.3.1 (Trivial Edge)**.** *We call an edge $e \in E$* trivial *if it is in only two near minimum cuts, which are the degree constraint of its endpoints.*

In Proposition 3.3.2 we showed that if x_e is bounded away from $1/2$ with constant probability both of its endpoints have an even degree. So, if e is a trivial edge and x_e is bounded away from $1/2$, e is good. If e is a trivial edge and x_e is very close to $1/2$, then e is not necessarily good as we discussed in Example 3.3.1. But, by Lemma 3.3.7, if e is adjacent to another trivial half edge at least one of them is good. This is summarized in the following proposition.

Proposition 6.3.2. *Any trivial edge e such that $x_e < \frac{1}{2} - \frac{1}{1000}$ or $x_e > \frac{1}{2} + \frac{1}{1000}$ is good. Furthermore, of any pair of adjacent trivial edges, one of them is good.*

Recall that a cut class $\mathcal{C}$ is trivial if it has exactly two atoms and one of them is a singleton. Also recall that an edge e is contained in a cut class $\mathcal{C}$ if its endpoints are in distinct atoms of this cut class. The second type of good edges are inside edges.

Definition 6.3.3 (Inside Edge)**.** *An edge $e = \{u, v\}$ is an inside edge of a non-trivial small cut class $\mathcal{C}$ if $\mathcal{C}$ is the only non-trivial cut class that contains e, and atoms of $\mathcal{C}$ containing u and v are singletons.*

We say a cut class $\mathcal{C}$ is *cyclic* if $20|\psi(\mathcal{C})|^2 \leq 1/\eta$. In the next any inside edge of a cyclic cut class is good.

Proposition 6.3.4. *For any inside edge $e = \{u, v\}$ of a cut class $\mathcal{C}$ that is not the root of Γ, if $20|\psi(\mathcal{C})|^2 \leq 1/\eta$, then*

$$\mathbb{P}\left[\mathcal{E}(e)\right] \geq \frac{1}{2000}.$$

The above proposition is proved in Section 6.4.

For many of the possible feasible solutions of LP (2.4.1) the sum of the fractions of good trivial and inside edges covered in the above lemmas add up to $\Omega(n)$. Unfortunately, there are graphs that do not have any trivial or inside edges, even though they do not have any large cut classes. Consider the fractional LP solution shown in the left of Figure 6.3.3. This graph does not have any inside edges and only the edges $\{1, 2\}$ and $\{n-1, n\}$ are trivial. So, we find to study a third group of good edges that we call "thread edges".

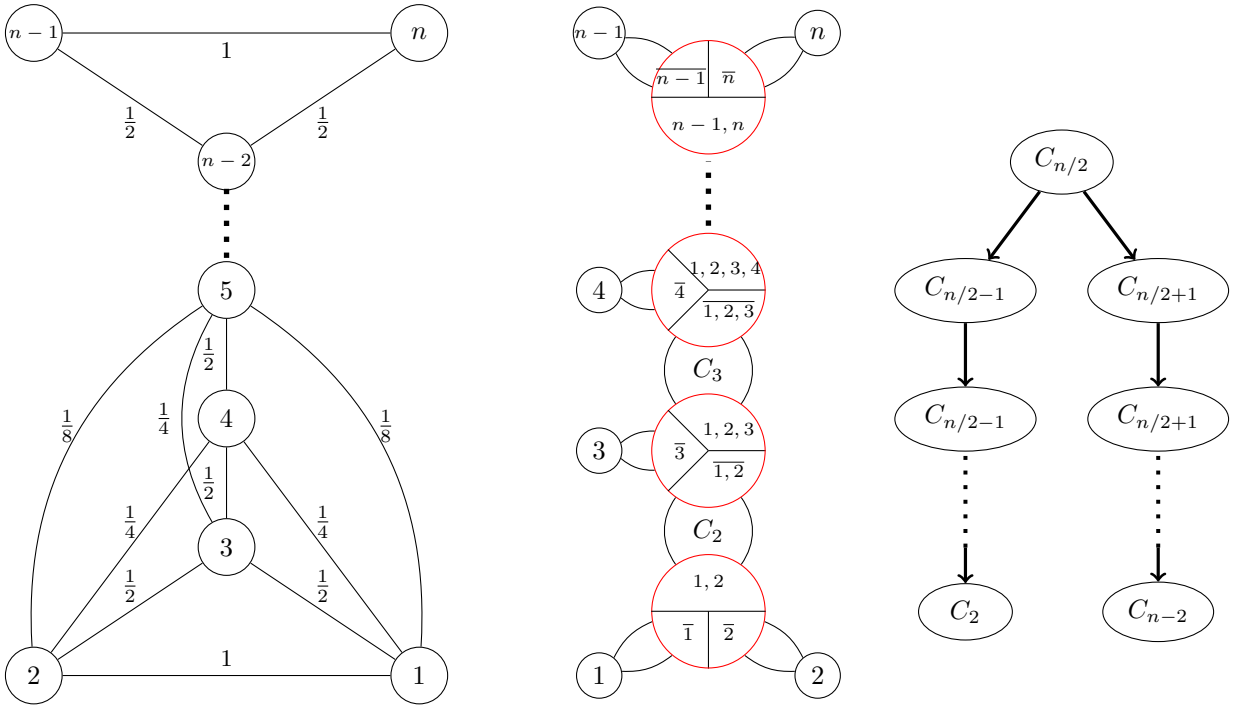

Figure 6.3.3: In the left we have drawn a feasible solution of Held-Karp relaxation (2.4.1). The graph is constructed by connecting each pair of vertices i, j by an edge $x_{\{i,j\}} := 2^{-|i-j|}$, unless i or j $= 1,$ n. For any $2 \leq j \leq n-1$, $x_{\{1,j\}} = 2^{j-2}$, and $x_{\{j,n\}} = 2^{n-j-1}$. Also, $x_{\{1,n\}} = 2^{-n-3}$. In the middle we illustrated the tree hierarchy of the cut classes, Here, $\mathcal{C}_2, \mathcal{C}_3, \ldots, \mathcal{C}_{n-2}$ are the non-trivial cut classes, and the rest are trivial. Each cut class $\mathcal{C}_i$ contains two atoms $\psi(\mathcal{C}_i) := \{\{1, 2, \ldots, i\}, \overline{1, \ldots, i}\}$, we abused notation and used $\overline{1, \ldots, i}$ instead of $\overline{\{1, \ldots, i\}}$ (see Section 2.6 for background on the tree hierarchy and cactus representation). In the right we show the tree of cut classes Γ. Note that Γ contains only the non-trivial cut classes. $\mathcal{C}_{n/2}$ is chosen to be the root, and the tree is made of two long threads. All of the edges of the graph except $\{1, 2\}, \{n-1, n\}$, are non-trivial, and the graph does not contain any inside edges.

Before defining thread edges we need to define a few terms. We define a rooted tree Γ as follows.

The nodes of this tree are non-trivial cut classes of $1+\eta$ near minimum cuts of G. Recall that in Lemma 2.6.13 we show that in all cut classes, except possibly one, say $\mathcal{C}_r$, there is an atom with more than $n/2$ vertices. If $\mathcal{C}_r$ does not exist we let it be any proper cut class if one exist and if not we let it be any non-trivial non-proper cut class.

We root the tree Γ at $\mathcal{C}_r$. Now for each cut class $\mathcal{C} \neq \mathcal{C}_r$ we need to define its father in Γ. By Lemma 2.6.3 there are unique atoms $A \in \psi(\mathcal{C}_r)$ and $B \in \psi(\mathcal{C})$ such that $A \cup B = V$. We call B be the *father connector* of $\mathcal{C}$. Note that by the definition of $\mathcal{C}_r$, $|B| \geq n/2$.

Now, let $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_l$ be the cut classes such that each $\mathcal{C}_i$ has an atom $A_i \in \psi(\mathcal{C}_i)$ such that $\overline{B} \subseteq A_i$ (note that $l \geq 1$ because $\mathcal{C}_r$ satisfies this). Since by Lemma 2.6.3 A_i, A_j does not cross perhaps after renaming we have

$$\overline{B} \subseteq A_1 \subseteq A_2 \subseteq \ldots \subseteq A_l = A. \tag{6.3.1}$$

First, assume $\overline{B} = A_1$. If $\mathcal{C}$ is non-proper (and so $\mathcal{C}_1$ is proper) we let $\mathcal{C}_1$ be the father of $\mathcal{C}$. Otherwise we throw out $\mathcal{C}_1$ in the above chain and rename $\mathcal{C}_i$ to $\mathcal{C}_{i-1}$. Now, assume $\overline{B} \neq A_1$. If $A_1 \neq A_2$ we let $\mathcal{C}_1$ be the father of $\mathcal{C}$, otherwise by part (iv) of Lemma 2.6.3 exactly one of $\mathcal{C}_1, \mathcal{C}_2$ is non-proper. We let the non-proper one be the father of $\mathcal{C}$. This completes the definition of Γ. See Figure 6.3.3 for an example of Γ.

We say an atom A of a non-trivial cut class $\mathcal{C}$ is a *connector* if there is a non-trivial cut class $\mathcal{C}'$ that is a child or the father of $\mathcal{C}$ such that for $B \in \psi(\mathcal{C}')$, $A \cup B = V$, and it is non-connector otherwise. For example, the father-connector is a connector atom. As another example all of the atoms $A_1, \ldots, A_l$ in the chain described in (6.3.1) are connectors. Observe that any connector of $\mathcal{C}$ coincides with at least one connector of a child or the father of $\mathcal{C}$ in a vertex of the tree hierarchy. So the number of connectors of any cut class $\mathcal{C}$ is at most the sum of the in-degree and the out-degree of $\mathcal{C}$ in Γ.

The non-connector atoms play an important role in the proof.

Fact 6.3.5. *For any non-connector atom A of a non-trivial cut class $\mathcal{C}$, all of the edges of $E(A)$ are trivial.*

Proof. Let $e = \{u, v\} \in E(A)$, and suppose there is a cut class $\mathcal{C}'$ that contains e, i.e., u, v belong to distinct atoms of $B, B' \in \psi(\mathcal{C}')$. By Lemma 2.6.3 there is an atom $A^* \in \psi(\mathcal{C})$ and $B^* \in \psi(\mathcal{C}')$ such that $A^* \cup B^* = V$. Since $B \neq B'$, one of B, B', say B, is different from B^*. So, by Lemma 2.6.3 $B \subseteq A^*$. Therefore, since $B \cap A \supset \{u\} \neq \emptyset$, $A \cap A^* \neq \emptyset$, and we get $A = A^*$. But by the construction of Γ, this implies that $\mathcal{C}$ is a grand father of $\mathcal{C}'$ in Γ, and A is a connector of $\mathcal{C}$ which is a contradiction. □

Using above fact we can define an assignment of trivial and inside edges. We say a trivial edge edge e is assigned to a cut class $\mathcal{C}$ if $\mathcal{C}$ has a non-connector atom A that contains the endpoints of e. Also, any inside edge of $\mathcal{C}$ is assigned to this cut class.

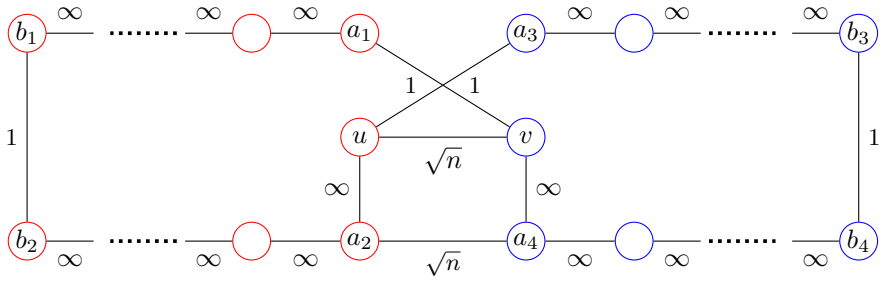

Figure 6.3.4: A random spanning tree distribution with a thread edge that is not good.

Definition 6.3.6 (Thread Edge)**.** *A thread is an induced directed path of cyclic cut classes in* Γ*. The length of a thread is the number of its vertices. Let* P *be a thread that is a directed path from a cut class* $\mathcal{C}$ *to* $\mathcal{C}'$*, and let* A *be the father connector of* $\mathcal{C}$*, and* A' *be the father connector of* $\mathcal{C}'$*. We say an edge* $e = \{u, v\}$ *is a* thread edge *of* P *if* $u, v \notin A$ *but* $u, v \in A'$*.*

Note that by above definition a trivial edge or an inside edge can be a thread edge as well.

Our goal is to show that every thread edge is good. But it turns out that in some special cases a thread edge is not even with high probability.

Example 6.3.7. *Consider the graph in Figure 6.3.4 and let* μ *be the* λ*-random spanning tree distribution corresponding to the* λ *values shown next to each edge. It turns out that the edge* $\{u, a_3\}$ *is highly negatively correlated with* $\{b_3, b_4\}$*, i.e., with high probability exactly one of them is in* $T \sim \mu$*. Similarly,* $\{v, a_1\}$ *is highly negatively correlated with* $\{b_1, b_2\}$ *and* $\{u, v\}$ *is highly negatively correlated with* $\{a_2, a_4\}$*. Let* A *be the set of blue vertices. It follows that*

$$\begin{aligned} \mathbb{E}\left[|T \cap \delta(A - \{v\})|\right] &\approx 2 \\ \mathbb{E}\left[|T \cap \delta(A)|\right] &\approx 2 \\ \mathbb{E}\left[|T \cap \delta(A \cup \{u\})|\right] &\approx 2. \end{aligned}$$

If $\mathcal{F}$ *is the collection of cuts* $\delta(A - \{v\}), \delta(A), \delta(A \cup \{u\})$*, then* Γ *is a thread of length three and* $\{u, v\}$ *is a thread edge. But it turns out that the probability of* $\mathcal{E}(\{u, v\})$ *is very close to zero.*

We note that the marginal vector of this graph does not correspond to a feasible solution of Held-Karp relaxation. Nonetheless the symmetry of red and blue vertices make it impossible to prove $\{u, v\}$ *is good. In Lemma 6.5.4 we show that under some technical assumptions thread edges of an "unbalanced" thread are good.*

We call a thread *unbalanced* with parameter $\epsilon \geq 0$ if the father-connector atoms of all of its cut classes contain at least $n(\frac{1}{2}+\epsilon)$ vertices. For $\eta, \epsilon > 0$ we use $\mathcal{U}_\epsilon$ to denote the set of unbalanced threads of $1 + \eta$ near minimum with parameter ϵ. In the next proposition we show that any unbalanced thread has a constant fraction of good edges.

Proposition 6.3.8. *For any $\epsilon < 1/100$ and $\eta < \epsilon/100$ any thread $P \in \mathcal{U}_\epsilon(\eta)$ of length at least 10 has good thread edges of fraction at least $\epsilon/4$ that are even with probability at least $\epsilon^2/33000$.*

The above proposition is proved in Section 6.5.

In the last part of the proof we show that if the sum of the fraction of good trivial or inside edges is small, then Γ contains many disjoint long unbalanced threads, so G has large fraction of good edges.

Proof of Theorem 6.2.2. Note that by definition τ is chosen such that the cyclic cut classes are the same as small cut classes. We fix ϵ later but throughout the proof we assume $|\operatorname{Large}(\tau)| \leq n$. Recall that E_{Small} is set of edges that not contained in any of the large cut classes. Let $E_T \subseteq E_{\text{Small}}$ be the set of the trivial edges and $E_T^* \subseteq E_T$ be the set of trivial good edges.

We prove the theorem in several steps. In the first step we lower bound $\mathbf{x}(E_T - E_T^*)$.

Claim 6.3.9. *If $\tau > 10000$, then $\mathbf{x}(E_T - E_T^*) \leq 2n/\tau + 2\mathbf{x}(E_{\text{Small}})/2.49$.*

Proof. Let $\tilde{E}_T := E_T - E_T^*$. By Proposition 6.3.2 for any $e \in \tilde{E}_T$, $|x_e - 1/2| \leq 1/1000$. Furthermore, no two edges in $\tilde{E}_T$ are adjacent. Now, let F be the set of edges adjacent to edges of $\tilde{E}_T$ that are contained in at least one large cut class,

$$F := \{\{u, v\} \in E - E_{\text{Small}} : \exists w \in V, \{u, w\} \in \tilde{E}_T\}.$$

It follows that

$$\mathbf{x}(F) \geq (3 - 1/500)|\tilde{E}_T| - 2(\mathbf{x}(E_{\text{Small}}) - \mathbf{x}(\tilde{E}_T)). \tag{6.3.2}$$

where we used the fact that each edge in $E_{\text{Small}} - \tilde{E}_T$ is adjacent to at most two edges of $\tilde{E}_T$.

Suppose there are l large cut classes. By Corollary 2.6.4 there is a set $S \subseteq \operatorname{Large}(\tau)$ of pairwise disjoint atoms such that $|S| \geq |\operatorname{Large}(\tau)| - 2l$. Let

$$S^* := \{A \in \operatorname{Large}(\tau) : \exists \{u, v\} \in \tilde{E}_T, |\{u, v\} \cap A| \geq 1\}.$$

Since $\tilde{E}_T \subseteq E_T$ is a set of trivial edges, for any edge $e \in \tilde{E}_T$ any atom of $\operatorname{Large}(\tau)$ contains either both endpoints of e or none of them. So, the endpoints of any edge $e \in \tilde{E}_T$ is contained in at most one atom of S. Therefore,

$$|S^*| \leq |\operatorname{Large}(\tau) - S| + |\tilde{E}_T| \leq 2l + |\tilde{E}_T| \leq 2n/\tau + |\tilde{E}_T|. \tag{6.3.3}$$

where the last inequality follows by Corollary 2.6.12. By Corollary 3.2.19,

$$\begin{aligned}\mathbf{x}(F) \leq \sum_{A\in S^*} \mathbf{x}(\delta(A)) &= \sum_{A\in \mathrm{Large}(\tau)} \mathbf{x}(\delta(A)) - \sum_{A\in \mathrm{Large}(\tau)-S^*} \mathbf{x}(\delta(A)) \\ &\leq 2(1+3\eta)|\mathrm{Large}(\tau)| - 2(\mathrm{Large}(\tau) - |S^*|) \\ &\leq 6\eta(1+2/\tau)n + 4n/\tau + 2|\tilde{E}_T|\end{aligned}$$

where last inequality uses the fact that $|\mathrm{Large}(\tau)| \leq n(1+2/\tau)$ and (6.3.3). Putting the above equation and (6.3.2) together finishes the proof.

$$2.49\mathbf{x}(\tilde{E}_T) \leq 2\mathbf{x}(E_{\mathrm{Small}}) + 4.1 \cdot n/\tau.$$

where we used $\tau > 10000$. □

Let $E_N = E_{\mathrm{Small}} - E_T$ and let $E_I^* \subset E_{S_N}$ be the set of inside edges of small cut classes (recall that any inside edge is good by Proposition 6.3.4). Next, we show that if $\mathbf{x}(E_T^* \cup E_I^*)$ is small then Γ has many long disjoint unbalanced threads.

Claim 6.3.10. *If $\tau \geq \max\{10000, 200/\mathbf{x}(E_N)\}$ $\mathbf{x}(E_I^*)+\mathbf{x}(E_T^*) \leq \mathbf{x}(E_N)/500$, then Γ has $\mathbf{x}(E_N)/260$ disjoint threads of length 10 with parameter $\mathbf{x}(E_N)/(520n)$.*

Proof. First, we lower bound the number of vertices of Γ. Let $\mathcal{C}_1, \ldots, C_s$ be the set of small non-trivial cut classes. For each $1 \leq i \leq s$, let d_i be the sum of in-degree and out-degree of $\mathcal{C}_i$ in Γ, and let y_i be the sum of the fraction of good edges in E_{S_I} or E_{S_T} assigned to $\mathcal{C}_i$. By Lemma 6.4.4,

$$\begin{aligned}\mathbf{x}(E_I^*) + \mathbf{x}(E_T^*) \geq \sum_{i=1}^{s} y_i \geq \sum_{i=1}^{s} 0.9 \left\lceil \frac{|\psi(\mathcal{C}_i)| - 2d_i}{2} \right\rceil &\geq 0.45 \sum_{i=1}^{s} |\psi(\mathcal{C}_i)| - 0.9 \sum_{i=1}^{s} d_i \\ &\geq 0.44\mathbf{x}(E_N) - 0.9 \sum_{i=1}^{s} d_i,\end{aligned}$$

where in the last inequality we used Corollary 3.2.3 and $\tau > 10000$, that is

$$\mathbf{x}(E_N) \leq \sum_{i=1}^{s} \mathbf{x}(E(G(\psi(\mathcal{C}_i)))) \leq \sum_{i=1}^{s} |\psi(\mathcal{C}_i)|(1+\eta|\psi(\mathcal{C}_i)|) \leq \sum_{i=1}^{s} |\psi(\mathcal{C}_i)|(1+1/(20\tau)).$$

Therefore, $\sum_{i=1}^{s} d_i \geq 0.48\mathbf{x}(E_N)$. So, Γ contains at least $0.24\mathbf{x}(E_N)$ edges, and at least this many nodes. In other words, we have at least $0.24\mathbf{x}(E_N)$ cut classes.

Let k be the number of leaves of Γ. By Corollary 6.4.5, any small cut class which is a leaf is assigned good edges of fraction at least 0.9. Since we have at most n/τ large cut classes,

$$k \leq (\mathbf{x}(E_I^*) + \mathbf{x}(E_T^*))/0.9 + n/\tau \leq \mathbf{x}(E_N)(1/400 + 1/200) \leq \mathbf{x}(E_N)/130.$$

Let Γ' be the tree obtained by contracting each small cut class of degree 2 in Γ except the root. Since Γ has k leaves, Γ' has at most $2k$ vertices. Since each edge of Γ' is corresponding to a thread in Γ and Γ contains at least $30k$ vertices, Γ must have at least k disjoint threads of length 10.

So, we just need to show many of these threads are unbalanced. It follows from Lemma 2.6.14 that at least $k/2 - O(1)$ of these threads are unbalanced with parameter $(k/4 - O(1))/n$. This completes the proof of Claim 6.3.10. □

Let $\epsilon = 1200/\tau$, $\epsilon_1 = \epsilon/10^8$. $\rho = \epsilon^2/10^{12}$. If $\mathbf{x}(E_T^*) + \mathbf{x}(E_I^*) \geq \epsilon_1 n$ then we are done by Proposition 6.3.2 Proposition 6.3.4.

So, we assume that is not the case. By Lemma 3.2.21 if $|\operatorname{Large}(\tau)| \leq (1-\epsilon)n$, then

$$\mathbf{x}(E_{\text{Small}}) \geq (\epsilon - 3\eta)n.$$

By Claim 6.3.9

$$\mathbf{x}(E_N) = \mathbf{x}(E_{\text{Small}}) - \mathbf{x}(E_T - E_T^*) - \mathbf{x}(E_T^*) - \mathbf{x}(E_I^*) \geq \mathbf{x}(E_{\text{Small}})(1 - 2/2.49) - \epsilon_1 n - 2n/\tau \geq 0.19\epsilon n.$$

where we used $1/\tau < \epsilon/1200$. Now, by Claim 6.3.10, Γ has $\epsilon n/1400$ disjoint threads of length 10 with parameter $\epsilon/2800$. By Proposition 6.3.8 each of these threads has $\epsilon/12000$ good edges that are even with probability $\epsilon^2/10^{12}$. Letting E^* be the set of trivial good edges, inside good edges and good thread edges. By the above argument

$$\mathbf{x}(E^*) \geq \frac{\epsilon n}{1400} \cdot \frac{\epsilon}{12000} \geq \epsilon_1 n.$$

□

6.4 Inside Good Edges

In this subsection we prove Proposition 6.3.4. We recommend the readers to see Section 3.2 for properties of cyclic cut classes.

In Section 3.3 we showed that, except certain special cases, subsets of vertices of size 1 or 2 look like a Hamiltonian path with a constant probability. In this section we want to prove a locally Hamiltonian property for cyclic cut classes. In Example 3.3.1 we observed that even an edge that is contained only in two near minimum cuts is not necessarily good. But, in this section we are looking at inside edges that may be contained in $\Omega(\sqrt{\tau})$ near minimum cuts, and it seems it should be significantly harder to prove an inside edge is good. But as we will show the structure of near minimum cuts and the cyclic cut classes help us to prove the claim.

For a cut class $\mathcal{C}$, recall that $G(\psi(\mathcal{C}))$ is the graph obtained by contracting the atoms of the cut class $\mathcal{C}$ (see Figure 3.2.6 for an example). We prove our claim by showing that $T \sim \mu$ is a Hamiltonian

cycle in $G(\psi(\mathcal{C}))$ with constant probability. This can be considered as a locally Hamiltonian property for cyclic cut classes. Note that if T is a Hamiltonian cycle in $G(\psi(\mathcal{C}))$, it does not mean that T has a cycle in G. This is because T does not necessarily include a spanning tree in the induced subgraphs $G[A]$ for $A \in \psi(\mathcal{C})$. For a spanning tree T and cut class $\mathcal{C}$ we use $\mathcal{H}_{\mathcal{C}}(T)$ to denote T is a Hamiltonian cycle in $G(\psi(\mathcal{C}))$.

Before getting into the details of the argument we prove a simple averaging argument that shows for any set $U \subset V$, if $(U, \overline{U})$ is a near minimum cut, and $|U| \leq (1-\epsilon)n$ for a constant ϵ, then there is a constant probability that $T \sim \mu$ is a spanning tree in $G[U]$ (see Corollary 6.4.2).

Lemma 6.4.1. *For any set $S \subseteq E$,*

$$\mathbb{P}\left[|T \cap S| = \text{rank}(S)\right] \geq 1 + \mathbf{z}(S) - \text{rank}(S).$$

Proof. Let $p := \mathbb{P}\left[|T \cap S| = \text{rank}(S)\right]$ Since for any spanning tree T, $|T \cap S| \leq \text{rank}(S)$,

$$\begin{aligned} \mathbf{z}(S) = \mathbb{E}\left[|T \cap S|\right] &\leq p \cdot \text{rank}(S) + (1-p) \cdot (\text{rank}(S) - 1) \\ &= \text{rank}(S) + p - 1. \end{aligned}$$

Note that in the above equation we specifically used the fact that μ preserves the marginal probability in $\mathbf{z}$ and not $\mathbf{x}$, this is because we are summing up $\Theta(n)$ fraction of edges of G. Therefore $p \geq 1 + \mathbf{z}(S) - \text{rank}(S)$. □

Suppose $U \subset V$ such that $|U| \leq o(n)$ and $(U, \overline{U})$ is a near minimum cut of G. The next corollary shows that with high probability $T \sim \mu$ is a spanning tree inside U. Even if $|U| \simeq n/2$, this probability is still roughly $1/2$ when η is small enough.

Corollary 6.4.2. *If $(U, \overline{U})$ is a $(1+\eta)$ near minimum cut with respect to $\mathbf{x}$, and $|U|/n + \eta < 1$, then*

$$\mathbb{P}\left[|T \cap E(U)| = |U| - 1\right] \geq 1 - \frac{|U|}{n} - \eta$$

Proof. Since the fractional degree of each vertex with respect to $\mathbf{x}$ is 2, and $\mathbf{x}(\delta(U)) \leq 2 + 2\eta$, we have

$$\mathbf{x}(E(U)) \geq \frac{1}{2}(2|U| - 2 - 2\eta) = |U| - 1 - \eta.$$

So,

$$\mathbf{z}(E(U)) \geq (1 - 1/n)(|U| - 1 - \eta) \geq |U| - \frac{|U|}{n} - 1 - \eta.$$

The conclusion follows from Lemma 6.4.1, by letting $S := E(U)$, and noting that $\text{rank}(S) \leq |U| - 1$. □

Now, we are ready to prove the main result of the section. We show that with constant probability, all the near minimum cuts of any cyclic cut class are even.

Lemma 6.4.3. *Let $\mathcal{C}$ be a non-trivial cut class of $(1+\eta)$ near minimum cuts of G that is not the root of Γ. If $20|\psi(\mathcal{C})|^2 \leq 1/\eta$, then, with probability at least $1/2000$, T is a Hamiltonian cycle in $G(\psi(\mathcal{C}))$, i.e.,*

$$\mathbb{P}\left[\mathcal{H}_{\mathcal{C}}(T)\right] \geq 1/2000.$$

So, with a constant probability all of the near minimum cuts in $\mathcal{C}$ have exactly two edges in a tree T.

Proof. Since $2|\psi(\mathcal{C})|\eta \leq 1$, by Corollary 3.2.3, $\mathcal{C}$ does not have any inside atoms. Let $\psi(\mathcal{C}) = \{A_1, A_2, \ldots, A_k\}$ and assume that they are placed around the polygon in the order of their labels. So $\mathcal{H}_{\mathcal{C}}(T)$ is the event that $|T \cap E(A_i, A_{i+1})| = 1$ for all consecutive pair of atoms.

Since all cuts of $\mathcal{C}$ the representing diagonals of the polygon, if $\mathcal{H}_{\mathcal{C}}(T)$ occurs, then T contains exactly two edges in each of the cuts in $\mathcal{C}$. Let A_1 be the father-connector of $\mathcal{C}$. Since $\mathcal{C}$ is not the root of Γ, $|A_1| >= \frac{n}{2}$ vertices. Let

$$\begin{aligned} F &:= \bigcup_{i=2}^{k} E(A_i), \\ I &:= \bigcup_{1\leq i,j\leq k} E(A_i, A_j) - \bigcup_{1\leq i\leq k} E(A_i, A_{i+1}) \end{aligned}$$

We compute the probability of $\mathcal{H}_{\mathcal{C}}(T)$ in the following steps: first we condition on $|T \cap I| = 0$. Then we condition on $|T \cap E(\overline{A_1})| = |\overline{A_1}| - 1$. Let μ'' be the resulting distribution. Observe that μ'' is no longer a random spanning tree distribution, it is a product of two random spanning tree distribution, one in the induced subgraph $G[\overline{A_1}]$ and the other in $G/\overline{A_1}$. So, any random variable that is a function of edges in $E(\overline{A_1})$ is independent of functions of edges in $E - E(\overline{A_1})$. So, we show that under μ'', with a constant probability $|T \cap F| = \text{rank}(F)$ and with a constant probability $|T \cap E(A_1, A_2)| = 1, |T \cap E(A_k, A_1)| = 1$. Since any tree $T \sim \mu''$ has no edges in I, these two imply that $\mathcal{H}_{\mathcal{C}}(T)$ occurs.

First by Corollary 3.2.3 (3), we have

$$\sum_{i=1}^{k} \mathbf{x}(E(A_i, A_{i+1})) \geq k(1-k\eta) \geq k - \frac{1}{20},$$

where we the fact that G is fractionally 2-connected and the lemma's assumption $k^2\eta \leq 1/20$. Since by Proposition 3.2.1, $\mathbf{x}(E(G(\psi(\mathcal{C})))) \leq k(1+\eta)$, we have

$$\mathbf{z}(I) \leq \mathbf{x}(I) \leq k(1+\eta) - \sum_{i=1}^{k} \mathbf{x}(E(A_i, A_{i+1})) \leq k\eta + \frac{1}{20} \leq \frac{1}{10}.$$

By Markov inequality,

$$\mathbb{P}\left[|T \cap I| = 0\right] \geq 9/10. \tag{6.4.1}$$

Let $\mu' := \{\mu \mid |T \cap I| = 0\}$. By Corollary 2.9.9 and Corollary 6.4.2

$$\mathbb{E}_{\mu'}\left[|T \cap E(\overline{A_1})|\right] \geq \mathbf{z}(E(\overline{A_1})) \geq 1 - |\overline{A_1}|/n - k\eta \geq 9/20$$

where we used $|\overline{A_1}| = n - |A_1| \leq n/2$, Let $\mu'' = \{\mu' \mid |T \cap E(\overline{A_1})| = |\overline{A_1}| - 1\}$. Then,

$$\begin{aligned}
\mathbb{P}_{\mu}\left[\mathcal{H}_{\mathcal{C}}(T)\right] &= \mathbb{P}_{\mu}\left[\mathcal{H}_{\mathcal{C}}(T) \mid |T \cap I| = 0\right] \mathbb{P}_{\mu}\left[|T \cap I| = 0\right] \\
&\geq \frac{9}{10}\mathbb{P}_{\mu'}\left[\mathcal{H}_{\mathcal{C}}(T) \mid |T \cap \overline{A_1}| = |\overline{A_1}| - 1\right] \mathbb{P}_{\mu'}\left[|T \cap \overline{A_1}| = |\overline{A_1}| - 1\right] \\
&\geq \frac{9}{10} \cdot \frac{9}{20}\mathbb{P}_{\mu''}\left[|T \cap F| = \operatorname{rank}(F)\right] \cdot \mathbb{P}_{\mu''}\left[|T \cap E(A_1, A_2)| = 1, |T \cap E(A_k, A_1)| = 1\right].
\end{aligned}$$

So, it remains to lower bound the two terms in the right hand side. First, we compute $\mathbb{P}_{\mu''}\left[|T \cap F| = \operatorname{rank}(F)\right]$. By Theorem 2.9.11 and Corollary 2.9.9,

$$\begin{aligned}
\mathbb{E}_{\mu''}\left[|T \cap F|\right] \geq \mathbb{E}_{\mu'}\left[|T \cap F|\right] \geq \mathbf{z}(F) &= (1 - 1/n)\mathbf{x}(F) \\
&\geq (1 - 1/n)\sum_{i=2}^{k}(|A_i| - 1 - k\eta) \\
&\geq \sum_{i=2}^{k}|A_i| - \frac{\sum_{i=2}^{k}|A_i|}{n} - (k-1)(1 + k\eta) \\
&\geq \sum_{i=2}^{k}(|A_i| - 1) - \frac{1}{2} - \frac{1}{20},
\end{aligned}$$

where the third inequality follows from Proposition 3.2.1, and the last inequality follows from the fact that $|A_1| \geq n/2$. So $\mathbb{P}_{\mu''}\left[|T \cap F| = \operatorname{rank}(F)\right] \geq 9/20$. It remains to lower bound the probability of the event $|T \cap E(A_k, A_1)| = 1$ and $|T \cap E(A_1, A_2)| = 1$. We use Lemma 3.3.3 to finish the proof. By Corollary 2.9.9

$$\begin{aligned}
9/20 \leq 1 - k\eta - 1/2 \leq \quad \mathbb{E}_{\mu''}\left[|T \cap E(A_1, A_k)|\right], \mathbb{E}_{\mu''}\left[|T \cap E(A_1, A_2)|\right] \quad &\leq 1 + \frac{1}{10} \\
29/20 \leq 2 - k\eta - 1/2 \leq \quad \mathbb{E}_{\mu''}\left[|T \cap E(A_1, A_k| + |T \cap E(A_1, A_2)|\right] \quad &\leq 2 + \frac{1}{10}.
\end{aligned}$$

By Proposition 2.9.16,

$$\mathbb{P}_{\mu''}\left[|T \cap E(A_1, A_2)| + |T \cap E(A_k, A_1)| = 2\right] \geq \min\{\mathrm{Ber}(29/20, 2), \mathrm{Ber}(2.1, 2)\} \geq \frac{1}{20}.$$

Let $A = E(A_1, A_2)$ and $B = E(A_1, A_k)$, $\alpha = 9/20$, $\beta = 9/20$, $\epsilon = 1/20$, we obtain in Lemma 3.3.3 we obtain

$$\mathbb{P}_{\mu''}\left[|T \cap E(A_1, A_2)| = 1, |T \cap E(A_1, A_k)| = 1\right] \geq 1/500.$$

This completes the proof of Lemma 6.4.3. □

Now, we are ready to prove Proposition 6.3.4.

Proof of Proposition 6.3.4. Since e is an inside edge of $\mathcal{C}$, u and v are singleton atoms of $\mathcal{C}$ and the only near minimum cuts containing e are the representing diagonals of the polygon representation of $\mathcal{C}$ and the trivial cuts $(\{u\}, \overline{\{u\}})$ and $(\{v\}, \overline{\{v\}})$. By the above lemma with probability $1/4000$, T is a Hamiltonian cycle in $G(\psi(\mathcal{C}))$. But since u and v are singleton atoms of $\mathcal{C}$, they are not contracted in $G(\psi(\mathcal{C}))$, and $|T \cap \delta(u)| = |T \cap \delta(v)| = 2$. □

Since each cut class has a large fraction of inside or trivial edges if its degree in Γ is small, we get the following bound on the good edges assigned to the cut class.

Lemma 6.4.4. *Let $\mathcal{C}$ be a non-trivial cyclic cut class such that $\mathcal{C} \neq \mathcal{C}_r$ and let d be the sum of in-degree and out-degree of $\mathcal{C}$ in Γ. Then the sum of the fraction of good edges assigned to $\mathcal{C}$ is at least* $0.9 \left\lceil \frac{|\psi(\mathcal{C})|-2d}{2} \right\rceil$.

Proof. Since $\mathcal{C}$ has degree d, it at most d connectors. Therefore, by Proposition 3.2.1 there are $k \geq |\psi(\mathcal{C})| - 2d$ pairs of atoms of $\mathcal{C}$,

$$\{\{A_1, B_1\}, \{A_2, B_2\}, \ldots, \{A_k, B_k\}\},$$

such that for all $1 \leq i \leq k$, $E(A_i, B_i) \geq 1 - \eta|\psi(\mathcal{C})|$, each atom is in at most two pairs and all of $A_1, \ldots, A_k, B_1, \ldots, B_k$ are non-connectors. We show that $\mathcal{C}$ is assigned a fraction $0.9\lceil k/2 \rceil$ of good edges.

Consider a pair $\{A_i, B_i\}$, if they both are singletons, say $A_i = \{u\}$ and $B_i = \{v\}$, then by Proposition 3.2.1 the inside edge $\{u, v\}$ has fraction at least $1 - \eta|\psi(\mathcal{C})| \geq 0.9$. Since $\mathcal{C}$ is cyclic, by Proposition 6.3.4 $\{u, v\}$ is a good edge and it is assigned to $\mathcal{C}$.

Otherwise, without loss of generality, suppose A_i is a non-singleton. Since A_i is non-connector by Fact 6.3.5 all of the edges of $E(A_i)$ are trivial edges and since A_i is non-connector they are assigned to $\mathcal{C}$. By Proposition 3.2.1, $\mathbf{x}(\delta(A_i)) \leq 2(1 + \eta|\psi(\mathcal{C})|)$. If $A_i = \{u, v\}$ we have

$$x_{\{u,v\}} \geq 1 - \eta|\psi(\mathcal{C})| \geq 0.9,$$

and by Proposition 6.3.2 it is a good edge. Otherwise, $|A_i| \geq 3$, so

$$\mathbf{x}(E(A_i)) \geq 3 - (1 + \eta|\psi(\mathcal{C})|) \geq 1.9$$

and by Proposition 6.3.2 at least a fraction $(1/2 - 1/1000)\mathbf{x}(E(A))$ of them are good. The lemma follows from the fact that each atom A_i is in at most two pairs. □

Corollary 6.4.5. *Let $\mathcal{C}$ be a non-trivial cyclic cut class that is a leaf of Γ, i.e., it has out-degree 0. Then $\mathcal{C}$ is assigned good edges of fraction at least* 0.9.

Proof. If $|\psi(\mathcal{C})| \geq 4$, then we are done by Lemma 6.4.4. Otherwise, $\mathcal{C}$ must have two atoms, say $\psi(\mathcal{C}) = \{A_1, A_2\}$, and suppose A_1 is the father-connector of $\mathcal{C}$. Since $\mathcal{C}$ is a non-trivial cut class, A_2 is not a singleton. Also since $\mathcal{C}$ has out-degree 0, A_2 is non-connector. Therefore, by Fact 6.3.5 the edges between the vertices of A_2 are trivial and similar to the proof of Lemma 6.4.4 at least 0.9 of them are good and assigned to $\mathcal{C}$. □

6.5 Thread Good Edges

In this section we prove Proposition 6.3.8. This is the most unclean part of the proof, we recommend the readers to fully understand the proofs in Sections 3.3 and 6.4 before reading this section.

Let $P \in \mathcal{U}_\epsilon(\eta)$. We show that P is assigned thread edges of fraction at least $\epsilon/4$. Recall that these thread edges can be inside edges or trivial edges assigned to one of the cut classes of P. If P has trivial or inside thread edges of total fraction at least $\epsilon/4$ we are done. Therefore, in the proof we assume P has less then $\epsilon/4$ trivial or inside good thread edges (it can be even zero). It turns out that this will eliminate many possible cases. First, we show that if P contains a cut class $\mathcal{C}_i$ where $|\psi(\mathcal{C}_i)| \neq 2, 4$, then it will be assigned trivial or inside good edges of fraction at least $3/4$.

By Lemma 6.4.4, any cut class with $|\psi(\mathcal{C}_i)| > 4$ of out-degree 1 is assigned good edges of fraction at least 0.9. But by definition any cut class of P is cyclic and have out-degree 1 in Γ. So, in the rest of this section we assume all of the cut classes of P have either 2 or 4 atoms. We consider two cases: (i) all cut classes of P are non-proper. In Lemma 6.5.5 we show that if P contains 6 consecutive non-proper cut classes then it is assigned good edges of fraction $\frac{\epsilon}{4}$. (ii) P contains at least one cut class with 4 atoms. In Lemma 6.5.7 we show that if P contains 3 consecutive cut classes such that the third one has 4 atoms then it is assigned good edges of fraction $\frac{\epsilon}{4}$. Putting them together it is straightforward that if P has at least 8 cut classes, then at least one of the two cases occurs.

Let us define some notations first: Let $\mathcal{C}_1, \mathcal{C}_1$ be two consecutive cut classes in P. We use the notation $\mathcal{C}_1 \to \mathcal{C}_2$ to show $\mathcal{C}_{i+1}$ is the unique child of $\mathcal{C}_i$. Furthermore, say B is the father connector of $\mathcal{C}_2$, by Lemma 2.6.3 $\mathcal{C}_1$ has a unique atom A such that $A \cup B = V$. We call A the *child-connector* of $\mathcal{C}_1$.

Before getting into the proof let us describe another classes of thread edges of P that are trivial. This special case would make the proofs of the two lemmas much simpler.

Lemma 6.5.1. *Let $P \in \mathcal{U}_0(\eta)$, and $\mathcal{C}_1 \to \mathcal{C}_2 \in P$ be cut classes with at most 4 atoms. Let $A \in \psi(\mathcal{C}_1)$ be the child connector of $\mathcal{C}_1$, and $B \in \psi(\mathcal{C}_2)$ be the father-connector of $\mathcal{C}_2$. If $|A - \overline{B}| > 2$ and $\eta < 1/20$, then P has good thread edges of fraction $\frac{3}{8}$. Otherwise, if $|A - \overline{B}| = 2$ and the two vertices are connected by an edge of fraction at least ϵ, then P has good edges of fraction at least ϵ.*

Proof. Let $D := A - \overline{B}$. By the definition of A, B, $A \cup B = V(G)$, so $\overline{B} \cup D = A$. By definition of P the edges of $E(D)$ are trivial thread edges assigned to P. So, we need to show a large fraction of

edges of $E(D)$ is good. Since $\mathcal{C}_1, \mathcal{C}_2$ have at most 4 atoms, by Proposition 3.2.1,

$$\mathbf{x}(\delta(A)), \mathbf{x}(\delta(\overline{B})) \leq 2(1+2\eta). \tag{6.5.1}$$

Since for each $v \in D$, $\mathbf{x}(\delta(v)) = 2$,

$$\mathbf{x}(E(D)) \geq \frac{2|D| - \mathbf{x}(\delta(A)) - \mathbf{x}(\delta(\overline{B}))}{2} \geq \frac{2|D| - 4(1+2\eta)}{2} = |D| - 2 - 4\eta.$$

If $|D| \geq 3$, $\mathbf{x}(E(D)) \geq 1 - 4\eta$ and by Proposition 6.3.2 at least half of these trivial edges are good. Therefore P has good thread edges of fraction at least $3/8$, and we are done. Otherwise, suppose $D = \{u, v\}$, and $x_{\{u,v\}} > 0$. If $x_{\{u,v\}}$ is bounded away from $1/2$, then by Proposition 6.3.2 it is a good edge. Otherwise $x_{\{u,v\}}$ is very close to $1/2$. In the following claim we use the fact that $\mathbf{x}(\delta(A)), \mathbf{x}(\delta(B)) \approx 2$ to show that such an edge is always good.

Claim 6.5.2. *If $D = \{u, v\}$ and the edge $e = \{u, v\}$ satisfies $|x_e - 1/2| < 1/200$ then $\mathbb{P}[\mathcal{E}(e)] \geq 1/40000$.*

Proof. Let $S_A := E(\{u, v\}, \overline{A})$, $S_B := E(\{u, v\}, \overline{B})$ and $S := S_A \cup S_B$ be the set of edges separating u and v from the rest of the graph. Recall that by Corollary 3.3.4, $\mathbb{P}[\mathcal{E}(e)]$ is a constant, unless $|T \cap S| = 3$ with a high probability. So, essentially we need to show that with a constant probability $|T \cap S| = 2$.

By (6.5.1) and the claim's assumption,

$$\begin{aligned} 2 \leq \quad & \mathbf{x}(\delta(A)) = \mathbf{x}(S_A) + \mathbf{x}(E(\overline{A}, \overline{B})) && \leq 2(1+2\eta) \\ 2 \leq \quad & \mathbf{x}(\delta(B)) = \mathbf{x}(S_B) + \mathbf{x}(E(\overline{A}, \overline{B})) && \leq 2(1+2\eta) \\ 3 - 1/100 \leq \quad & \mathbf{x}(S_A) + \mathbf{x}(S_B) && \leq 3 + 1/100. \end{aligned}$$

.Therefore,

$$1.49 \leq 1.5 - 1/200 - 2\eta \leq \mathbf{x}(S_A), \mathbf{x}(S_B) \leq 1.5 + 1/200 + 2\eta \leq 1.51,$$

where we used the assumption that $\eta < 1/400$.

Let $\mu' = \{\mu \mid I_e = 1\}$. By Fact 2.9.8,

$$\begin{aligned} \mathbb{E}_{\mu'}\left[|T \cap (E(\overline{A}) \cup E(\overline{B}))|\right] \quad & \geq \quad \mathbf{z}(E(\overline{A})) + \mathbf{z}(E(\overline{B})) - 0.5 - \frac{1}{200} \\ & \geq \quad (1 - \frac{1}{n})(|\overline{B} \cup \overline{A}| - 2 - 4\eta) - 0.5 - \frac{1}{200} \\ & \geq \quad |\overline{B}| + |\overline{A}| - 3.5 - 4\eta - \frac{1}{200} \end{aligned}$$

where in the second inequality we used (6.5.1). So, we either have $\mathbb{E}_{\mu'}\left[|T \cap E(\overline{A})|\right] \geq |\overline{A}| - 1.74$ or $\mathbb{E}_{\mu'}\left[|T \cap E(\overline{B})|\right] \geq |\overline{B}| - 1.74$. Wlog suppose the former is the case. By Lemma 6.4.1 we have

$$\mathbb{P}_{\mu'}\left[|T \cap E(\overline{A})| = |\overline{A}| - 1\right] \geq 0.24.$$

Let $\mu'' = \{\mu' \mid |T \cap \overline{A}| = |\overline{A}| - 1\}$. Since $T \sim \mu''$ is connected in the induced subgraphs $G[\overline{A}]$ and $G[\{u, v\}]$ it must have at most one edge in S_A. So,

$$\mathbb{E}_{\mu''}\left[|T \cap S_A|\right] \leq 1.$$

Therefore, by Corollary 2.9.9

$$1.72 \leq \mathbb{E}_{\mu''}\left[|T \cap S_A|\right] + \mathbb{E}_{\mu''}\left[|T \cap S_B|\right] \leq \mathbf{x}(S_B) \leq 2.51,$$

Therefore, by Proposition 2.9.16,

$$\epsilon \geq \min\{\mathrm{Ber}(1.72, 2), \mathrm{Ber}(2.51, 2)\} \geq 1/52.$$

This shows that there are exactly two edges in the cut $\{u, v\}, \overline{\{u, v\}}$ with a constant probability.

Now, we are ready to prove the claim using Lemma 3.3.3. Let $X = |T \cap \delta(u)|$, $Y = |T \cap \delta(v)|$. Then, by the above equation and Corollary 2.9.9,

$$\begin{aligned} 0.23 \leq \quad & \mathbb{E}_{\mu''}[X], \mathbb{E}_{\mu''}[Y] \quad \leq 1.51 \\ 1.72 \leq \quad & \mathbb{E}_{\mu''}[X + Y] \quad \leq 2.51. \end{aligned}$$

So, we can let $\alpha = 0.24$, $\beta = 0.23$ and $\epsilon = 1/52$ we get

$$\mathbb{P}_{\mu''}[X = 1, Y = 1] \geq 1/4000.$$

Therefore,

$$\begin{aligned} \mathbb{P}[\mathcal{E}(e)] &= \mathbb{P}[\mathcal{E}(e) | I_e = 1]\,\mathbb{P}[i_e = 1] \\ &\geq 0.49\mathbb{P}_{\mu'}\left[\mathcal{E}(e) \mid |T \cap E(\overline{A})| = |\overline{A}| - 1\right]\mathbb{P}_{\mu'}\left[|T \cap E(\overline{A})| = |\overline{A}| - 1\right] \\ &\geq 0.1\mathbb{P}_{\mu''}[X = 1, Y = 1] \geq 1/40000. \end{aligned}$$

This completes the proof of Claim 6.5.2. □

This completes the proof of Lemma 6.5.1. □

It follows from the above lemma that if a thread $P \in \mathcal{U}_\epsilon(\eta)$ has less than $\epsilon/4$ good thread edges,

then, for any two cut classes $\mathcal{C}_1 \to \mathcal{C}_2 \in P$ where A is the child connector of $\mathcal{C}_1$ and B is the father connector of $\mathcal{C}_2$, either $|A - \overline{B}| = 1$, or $|A - \overline{B} = 2|$, but $\mathbf{x}(E(A - \overline{B})) < \epsilon/4$. This property simplifies many of the cases that we consider in the rest of this section.

In the rest of this section, for a vertex $u \in V$, and $A \subseteq V$ we use

$$\mathbf{x}(u, A) := \sum_{\{u,v\} \in E : v \in A} x_{\{u,v\}}.$$

Note that u may be a vertex of A.

Corollary 6.5.3. *Let $P \in \mathcal{U}_0(\eta)$ for some $\epsilon \geq 0$, $\mathcal{C}_1 \to \mathcal{C}_2 \in P$ be cut classes with at most 4 atoms, A be the child-connector of $\mathcal{C}_1$, B be the father connector of $\mathcal{C}_2$, and $D = A - \overline{B}$. If $D = \{u\}$, then $1 - 4\eta \leq \mathbf{x}(u, \overline{A}), \mathbf{x}(u, \overline{B}) \leq 1 + 4\eta$, and if $D = \{u, v\}$, then*

$$1 - 4\eta - x_{\{u,v\}} \leq \mathbf{x}(u, \overline{A}), \mathbf{x}(u, \overline{B}), \mathbf{x}(v, \overline{B}), \mathbf{x}(v, \overline{B}) \leq 1 + 4\eta$$

Proof. We start by proving the case that $|D| = 1$. Let $D = \{u\}$. By Proposition 3.2.1

$$\begin{aligned} 2 \leq \quad & \mathbf{x}(\delta(\overline{A})) = \mathbf{x}(u, \overline{A}) + \mathbf{x}(E(\overline{A}, \overline{B})) \quad \leq 2(1 + 2\eta) \\ 2 \leq \quad & \mathbf{x}(\delta(\overline{B})) = \mathbf{x}(u, \overline{B}) + \mathbf{x}(E(\overline{A}, \overline{B})) \quad \leq 2(1 + 2\eta) \end{aligned}$$

Adding up the above inequality and using $\mathbf{x}(\delta(u)) = 2$ we get

$$4 \leq \mathbf{x}(\delta(u)) + 2\mathbf{x}(E(\overline{A}, \overline{B})) = 2 + 2\mathbf{x}(E(\overline{A}, \overline{B})) \leq 4(1 + 2\eta).$$

Therefore, $1 \leq \mathbf{x}(E(\overline{A}, \overline{B})) \leq 1 + 4\eta$. So,

$$\mathbf{x}(u, \overline{A}), \mathbf{x}(u, \overline{B}) \geq 1 - 4\eta.$$

Now, suppose $D = \{u, v\}$. Since $\mathbf{x}(\delta(\{u, v\})) = 4 - 2x_{\{u,v\}}$, similar to above we get

$$\begin{aligned} x_{\{u,v\}} \leq \quad & \mathbf{x}(E(\overline{A}, \overline{B})) \quad \leq 4\eta + x_{\{u,v\}}, \\ 2 - 4\eta - x_{\{u,v\}} \leq \quad & \mathbf{x}(E(D, \overline{A})), \mathbf{x}(E(D, \overline{B})) \quad \leq 2 + 4\eta - x_{\{u,v\}}. \end{aligned}$$

For the sake of contradiction, assume $\mathbf{x}(u, \overline{A}) < 1 - 4\eta - x_{\{u,v\}}$ (the other cases are similar). Then, by the above equation $\mathbf{x}(v, \overline{A}) > 1$. Therefore,

$$2 \leq \mathbf{x}(\delta(\{v\} \cup \overline{A})) = \mathbf{x}(u, \overline{A}) + \mathbf{x}(v, \{u\} \cup \overline{B}) + \mathbf{x}(E(\overline{A}, \overline{B})) < 1 - 4\eta - x_{\{u,v\}} + 1 + 4\eta + x_{\{u,v\}} = 2,$$

which is a contradiction. □

The following technical lemma is the key lemma of this section.

Lemma 6.5.4. *Let $A \subseteq V$ and $v \in \overline{A}$, let μ be any λ-random spanning tree distribution, $\delta(A) = S_X \cup S_Y$, and $\mu' = \{\mu \mid |T \cap E(A)| = |A| - 1\}$. For $T \sim \mu$ let*

$$X := |T \cap S_X|, Y := |T \cap S_Y|, I := |T \cap E(\{v\}, A)|, Z := |T \cap E(\{v\}, \overline{A})|.$$

If for $0 < \epsilon < 1/50$, $\mathbb{E}\left[|T \cap E(A)\right] \geq |A| - 1 - (1/2 - \epsilon)$ and

$$1 - \epsilon/2 \leq \mathbb{E}\left[X\right], \mathbb{E}\left[Y\right], \mathbb{E}\left[I\right], \mathbb{E}\left[Z\right] \leq 1 + \epsilon$$

Then,

$$\mathbb{P}_{\mu'}\left[X = Y = Z = I = 1\right] \geq \epsilon^2/2000.$$

Proof. Let

$$I_X := |T \cap E(\{v\}, A) \cap S_X|, I_Y := |T \cap E(\{v\}, A) \cap S_Y|.$$

Note that $I_X + I_Y = I$. First, observe that $\mathbb{P}_{\mu'}\left[I \leq 1\right] = 1$, this is because any tree $T \sim \mu'$ with $|T \cap E(\{v\}, A)| = 1$ has a cycle. We consider two cases.

Case 1: $\mathbb{E}_{\mu'}\left[I_X\right], \mathbb{E}_{\mu'}\left[I_Y\right] \geq \epsilon/4$. Let $\mu'' = \{\mu' \mid I = 1\}$. The important observation is that, for $T \sim \mu''$, any function of edges of $E(\{v\}, A)$ is independent of any function of the rest of the edges of G. Therefore,

$$\mathbb{P}_{\mu'}\left[X = Y = Z = I = 1\right] = \min\{\mathbb{P}_{\mu''}\left[I_X = 1\right], \mathbb{P}_{\mu''}\left[I_Y = 1\right]\} \cdot \mathbb{P}_{\mu''}\left[X + Y - I = 1, Z = 1\right]. \tag{6.5.2}$$

We lower bound both of the terms in the right hand side. First, observe that

$$\mathbb{P}_{\mu''}\left[I_X = 1\right] = 1 - \mathbb{P}_{\mu''}\left[I_Y = 1\right] = \frac{\mathbb{E}_{\mu'}\left[I_X\right]}{\mathbb{E}_{\mu'}\left[I\right]}.$$

So, $\mathbb{P}_{\mu''}\left[I_X\right], \mathbb{P}_{\mu''}\left[I_Y\right] \geq \epsilon/2$, and the first term of the RHS of (6.5.2) is at least $\epsilon/4$.

It remains to lower bound the second term. We will use Lemma 3.3.3. By Corollary 2.9.9

$$\mathbb{E}_{\mu''}\left[Z\right] \geq \mathbb{E}_{\mu}\left[Z\right] - (|A| - \mathbb{E}_{\mu}\left[I + |T \cap E(A)|\right]) \geq \mathbb{E}_{\mu}\left[Z\right] - 1/2 + \epsilon/2 \geq 1/2.$$

Similarly,

$$\mathbb{E}_{\mu''}\left[X + Y - I\right] = \mathbb{E}_{\mu''}\left[X + Y\right] - 1 \geq \mathbb{E}_{\mu'}\left[X + Y\right] - 1 \geq \mathbb{E}_{\mu}\left[X + Y\right] - 1 - (1/2 - \epsilon) \geq 1/2,$$

and $\mathbb{E}_{\mu''}[X+Y+Z-I] \geq 1.5-\epsilon/2$. Putting together,

$$\begin{aligned} 1/2 \leq \quad & \mathbb{E}_{\mu''}[Z], \mathbb{E}_{\mu''}[X+Y-I] \quad \leq 1+2\epsilon \leq 1.1 \\ 1.5-\epsilon/2 \leq \quad & \mathbb{E}_{\mu''}[X+Y-I+Z] \quad \leq 2+3\epsilon \leq 2.1. \end{aligned}$$

where we used the assumption that $\epsilon < 1/50$. Therefore, by Proposition 2.9.16

$$\mathbb{P}_{\mu''}[X+Y+Z-I=2] \geq \min\{\mathrm{Ber}(1.4,2), \mathrm{Ber}(2.1,3)\} \geq 1/20.$$

Letting $\alpha = 0.45$, $\beta = 0.5$ and using Lemma 3.3.3

$$\mathbb{P}_{\mu'}[X=Y=Z=I=1] \geq \frac{\epsilon}{4}\mathbb{P}_{\mu''}[X+Y-I=Z=1] \geq \frac{\epsilon}{2000}.$$

Case 2: $\mathbb{E}_{\mu'}[|T \cap I_X|] < \epsilon/4$**.** First, observe that

$$\mathbb{E}_{\mu'}[I_Y] \geq \mathbb{E}_{\mu'}[I] - \epsilon/4 \geq \mathbb{E}_{\mu}[I] - (1/2-\epsilon) - \epsilon/4 = 1/2 + \epsilon/4. \tag{6.5.3}$$

where we used the lemma's assumption that $\mathbb{E}_{\mu}[I] \geq 1-\epsilon/2$. So, $\mathbb{E}_{\mu'}[Y-I_Y] \leq 1/2+\epsilon$. Let $\mu'' = \{\mu | I_Y = 1, Y - I_Y = 0\}$. Note that $I_X = 0$ for any $T \sim \mu''$. Therefore, it is sufficient to lower bound $\mathbb{P}_{\mu''}[X=1, Z=1]$. We will use Lemma 3.3.3.

By Corollary 2.9.9, it follows that any of the random variables $X, Z, X+Z$ is decreased no more than $1-5\epsilon/4$. For example,

$$\mathbb{E}_{\mu''}[X] \geq \mathbb{E}_{\mu'}[X] - (1-\mathbb{E}_{\mu'}[I_Y]) \geq \mathbb{E}_{\mu}[X] - (1/2-\epsilon) - (1/2-\epsilon/4) \geq \mathbb{E}_{\mu}[X] - (1-5\epsilon/4) \geq 3\epsilon/4.$$

where we used (6.5.3), and the lemma's assumption that $\mathbb{E}[X] \geq 1-\epsilon/2$. Therefore,

$$\begin{aligned} \epsilon/2 \leq \quad & \mathbb{E}_{\mu''}[X], \mathbb{E}_{\mu''}[Z] \quad \leq 1+\epsilon+1.5+2\epsilon \leq 1.6 \\ 1+\epsilon/4 \leq \quad & \mathbb{E}_{\mu''}[X+Z] \quad \leq 2.5+3\epsilon \leq 2.6 \end{aligned}$$

where we used $\epsilon < 1/50$. Therefore, by Proposition 2.9.16,

$$\mathbb{P}_{\mu''}[X+Z=2] \geq \min\{\mathrm{Ber}(1+\epsilon/4, 2), \mathrm{Ber}(2.6, 2)\} \geq \epsilon/4(1-\epsilon/4).$$

where we used $\epsilon < 1/50$. Letting $\alpha = 0.2$, $\beta = \epsilon/2$, by Lemma 3.3.3 we get

$$\mathbb{P}_{\mu''}[X=1, Z=1] \geq \frac{\epsilon^2}{130}.$$

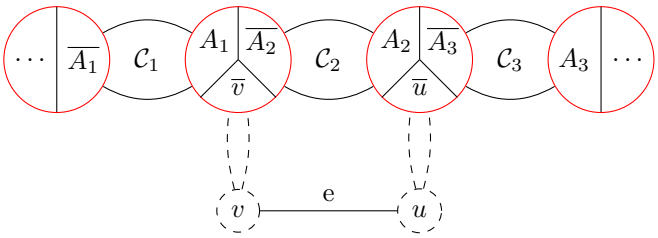

Figure 6.5.5: Setting in Lemma 6.5.6. Dashed edges represent deleted trivial cut classes. Connecting atoms are represented in the same circle.

Therefore,

$$\mathbb{P}_{\mu'}\left[X = Y = Z = I = 1\right] \geq \mathbb{P}_{\mu'}\left[I_Y = 1\right] \cdot \mathbb{P}_{\mu'}\left[Y - I_Y = 0 \mid I_Y = 1\right] \cdot \mathbb{P}_{\mu''}\left[X = 1, Z = 1\right] \geq \frac{\epsilon^2}{1400}.$$

This completes the proof of Lemma 6.5.4. □

6.5.1 Threads with non-proper Cut Classes

In the part we discuss threads where all of their cut classes are non-proper, i.e., have exactly 2 atoms. The following lemma is the main result of this subsection.

Lemma 6.5.5. *Let $P \in \mathcal{U}_\epsilon(\eta)$ for $\epsilon < 1/100$, $\eta < \epsilon/100$, $\mathcal{C}_1 \to \mathcal{C}_2 \to \ldots \to \mathcal{C}_6 \in P$ be non-proper cut classes with child-connectors $A_1, \ldots, A_6$ respectively. Then P had good thread edges of fraction at least $\epsilon/4$ that are even with probability at least $\frac{\epsilon^2}{33000}$.*

Before proving this lemma we show that if there is an edge $e = \{u, v\}$ for $v \in A_i - A_{i+1}$, and $u \in A_{i+1} - A_{i+2}$, where $1 \leq i \leq 4$, then e is a good.

Lemma 6.5.6. *Let $P \in \mathcal{U}_\epsilon(\eta)$ for $\epsilon < 1/100$ and $\eta < \epsilon/100$, $\mathcal{C}_1 \to \mathcal{C}_2 \to \mathcal{C}_3 \in P$ be non-proper cut classes with child-connectors A_1, A_2, A_3 respectively. For any $u \in A_2 - A_3$, $v \in A_1 - A_2$, If $e = \{u, v\} \in E$ and P has less than $\epsilon/4$ good thread edges then*

$$\mathbb{P}\left[\mathcal{E}(e)\right] = \frac{\epsilon^2}{33000}.$$

Proof. As shown in Figure 6.5.5 e is contained in 3 near minimum cuts: 2 trivial degree cuts, and the near minimum cut defined by the cut class $\mathcal{C}_2$.

Our proof is simply an application of Lemma 6.5.4. Since $(A_2, \overline{A_2})$ is a $(1+\eta)$ near minimum cut,

$$\mathbb{E}_\mu\left[|T \cap E(A_2)|\right] = \mathbf{z}(E(A_2)) \geq |A_2| - 1 - \eta - |A_2|/n \geq |A_2| - 1 - (1/2 - \epsilon) - \eta.$$

where in the last equation we used the assumption that $P \in \mathcal{U}_\epsilon(\eta)$, i.e., $|B_2| \geq (1/2 + \epsilon)n$. Since P

has less than $\epsilon/4$ good thread edges, by Lemma 6.5.1 and Corollary 6.5.3,

$$\begin{aligned} 1-\epsilon/4-4\eta &\leq \mathbf{x}(v,A_2), \mathbf{x}(u,\overline{A_2}) &&\leq 1+4\eta, \\ 1-4\eta &\leq \mathbf{x}(v,\overline{A_2}) &&\leq 1+4\eta+\epsilon/4, \\ 1-4\eta &\leq \mathbf{x}(A_2-\{u\},\overline{A_2}) &&\leq 1+\epsilon/4+6\eta. \end{aligned}$$

Now, let $\mu' = \{\mu \mid |T \cap E(A_2)| = |A_2|-1\}$. Also, let $S_X := E(\{u\},\overline{A_2}), S_Y := \delta(A_2)-S_X$, $\epsilon' = \epsilon-\eta$. Since $\eta < \epsilon/100$, $\epsilon/4+4\eta < \epsilon'/2$. Therefore, for $A = A_2$ and $\epsilon = \epsilon'$ Lemma 6.5.4 implies

$$\mathbb{P}_{\mu'}\left[|S_X \cap T| = 1, |S_Y \cap T| = 1, |\delta(v) \cap T| = 2\right] \geq \frac{{\epsilon'}^2}{2000}.$$

Observe that for $T \sim \mu'$ random variables which are functions of edges inside A_2 are independent of functions of edges outside A_2. Therefore,

$$\begin{aligned} \mathbb{P}_{\mu}\left[\mathcal{E}(e)\right] \geq \frac{1}{2}\mathbb{P}_{\mu'}\left[\mathcal{E}(e)\right] &= \frac{1}{2}\mathbb{P}_{\mu'}\left[|S_X \cap T| = |S_Y \cap T| = 1, |\delta(v) \cap T| = 2\right] \cdot \mathbb{P}_{\mu'}\left[|T \cap E(A_2) \cap \delta(u)| = 1\right] \\ &\geq \frac{{\epsilon'}^2}{4000} \cdot \mathbb{P}_{\mu'}\left[|T \cap E(A_2) \cap \delta(u)| = 1\right]. \end{aligned}$$

where the first inequality follows by Corollary 6.4.2. But, by Corollary 6.5.3,

$$1-\epsilon/4-4\eta \leq \mathbf{x}(u,A_2) \leq 1+4\eta,$$

so $0.9 \leq \mathbb{E}_{\mu'}\left[|T \cap E(A_2) \cap \delta(u)|\right] \leq 1.5$. Therefore, the lemma follows by an application of Proposition 2.9.16,

$$\mathbb{P}_{\mu'}\left[|T \cap E(A_2) \cap \delta(u)| = 1\right] \geq \min\{\mathrm{Ber}(0.9,1), \mathrm{Ber}(1.5,1)\} \geq 1/8.$$

□

Now we are ready to prove Lemma 6.5.5:

Proof of Lemma 6.5.5. Let $D_i := A_i - A_{i+1}$, for $1 \leq i \leq 5$. By Lemma 6.5.1 if $|D_i| > 2$ for some $1 \leq i \leq 5$, then we are done, so assume $|D_i| \leq 2$ for all $i \leq 5$. Also, if $|D_i| = 2$, for some i, then there is edge of fraction at least $1/3$ between a pair of vertices in D_i and D_{i+1}; so by Lemma 6.5.6 P has good thread edges of fraction at least $\frac{1}{3} \geq \epsilon/4$ and we are done. So assume $D_i = \{u_i\}$ for $1 \leq i \leq 5$ (see Figure 6.5.6).

By Lemma 6.5.6 if two consecutive vertices are connected by an edge of fraction $\frac{1}{100} \geq \epsilon/4$, then that edge is good and we are done. So, suppose $\forall 1 \leq i \leq 5 : x_{\{u_i,u_{i+1}\}} \leq \frac{1}{100}$ (note that this can be even zero). We show that $x_{\{u_2,u_4\}} \geq \frac{19}{20}$, and it is even with a constant probability. Edge (u_2,u_4) is included in 4 near minimum cuts: 2 degree cuts, and the cuts corresponding to the cut classes $\mathcal{C}_3$ and $\mathcal{C}_4$ (i.e., $(A_3,\overline{A_3})$ and $(A_4,\overline{A_4})$); so we need to show all these 4 cuts are even with

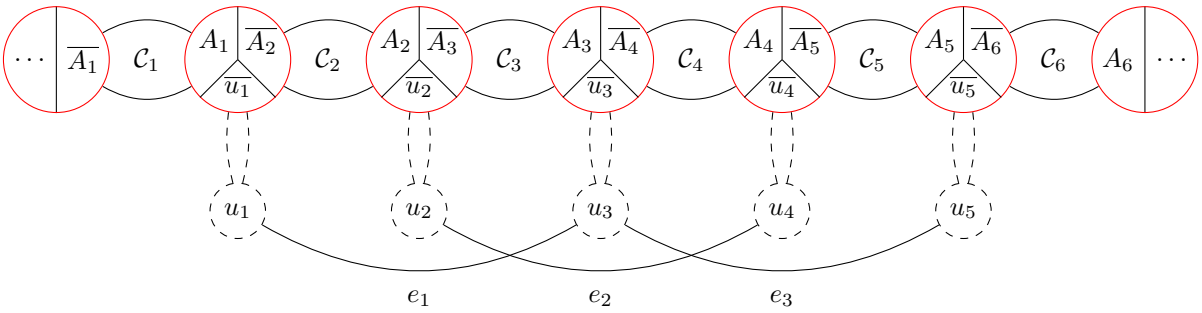

Figure 6.5.6: Setting in Lemma 6.5.5.

a constant probability. We prove that the 3 edges $\{u_1, u_3\}, \{u_2, u_4\}, \{u_3, u_5\} \in E$ and each have a large fraction, then by union bound all of them will be sampled in spanning tree $T \sim \mu$, and thus the 2 cuts $(A_3, \overline{A_3})$ and $(A_4, \overline{A_4})$ are even. Finally, we use Lemma 3.3.3 to show that the degree of u_2 and u_4 is even with a constant probability.

By Corollary 6.5.3 we have

$$\begin{aligned} 1 - 4\eta - \frac{2\epsilon}{4} &\leq \mathbf{x}(u_3, \overline{A_2}), \mathbf{x}(u_2, \overline{A_1}), \\ 1 - 4\eta - \frac{\epsilon}{4} &\leq \mathbf{x}(u_1, A_2) \end{aligned}$$

Thus,

$$x_{\{u_1,u_3\}} \geq \mathbf{x}(u_1, A_2) - (\mathbf{x}(\delta(A_2)) - \mathbf{x}(u_2, \overline{A_1}) - \mathbf{x}(u_3, \overline{A_2})) \geq 1 - 14\eta - 5\epsilon/4 \geq \frac{19}{20}.$$

Similarly, it can be shown that $x_{\{u_3,u_5\}}, x_{\{u_2,u_4\}} \geq \frac{19}{20}$. Let $e_1 = \{u_1, u_3\}, e_2 = \{u_2, u_4\}, e_3 = \{u_3, u_5\}$. Define

$$X := |T \cap \{E(A_3, \overline{A_3}) \cup E(A_4, \overline{A_4})\} - \{e_1, e_2, e_3\}|,$$

and let I_1, I_2, I_3 be the indicator random variables for the edges e_1, e_2, e_3 respectively. Let $\mu' := \{\mu \mid X = 0\}$. Since

$$\mathbb{E}_\mu [X] \leq 4(1 + \eta) - x_{e_1} - 2x_{e_2} - x_{e_3} \leq \frac{1}{5} + 4\eta,$$

by Corollary 2.9.9, this can only increase the probability of other edges by at most $\frac{1}{5} + 4\eta$. Now let $\mu'' = \{\mu' \mid I_1 = 1, I_2 = 1, I_3 = 1\}$. By Fact 2.9.8,

$$\mathbb{E}_{\mu'} [I_1 + I_2 + I_3] \geq \mathbb{E}_\mu [I_1 + I_2 + I_3] \geq 3 - \frac{3}{20}.$$

Therefore,

$$\mathbb{P}_\mu [\mathcal{E}(e_2)] = \mathbb{P}_{\mu'} [\mathcal{E}(e_2)] \mathbb{P}_\mu [X = 0] \geq (4/5 - 4\eta) \mathbb{P}_{\mu''} [\mathcal{E}(e_2)] \mathbb{P}_{\mu'} [I_1 = 1, I_2 = 1, I_3 = 1] \geq 0.7 \mathbb{P}_{\mu''} [\mathcal{E}(e_2)]$$

Thus it is sufficient to show $\mathbb{P}_{\mu''}[\mathcal{E}(e_2)]$ is a constant. In any tree $T \sim \mu''$, cuts $(A_3, \overline{A_3})$ and $(A_4, \overline{A_4})$ are even. So, it is sufficient to show u_2 and u_4 have even degree. By Corollary 2.9.9,

$$\begin{aligned} 4/5 \leq \quad & \mathbb{E}_{\mu''}[|T \cap \delta_{u_2}(u_4)|], \mathbb{E}_{\mu''}[|T \cap \delta_{u_4}(u_2)|] \quad \leq 7/5 \\ 9/5 \leq \quad & \mathbb{E}_{\mu''}[|T \cap \delta_{u_4}(u_2)| + |T \cap \delta_{u_2}(u_4)|] \quad \leq 12/5 \end{aligned}$$

Therefore, by Proposition 2.9.16,

$$\mathbb{P}_{\mu''}[|T \cap \delta_{u_4}(u_2)| + |T \cap \delta_{u_2}(u_4) = 2] \geq \min\{\text{Ber}(9/5, 2), \text{Ber}(12/5, 2)\} \geq 0.025.$$

Letting $A := \delta_{u_4}(u_2), B := \delta_{u_2}(u_4), \alpha = \frac{3}{10}, \beta := \frac{4}{5}$, by Lemma 3.3.3,

$$\mathbb{P}_{\mu''}[\mathcal{E}(e_2)] \geq 0.7 \cdot \mathbb{P}_{\mu''}[|T \cap \delta_{u_4}(u_2)| = 1, |T \cap \delta_{u_2}(u_4)| = 1] \geq 1/2000.$$

□

6.5.2 Threads with Proper Cut Classes

It remains to consider the cases where a thread P contains cut classes with 4 atoms. Let $\mathcal{C} \in P$ be a cut class with 4 atoms. Similar to the proof of Lemma 6.4.4, observe that if $\mathcal{C}_i$ has a non-connector non-singleton atom, or a consecutive pair of singletons, then P has good thread edges of fraction at least 3/4. We say a cut class is an *exceptional 4-cut class* if it doesn't have a non-connector non-singleton atom or a consecutive pair of singletons (see Figure 6.5.7 for an example). It follows that an exceptional 4-cut class $\mathcal{C}$ has a very simple structure: say $\psi(\mathcal{C}) = \{A_1, A_2, A_3, A_4\}$ in the order that atoms are placed in the polygon representation and A_1 is the father-connector, then A_2 and A_4 are (non-connecting) singletons and A_3 is the unique child-connector. It follows from Corollary 3.2.3 that $\mathbf{x}(E(A_i, A_{i+1})) \geq 1 - \eta$ for all $1 \leq i \leq 4$.

The following is the main result of this subsection.

Lemma 6.5.7. *Let $P \in \mathcal{U}_\epsilon(\eta)$ for $\epsilon < 1/100$ and $\eta < \epsilon/100$. Let $\mathcal{C}_1 \to \mathcal{C}_2 \to \mathcal{C}_3 \in P$ be 3 cut classes with child-connectors A_1, A_2, A_3 and father-connectors B_1, B_2, B_3 such that $\mathcal{C}_3$ has 4 atoms. If all of classes with 4 atoms in P are exceptional 4-cut class, then P had good thread edges of fraction at least $\epsilon/4$, that are even with probability at least $\frac{\epsilon^2}{10000}$.*

Recall that for a cut class $\mathcal{C}$, $\mathcal{H}_{\mathcal{C}}(T)$ is the event that T is a Hamiltonian cycle in $G(\psi(\mathcal{C}))$.

Lemma 6.5.8. *Let $P \in \mathcal{U}_\epsilon(\eta)$ for $\epsilon > 1/100$, and $\eta < \epsilon/100$, $\mathcal{C}_1 \to \mathcal{C}_2 \in P$, be 2 cut classes with child connectors A_1, A_2, father connectors B_1, B_2 such that $\mathcal{C}_2$ is an exceptional 4-cut class and $\{u\}$ and $\{v\}$ are the two singletons of $\mathcal{C}_2$. If P has less than $\epsilon/4$ fraction of good thread edges, then for*

any $w \in A_1 - A_2 - \{u, v\}$,

$$\mathbb{P}\left[\mathcal{H}_{\mathcal{C}_2}(T), |T \cap \delta(w)| = 2\right] \geq \frac{\epsilon^2}{10000}.$$

Proof. The proof is essentially an application of Lemma 6.5.4. Let $Z := |T \cap (\{u, v\} \cup E(A_2, B_2))|$. By Corollary 3.2.3,

$$\mathbb{E}_\mu[X] = \mathbf{x}(E(\overline{A_2}, \overline{B_2})) + x_{\{u,v\}} \leq 4(1 + 2\eta) - 4(1 - \eta) = 12\eta.$$

Let $\mu' = \{\mu \mid Z = 0\}$. Let $A := A_2 \cup \{u, v\} = \overline{B}$. Since $\mathbf{x}(A, \overline{A}) \leq 2 + 4\eta$,

$$\mathbb{E}_{\mu'}[|T \cap E(A)|] \geq \mathbf{z}(E(A)) \geq |A| - 4\eta - |A|/n \geq |A| - 1 - (1/2 - \epsilon) - 4\eta.$$

where we used the assumption that $P \in \mathcal{U}_\epsilon(\eta)$, i.e., $|B_2| \geq (1/2 + \epsilon)n$. Since P has less than $\epsilon/4$ good thread edges, by Lemma 6.5.1 and Corollary 6.5.3

$$\begin{aligned} 1 - \epsilon/4 - 4\eta \leq \quad & \mathbf{x}(w, A) && \leq 1 + 16\eta, \\ 1 - \eta \leq \quad & \mathbf{x}(u, \overline{A}), \mathbf{x}(v, \overline{A}) && \leq 1 + 11\eta. \\ 1 - 4\eta \leq \quad & \mathbf{x}(w, \overline{A}) && \leq 1 + \epsilon/4 + 16\eta \end{aligned}$$

Now, let $\mu'' = \{\mu \mid |T \cap E(A)| = |A| - 1\}$. Also, let $S_X := \delta(u) \cap \delta(A)$, $S_Y := \delta(A) \cap \delta(v) = \delta(A) - S_X$, $\epsilon' = \epsilon - 4\eta$. Since $\eta < \epsilon/100$ and $\epsilon/4 + 16\eta \leq \epsilon'/2$. Therefore, for $\mu = \mu'$ and $\epsilon = \epsilon'$ Lemma 6.5.4 implies

$$\mathbb{P}_{\mu''}[|E(\{u\}, B_2) \cap T| = 1, |E(\{v\}, B_2) \cap T| = 1, |\delta(w) \cap T| = 2] \geq \frac{{\epsilon'}^2}{2000}.$$

To prove the lemma it remains to show for $T \sim \mu''$ $|T \cap E(\{u\}, A_2)| = 1$ and $|T \cap E(\{v\}, A_2)| = 1$ occur with a constant probability. Since $T \sim \mu''$ is a spanning tree in the induced subgraph $G[A] = G[A_2 \cup \{u, v\}$, it is sufficient to show $T \sim \mu''$ is also a spanning tree in the induced subgraph $G[A_2]$ with a constant probability. Furthermore, since for $T \sim \mu''$ random variables which are functions of edges inside A are independent edges outside A_2,

$$\begin{aligned} \mathbb{P}_\mu[\mathcal{H}_{\mathcal{C}_2}(T)] \quad &\geq \quad (1 - 12\eta)\mathbb{P}_{\mu'}[\mathcal{H}_{\mathcal{C}_2}(T)] \\ &\geq \quad (1 - 12\eta)\mathbb{P}_{\mu''}[\mathcal{H}_{\mathcal{C}_2}(T)] \cdot \mathbb{P}_{\mu'}[|T \cap E(A)| = |A| - 1] \\ &\geq \quad 0.5 \cdot \mathbb{P}_{\mu''}[|T \cap E(\{u\}, B_2)| = 1, |T \cap E(\{v\}, B_2)| = 1, |T \cap \delta(w)| = 2] \\ &\qquad \cdot \mathbb{P}_{\mu''}[|T \cap E(A_2)| = |A_2| - 1] \end{aligned}$$

Finally by Theorem 2.9.11

$$\mathbb{E}_{\mu''}[|T \cap E(A_2)|] \geq \mathbb{E}_{\mu'}[|T \cap E(A_2)|] \geq \mathbb{E}_\mu[|T \cap E(A_2)|] = \mathbf{z}(E(A_2)) \geq |A_2| - 1 - (1/2 - \epsilon) - 4\eta - O(1/n).$$

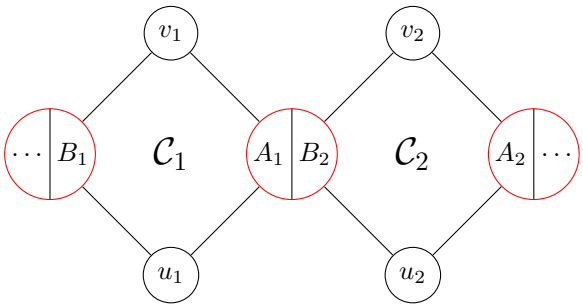

Figure 6.5.7: Setting in Lemma 6.5.9

where the second inequality follows by Corollary 2.9.9. So, $\mathbb{P}_{\mu''}\left[|T \cap E(A_2)| = |A_2| - 1\right] \geq 1/2$. □

The lemma implies the following Corollary.

Lemma 6.5.9. *Let $P \in \mathcal{U}_0(\eta)$ for $\eta < 1/1000$, $\mathcal{C}_1 \to \mathcal{C}_2 \in P$ be exceptional 4-cut classes with child connectors A_1, A_2 and father-connectors B_1, B_2. If $|A_1 - A_2| = 2$, then*

$$\mathbb{P}\left[\mathcal{H}_{\mathcal{C}_1}(T), \mathcal{H}_{\mathcal{C}_2}(T)\right] \geq 1/2000.$$

So, P has good thread edges of fraction at least $3/2$.

Proof. Let u_1, v_1 be the singletons of $\mathcal{C}_1$, and u_2, v_2 be the two singletons of $\mathcal{C}_2$ (see Figure 6.5.7). The proof is very similar to Lemma 6.4.3. Let $F := E(A_2)$, $I := E(A_1, B_1) \cup E(A_2, B_2) \cup \{\{u_1, v_1\}, \{u_2, v_2\}\}$. First we condition on $|T \cap I| = 0$. Then we condition on $|T \cap E(\overline{B_1})| = |\overline{B_1}| - 1$. Let μ'' be the resulting distribution. The important observation is that for any $T \sim \mu''$, $\mathcal{H}_{\mathcal{C}_1}(T), \mathcal{H}_{\mathcal{C}_2}(T)$ occur if and only if $|T \cap F| = |A_2| - 1 = \text{rank}(F)$ and $|T \cap E(\{u_1\}, B_1)| = |T \cap E(\{v_1\}, B_1)| = 1$. So,

$$\begin{aligned}\mathbb{P}\left[\mathcal{H}_{\mathcal{C}_1}(T), \mathcal{H}_{\mathcal{C}_2}(T)\right] &= \mathbb{P}\left[|T \cap I| = 0, |T \cap E(\overline{B_1})| = |\overline{B_1}| - 1\right] \cdot \mathbb{P}_{\mu''}\left[\mathcal{H}_{\mathcal{C}_1}(T), \mathcal{H}_{\mathcal{C}_2}(T)\right] \\ &\geq 2/5 \cdot \mathbb{P}_{\mu''}\left[|T \cap E(A_2)| = |A_2| - 1\right] \cdot \mathbb{P}_{\mu''}\left[|T \cap E(\{v_1\}, B_1)| = |T \cap E(\{u_1\}, B_1)| = 1\right]\end{aligned}$$

where we used $\mathbb{E}_\mu\left[|T \cap I|\right] \leq 24\eta \leq 1/20$, $\mathbb{E}_\mu\left[|T \cap E(\overline{B_1})|\right] \geq 1/2 - 2\eta$. Similar to Lemma 6.4.3 the first term of the RHS occur with probability at least $1/2 - 1/20$ and the second term occur with probability at least $1/20$. □

Now we are ready to prove Lemma 6.5.7:

Proof of Lemma 6.5.7. Let $\{u\}, \{v\}$ be the singletons of $\mathcal{C}_3$. If $|A_2 - \overline{B_3}| > 2$ then by Lemma 6.5.1 P has good thread edges of fraction at least $3/8$ and we are done. If $A_2 - \overline{B_3} = \{u', v'\}$ then if $x_{\{u',v'\}} \geq \epsilon/4$ then we are done by Lemma 6.5.1. If $x_{\{u',v'\}} < \epsilon/4$ then we must have $x_{\{u',u\}} \geq 1/3$

or $x_{\{u',v\}} \geq 1/3$. But by Lemma 6.5.8 either of these edges are good and we are done. Similarly, if $A_2 - \overline{B_3} = \{u'\}$ then P has a good edge $\{u', u\}$ or $\{u', v\}$ of fraction at least $1/3$.

So, it must be the case that $A_2 = \overline{A_3}$. If $\mathcal{C}_2$ has four atoms then we are done by Lemma 6.5.9. So, $\mathcal{C}_2$ has two atoms. Note that $\mathcal{C}_2$ shows that the cut $(B_3, \overline{B_3})$ is a $(1+\eta)$ near minimum cut.

Now, we consider two cases. FIrst assume $|A_1 - A_2| = 0$. Since $\mathcal{C}_1 \neq \mathcal{C}_2$, we have $\mathcal{C}_1$ is an exceptional 4-cut class and $A_1 = \overline{B_3}$. Then, by Lemma 6.5.9 $\mathcal{H}_{\mathcal{C}_1}(T), \mathcal{H}_{\mathcal{C}_2}(T)$ occur with a constant probability. And, any edge between the singletons of $\mathcal{C}_1$ and $\mathcal{C}_3$ is good. So, we are done.

Finally, suppose $1 \leq |A_1 - A_2| \leq 2$. Similar to above if $\mathbf{x}(E(A_1 - A_2)) \geq \epsilon/4$ then we are done by Lemma 6.5.1, and otherwise there is an edge of fraction at least $1/3$ from a vertex in $A_1 - A_2$ to u or v which is a good edge by Lemma 6.5.8. □

Part II

New Analysis of Spectral Graph Algorithms through Higher Eigenvalues

Chapter 7

Background

In this chapter we set up notations for all parts of part 2 of the thesis. We also review several technical backgrounds and theorems that we will use to prove our theorems.

Let $G = (V, E)$ be an undirected (weighted) graph with $n := |V|$ vertices. In this thesis we only work with finite graphs, although some our results naturally generalize to infinite graphs. FIrst, we define some functions and operators on graph G. Note that when the graph is clear in the context we directly work with these operators, but in some of the proof we work subgraphs of G in that case we use subscripts to differentiate between the value of a function on G or its subgraphs.

For all pair of vertices $u, v \in V$ let $w_{u,v}$ be the weight of the edge between u and v. We assume G is undirected, therefore, $w(u, v) = w(v, u)$ for all pair of vertices. Also, for each vertex $v \in V$ let $w(v) := \sum_{u \in V} w(v, u)$. The *volume* of a set $S \subseteq V$ is the summation of the weighted degree of vertices in S,

$$\mathrm{vol}(S) := \sum_{v \in S} w(v).$$

We write $\ell^2(V)$ for the Hilbert space of functions $f : V \to \mathbb{R}$. We use $\mathbf{0}$ to denote the all zero function, $\mathbf{0}(v) = 0$ for all $v \in V$ and $\mathbf{1}$ to denote the all 1 function. For two functions $f, g : V \to \mathbb{C}$ we define the ordinary inner product

$$\langle f, g \rangle := \sum_{v \in V} f(v) \cdot \overline{g(v)}.$$

where $\overline{g(v)}$ is the conjugate of $g(v)$. We say f and g are orthogonal if $\langle f, g \rangle = 0$. If $f, g \in \ell^2(V)$ then the above is simply $\langle f, g \rangle = \sum_{v \in V} f(v) \cdot g(v)$. The norm of a function $f \in \ell^2(V)$ is simply $\|f\| := \sqrt{\langle f, f \rangle}$.

We also write $\ell^2(V, w)$ for the Hilbert space of functions $f : V \to \mathbb{R}$ with inner product

$$\langle f, g \rangle_w := \sum_{v \in V} w(v) \cdot f(v) \cdot g(v),$$

and norm $\|f\|_w^2 = \langle f, f \rangle_w$. This space will be useful when we deal with non-regular graphs. It is recommended that readers assume that all of the graphs are regular and unweighted, thus all inner products are ordinary.

Unless otherwise specified, we use $A : \mathbb{C}^V \to \mathbb{C}^V$ to denote the adjacency operator of G, where for any function $f \in \mathbb{C}^V$ and $v \in V$, $Af(v) := \sum_{u \sim v} w(u, v) f(u)$, matrix of G. We also use D to denote the diagonal degree operator, where for any $v \in V$ $Df(v) := w(v) f(v)$. In this thesis, for the sake of brevity, we may use matrices and operators interchangeably. Since we only work with linear operators this translation is always possible. For example, the entry in row u and column v of A is defined as

$$A(u, v) = \langle A\mathbf{1}_v, \mathbf{1}_u \rangle = w(u, v) \cdot \mathbb{I}\left[(u, v) \in E\right],$$

where $\mathbb{I}[.]$ is the indicator function, and $\mathbf{1}_u(v) := 1$ if $v = u$ and $\mathbf{1}_u(v) = 0$ otherwise.

The *support* of a function $f : V \to \mathbb{R}$ is the set of vertices with non-zero value in f.

$$\text{supp}(f) := \{v : f(v) \neq 0\}.$$

We say two functions $f, g : V \to \mathbb{R}$ are *disjointly supported* if $\text{supp}(f) \cap \text{supp}(g) = \emptyset$. We say f is r-Lipschitz with respect to g if for all $u, v \in V$,

$$|f(u) - f(v)| \leq r \cdot |g(u) - g(v)|.$$

7.1 Spectral Graph Theory

Let $M \in \mathbb{C}^{V \times V}$ be a square matrix. The *kernel* of M, $\ker(M)$ is the set of functions $f : V \to \mathbb{R}$ such that $Mf = 0$.

A scalar λ is an *eigenvalue* of M iff there is a non-zero function $f : V \to \mathbb{C}$ such that

$$Mf = \lambda f.$$

In that case we call f an eigenfunction of M. The *eigen-space* of λ is the linear space of all function f where $Mf = \lambda_f$. For example, $\ker(M)$ is the same as the eigen-space of eigenvalue 0.

The above equation is equivalent to

$$(M - I\lambda)f = 0.$$

Since f is a non-zero function, the above equation is equivalent to

$$\det(M - I\lambda) = 0.$$

For a fixed matrix M, the function $\lambda \to \det(M - I\lambda)$ is a univariate polynomial of degree $|V| = n$. Therefore, it has n roots, or equivalently, M has n eigenvalues with multiplicities.

The adjoint of an operator M is the operator M^T which satisfies the following. For any $f, g \in V \to \mathbb{C}^n$, $\langle Mf, g\rangle = \langle f, M^T g\rangle$. If M is a real matrix then M^T is just the transpose of M. We say M is *self-adjoint* with respect to the inner product $\langle ., .\rangle$, if $M^T = M$. Therefore, any real symmetric matrix is self-adjoint. For example, the adjacency operator of A, or the degree operator D are self adjoint.

Self-adjoint operators have several important properties that will be crucial in many proofs in this thesis. First of all, all of their eigenvalues are real.

Lemma 7.1.1. *All of the eigenvalues of any self-adjoint operator are real.*

Proof. Let λ be an eigenvalue of a self-adjoint operator M with a non-zero eigenfunction f. Then,

$$\lambda \|f\|^2 = \langle \lambda f, f\rangle = \langle Mf, f\rangle = \langle f, Mf\rangle = \langle f, \lambda f\rangle = \overline{\lambda} \|f\|^2 .$$

Since $f \neq 0$, $\lambda = \overline{\lambda}$, and $\lambda \in \mathbb{R}$. □

Observe that since the eigenvalues of real symmetric matrices are real, their corresponding eigenfunction can be assumed to be real. In fact, for any function $f : V \to \mathbb{C}$, if it has non-zero real/imaginary parts, then the imaginary/real parts of f provide an eigenfunctions of the same eigenvalue.

Lemma 7.1.2. *For any two eigenfunctions f, g of a self-adjoint operator M with corresponding eigenvalues λ, λ', if $\lambda \neq \lambda'$, or $\langle f, g\rangle = 0$.*

Proof.

$$\lambda\langle f, g\rangle = \langle \lambda f, g\rangle = \langle Mf, g\rangle = \langle f, Mg\rangle = \langle f, \lambda' g\rangle = \lambda'\langle f, g\rangle.$$

Therefore, $\langle f, g\rangle = 0$. □

From now on we assume M is a symmetric real matrix. Since the eigenvalues of M are real, without loss of generality, we can assume the eigenfunctions are also real. So, all of the eigenfunctions of M lie in the space $\ell^2(V)$.

The next theorem is the fundamental theorem of spectral graph theory:

Theorem 7.1.3 (Spectral Theorem)**.** *Let $M \in \mathbb{R}^{V\times V}$ be a symmetric matrix. There are n real eigenvalues $\lambda_1, \ldots, \lambda_n$ together with n orthonormal functions $f_1, \ldots, f_n : V \to \mathbb{R}$ such that for all $1 \leq i \leq n$, $Mf_i = \lambda_i f_i$.*

The next proof is based the lecture notes of Trevisan in the course of Spectral Graph Theory.

Proof. We prove this by induction. First, from the above arguments, M has at least one eigenfunction f_1 corresponding to an eigenvalue λ_1. Suppose, we are given k orthonormal functions $f_1, \ldots, f_k$ corresponding to $\lambda_1, \ldots, \lambda_k$. Let $\mathcal{S} \subseteq \mathbb{R}^n$ be the space of all functions that are orthogonal to $f_1, \ldots, f_k$. We show that there is a function $f_{k+1} \in \mathcal{S}$ such that $Mf_{k+1} = \lambda_{k+1} f_{k+1}$ for some $\lambda_{k+1} \in \mathbb{R}$. This completes the proof of theorem.

First, observe that for any $f \in \mathcal{S}$, $Mf \in \mathcal{S}$. This is because for any $1 \leq i \leq k$,

$$\langle Mf, f_i \rangle = \langle f, Mf_i \rangle = \lambda_i \langle f, f_i \rangle = 0. \tag{7.1.1}$$

Let $g_1, \ldots, g_{n-k}$ be a basis for $\mathcal{S}$. Let $B : \mathbb{R}^{n-k} \to \mathcal{S}$ be the following operator: $Bh := \sum_{i=1}^{n-k} h(i) \cdot g_i$. Also, let B^T be the adjoint of B; in this case for $g \in \mathcal{S}$, and $1 \leq i \leq n-k$, $B^T g(i) = \langle g, g_i \rangle$. Let $M' := B^T M B$. Since for any $g, h \in \mathbb{R}^{n-k}$, $\langle B^T M B g, h \rangle = \langle g, B^T M B h \rangle$, M' is self-adjoint. Therefore, it has an eigenfunction h with an eigenvalue $\lambda \in \mathbb{R}$. Thus,

$$B^T M B h = \lambda h,$$

and

$$B B^T M B h = \lambda B h. \tag{7.1.2}$$

Since $Bh \in \mathcal{S}$, by equation (7.1.1), $MBh \in \mathcal{S}$. On the other hand, BB^T is an identity on $\mathcal{S}$. This is because for any $g \in \mathcal{S}$, and $1 \leq i \leq n-k$,

$$\langle g_i, BB^T g \rangle = \langle B^T g_i, B^T g \rangle = \langle g_i, g \rangle.$$

Therefore, $BB^T MBh = MBh$, and by equation (7.1.2) $MBh = \lambda Bh$. Thus, $f_{k+1} = Bh$ is an eigenfunction of M with eigenvalue λ. □

One of the important consequence of the above theorem is that we can provide an explicit formula for the $M(u, v)$ in terms of the eigenvalues/eigenfunctions of M.

Corollary 7.1.4. *Let $M \in \mathbb{R}^{V \times V}$ be a symmetric matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$ and the corresponding orthonormal eigenfunctions $f_1, \ldots, f_n$. For any two functions $f, g \in \ell^2(V)$,*

$$\langle Mf, g \rangle = \sum_{i=1}^{n} \lambda_i \cdot \langle f_i, g \rangle \cdot \langle f_i, f \rangle.$$

So, for any $u, v \in V$

$$M(u, v) = \langle M \mathbf{1}_u, \mathbf{1}_v \rangle = \sum_{i=1}^{n} \lambda_i \cdot f_i(u) \cdot f_i(v).$$

Proof. Let $\alpha_i = \langle f_i, f\rangle$, $\beta_i = \langle f_i, g\rangle$. Since $f_1, \ldots, f_n$ form a basis of $\ell^2(V)$,

$$\langle Mf, g\rangle = \left\langle M\sum_{i=1}^n \alpha_i f_i, \sum_{i=1}^n \beta_i f_i \right\rangle = \left\langle \sum_{i=1}^n \lambda_i \alpha_i f_i, \sum_{i=1}^n \beta_i f_i \right\rangle = \sum_{i=1}^n \lambda_i \cdot \alpha_i \cdot \beta_i.$$

□

We say M is *positive semidefinite* if all eigenvalues of M are non-negative. We also write $M \preceq M'$ if $M' - M$ is positive semidefinite. We say M is *non-singular* if all eigenvalues of M are non-zero, and it singular otherwise. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of M with corresponding eigenfunctions $f_1, \ldots, f_n$. If M is non-singular, then the *inverse* of M, M^{-1} is defined as follows: the eigenfunctions of M^{-1} are $f_1, \ldots, f_n$ with corresponding eigenvalues $1/\lambda_1, \ldots, 1/\lambda_n$. In other words, the u, v entry of M^{-1} is,

$$M^{-1}(u, v) = \sum_{i=1}^n \frac{1}{\lambda_i} \cdot f_i(u) \cdot f_i(v).$$

If M is singular, then we can define the *pseudo-inverse* of M, $M^\dagger$, as follows: the eigenfunctions of $M^\dagger$ are $f_1, \ldots, f_n$, and for any $1 \le i \le n$, if $\lambda_i \neq 0$, then the corresponding eigenvalue of f_i is $1/\lambda_i$, otherwise the corresponding eigenvalue of f_i is 0. In other words, the u, v entry of $M^\dagger$ is,

$$M^\dagger(u, v) = \sum_{i:\lambda_i \neq 0} \frac{1}{\lambda_i} \cdot f_i(u) \cdot f_i(v).$$

Next, we describe an equivalent definition of eigenvalues and eigenfunctions as the optimizers of a mathematical quantity.

Theorem 7.1.5 (Variational Characterization of Eigenvalues)**.** *Let $M \in \mathbb{R}^{V \times V}$ be a symmetric matrix with real eigenvalues $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$, counted with multiplicities. For any $k \ge 1$, and any set of orthonormal functions $g_1, \ldots, g_k$*

$$\lambda_{k+1} \ge \min_{\substack{f \in \ell^2(V) - \mathbf{0} \\ f \perp g_1, \ldots, g_k}} \frac{\langle Mf, f\rangle}{\langle f, f\rangle}. \tag{7.1.3}$$

Furthermore, if $Mg_i = \lambda_i g_i$ for all $1 \le i \le k$, then we have an equality and the minimizer is an eigenfunction of λ_{k+1}.

Proof. Let $f_1, \ldots, f_{k+1}$ be orthonormal eigenfunctions corresponding to $\lambda_1, \ldots, \lambda_{k+1}$. Since $f_1, \ldots, f_{k+1}$ are orthogonal, there is a vector $f \in \text{span}\{f_1, \ldots, f_{k+1}\}$ such that $f \perp g_1, \ldots, g_k$. But, by Corollary 7.1.4,

$$\frac{\langle Mf, f\rangle}{\langle f, f\rangle} = \frac{\sum_{i=1}^{k+1} \lambda_i \langle f, f_i\rangle^2}{\sum_{i=1}^{k+1} \langle f, f_i\rangle^2} \le \lambda_{k+1}.$$

Now, assume that $Mg_i = \lambda_i g_i$. By the inductive proof of Theorem 7.1.3, there are orthonormal

functions $f_{k+1}, \ldots, f_n$ such that $Mf_i = \lambda_i f_i$ for $k+1 \leq i \leq n$, and such that for all $1 \leq i \leq k$, $k+1 \leq j \leq n$, $\langle f_i, f_j \rangle = 0$. Since $g_1, \ldots, g_k, f_{k+1}, \ldots, f_n$ make a basis for the space of functions in $\ell^2(V)$, for any non-zero function f that is orthogonal to $g_1, \ldots, g_k$, by Corollary 7.1.4, we have

$$\frac{\langle Mf, f \rangle}{\langle f, f \rangle} = \frac{\sum_{i=k+1}^{n} \lambda_i \langle f, f_i \rangle^2}{\sum_{i=k+1}^{n} \langle f, f_i \rangle^2} \geq \lambda_{k+1}$$

The last inequality in the above equation is an equality only if $\langle f, f_i \rangle = 0$ for any $i > k$ where $\lambda_i < \lambda_k$. In other words, when f is in eigen-space of λ_{k+1}. □

The following is a simple corollary of the above theorem.

Corollary 7.1.6. *For any symmetric matrix $M \in \mathbb{R}^{V \times V}$ and $1 \leq k \leq n$, the k-th eigenvalue of M, λ_k, satisfies*

$$\lambda_k = \min_{f_1, \ldots, f_k \in \ell^2(V)} \max_{f \neq \mathbf{0}} \left\{ \frac{\langle Mf, f \rangle}{\langle f, f \rangle}, f \in \operatorname{span}\{f_1, \ldots, f_k\} \right\}.$$

where the minimum is over k non-zero linearly independent functions in $\ell^2(V)$.

Next we describe several important properties of eigenvalues.

Lemma 7.1.7. *For any symmetric matrix $M \in \mathbb{R}^{V \times V}$, with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, we have*

i) For any integer $k \geq 1$, the eigenvalues of M^k are $\lambda_1^k, \ldots, \lambda_n^k$.

ii) Let $\operatorname{trace}(M) := \sum_{v \in V} M(v, v)$. *Then,*

$$\operatorname{trace}(M) = \sum_{i=1}^{n} \lambda_i.$$

Proof. Let $f_1, \ldots, f_n$ be an orthonormal set of eigenfunctions corresponding to $\lambda_1, \ldots, \lambda_n$. (i) is trivial,

$$M^k f_i = \lambda M^{k-1} f_i = \ldots = \lambda_i^k f_i.$$

To prove (ii) we use the spectral theorem. By Corollary 7.1.4,

$$\begin{aligned} \operatorname{trace}(M) = \sum_{v \in V} \langle M\mathbf{1}_v, \mathbf{1}_v \rangle = \sum_{v \in V} \sum_{i=1}^{n} \lambda_i f_i(v)^2 &= \sum_{i=1}^{n} \lambda_i \sum_{v \in V} f_i(v)^2 \\ &= \sum_{i=1}^{n} \lambda_i \left\| f_i \right\|^2 = \sum_{i=1}^{n} \lambda_i. \end{aligned}$$

□

7.2 Laplacian Matrix

The *combinatorial Laplacian* of G is defined by $L = D - A$. This matrix has several important properties. First of all, for any function $f \in \ell^2(V)$,

$$\langle fL, f\rangle = \sum_{(u,v)\in E} w(u,v)\cdot|f(u)-f(v)|^2. \tag{7.2.1}$$

Therefore, by Corollary 7.1.6 all of the eigenvalues of L are non-negative, and L is a positive semi-definite matrix. Furthermore, the first eigenvalue of L is always 0 and the corresponding eigenfunction is any non-zero constant function.

When we are working with non-regular graphs it is straightforward to work with a normalized variant of the Laplacian matrix. The *normalized Laplacian* matrix is given by

$$\mathcal{L} := I - D^{-1/2}AD^{-1/2}.$$

Observe that for an unweighted, d-regular graph, we have $\mathcal{L} = \frac{1}{d}L$, thus, for any eigenfunction f of A with eigenvalue λ, f is also an eigenfunction of $\mathcal{L}$ with corresponding eigenvalue of $1-\lambda/d$.

$$\mathcal{L}f = \frac{1}{d}(D-A)f = If - Af/d = (1-\lambda/d)f$$

As we will show in Section 7.4 $\mathcal{L}$ is closely related to the transition probability matrix of simple random walk on G. Therefore, analyzing the spectrum of $\mathcal{L}$ provides bounds on the spectrum of the random walk matrix. As we will show in Chapter 9 these provide bounds on mixing time/return probabilities of random walks.

Let $g \in \ell^2(V)$ be a non-zero function and let $f = D^{-1/2}g$. Then,

$$\begin{aligned}\frac{\langle g, \mathcal{L}\, g\rangle}{\langle g, g\rangle} = \frac{\langle g, D^{-1/2}LD^{-1/2}g\rangle}{\langle g, g\rangle} &= \frac{\langle f, Lf\rangle}{\langle D^{1/2}f, D^{1/2}f\rangle} \\ &= \frac{\sum_{u\sim v} w(u,v)|f(u)-f(v)|^2}{\sum_{v\in V} w(v)f(v)^2} =: \mathcal{R}(f),\end{aligned} \tag{7.2.2}$$

where we used (7.2.1). The latter value is referred to as the *Rayleigh quotient of f (with respect to G)*.

Unless otherwise specified, we use $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$ to denote the eigenvalues of $\mathcal{L}$. By Corollary 7.1.6, for any $1 \le k \le n$,

$$\lambda_k = \min_{f_1,\ldots,f_k\in\ell^2(V,w)} \max_{f\neq 0} \left\{\mathcal{R}(f) : f \in \operatorname{span}\{f_1,\ldots,f_k\}\right\}, \tag{7.2.3}$$

where the minimum is over sets of k non-zero linearly independent functions in the Hilbert spaces $\ell^2(V, w)$.

Next, we prove several properties of these eigenvalues.

$\lambda_1 = 0$: First, $\mathcal{R}(f) \geq 0$ for any $f \in \ell^2(V, w)$, by (7.2.3), $\lambda_1 \geq 0$. On the other hand, $\mathcal{R}(\mathbf{1}) = 0$. Therefore, $\lambda_1 = 0$, and the first eigenfunction is $g = c \cdot D^{1/2}\mathbf{1}$ for any constant $c \in \mathbb{R}$. Furthermore, $\mathcal{L}$ is a positive semidefinite operator.

$\lambda_k = 0$ **iff** G **has at least** k **connected components:** First observe that if G has k connected components then we can define k linearly independent functions $f_1, \ldots, f_k$ where $\mathcal{R}(f_i) = 0$. We just define each f_i to be the indicator function of one of the connected components of G.

Conversely, suppose $\lambda_k = 0$, and let $f_1, f_2, \ldots, f_k$ be the eigenfunctions corresponding $\lambda_1, \ldots, \lambda_k$. Let $F : V \to \mathbb{R}^k$ where for every $v \in V$, $F(v) := (f_1(v), \ldots, f_k(v))$. This function is known as the spectral embedding of G and we prove several of its properties in Section 8.1. Since $\mathcal{R}(f_1) = \mathcal{R}(f_2) = \ldots = \mathcal{R}(f_k) = 0$, $\mathcal{R}(F) = 0$. Therefore, the vertices in each connected component of G are mapped to same vector in F. On the other hand, since $f_1, \ldots, f_k$ are linearly independent, we have

$$|\{F(v) : v \in V\}| \geq \dim\{F(v) : v \in V\} \geq k.$$

Henceforth, G has at least k connected components.

$\lambda_n \leq 2$ **and** $\lambda_n = 2$ **iff** G **is bipartite:** First, we show $\lambda_n \leq 2$. By (7.2.3),

$$\lambda_n = \max f \in \ell^2(V, w)\mathcal{R}(f) \leq \max_{f \in \ell^2(V,w)} \max_{(u,v) \in E} \frac{|f(u) - f(v)|^2}{f(u)^2 + f(v)^2} \leq 2.$$

The last inequality in above equation is an equality iff for all $(u, v) \in E$, $f(u) = -f(v)$, i.e., G is bipartite. On the other hand, if $G = (X, Y, E)$ is a bipartite graph, we can simply let $f(v) = 1$ for $v \in X$ and $f(v) = -1$ for any $v \in Y$. It follows that $\mathcal{R}(f) = 2$, thus $\lambda_n = 2$.

As we proved above just by knowing the second eigenvalue of $\mathcal{L}$ one can understand whether G is connected, without executing any graph search algorithm. Although this relation does not give any faster algorithm for detecting the connectivity of a graph, it is a simple example of relating an algebraic property of the normalized Laplacian matrix to a combinatorial property of G. Another example is the third property, that is just by knowing the last eigenvalue of $\mathcal{L}$ one can understand whether G is bipartite. We will provide robust versions of several of above arguments together with general versions of these relations later (see e.g. Chapter 10).

Although (7.2.3) provides an explicit characterization of eigenvalues, for many of the cases it is hard to find the exact or approximate value of RHS. The following lemma that we proved in

[KLL+13] provides an easy way to upper bound λ_k. This will be very useful when we analyze the eigenvalues of graphs and relate them to cuts.

Lemma 7.2.1. *For any graph G, $k \geq 1$, and any k disjointly supported functions $f_1, f_2, \ldots, f_k \in \ell^2(V, w)$, the k^{th} smallest eigenvalue of normalized Laplacian matrix $\mathcal{L}$ satisfies,*

$$\lambda_k \leq 2 \max_{1 \leq i \leq k} \mathcal{R}(f_i).$$

Proof. By Corollary 7.1.6, it is sufficient to show that for any function $f \in \text{span}\{f_1, \ldots, f_k\}$, $\mathcal{R}(f) \leq \max_i \mathcal{R}(f_i)$. Note that $\mathcal{R}(f_i) = \mathcal{R}(cf_i)$ for any constant c, so we can assume $f := \sum_{i=1}^k f_i$. Since $f_1, \ldots, f_k$ are disjointly supported, for any $u, v \in V$, we have

$$|f(u) - f(v)|^2 \leq \sum_{i=1}^k 2|f_i(u) - f_i(v)|^2.$$

Therefore,

$$\begin{aligned} \mathcal{R}(f) = \frac{\sum_{u \sim v} w(u,v)|f(u) - f(v)|^2}{\sum_{v \in V} w(v) \cdot f(v)^2} &\leq \frac{2\sum_{u \sim v}\sum_{i=1}^k w(u,v)|f_i(u) - f_i(v)|^2}{\|f\|_w^2} \\ &= \frac{2\sum_{i=1}^k \sum_{u \sim v} w(u,v)|f_i(u) - f_i(v)|^2}{\sum_{i=1}^k \|f_i\|_w^2} \leq 2 \max_{1 \leq i \leq k} \mathcal{R}(f_i). \end{aligned}$$

□

The next lemma shows that for any graph G there is a non-negative function $f \in \ell^2(V, w)$ such that $\mathcal{R}(f) \leq \lambda_2$ and $\text{vol}(\text{supp}(f)) \leq \text{vol}(V)/2$. We will prove generalizations of this lemma in Chapter 10.

Lemma 7.2.2. *For any non-constant function $g \in \ell^2(V)$ there are two non-negative disjointly supported functions $f_+, f_- \in V \to \mathbb{R}$ such that $\mathcal{R}(f_+), \mathcal{R}(f_-) \leq 4\mathcal{R}(D^{-1/2}g)$.*

Proof. W.l.o.g. we assume that $\langle g, \mathbf{1} \rangle = 0$, otherwise we can work with $g - \mathbb{E}_{v \in V}[g(v)]$ instead. Also, we assume $\|g\| = 1$. Let $g_+(v) := \max\{g(v), 0\}$, $g_-(v) := \min\{g(v), 0\}$. First we show that there are two disjointly supported functions h_+ and h_- such that $\|h_+\|^2, \|h_-\|^2 \geq 1/4$. Then, we construct f_+, f_- from h_+, h_-. W.l.o.g. we also assume that $\|g_+\|^2 < 1/4$ (otherwise, if $\|g_+\|^2, \|g_-\|^2 \geq 1/4$ we simply let $f_+ = D^{-1/2}g_+, f_- = -D^{-1/2}g_-$, and it is easy to see f_+, f_- satisfy lemma's conclusion). Now, let $h = g + \frac{1}{2\sqrt{n}}\mathbf{1}$, and similarly let $h_+(v) := \max\{h(v), 0\}$, and $h_-(v) := \min\{h(v), 0\}$. Then,

$$\|h\|^2 = \|g\|^2 + \frac{1}{4n}\|\mathbf{1}\|^2 = 5/4.$$

where we used the fact that $\langle g, \mathbf{1} \rangle = 0$. Therefore,

$$\|h_+\|^2 = 5/4 - \|h_-\|^2 \geq 5/4 - \|g_-\|^2 \geq \frac{1}{4}$$

On the other hand,

$$\|h_+\|^2 \leq \|g_+\|^2 + \frac{1}{4n}\|\mathbf{1}\|^2 + \frac{1}{2\sqrt{n}}\sum_{v\in S} g(v) \leq \frac{1}{2} + \frac{1}{2}\|g_+\| \leq 1.$$

Thus, $\|h-\|^2 \geq 1/4$. Now, since h is just a linear transformation of g, $\mathcal{R}(D^{-1/2}h) = \mathcal{R}(D^{-1/2}g)$. Now, let $f_+ = D^{-1/2}h_+$, $f_- = -D^{-1/2}h_-$. Then,

$$\mathcal{R}(f_+) = \frac{\langle \mathcal{L}h_+, h_+\rangle}{\|h_+\|^2} \leq 4\frac{\langle \mathcal{L}h, h\rangle}{\|h\|^2} = \frac{\langle \mathcal{L}g, g\rangle}{\|g\|^2} = \mathcal{R}(D^{-1/2}g).$$

The same also holds for $\mathcal{R}(f_-)$. □

If in the previous Lemma g is an actual eigenfunction of $\mathcal{L}$ then a very tighter bound can be proved.

Lemma 7.2.3 (Horry, Linial and Widgerson [HLW06])**.** *Given a function $g \in \ell^2(V)$ such that $\mathcal{L}g = \lambda g$ for some $\lambda > 0$, there are two non-negative disjointly supported functions $f_+, f_- \in \ell^2(V, w)$ such that $\mathcal{R}(f_+), \mathcal{R}(f_-) \leq \lambda$.*

Proof. Let $g_+ \in \ell^2(V)$ be the function with $g_+(u) = \max\{g(u), 0\}$ and $g_- \in \ell^2(V)$ be the function with $g_-(u) = \min\{g(u), 0\}$. Then, for any vertex $u \in \mathrm{supp}(g_+)$,

$$(Łg_+)(u) = g_+(u) - \sum_{v:v\sim u} \frac{w(u,v)g_+(v)}{\sqrt{w(u)w(v)}} \leq g(u) - \sum_{v:v\sim u} \frac{w(u,v)g(v)}{\sqrt{w(u)w(v)}} = (Łg)(u) = \lambda \cdot g(u).$$

Therefore,

$$\langle g_+, Łg_+\rangle = \sum_{u\in\mathrm{supp}(g_+)} g_+(u)\cdot(Łg_+)(u) \leq \sum_{u\in\mathrm{supp}(g_+)} \lambda \cdot g_+(u)^2 = \lambda \cdot \|g_+\|^2.$$

Letting $f_+ = D^{-1/2}g_+$, we get

$$\lambda \geq \frac{\langle g_+, Łg_+\rangle}{\|g_+\|^2} = \frac{\langle f_+, Lf_+\rangle}{\|f_+\|_w^2} = \mathcal{R}(f_+).$$

Similarly, we can define $f_- = -D^{-1/2}g_-$, and show that $\mathcal{R}(f_-) \leq \lambda$. □

7.3 Cayley Graphs and Their Spectrum

In this section we compute eigenvalues/eigenfunctions of several useful graphs including Cycle, Hypercube, Grid, etc. These examples will be useful when we show tightness of our theorems/analysis. The material of this section is mainly based on Luca Trevisan's lecture notes.

Let Γ be a finite group. We will use additive notation, although the following definition applies to non-commutative groups as well. A subset $S \subseteq \Gamma$ is *symmetric* if $a \in S \iff -a \in S$.

Definition 7.3.1. *For a group Γ and a symmetric subset $S \subseteq \Gamma$, the Cayley graph $\mathrm{Cay}(\Gamma, S)$ is the graph whose vertex set is Γ, and such that (a, b) is an edge if and only if $b - a \in S$. Note that the graph is undirected and $|S|$-regular.*

We can also define Cayley weighted graphs: if $w : \Gamma \to \mathbb{R}$ is a function such that $w(a) = w(-a)$ for every $a \in \Gamma$, then we can define the weighted graph $\mathrm{Cay}(G, w)$ in which the edge (a, b) has weight $w(b - a)$. We will usually work with unweighted graphs.

For example, the n-vertex cycle can be realized as the Cayley graph $\mathrm{Cay}(\mathbb{Z}/n\mathbb{Z}, \{-1, 1\})$. The $\sqrt{n} \times \sqrt{n}$ torus can be realized as the Cayley graph

$$\mathrm{Cay}((\mathbb{Z}/\sqrt{n}\mathbb{Z}) \times (\mathbb{Z}/\sqrt{n}\mathbb{Z}), \{(-1, 0), (1, 0), (0, -1), (0, 1)\}).$$

Also, the h-dimensional hypercube can be realized as the Cayley graph

$$\mathrm{Cay}((\mathbb{Z}/2\mathbb{Z})^h, \{(1, 0, ..., 0), (0, 1, ..., 0), ..., (0, 0, ..., 1)\})$$

where the group is the set $\{0, 1\}^h$ with the operation of bit-wise xor, and the set S is the set of bit-vectors with exactly one 1.

Definition 7.3.2 (Character)**.** *For a group Γ, a function $f : \Gamma \to \mathbb{C}$ is a character of Γ if*

- *f is a group homomorphism of Γ into the multiplicative group $\mathbb{C} - \{0\}$.*
- *for every $a \in \Gamma$, $|f(a)| = 1$.*

It turns out that any finite abelian group, Γ, has exactly $|\Gamma|$ characters, and these characters are exactly the eigenfunctions of the graph $\mathrm{Cay}(\Gamma, S)$ for any symmetric $S \subseteq \Gamma$. Next we describe the characters of an abelian group Γ and the corresponding eigenvalues.

Recall that every finite abelian group is isomorphic to a product of cyclic groups

$$(\mathbb{Z}/n_1\mathbb{Z}) \times (\mathbb{Z}/n_2\mathbb{Z}) \times \ldots \times (\mathbb{Z}/n_k\mathbb{Z}).$$

First of all, for all $0 \leq r \leq n - 1$, the cyclic group $(\mathbb{Z}, n\mathbb{Z})$ has a character,

$$\chi_r(a) = e^{2\pi i a r/n}.$$

Now, if χ_1 is a character of Γ_1 and χ_2 is a character of Γ_2, then $\chi(a,b) := \chi_1(a) \cdot \chi_2(b)$ is a character of $\Gamma_1 \times \Gamma_2$.

Theorem 7.3.3. *For any finite abelian group, Γ, a symmetric set $S \subseteq \Gamma$, and a character $\chi : V \to \mathbb{C}$, χ is a eigenfunction of the adjacency operator of the graph $G = \mathrm{Cay}(\Gamma, S)$. with the corresponding eigenvalue of*

$$\lambda = \sum_{a \in S} \chi(a).$$

Now we are ready to study the spectrum of Cycle,Grid, and the hypercube.

7.3.1 The Cycle

A cycle with n vertices is the Cayley graph $\mathrm{Cay}(\mathbb{Z}/n\mathbb{Z}, \{1, -1\})$ with characters $\chi_r(a) = e^{2\pi i r a/n}$ for all $0 \le r \le n-1$. By Theorem 7.3.3 these are the eigenfunctions of the normalized Laplacian matrix of a cycle with corresponding eigenvalues,

$$\lambda_r = 1 - \frac{1}{2}(e^{2\pi i r/n} + e^{-2\pi i r/n}) = 1 - \cos(2\pi r/n)$$

where we used the fact that cycle is a 2-regular graph, and that $e^{ix} = \cos(x) + i\sin(x)$. Consequently, for $k \le n/2\pi$,

$$\lambda_k = \frac{1}{2!}\Big(\frac{2\pi k}{n}\Big)^2 - \frac{1}{4!}\Big(\frac{2\pi k}{n}\Big)^4 + \ldots = \Theta(k^2/n^2).$$

7.3.2 The Grid

Let G be a $n = l \times l$ torus with characters

$$\chi_{r_1,r_2}(a,b) = e^{2\pi i r_1 a/l} \cdot e^{2\pi i r_2 a/l}$$

for all $0 \le r_1, r_2 \le l-1$. By Theorem 7.3.3, these are the eigenfunctions of the normalized Laplacian matrix of G with corresponding eigenvalues,

$$\lambda_{r_1,r_2} = 1 - \frac{1}{4}(e^{2\pi i r_1/l} + e^{-2\pi i r_1/l} + e^{2\pi i r_2/l} + e^{-2\pi i r_2/l}) = 1 - \cos(2\pi r_1/l)/2 - \cos(2\pi r_2/l)/2.$$

This implies that for any $k \le n/4\pi^2$,

$$\lambda_k \le \max_{0 \le r_1, r_2 \le \sqrt{k}} \lambda_{r_1,r_2} \le 1 - \cos(2\pi k/l) = \Theta(k^2/l^2) = \Theta(k/n).$$

On the other hand,

$$\lambda_k \ge \min_{r_1 \sqrt{k} \text{ or } r_2 \ge \sqrt{k}} \lambda_{r_1,r_2} \ge 1 - \cos(2\pi k/l)/2 = \Theta(k/n).$$

Therefore, the k^{th} smallest eigenvalue of the normalized Laplacian of G is $\lambda_k = \Theta(k/n)$.

7.3.3 The Ladder

Let $\Gamma = (\mathbb{Z}/n\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$. For any $0 \leq r \leq n-1$ and $s \in \{0,1\}$, this group has the characters

$$\chi_{r,s}(a,b) = e^{2\pi i r a/l}(-1)^b,$$

Let $S = \{(1,0),(-1,0),(0,1)\}$ with the following weights, $w((1,0)) = w((-1,0)) = 1$ and $w(0,1) = 100/l^2$. The ladder is the Cayley graph $G = \mathrm{Cay}(\Gamma, S)$ (see Figure 7.8.1 for an illustration). The eigenfunctions of the normalized Laplacian matrix are $\chi_{r,s}$ with corresponding eigenvalues,

$$\lambda_{r,s} = 1 - \frac{1}{2+100/l^2}(e^{2\pi i r/l} + e^{-2\pi i r/l} + (-1)^s \cdot 100/l^2) = 1 - \frac{2\cos(2\pi r/l) + (-1)^s \cdot 100/l^2}{2+100/l^2}$$

Therefore, for $r > 0$, we get

$$\lambda_{r,0} \leq \frac{2\pi^2 r^2}{l^2+50}, \quad \text{and} \quad \lambda_{0,1} = \frac{100}{l^2+50}$$

Consequently, $\lambda_2 = \lambda_{1,0}$ and $\lambda_3 = \lambda 2, 0$, and the eigenfunction corresponding to λ_2 is $\chi_{1,0}$.

7.3.4 The Hypercube

The group $(\mathbb{Z}/2\mathbb{Z})^h = \{0,1\}^h$ with bitwise xor has 2^h characters; for every $\mathbf{r} \in \{0,1\}^h$ there is a character $\chi_{\mathbf{r}} : \{0,1\}^d \to \{-1,1\}$ defined as

$$\chi_{\mathbf{r}}(\mathbf{a}) = (-1)^{\sum_{i=1}^h \mathbf{r}(i)\mathbf{a}(i)}$$

Let us denote the set S by $\{\mathbf{1}_1, \ldots, \mathbf{1}_h\}$, where we let $\mathbf{1}_j \in \{0,1\}^h$ denote the bit-vector that has a 1 in the j^{th} position, and zeroes everywhere else. Recall that a hypercube is the Cayley graph $G = \mathrm{Cay}(\{0,1\}^h, S)$. This means that, for every $\mathbf{r} \in \{0,1\}^h$, the corresponding eigenvalue of the normalized Laplacian matrix is

$$\lambda_{\mathbf{r}} = 1 - \frac{1}{h}\sum_{i=1}^h \chi_{\mathbf{r}}(\mathbf{1}_i) = 1 - \frac{1}{h}\sum_{i=1}^h (-1)^{\mathbf{r}(i)} = 1 - \frac{1}{h}(h - 2\left\|\mathbf{r}\right\|_1) = 2\left\|\mathbf{r}\right\|_1/h$$

where we used the fact that G is h-regular. For example, $\lambda_2 = \lambda_3 = \ldots = \lambda_{h+1} = 2/h$. In general, for any $0 \leq j \leq h$, and $k = \binom{h}{0} + \binom{h}{1} + \ldots + \binom{h}{j}$ we have

$$\lambda_{k+1} = \lambda_{k+2} = \ldots = \lambda_{k+\binom{h}{j+1}} = \frac{2(j+1)}{h}.$$

7.4 Random Walks

We shall consider *lazy random walks* on G where from each vertex u, with probability $1/2$ we stay at u, and with probability $w(u,v)/(2w(u))$ we jump to the neighbor v of u. We write $P := 1/2(I + D^{-1}A)$ for the transition probability operator of the lazy random walk on G, where I is the identity operator. Observe that in the matrix representation, $P(u,v) = w(u,v)/(2w(u))$, and $P(v,v) = 1/2$. We may also work with adjoint of P, $P^T = 1/2(I + AD^{-1})$. Observe that for a function $f \in \mathbb{R}^V$,

$$P^T f(u) = \sum_{v \in V} w(v,u) f(v)/2w(v).$$

Although P is not a symmetric matrix (or a self-adjoint operator), it features many properties of the symmetric matrices. First of all, P can be transformed to a symmetric matrix simply by considering $D^{1/2}PD^{-1/2}$. Furthermore, for any eigenfunction f of $D^{1/2}PD^{-1/2}$ with eigenvalue λ, $D^{-1/2}f$ is an eigenfunction of P and $D^{1/2}f$ is an eigenfunction of P^T both with eigenvalue λ. This is because,

$$D^{1/2}PD^{-1/2}f = \lambda f \Rightarrow \begin{cases} PD^{-1/2}f = \lambda D^{-1/2}f. \\ DPD^{-1}D^{1/2}f = P^T D^{1/2}f = \lambda D^{1/2}f. \end{cases}$$

Note that for any eigenfunction f of P, Df is an eigenfunction of P^T. Therefore, eigenvalues of P and P^T are exactly the same and they are all real. On the other hand, since

$$D^{1/2}PD^{-1/2} = 1/2(I + D^{-1/2}AD^{-1/2}) = I - 1/2(I - D^{-1/2}AD^{-1/2} = I - \mathcal{L}/2,$$

the eigenfunctions of $D^{1/2}PD^{-1/2}$ and $\mathcal{L}$ are the same, and the k^{th} largest eigenvalue of P is equal to 1 minus half of the k^{th} smallest eigenvalue of $\mathcal{L}$. That is, the eigenvalues of P are

$$1 = 1 - \lambda_1/2 \geq 1 - \lambda_2/2 \geq \ldots \geq 1 - \lambda_n/2 \geq 0.$$

Thus, P is also a positive semi-definite operator. The following lemma summarizes the above discussion

Lemma 7.4.1. *Matrix P has eigenvalues $1 - \lambda_1/2, \ldots, 1 - \lambda_n/2$ with corresponding orthonormal eigenfunctions $f_1, \ldots, f_n \in \ell^2(V,w)$. That is $\langle f_i, f_j \rangle_w = 0$ for any $i \neq j$ and $\|f_i\|_w = 1$ for all i. Here $\lambda_1, \ldots, \lambda_n$ are eigenvalues of $\mathcal{L}$.*

Recall that the eigenfunction of $\mathcal{L}$ corresponding to $\lambda_1 = 0$ is $D^{1/2}\mathbf{1}$. Therefore, the eigenfunction corresponding to eigenvalue of 1 of P is the constant function, and the corresponding eigenfunction of P^T is $c \cdot D\mathbf{1}$ for any $c \in \mathbb{R}$. Let $\pi := D\mathbf{1}/\mathrm{vol}(V)$, i.e., $\pi(v) := w(v)/\mathrm{vol}(V)$, for all $v \in V$. We call π the *stationary distribution* of the random walk.

Unless otherwise specified, we use X_t as a random variable to denote the t^{th} step of a lazy

random walk. For example, X_0 shows the starting vertex. We say a non-negative function $q \in \mathbb{R}^V$ is a *probability distribution* on G, if $\sum_{v \in V} q(v) = 1$. Observe that if q is a probability distribution on G, then so is $P^T q$. Indeed, $P^T q$ is the probability distribution of the random walk after one step,

$$\mathbb{P}\left[X_1 = v \mid X_0 \sim q\right] = P^T q(v),$$

where $X_0 \sim q$ means that X_0 is distributed with q. Similarly $(P^T)^t q$ is the distribution of the walk after t steps. The fundamental theorem of Markov chains implies that if G is connected, then $\lim_{t\to\infty} (P^T)^t q = \pi$. This is the reason that π is called the stationary distribution of the walk. The *mixing time* of a random walk is the time it takes to get to the stationary distribution. More formally, the ϵ-mixing time in norm p is defined as follows,

$$\tau_p(\epsilon) := \min\Big\{t : \forall u \in V, \Big(\sum_{v\in V}\Big|\frac{P^t(u,v)}{\pi(v)} - 1\Big|^p \pi(v)\Big)^{1/p} \le \epsilon\Big\}.$$

In particular, $\tau_1(\epsilon)$ is called, mixing time in *total variation distance*, and $\tau_\infty(\epsilon)$ is called the *uniform* mixing time. Note that for $\epsilon \approx 0$, the above equation implies the for all $u \in V$, the u^{th} row of P^t is approximately equal to $\pi(.)$. This is the same as $(P^T)^t q$ is approximately equal to $\pi(.)$ for any distribution q on G.

In the next lemma we upper bound the mixing time of the lazy random walk in terms of λ_2.

Lemma 7.4.2. *Let $\pi_{\min} := \min_{v\in V} \pi(v)$. For any connected graph G, and $p \ge 1$ and $\epsilon < 1$,*

$$\tau_p(\epsilon) \le \frac{-2\log(\pi_{\min})\log(1/\epsilon)}{\lambda_2}.$$

The following proof is based on [LPW06, Thm 12.3].

Proof. Let $f_1, f_2, \ldots, f_n \in \ell^2(V, w)$ be orthonormal eigenfunctions of P corresponding to eigenvalues $1 - \lambda_1/2, 1 - \lambda_2/2, \ldots, 1 - \lambda_n/2$. By the above argument f_1 is a constant function, since $\|f_1\|_w = 1$, we must have

$$f_1 = \mathbf{1}/\sqrt{\mathrm{vol}(V)}. \tag{7.4.1}$$

Since $f_1, \ldots, f_n$ form a basis in $\ell^2(V, w)$, for any $v \in V$ we have,

$$\mathbf{1}_v = \sum_{i=1}^n \langle \mathbf{1}_v, f_i\rangle_w \cdot f_i = \sum_{i=1}^n f_i(v)w(v) \cdot f_i.$$

Therefore, for any $v \in V$,

$$P^t \mathbf{1}_v = P^t \sum_{i=1}^n f_i(v)w(v)f_i = \sum_{i=1}^n f_i(v)w(v)\lambda_i^t f_i.$$

Hence, for any pair of vertices $u, v \in V$,

$$\left|\frac{P^t(u,v)}{\pi(v)} - 1\right| = \left|\mathrm{vol}(V) \cdot \sum_{i=1}^{n} f_i(v)(1-\lambda_i/2)^t f_i(u) - 1\right| = \mathrm{vol}(V) \sum_{i=2}^{n} |f_i(v)(1-\lambda_i/2)^t f_i(u)|$$
$$\leq \mathrm{vol}(V)(1-\lambda_2/2)^t \sum_{i=2}^{n} |f_i(v) f_i(u)|$$
$$\leq \mathrm{vol}(V) e^{-t\lambda_2/2} \sqrt{\sum_{i=1}^{n} f_i(v)^2 \sum_{i=1}^{n} f_i(u)^2} \tag{7.4.2}$$

where the second equality follows by equation (7.4.1), and that $\lambda_1 = 0$, and the last inequality follows by the Cauchy-Schwarz inequality. On the other hand, for any $v \in V$,

$$\sum_{i=1}^{n} f_i(v)^2 = \sum_{i=1}^{n} \langle f_i, \mathbf{1}_v \rangle^2 = \sum_{i=1}^{n} \langle f_i, D^{-1}\mathbf{1}_v \rangle_w^2 = \frac{1}{w(v)^2} \|\mathbf{1}_v\|_w^2 = \frac{1}{w(v)} \tag{7.4.3}$$

Putting equations (7.4.2) and (7.4.3) together we get,

$$\left|\frac{P^t(u,v)}{\pi(v)} - 1\right| = \frac{e^{-t\lambda_2/2}}{\sqrt{\pi(v)\pi(u)}} \leq \frac{e^{-t\lambda_2/2}}{\pi_{\min}}$$

Therefore, $\tau_p(\epsilon) \leq -2\log(\pi_{\min})\log(1/\epsilon)/\lambda_2$ for any $p \geq 1$. □

Suppose we run a random walk from a vertex $u \in V$, and let q be the probability d In the next lemma we show that if G is a connected graph

7.5 Gaussian Distribution

The *normal distribution* with mean μ and variance σ^2, $\mathcal{N}(\mu, \sigma^2)$ has probability density function

$$\frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

A random l dimensional Gaussian vector is a vector $\boldsymbol{\zeta} = (\zeta_1, \zeta_2, \ldots, \zeta_l)$ where each $\zeta_1, \ldots \zeta_l$ is chosen independently from $\mathcal{N}(0,1)$.

Next, we recall several important properties of the normal distribution.

P1) If $\zeta_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $\zeta_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ are chosen independently then $\zeta_1 + \zeta_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. Consequently, Gaussian vectors are spherically symmetric in the sense that for any vector $\mathbf{x} \in \mathbb{R}^h$, the distribution of $\langle \mathbf{x}, \boldsymbol{\zeta} \rangle$ is the same as a $\mathcal{N}(0, \|\mathbf{x}\|^2)$. Therefore, for $l \in \mathbb{N}$, the expected

norm of the vector $\boldsymbol{\zeta}$ is

$$\mathbb{E}\left[\|\boldsymbol{\zeta}\|^2\right] = \sum_{i=1}^{l} \mathbb{E}\left[\zeta_i^2\right] = l.$$

P2) For any $s \in \mathbb{R}$,

$$\frac{1}{\sqrt{2\pi}}(\frac{1}{s} - \frac{1}{s^3})e^{-s^2/2} \leq \mathbb{P}_{\zeta\sim\mathcal{N}(0,1)}\left[\zeta \geq s\right] \leq \frac{1}{s\cdot\sqrt{2\pi}}e^{-s^2/2}.$$

This property is proved in [Fel68, Ch. VII].

In the next lemma we prove a Chernoff bound on sum of squares of independent Gaussian random variables. Similar bounds can be found in [Mat02, Ch. 15] or [LT11, Ch. 1]).

Lemma 7.5.1. *Let $\zeta_1, \ldots, \zeta_l \sim \mathcal{N}(0,1)$ be l independent Gaussian random variables. For any $r \geq 2$,*

$$\mathbb{P}\left[\sum_{i=1}^{l} \zeta_i^2 > r\cdot l\right] \leq e^{-rl/14}.$$

Also, for any $0 < \delta < 1$,

$$\mathbb{P}\left[\sum_{i=1}^{l} \zeta_i^2 > (1+\delta)\cdot l \text{ or } \sum_{i=1}^{l} \zeta_i^2 < (1+\delta)\cdot l\right] \leq 2e^{-\delta^2 l/8}.$$

Proof. First, observe that, for $\zeta \sim \mathcal{N}(0,1)$, and $\alpha < 1/2$, and $\sigma^2 = 1/(1-2\alpha)$,

$$\mathbb{E}\left[\exp(\alpha\cdot\zeta)\right] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2(1/2-\alpha)} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} dx = (1-2\alpha)^{-1/2}$$

Let $X = \zeta_1^2 + \ldots + \zeta_l^2$. For $\alpha < 1/2$, and $s > 0$ we get

$$\mathbb{P}\left[X > s\right] = \mathbb{P}\left[\exp(\alpha\cdot X) > \exp(\alpha\cdot s)\right] \leq \frac{\mathbb{E}\left[\exp(\alpha\cdot X)\right]}{\exp(\alpha\cdot s)} = \frac{\prod_{i=1}^{l}\mathbb{E}\left[\exp(\alpha\cdot\zeta_i^2)\right]}{\exp(\alpha\cdot s)} \leq \frac{(1-2\alpha)^{-l/2}}{\exp(\alpha\cdot s)}$$

where the first inequality follows by the Markov inequality and the second inequality follows by the fact that $\zeta_1, \ldots, \zeta_l$ are independent random variables. Now, if $s \geq 2l$, for $\alpha = 1/4$ we get

$$\mathbb{P}\left[X > s\right] \leq e^{-\alpha s}(1-2\alpha)^{-l/2} \leq e^{-s/4+0.35l} \leq e^{-s/14}.$$

Otherwise, let $\alpha = (s-l)/(4l)$. Since $2\alpha \leq 1/2$,

$$\mathbb{P}\left[X > s\right] \leq e^{-\alpha\cdot s+\alpha\cdot l+2l\alpha^2} = e^{-\frac{(s-l)^2}{4l}+\frac{(s-l)^2}{8l}} = e^{-\frac{(s-l)^2}{8l}}. \tag{7.5.1}$$

where we used the taylor series of $(1-2\alpha)^{-1}$.

Similarly, for $0 < s < l$,

$$\mathbb{P}[X < s] = \mathbb{P}[\exp(-\alpha \cdot X) > \exp(-\alpha \cdot s)] \leq \frac{\mathbb{E}[\exp(-\alpha \cdot X)]}{\exp(-\alpha \cdot s)} \leq \frac{\prod_{i=1}^{l} \mathbb{E}\left[\exp(-\alpha \zeta_i^2)\right]}{\exp(-\alpha \cdot s)} \leq \frac{(1+2\alpha)^{-l/2}}{\exp(-\alpha \cdot s)}$$

For $\alpha = (s-l)/(4l)$ we can argue similar to equation (7.5.1), and we get $\mathbb{P}[X < s] \leq \exp(-(s-l)^2/(8l))$. □

Let $\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \ldots, \boldsymbol{\zeta}_l$ be i.i.d. k-dimensional Gaussians, and consider the random mapping $\Gamma_{k,l} : \mathbb{R}^k \to \mathbb{R}^l$ defined by

$$\Gamma_{k,l}(\mathbf{x}) = l^{-1/2}(\langle \boldsymbol{\zeta}_1, \mathbf{x} \rangle, \langle \boldsymbol{\zeta}_2, \mathbf{x} \rangle, \ldots, \langle \boldsymbol{\zeta}_l, \mathbf{x} \rangle). \tag{7.5.2}$$

From P1, for every $\mathbf{x} \in \mathbb{R}^k$,

$$\mathbb{E}\left[\|\Gamma_{k,l}(\mathbf{x})\|^2\right] = \|\mathbf{x}\|^2. \tag{7.5.3}$$

Since Gaussians are spherically symmetric, up to a change of basis, we can assume $\mathbf{x} = (1, 0, 0, \ldots, 0)$. Therefore, the following basic estimates follow from Lemma 7.5.1. For every $0 < \delta < 1$,

$$\mathbb{P}\left[\|\Gamma_{k,l}(\mathbf{x})\|^2 \notin [(1-\delta)\|\mathbf{x}\|^2, (1+\delta)\|\mathbf{x}\|^2]\right] \leq 2e^{-\delta^2 l/8}, \tag{7.5.4}$$

and for every $\lambda \geq 2$,

$$\mathbb{P}\left[\|\Gamma_{k,l}(x)\|^2 \geq r\|x\|^2\right] \leq e^{-rl/14}. \tag{7.5.5}$$

Corollary 7.5.2. *Let $\boldsymbol{\zeta} \in \mathbb{R}^l$ be a random l-dimensional Gaussian vector. For a sufficiently large l, and any unit vector $\mathbf{x} \in \mathbb{R}^l$,*

$$\mathbb{P}\left[\frac{|\langle \boldsymbol{\zeta}, \mathbf{x} \rangle|}{\|\boldsymbol{\zeta}\|} \geq \frac{2}{3\sqrt{l}}\right] \geq 1/2.$$

Proof. By P1 we can assume that $\mathbf{x} = (1, 0, 0, \ldots, 0)$. Also let $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_l)$. Therefore,

$$\mathbb{P}[|\langle \boldsymbol{\zeta}, \mathbf{x} \rangle| > 0.67] = \mathbb{P}[|\boldsymbol{\zeta}_1| > 0.67] > 1/2.$$

where the last equation follows by computing the cumulative distribution of the standard normal distribution. On the other hand, by Lemma 7.5.1, for a sufficiently large l, $\|\boldsymbol{\zeta}\|^2 < l(1+0.01)$ with high probability. Putting these together the claim follows by the union bound. □

7.6 Eigenfunction Computation

In general, it is not possible to exactly compute eigenfunctions of a matrix. This is because the values assigned to vertices in an eigenfunction are not necessarily rational (e.g. the cycle graph). Therefore, spectral algorithms cannot leverage the exact value of eigenfunctions, they should rather be stable and work with approximate eigenfunctions. All of spectral algorithms designed in this thesis work with any arbitrary functions as long as they have sufficiently small Rayleigh quotient.

In this section we provide fast algorithms for approximating eigenfunctions of the normalized Laplacian matrix. The materials of this section is mainly based on [Vis13, Ch 8].

Proposition 7.6.1. *For any symmetric positive semidefinite matrix $M \in \mathbb{R}^{V \times V}$, an error parameter $\epsilon > 0$, sufficiently large n and a positive integer $k > \frac{1}{\epsilon} \cdot \log \frac{18n}{4\epsilon}$, a random Gaussian function $g \in \mathcal{N}(0,1)^V$ (i.e., $g(v) \sim \mathcal{N}(0,1)$ for all $v \in V$), and $f = M^k g / \left\| M^k g \right\|$ we have*

$$\mathbb{P}_{g \sim \mathcal{N}(0,1)^V} \left[\frac{\langle Mf, f \rangle}{\langle f, f \rangle} \geq (1-\epsilon)\lambda_n \right] \geq 1/2,$$

where λ_n is the largest eigenvalue of M in absolute value.

Proof. Let $f_1, \ldots, f_n$ be the eigenfunctions of M with corresponding eigenvalues $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$. Let $\bar{g} = g / \|g\|$, and let $\alpha_i := \langle \bar{g}, f_i \rangle$. By Corollary 7.5.2, with probability $1/2$, $\alpha_n = |\langle \bar{g}, f_n \rangle| \geq 2/3\sqrt{n}$. Let $f = M^k \bar{g}$ be such a function. Let l be the largest index such that $\lambda_l \geq (1-\epsilon/2)\lambda_n$. By Corollary 7.1.4

$$\langle f, f \rangle = \sum_{i=1}^{n} \alpha_i^2 \lambda_i^{2k} \leq (1-\epsilon/2)^{2k} \lambda_n^{2k} + \sum_{i=l}^{n} \alpha_i^2 \lambda_i^{2k} \leq \exp(-\epsilon k/2) + \sum_{i=l}^{n} \alpha_i^2 \lambda_i^{2k} \leq \sum_{i=l}^{n} (1+\epsilon/2) \alpha_i^2 \lambda_i^2 k.$$

where the second inequality follows by the fact that $(1-\epsilon/2) \leq \exp(-\epsilon/2)$. Therefore,

$$\frac{\langle Mf, f \rangle}{\langle f, f \rangle} = \frac{\sum_{i=1}^{n} \alpha_i^2 \lambda_i^{2k+1}}{\sum_{i=1}^{n} \alpha_i^2 \lambda_i^{2k}} \geq \frac{\sum_{i=l}^{n} \alpha_i^2 \lambda_i^{2k+1}}{\sum_{i=l}^{n} (1+\epsilon/2) \alpha_i^2 \lambda_i^{2k}} \geq \frac{\lambda_l}{1+\epsilon/2} \geq \frac{(1-\epsilon/2)\lambda_n}{1+\epsilon/2} \geq (1-\epsilon)\lambda_n.$$

□

Let us describe an application of above proposition in approximating the second eigenfunctions of $\mathcal{L}$. Equivalently, we can find the eigenfunction corresponding to the second largest eigenvalue of $I - \mathcal{L}/2$. Since the largest eigenfunction of $I - \mathcal{L}/2$ is $g_1 = D^{1/2}\mathbf{1}$ we can run the above power method in the space of functions orthogonal to $g_1 = D^{1/2}\mathbf{1}$ by removing the component along g_1, i.e., letting $f = f - \langle f, g_1 \rangle g_1$. Now, if we run the power method, f converges to the second eigenfunction of $\mathcal{L}$, g_2. So, by the above proposition, we need $\epsilon = O(\lambda_2)$ to get a constant factor approximation of λ_2. But this means that the running the of the power method is $\tilde{O}((m+n)/\lambda_2)$ which can become worst than quadratic even on a cycle.

So, in the rest of this section we describe an alternative approach that is based on recent developments on fast Laplacian solvers [ST04, KMP10, KMP11, KOSZ13].

The following is the main theorem of this section [ST04].

Theorem 7.6.2. *For any undirected graph G on n vertices and m edges and $\epsilon < 1/2$, there is an algorithm that outputs a function a function $g_2 : V \to \mathbb{R}$ such that $\langle g_2, \mathbf{1} \rangle = 0$ and $\mathcal{R}(D^{-1/2} g_2) \leq (1+5\epsilon)\lambda_2$, where λ_2 is second smallest eigenvalue of $\mathcal{L}$ and the algorithm runs in $\tilde{O}(m+n)$ time where we avoid the* polylog(n) *factors.*

Instead of applying the above power method to $I - 2\mathcal{L}$ we can apply it to the psuedo-inverse of $\mathcal{L}$, $\mathcal{L}^\dagger$. Since the largest eigenvalue of $\mathcal{L}^\dagger$ is $1/\lambda_2$ we can get a constant factor approximation by running the power method only for $\Theta(\log n)$ iterations. But the problem is that we are not aware of any near linear time algorithm that computes $\mathcal{L}^\dagger$.

The idea is that we can compute $\mathcal{L}^\dagger f$ for a given function f by solving the linear system of equations $\mathcal{L}g = f$. We can use the following theorem that is based on recent developments on fast Laplacian solvers [ST04, KMP10, KMP11, KOSZ13] to approximately solve this linear system.

Theorem 7.6.3 ([Vis13, Section 3])**.** *There is an algorithm which takes as input L, the Laplacian of a graph G, a function $g \in \ell^2(V)$, and an error parameter $\epsilon > 0$, and returns the function Zg where Z is a symmetric linear operator such that*

$$(1-\epsilon)Z^\dagger \preceq L \preceq (1+\epsilon)Z^\dagger.$$

The algorithm runs in $O(m \log n \log 1/\epsilon)$ time where m is the number of edges of G.

Here, we do not prove but we refer the interested reader to [Vis13].

Note that the above equation means that the matrices $Z^\dagger$ and L are co-spectral, mathematically, for any function $f : V \to \mathbb{R}$,

$$(1-\epsilon)\langle f, Z^\dagger f\rangle \leq \langle f, Lf\rangle \leq (1+\epsilon)\langle f, Z^\dagger f\rangle.$$

Let $\mathcal{Z} = D^{-1/2} Z D^{-1/2}$. It follows from the above equation that so for any $g : V \to R$ and $f = D^{-1/2} g$,

$$(1-\epsilon)\langle g, \mathcal{Z}^\dagger g\rangle = (1-\epsilon)\langle f, Z^\dagger f\rangle \leq \langle f, Lf\rangle = \langle g, \mathcal{L}g\rangle \leq (1+\epsilon)\langle f, Z^\dagger f\rangle \leq (1+\epsilon)\langle g, \mathcal{Z}^\dagger g\rangle.$$

Therefore,

$$(1-\epsilon)\mathcal{Z}^\dagger \preceq \mathcal{L} \preceq (1+\epsilon)\mathcal{Z}^\dagger. \tag{7.6.1}$$

Now, we are ready to prove Theorem 7.6.2.

Proof of Theorem 7.6.2. Let $f_1, f_2, \ldots, f_n$ be the eigenfunctions of $\mathcal{Z}^\dagger$ corresponding to the eigenvalues $\theta_1 \geq \theta_2 \geq \ldots \theta_n = 0$ (note that for this proof we label the eigenvalues in decreasing order). Let $g \sim \mathcal{N}(0,1)^V$, $\bar{g} = g/\|g\|$, and $\alpha_i := \langle \bar{g}, f_i\rangle$. Again by Corollary 7.5.2 with probability at least $1/2$, $|\alpha_1| \geq 2/3\sqrt{n}$; assume that $\bar{g}$ satisfies this.

Let $f = \mathcal{Z}^k \bar{g}$ for $k = O(\frac{1}{\epsilon}\log(n/\epsilon))$, note that k does not have any dependency to the eigenvalues

of $\mathcal{Z}$. Let j be the smallest such that $\theta_j \leq (1+\epsilon)\theta_{n-1}$. Then,

$$\begin{aligned}
\frac{\langle f, \mathcal{Z}^\dagger f\rangle}{\langle f, f\rangle} = \frac{\sum_{i=1}^{n-1} \alpha_i^2 \theta_i^{(2k-1)}}{\sum_{i=1}^{n-1} \alpha_i \theta_i^{-2k}} &\leq \frac{\sum_{i=1}^{j} \alpha_i^2 \theta_i^{-(2k-1)}}{\sum_{i=1}^{n-1} \alpha_i^2 \theta_i^{-2k}} + \frac{\sum_{i>j}^{n-1} \alpha_i^2 \theta_i^{-(2k-1)}}{\sum_{i>j}^{n-1} \alpha_i^2 \theta_i^{-2k}} \\
&\leq \frac{\sum_{i=1}^{j} \alpha_i^2 \theta_i^{-(2k-1)}}{\alpha_{n-1}^2 \theta_{n-1}^{-2k}} + \theta_j \\
&\leq \theta_{n-1} \frac{\sum_{i=1}^{j} \alpha_i^2 (1+\epsilon)^{-(2k-1)}}{\alpha_{n-1}^2} + (1+\epsilon)\theta_{n-1} \\
&\leq \theta_{n-1} \cdot \frac{9n}{4} \sum_{i=1}^{j} \alpha_i^2 e^{-\frac{\epsilon(2k-1)}{2}} (1+\epsilon)\theta_{n-1} \\
&\leq \theta_{n-1} \cdot \epsilon + (1+\epsilon)\theta_{n-1} \leq (1+2\epsilon)\theta_{n-1}.
\end{aligned}$$

where the third last inequality follows by $1/(1+x) \leq e^{-x/2}$ for $x \leq 1$, and the second to last inequality follows by choosing $k \geq (\log(9n/4\epsilon) + 1/2)/\epsilon$. Therefore, by (7.6.1)

$$\frac{\langle f, \mathcal{L} f\rangle}{\langle f, f\rangle} \leq (1+\epsilon)\frac{\langle f, \mathcal{Z}^\dagger f\rangle}{\langle f, f\rangle} \leq (1+\epsilon)(1+2\epsilon)\theta_{n-1} \leq \frac{(1+\epsilon)(1+2\epsilon)}{(1-\epsilon)}\lambda_2 = (1+5\epsilon)\lambda_2.$$

Letting $g_2 = f/\|f\|$ completes the proof. □

The following corollary proved by Sachdeva and Vishnoi generalizes Theorem 7.6.2 to multiple eigenfunctions.

Corollary 7.6.4 ([SV]). *For any undirected graph G on n vertices and m edges and any $k \geq 1$, there is an algorithm that outputs orthonormal functions $f_1, f_2, \ldots, f_k \in \ell^2(V, w)$ such that for all $1 \leq i \leq k$, $\mathcal{R}(f_i) \leq 2\lambda_i$. and runs in $\tilde{O}(k(m+n))$ time where we avoid the* polylog(n) *factors.*

The proof of the above corollary essentially follows from a repeated application of Theorem 7.6.2 and Theorem 7.1.5.

7.7 Expansion, Conductance and Sparsest Cut

For a set $S \subseteq V$, we write

$$E(S, \overline{S}) := \{(u, v) : u \in S, v \notin S\}$$

to denote the set of edges in the cut $(S, \overline{S})$. We also use

$$w(S, \overline{S}) := \sum_{(u,v) \in E(S,\overline{S})} w(u, v)$$

to denote the sum of the weight of edges in the cut $(S, \overline{S})$.

The *conductance* of a set S is the ratio of the fraction of edges in the cut $(S, \overline{S})$ w.r.t. the volume of S,

$$\phi(S) := \frac{w(S, \overline{S})}{\text{vol}(S)}.$$

Observe that for any set $S \subseteq V$, $0 \leq \phi(S) \leq 1$. If $\phi(S) \approx 0$, S may represent a cluster in G. Conductance is a very well studied measure for graph clustering in the literature (see e.g. [SM00, KVV04, TM06]). Let us motivate this definition by providing some examples.

Suppose that G is an unweighted graph that represents friendships in a social network. In particular, each vertex represents an individual and an edge (u, v) shows that u is a friend of v and vice versa. Then, If $\phi(S) = \epsilon$ for $S \subseteq V$ and ϵ very close to zero, it means that at least $1 - 2\epsilon$ fraction of all of friendship relations of individuals of S are with other individuals in S. In other words, those in S are mostly friend with each other than the rest of the world.

One way to design a graph clustering algorithm is to find sets with smallest conductance in G. Observe that for any graph G, $\phi(V) = 0$. To avoid these special cases, the problem is formulated as finding a sufficiently small set with the least conductance. The conductance of G, $\phi(G)$ is the smallest conductance among all sets that have at most half of the total volume,

$$\phi(G) := \min_{S:\text{vol}(S) \leq \text{vol}(V)/2} \phi(S).$$

There are several other similar objectives (e.g. edge expansion, sparsest cut, etc.) that one can use to compare a cut $(S, \overline{S})$ with other cuts in G (see below for the some examples). Typically these parameters are closely related in regular graphs but they are incomparable in non-regular graphs. In this thesis we mostly focus on conductance. The main reason for that is conductance is closely related to the eigenvalues of the normalized Laplacian matrix (see e.g. Section 7.8). Because of that analyzing conductance of sets leads to analysis of random walks or Markov Chains, simple and fast clustering algorithms, etc.

The *edge expansion* of a set S is defined as the ratio of the edges leaving S to the number of vertices in S,

$$\text{expansion}(S) := \frac{w(S, \overline{S})}{|S|}.$$

In a non-regular graph the edge expansion is incomparable with the conductance. But, if G is a d-regular graph, that is $w(v) = d$ for all $v \in V$, then for any $S \subseteq V$, $\phi(S) = \text{expansion}(S)/d$. The edge expansion of G is the smallest edge expansion of any subset with at most half of the vertices,

$$\text{expansion}(G) = \min_{S:|S| \leq |V|/2} \text{expansion}(S).$$

Edge expansion is closely related to the *sparsest cut* problem. Let the sparsity of a cut $(S, \overline{S})$ be the

following:

$$\text{sparsity}(S) := \frac{w(S,\overline{S})}{|S| \cdot |V-S|}.$$

In the uniform sparsest cut problem we want to find a cut with the minimum sparsity,

$$\text{sparsity}(G) := \min_{\emptyset \subset S \subset V} \text{sparsity}(S).$$

Observe that for any set $S \subseteq V$ such that $|S| \leq |V|/2$, we have

$$\text{expansion}(S) \leq n \cdot \text{sparsity}(S) \leq 2\,\text{expansion}(S).$$

Therefore, any approximation algorithm for expansion(G) is also an approximation algorithm for sparsity(G) (up to a factor of 2), and vice versa. Furthermore, if G is a d-regular graph, then

$$\phi(S) \leq \frac{n}{d} \cdot \text{sparsity}(S) \leq 2\phi(S).$$

Therefore, in regular graphs these 3 parameters are closely related, but in non-regular graphs $\phi(G)$ is incomparable to both sparsity(G) and expansion(G).

In the general version of the sparsest cut problem (a.k.a. non-uniform sparsest cut) we are given a set of demands, $\text{dem}(u,v) \geq 0$ for any $u, v \in V$ and the non-uniform sparsity of a cut $(S,\overline{S})$ is defined as

$$\text{sparsity}(S, \text{dem}) := \frac{w(S,\overline{S})}{\sum_{u \in S, v \notin S} \text{dem}(u,v)}.$$

The goal is to find a cut with the smallest non-uniform sparsity,

$$\text{sparsity}(G, \text{dem}) := \min_{\emptyset \subset S \subset V} \text{sparsity}(S, \text{dem}).$$

For example, the uniform sparsest cut problem correspond to letting $\text{dem}(u,v) = 1$ for all $u, v \in V$. As another example, if $\text{dem}(u,v) = w(u)w(v)$ for all $u, v \in V$, then

$$\phi(G) \leq \text{vol}(V) \cdot \text{sparsity}(G, \text{dem}) \leq 2\phi(G).$$

Intuitively, the function dem$(.,.)$ specifies which pairs of vertices we would like to see disconnected in G, and our goal is to find a set of edges to remove such that their number is small compared to the sum of the pairwise demands of vertices that are disconnected by their removal.

7.7.1 Continuous Relaxations of Conductance

There are several continuous relaxations for the conductance of a graph G. The first one is the Rayleigh quotient. Recall that for a function $f : V \to \mathbb{R}$,

$$\mathcal{R}(f) = \frac{\sum_{\{u,v\}\in E} w(u,v)|f(u) - f(v)|^2}{\sum_{v\in V} w(v)f^2(v)}.$$

Observe that if $f : V \to \{0, 1\}$ is not a constant function, then

$$\mathcal{R}(f) = \phi(\mathrm{supp}(f)).$$

So, if f is a zero-one function where $\mathrm{vol}(\mathrm{supp}(f)) \leq \mathrm{vol}(V)/2$, and f minimizes the Rayleigh quotient among all such functions, then the support of f is the sparsest cut of G. Therefore,

$$\min_{f_1,f_2\in\ell^2(V,w),f_1\perp f_2} \max\{\mathcal{R}(f_1), \mathcal{R}(f_2)\}$$

is a lower bound on $\phi(G)$ (see Claim 7.8.2 for a rigorous proof).

The importance of this relaxation is that we exactly know the optimizers and we can compute them efficiently. By (7.2.3), the minimizers of the Rayleigh quotient over all functions $f : V \to \mathbb{R}$ are the normalizations of the eigenfunctions of $\mathcal{L}$. Let $g_1, \ldots, g_n$ be the orthonormal eigenfunctions of $\mathcal{L}$. Then $f_1 = D^{1/2}g_1$ the minimizer of the Rayleigh quotient, $f_2 = D^{1/2}g_2$ is the minimizer among all functions that are orthogonal to f_2 with in the space $\ell^2(V, w)$, and so on. Therefore, $\max\{\mathcal{R}(f_1), \mathcal{R}(f_2)\}$ is a lower bound on $\phi(G)$. As we will show in Section 7.8, the optimum of the Rayleigh quotient gives a square root approximation of $\phi(G)$.

The second continuous relaxation of conductance is the Linear Programming relaxation of Leighton and Rao [LR99]. We do not use this relaxation explicitly in this thesis but we include it for the sake of completeness. For the simplicity of notation let us assume G is unweighted and w-regular.

$$\begin{aligned}
\min \quad & \frac{n}{2w} \cdot \sum_{\{u,v\}\in E} d(u,v) \\
\text{subject to} \quad & \sum_{u,v} d(u,v) = w \cdot 1, \\
& d(u,v) \geq d(u,u') + d(u',v) && \forall u, v, u' \in V \\
& d(u,v) \geq 0 && \forall u, v \in V.
\end{aligned}$$

If $\phi(G) = \phi(S)$ for $S \subseteq V$, then we can let $d(u,v) = 1/(|S| \cdot (n - |S|))$ if u, v are in different sides of the cut $(S, \overline{S})$ and $d(u,v) = 0$ otherwise. This shows that the value of the above linear program lower bounds $\phi(G)$.

The above relaxation is not necessarily stronger than the Rayleigh quotient. Leighton and

Rao [LR99] show that the optimum of above linear program is at least $\Omega(1/\log(n))$ fraction of $\phi(G)$, i.e., the integrality gap of this continuous relaxation is $O(\log n)$. Therefore, for graphs where $\phi(G) \ll \log(n)$ the above relaxation is stronger than the Rayleigh quotient.

The third relaxation is the Arora, Rao and Vazirani's Semidefinite programming relaxation [ARV09].

$$\begin{aligned} \min \quad & \frac{n}{2w} \sum_{\{u,v\} \in E} \|\mathbf{x}_u - \mathbf{x}_v\|^2 \\ \text{subject to} \quad & \sum_{u,v \in V} \|\mathbf{x}_u - \mathbf{x}_v\|^2 = 1 \\ & \|\mathbf{x}_u - \mathbf{x}_v\|^2 \leq \|\mathbf{x}_u - \mathbf{x}_{u'}\|^2 + \|\mathbf{x}_{u'} - \mathbf{x}_v\|^2 \qquad \forall u, v, u' \in V \end{aligned}$$

If $\phi(G) = \phi(S)$ for $S \subseteq V$, then we can let

$$\mathbf{x}_u = \begin{cases} \sqrt{\frac{1}{|S| \cdot (n - |S|)}} & \text{if } u \in S, \\ 0 & \text{otherwise.} \end{cases}$$

This shows that the value of the above semidefinite program lower bounds $\phi(G)$.

It is an easy exercise that the above relaxation is always as strong as the Rayleigh quotient. In fact if we drop the third constraint, the triangle inequality constraint, then the optimum of the above program (up to a factor 2) is exactly the same as λ_2. Arora, Rao and Vazirani [ARV09] show that the optimum of above program is at least $\Omega(1/\sqrt{\log(n)})$ of $\phi(G)$, i.e., the integrality gap of the above relaxation is $O(\sqrt{\log(n)})$.

7.7.2 Computability of Conductance

The problem of finding the set with the smallest conductance in a given graph is one of the fundamental problems in the field of computing. The problem is NP-hard [MS90], and although we still don't know if this problem is APX hard, it is conjectured that it is NP-hard to find a constant factor approximation algorithm for this problem. Chawla, Krauthgamer, Kumar, Rabani and Sivakumar [CKK+05] show that it is unique games hard to find a constant factor approximation algorithm for the non-uniform sparsest cut.

There are several approximation algorithm for approximating the (non-uniform) sparsest cut of a graph. Alon and Milman [AM85] proved a discrete version of Cheeger's inequality and designed a spectral algorithm that provides a $O(1/\sqrt{\phi(G)})$ approximation to $\phi(G)$ (see Section 7.8 for more details). Leighton and Rao [LR99] designed an $O(\log n)$ approximation algorithm to the sparsest cut problem (see also [LLR95] for different proof techniques). Arora Rao and Vazirani [ARV09] designed an $O(\sqrt{\log n})$ approximation algorithm to the uniform sparsest cut problem. Chawla, Gupta and Racke [CGR05] and Arora, Lee and Naor [ALN08] designed an $O(\log^{3/4} n)$ and $O(\sqrt{\log n} \log\log n)$

approximation algorithm to the non-uniform sparsest cut problem.

Within the long-term research program of developing better approximation algorithms for uniform sparsest cut in general graphs, there has been much success in the last couple of years toward developing better algorithms for restricted classes of graphs.

The technique of subspace enumeration [KT07, Kol11, ABS10] applies to the special class of graphs known as "low threshold rank" graphs. A graph G has a low threshold rank, if $\lambda_k = \Omega(1)$ for a small number k. Low threshold rank graphs can be considered as a generalization of expander graphs.

Arora, Barak and Steurer [ABS10] show that the technique of *subspace enumeration* developed in the work of Kolla and Tulsiani [KT07, Kol11], achieves an $O(1/\lambda_k)$ approximation in time $2^{O(k)}\,\mathrm{poly}(n)$. Later, Barak, Raghavendra and Steurer [BRS11] and Guruswami and Sinop [GS11] match this $O(1/\lambda_k)$ approximation factor in time $n^{O(k)}$ by using an SDP relaxation that is derived from the Arora-Rao-Vazirani relaxation by k "rounds" of a procedure defined by Lasserre. The procedure starts from a SDP relaxation of a combinatorial problem which can be formulated as a 0/1 integral program, and defines a family of relaxations with additional variables and constraints. The k-th relaxation in this family has size $n^{O(k)}$. We will refer to the k-th relaxation in this family as Lass_k-ARV.

These techniques are very powerful, and they lead to approximation algorithms for many constraint satisfaction problems including maximum cut, sparsest cut, minimum uncut, graph coloring, etc [AG11, GS12, GS13, OT12]. These algorithms run in time that is exponential in k, and they typically provide an approximation ratio of $1/\,\mathrm{poly}(\lambda_k)$. A notable exception is the work of Guruswami and Sinop [GS13], who show that even if $\lambda_k \ll 1$, but there is a gap between λ_k and $\phi(G)$, such as, say, $\lambda_k > 2\phi(G)$, then a constant-factor approximation can be derived from Lass_k-ARV. The approximation ratio can be made arbitrarily close to one if the ratio $\lambda_k/\phi(G)$ is sufficiently large. Because of the exponential dependency on k in the running time of all of the above algorithms, we may obtain a polynomial time algorithm only for graphs with a *fast growing spectrum*, i.e., if $k = O(\log n)$.

7.7.3 The Small Set Expansion Problem

Lovasz and Kannan [LK99] defined the *conductance profile* of a graph G as follows:

$$\phi(k) := \min_{S:\mathrm{vol}(S)\leq k} \phi(S). \tag{7.7.1}$$

Lovasz and Kannan used the conductance profile as a parameter to prove strong upper bounds on the mixing time of random walks. The notion of conductance profile recently received significant attention in the literature because of its close connection to the small set expansion problem and the unique games conjecture [RS10].

Let

$$\phi^c(k) := \min_{S:\mathrm{vol}(S)\leq \mathrm{vol}(V)/k} \phi(S).$$

In words, $\phi^c(k) = \phi(\mathrm{vol}(V)/k)$. The small set expansion problem is defined as follows:

Problem 7.7.1 (GAP Small Set Expansion sse(k,ϵ) [RS10])**.** *Given a graph G, and constants $k, \epsilon > 0$, distinguish whether,*

$$\phi^c(k) \geq 1-\epsilon \quad \text{or} \quad \phi^c(k) \leq \epsilon.$$

Raghavendra and Steurer conjectured that the above problem is hard for sufficiently small values of δ,

Conjecture 7.7.2 (Gap-Small-Set Expansion Conjecture)**.** *For every $\epsilon > 0$, there exists k such that the problem Gap- Small-Set Expansion (ϵ, k) is NP-hard.*

Raghavendra and Steurer in [RS10] show that Gap-Small-Set-Expansion problem is easier than the Unique Games conjecture.

Theorem 7.7.3 (Raghavendra,Steurer [RS10])**.** *The Gap-Small-Set Expansion conjecture implies the Unique Games Conjecture.*

The above theorem shows that if someone wants to design an algorithm to refute Unique Games conjecture she must start by designing an algorithm for the small set expansion problem. Therefore, in this thesis we mainly pursue algorithms for the small set expansion problem.

Let the *threshold rank* of G, denoted by $\mathrm{rank}_\eta(\mathcal{L})$, be the number (with multiplicities) of eigenvalues λ of $\mathcal{L}$, satisfying $\lambda < \eta$. Arora, Barak and Steurer [ABS10] solved the small set expansion problem on regular graphs G, assuming that $\mathrm{rank}_\eta(\mathcal{L}) \geq n^{\mathrm{poly}(\eta/\epsilon)}$, for some constant $\eta > 0$. Unfortunately, their algorithm does not provide any approximation of $\phi^c(\delta)$ for general graphs, since they may have much fewer large eigenvalues.

Raghavendra, Steurer, Tetali [RST10], and Bansal et al. [BFK+11] used the semidefinite programming relaxation of the problem, and designed algorithms that approximate $\phi^c(k)$ within factors $O(\phi^c(k)^{-1/2}\sqrt{\log k})$, and $O(\sqrt{\log n \log k})$ of the optimum, respectively. However, in the interesting regime of $k = \omega(1)$, which is of interests to the small set expansion problem, the quality of both approximation algorithms is not independent of k.

In Chapter 12 we solve small set expansion problem when $k = \mathrm{poly}(n)$. Unfortunately, this would not refute Gap-Small-Set Expansion conjecture because there both k and ϵ are independent of the size of the graph.

7.8 Cheeger's Inequality

In Riemannian geometry, the Cheeger isoperimetric constant of a compact Riemannian manifold is a positive real number defined in terms of the minimal area of a surface that divides the manifold

into two disjoint pieces of equal volume. Cheeger [Che70] showed that the first nontrivial eigenvalue of the Laplace-Beltrami operator on the manifold is related to the isoperimetric constant. In this thesis we will not give more details about this, instead we discuss an analogue of this result on graphs. Interested readers are referred to a recent survey by Trevisan [Tre13].

In Section 7.2 we show that for any graph G, $\lambda_2 = 0$, if and only if G is disconnected. The discrete version of Cheeger's inequality provides a robust version of this fact. Roughly speaking, G is barely connected if and only if λ_2 is very close to zero. We use the notion of conductance to quantify graphs that are barely connected.

Theorem 7.8.1 (Discrete Cheeger's Inequality [AM85, Alo86, SJ89])**.** *For any graph G,*

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

Furthermore, there is a simple near linear time algorithm (the Spectral Partitioning algorithm) that given a function $f : V \to \mathbb{R}$ such that $\mathcal{R}(f) \leq c \cdot \lambda_2$ for a constant $c \geq 1$, finds a set S such that $\mathrm{vol}(S) \leq \mathrm{vol}(V)/2$ *and*

$$\phi(S) = O(\sqrt{c \cdot \phi(G)}).$$

Cheeger's inequality is one of the most influential results in spectral graph theory with significant applications in spectral clustering [ST96, KVV04], explicit construction of expander graphs [JM85, HLW06, Lee12], approximate counting [SJ89, JSV04], and image segmentation [SM00].

Next we provide several applications of this inequality.

Uniform Sparsest Cut Problem. Cheeger's inequality provides an $O(1/\sqrt{\phi(G)})$ approximation algorithm to $\phi(G)$. This is very important because the approximation factor does not depend on the size of the graph, it rather depends on the value of the optimum. Consequently, this inequality provides an efficient algorithm to test if a given graph is an expander. In addition, Cheeger's inequality provides a very simple near linear time algorithm for approximating $\phi(G)$. As we describe in the proof of Theorem 7.8.1 given an approximation of the second eigenfunction of $\mathcal{L}$, one can find a set S of conductance $O(\sqrt{\lambda_2})$ in linear time. Since by Theorem 7.6.2 we can find a function f such that $\mathcal{R}(f) \leq 2\lambda_2$ in near linear time we obtain a near linear time approximation algorithm to $\phi(G)$.

Graph Separator Problem. Planar Separator theorem states that any planar graph can be split into smaller pieces by removing a small number of vertices. In particular, by removing $O(\sqrt{n})$ vertices we can divide a planar graph into two sets each of at most $2n/3$ vertices. This theorem is very essential in several of divide and conquer algorithms on planar graphs. Spielman and Teng [ST96] show that if G is a planar bounded degree graph, then Spectral partitioning algorithm to produce a separator of G. This is later extended to bounded genus graphs or minor free graphs ([Kel04, BLR08, KLPT11]).

Analyzing Mixing time of Random Walks. Jerrum and Sinclair [SJ89] prove that the ϵ-uniform mixing time of any *lazy* random walk (equivalently, reversible Markov Chain) is bounded from above by,

$$\tau_\infty(\epsilon) \leq \frac{2}{\phi(G)^2}\left(\log\frac{1}{\min_v \pi(v)} + \log\frac{1}{\epsilon}\right). \tag{7.8.1}$$

The above inequality simply follows from an application of Cheeger's inequality to Lemma 7.4.2. This result is extended to non-reversible chains by Mihail [Mih89].

7.8.1 Proof of the Cheeger's Inequality

The left side of the Cheeger's inequality is also known as the easy direction and the right side the hard direction. We start by proving the easy direction.

Claim 7.8.2. *For any graph G, $\lambda_2/2 \leq \phi(G)$.*

Proof. For $S \subseteq V$, let $\mathbf{1}_S$ be the indicator of S, that is $\mathbf{1}_S(v) = 1$ if $v \in S$ and $\mathbf{1}_S(v) = 0$ otherwise. Then,

$$\mathcal{R}(\mathbf{1}_S) = \frac{\sum_{(u,v)\in E} w(u,v)|\mathbf{1}_S(u) - \mathbf{1}_S(v)|^2}{\sum_{v\in S} \mathbf{1}_S(v)\cdot w(v)} = \frac{\sum_{u\in S, v\notin S} w(u,v)}{\sum_{v\in S} w(v)} = \phi(S).$$

Let $S = \text{argmin}_{\text{vol}(S)\leq \text{vol}(V)/2}\,\phi(S)$. Then, by Lemma 7.2.1,

$$\lambda_2 \leq 2\max\{\mathcal{R}(\mathbf{1}_S), \mathcal{R}(\mathbf{1}_{\overline{S}})\} = 2\max\{\phi(S), \phi(\overline{S})\} = 2\phi(S) = 2\phi(G).$$

□

Before proving the right side of the Cheeger's inequality we show several lemmas that will be used later in this thesis. For any function $f : V \to \mathbb{R}$, and any threshold $t \in \mathbb{R}$, we define the *threshold cut* $S_f(t)$ as follows

$$S_f(t) := \{v : f(v) \geq t\}.$$

Many variants of the following lemma are known; see, e.g. [Chu96].

Lemma 7.8.3. *For every non-negative $f : V \to \mathbb{R}$, the following holds*

$$\min_{t>0} \phi(S_f(t)) \leq \frac{\sum_{(u,v)\in E} w(u,v)|f(v) - f(u)|}{\sum_v w(v)f(v)}.$$

Proof. First we recall a simple inequality that is very crucial. Let $a_1, a_2, \ldots, a_m, b_1, b_2 \ldots, b_m \geq 0$. Then,

$$\min_{1\leq i\leq m} \frac{a_i}{b_i} \leq \frac{a_1 + \ldots + a_m}{b_1 + \ldots + b_m}. \tag{7.8.2}$$

Since the right hand side is homogeneous in f, we may assume that $\max_{v\in V} f(v) \leq 1$. Let

$0 < t \leq 1$ be chosen uniformly at random. Then, by linearity of expectation,

$$\frac{\mathbb{E}\left[w(S_f(t), \overline{S_f(t)})\right]}{\mathbb{E}\left[\mathrm{vol}(S_f(t))\right]} = \frac{\sum_{(u,v)\in E} w(u,v)|f(u) - f(v)|}{\sum_v w(v)f(v)}.$$

Therefore, by equation (7.8.2) there exists a $0 < t$ such that $\phi(S_f(t)) \leq \frac{\sum_{(u,v)\in E} w(u,v)|f(v)-f(u)|}{\sum_v w(v)f(v)}$. □

Lemma 7.8.4. *For any non-negative function $f : V \to \mathbb{R}$, the following holds*

$$\min_{t>0} \phi(S_f(t)) \leq \sqrt{2\mathcal{R}(f)}.$$

Proof. Let $g(v) := f(v)^2$ for all $v \in V$. Observe that $g(v) < g(u)$ if and only if $f(u) < f(v)$. Therefore, by Lemma 7.8.3,

$$\begin{aligned}
\min_{t>0} \phi(S_f(t)) = \min_{t>0} \phi(S_g(t)) &\leq \frac{\sum_{(u,v)\in E} w(u,v)|g(u) - g(v)|}{\sum_{v\in V} w(v)g(v)} \\
&= \frac{\sum_{(u,v)\in E} w(u,v)\left|f(u)^2 - f(v)^2\right|}{\sum_{v\in V} w(v)f(v)^2} \\
&= \frac{\sum_{(u,v)\in E} w(u,v)\|f(u) - f(v)\| \cdot \|f(u) + f(v)\|}{\sum_{v\in V} w(v)f(v)^2} \\
&\leq \frac{\sqrt{\sum_{(u,v)\in E} w(u,v)\|f(u) - f(v)\|^2} \cdot \sqrt{\sum_{(u,v)\in E} w(u,v)\|f(u) + f(v)\|^2}}{\sum_{v\in V} w(v)f(v)^2} \\
&\leq \sqrt{\mathcal{R}(f)}\sqrt{\max_{(u,v)\in E} \frac{|f(u) + f(v)|^2}{f(u)^2 + f(v)^2}} \leq \sqrt{2\mathcal{R}(f)}\,.
\end{aligned}$$

where the second inequality follows by the CauchySchwarz inequality. □

Now, we are ready to prove the hard direction of Cheeger's inequality.

Proof of Theorem 7.8.1. Let $g : V \to \mathbb{R}$ be a non-constant function. By Lemma 7.2.2 and Lemma 7.2.3 there are two disjointly supported functions $f_+, f_- \in \ell^2(V, w)$ such that $\mathcal{R}(f_+), \mathcal{R}(f_-) \leq c \cdot \mathcal{R}(D^{-1/2}g)$, where $c = 1$ if g is an eigenfunction of $\mathcal{L}$ and $c = 4$ otherwise.

Wlog assume that $\mathrm{vol}(\mathrm{supp}(f_+)) \leq \mathrm{vol}(V)/2$. Then, by Lemma 7.8.4,

$$\phi(G) \leq \min_{t>0} \phi(S_{f_+}(t)) \leq \sqrt{2c \cdot \mathcal{R}(f)}.$$

□

Above proof shows that (up to linear normalizations) a simple partitioning algorithm that finds the best threshold cut defined by a function $D^{-1/2}g$ gives a $1/\sqrt{\phi(G)}$ approximation to $\phi(G)$. The details are described in Algorithm 10.

Algorithm 10 Spectral Partitioning Algorithm

Input: A non-constant function $g : V \to \mathbb{R}$.

Output: A set $S \subseteq V$ such that $\mathrm{vol}(S) \leq \mathrm{vol}(V)/2$ and $\phi(S) \leq O(\sqrt{\mathcal{R}(D^{-1/2}g)})$.

Let f_+ and f_- be the functions defined in Lemma 7.2.2. If g is an eigenfunction of $\mathcal{L}$, then we can simply let $f_+(v) = g(v)/\sqrt{w(v)}, f_-(v) = 0$ if $g(v) > 0$ and $f_+(v) = 0, f_-(v) = g(v)/\sqrt{w(v)}$ otherwise.

Wlog assume that $\mathrm{vol}(\mathrm{supp}(f_+)) \leq \mathrm{vol}(V)/2$,

Sort the vertices in non-decreasing order of values in f_+, that is suppose $f_+(v_1) \leq f_+(v_2) \leq \ldots \leq f_+(v_n)$.

Choose i such that $\phi(\{v_1, \ldots, v_i\}) = \min_{1 \leq j \leq n} \phi(\{v_1, \ldots, v_j\})$.

return $\{v_1, \ldots, v_i\}$.

Observe that for a given function g the algorithm can be implemented in near linear time, and by above proof it returns a set of conductance $O(\sqrt{\mathcal{R}(D^{-1/2}g)})$. We remark that, usually one would write the spectral partitioning algorithm by computing the best threshold cut of the function $D^{-1/2}g$. By Lemma 7.2.3, if g is a non-constant eigenfunction of $\mathcal{L}$, this algorithm would work since the support of $f_+(f_-)$ is simply the vertices with positive (negative) value in $D^{-1/2}g$, respectively. However, if g is any arbitrary non-constant function, and G is not a regular graph, we need to choose a particular normalization of g as described in Lemma 7.2.2.

7.8.2 Tightness of Cheeger's inequality

In this part we provide several examples to show that the all parts of the analysis in this section are tight. Nonetheless, we manage to improve this analysis in Chapter 11.

Example 7.8.5 (Hypercube)**.** *Let G be a h-dimensional hypercube. By our discussion in Subsection 7.3.4, $\lambda_2 = 2/h$. On the other hand, the sparsest cut of a hypercube is a dimensional cut, i.e., for $S = \{\mathbf{a} : \mathbf{a}(1) = 0\}$, we have*

$$\phi(S) = \frac{|E(S, \overline{S})|}{h \cdot |S|} = \frac{|S|}{h \cdot |S|} = 1/h.$$

Therefore, the left side of Cheeger's inequality is exactly tight.

Example 7.8.6 (Cycle)**.** *Let G be a cycle of length n. By our discussion in Subsection 7.3.1, $\lambda_2 = \cos(2\pi/n) = 4\pi^2/n^2 + O(1/n^4)$. On the other hand, let S be a path of length $n/2$, we get $\phi(S) = 1/|S| = n/2$. So, the right side of Cheeger's inequality is tight up to constant factors. Note that although the right side of the Cheeger's inequality is tight, spectral partitioning algorithm finds a cut of conductance $O(\phi(G))$, so our analysis is not tight for the cycle graph.*

Example 7.8.7 (Ladder)**.** *Our last example is the ladder graph described in Subsection 7.3.3. Suppose that $l > 10^6$, and G is a ladder graph made by connecting two cycles of length l with a perfect*

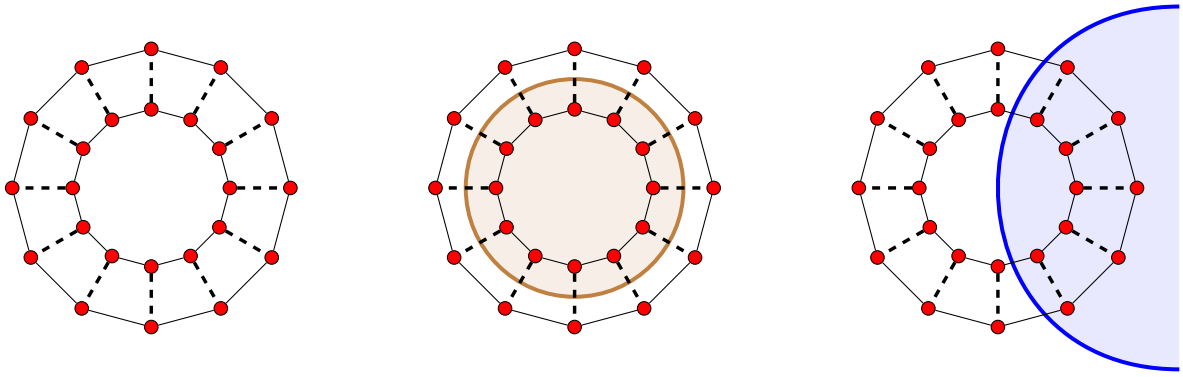

Figure 7.8.1: An illustration of the Ladder graph. Here each solid edge has weight 1 and each dashed edge has weight $100/l^2$ where l is the length of the cycles. The middle figure shows the set with the optimum conductance for a sufficiently large l, and the right figure shows the output of the spectral partitioning algorithm.

matching of size l (see Figure 7.8.1). Since the edges of the matching have weight $100/l^2$, the set with the smallest conductance is $S = \{(r,0) : 0 \leq 0 \leq n-1\}$ the set of vertices in one of the cycles,

$$\phi(S) = \frac{100|S|/l^2}{(2+100/l^2)|S|} = O(1/l^2).$$

Note that for any other set S such that $\mathrm{vol}(S) \leq \mathrm{vol}(V)/2$, $\phi(S) = \Omega(1/l)$.

We show that the spectral partitioning algorithm indeed divides the cycles and its output has conductance $\Omega(1/l)$. By our discussion in Subsection 7.3.3 $\lambda_2 = O(1/l^2)$ and the corresponding eigenfunction is $\chi_{1,0}$ where for each pair of vertices $(r,0)$ and $(r,1)$ that are matched $\chi_{1,0}(r,0) = \chi_{1,0}(r,0)$. This shows that the second eigenfunction doesn't discriminate between the vertices of the two cycles. Thus, the spectral partitioning algorithm divides the cycles and its output has conductance $\Omega(1/l)$. Since $\lambda_2 = O(1/l^2$, this implies that our analysis of the spectral partitioning algorithm is tight up to a constant factor.

7.9 Random Partitioning of Metric Spaces

A *distance* function is a function $d : V \times V \to \mathbb{R}_+$. such that for all $u, v \in V$, $d(u,v) = d(v,u)$ and for all $v \in V$, $d(v,v) = 0$. For a set $S \subseteq V$, the diameter of S with respect to a distance function $d(.,.)$ is the maximum distance between the vertices of S,

$$\mathrm{diam}(S,d) := \max_{u,v \in S} d(u,v).$$

The distance of a vertex u to a set S is the distance to the closest vertex in S,

$$d(u, S) := \min_{v \in S} d(u, v).$$

Also, for $S, T \subseteq V$, $d(S, T) := \min_{v \in T} d(v, S)$. For $S \in V$ and $r \geq 0$ we write

$$B_d(S, r) := \{v \in V : d(u, v) \leq r\}$$

to denote the closed ball of radius r about S (if the distance function is clear in the context we may drop the subscript). We also abuse the notation and use $B_d(v, r)$ if $S = \{v\}$. For any mapping $F : V \to \mathbb{R}^k$ we use $B_F(S, r)$ to denote the ball of radius r about S with respect to the Euclidean distance function $d(u, v) = \|F(u) - F(v)\|$.

We say $d(.,.)$ defines a metric if for any three vertices $u, v, v' \in V$,

$$d(u, v) \leq d(u, v') + d(v, v').$$

Also, we say d is a *pseudo metric* if the distance between two distinct vertices can be zero. Sometimes, it will be useful to consider the largest metric on the graph G which agrees with a given metric $d(.,.)$ on edges. This is the induced shortest-path (extended pesudo-) metric on G, where the length of an edge $(u, v) \in E$ is given by $d(u, v)$. We will use the notation $\hat{d}$ for this metric. Observe that $\hat{d} \geq d$ whenever $d(.,.)$ is a (pseudo-) metric.

In many parts of this thesis we would like to partition V into sets of small diameter such that the endpoints of most of the edges map to the same set. Consequently, one can argue that most of the sets in such a partitioning has a small conductance since they cover a large fraction of vertices while cutting only a few edges. So, in this subsection we focus on the literature of random partitioning on metric spaces. These provide a useful tool in analyzing high dimensional embedding of G.

We write a partition P of V as a function $P : V \to 2^V$ mapping a vertex $v \in V$ to the unique set in P that contains v. For $\Delta > 0$, we say that P is Δ-*bounded* if $\text{diam}(S, d) \leq \Delta$ for every $S \in P$. We will also consider distributions over random partitions. If $\mathcal{P}$ is a random partition of V, we say that $\mathcal{P}$ is Δ-bounded if this property holds with probability one.

A random partition $\mathcal{P}$ is (Δ, α, δ)-*padded* if $\mathcal{P}$ is Δ-bounded, and for every $v \in V$, we have

$$\mathbb{P}\left[B(v, \Delta/\alpha) = \mathcal{P}(v)\right] \geq \delta.$$

A random partition is (Δ, L)-*Lipschitz* if $\mathcal{P}$ is Δ-bounded, and, for every pair $u, v \in V$, we have

$$\mathbb{P}\left[\mathcal{P}(u) \neq \mathcal{P}(v)\right] \leq L \cdot \frac{d(u, v)}{\Delta}.$$

Before reviewing the literature on partitioning of metric spaces we give a simple example by

partitioning a mapping of vertices of G into $\mathbb{R}^k$. Suppose that we are given a function $F : V \to \mathbb{R}^k$ and we let $d(u, v) = \|F(u) - F(v)\|$ for all $u, v \in V$. The goal is to partition the vertices into sets of diameter Δ such that most of the endpoints of edges of G map to the same set. Let P be a partition of $\mathbb{R}^k$ into axis-parallel cubes of side length $\Delta/\sqrt{k}$. Every set S of P is the vertices inside one such cube. Observe that the diameter of any set in P is at most Δ (this is because the diameter of a k dimensional cube is $\sqrt{k}$ times its side length). We choose a Δ-bounded random partition $\mathcal{P}$ by choosing a uniformly random axis-parallel translation of P. It turns out that for any edge $(u, v) \in E$,

$$\mathbb{P}\left[\mathcal{P}(u) \neq \mathcal{P}(v)\right] \leq \sqrt{k} \cdot \frac{\|F(u) - F(v)\|}{\Delta/\sqrt{k}}.$$

Thus, $\mathcal{P}$ is $(\Delta, O(k))$-Lipschitz. Note the polynomial dependency to the size of the dimension. In other words, this shows that partitioning is much easier in lower dimensional space in the sense that larger fraction of adjacent vertices map to the same region.

It turns out that the above construction is not optimal because the diameter of a cube is $\sqrt{k}$ times larger than its side length. It is more efficient to use a partitioning of $\mathbb{R}^k$ into balls as opposed to cubes. But, unfortunately, there is no covering of $\mathbb{R}^k$ with disjoint balls of equal radii. So the above construction doesn't naturally extend to a partitioning by balls. Nonetheless, this extension is studied in [CCG+98] through a more sophisticated construction (see also [LN05, Lem 3.16]).

Theorem 7.9.1. *If vertices are mapped to $\mathbb{R}^k$, and $d(.,.)$ is the Euclidean distance between the vertices, then for every $\Delta > 0$, (V, d) admits a $(\Delta, O(\sqrt{k}))$-Lipschitz random partition.*

The next theorem is proved in [GKL03] and provides a padded partitioning of Euclidean metrics (see also [LN05, Lem 3.11]).

Theorem 7.9.2. *If vertices are mapped to $\mathbb{R}^k$, and $d(.,.)$ is the Euclidean distance between the vertices, then for every $\Delta > 0$ and $\delta > 0$, (V, d) admits a $(\Delta, O(k/\delta), 1-\delta)$-padded random partition.*

If $d(.,.)$ is any general metric space, then Lipschitz parameter must depend on the number of vertices of G. The following theorem can be derived from work of Leighton and Rao [LR99], or Linial and Saks [LS93]. The form stated below comes from work of Bartal [Bar98].

Theorem 7.9.3. *For any distance function $d(.,.)$, (V, d) admits a $(\Delta, O(\log n))$-Lipschitz random partition.*

Finally, we include $d(.,.)$ is a shortest path metric of a (weighted) low-dimensional graph then we can avoid any dependency to the number of vertices of G. A partitioning theorem for excluded-minor graphs is presented in [KPR93], with an improved quantitative dependence coming from [FT03].

Theorem 7.9.4. *If $d(.,.)$ is the shortest-path metric on a graph excluding K_c as a minor, then for every $\Delta > 0$ and $\delta > 0$, (V, d) admits a $(\Delta, O(c^2/\delta), 1 - \delta)$-padded random partition and a $(\Delta, O(c^2))$-Lipschitz random partition.*

For example, any shortest path metric of defined on a planar graph admits a $(\Delta, O(1))$-Lipschitz random partition. For the special case of bounded-genus graphs, a better bound is known [LS10].

Theorem 7.9.5. *If $d(.,.)$ is the shortest-path metric on a graph of genus g, then for every $\Delta > 0$ and $\delta > 0$, (V, d) admits a $(\Delta, O((\log g)/\delta), 1-\delta)$-padded random partition, and a $(\Delta, O(\log g))$-Lipschitz random partition.*

Chapter 8

New Machineries

The goal of this chapter is to describe several of the new machineries that are developed in our works. Many of these machineries have been used in several contexts including probability theory, spectral graph theory, and algorithm design. In this chapter we define these tools in an abstract sense and we try to motivate them by giving some examples.

8.1 Spectral Embedding

Spectral embedding of graphs uses the bottom k eigenfunctions of the normalized Laplacian matrix to embed the graph into $\mathbb{R}^k$. The primary use of this embedding has been in practical spectral clustering algorithms [SM00, NJW02] (see 1.2.1 for more details). In this section we formally define this embedding and prove several of its important properties. As we discuss in following chapters this embedding plays an important roles in many of our results on analyzing higher order eigenvalues of graphs.

Spectral embedding for finite graphs is easy to describe. Let $g_1, \ldots, g_k$ be orthonormal eigenfunctions of $\mathcal{L}$ corresponding to $\lambda_1, \ldots, \lambda_k$, and let $f_i = D^{-1/2} g_i$ for $1 \leq i \leq k$. Then the spectral embedding is the function $F \colon V \to \mathbb{R}^k$ defined by

$$F(v) := \big(f_1(v), f_2(v), \ldots, f_k(v)\big)\,. \tag{8.1.1}$$

In Figure 8.1.1 we plotted the spectral embedding of a cycle based on its first 3 eigenfunctions. Note that although the spectral embedding doesn't know the labeling of the vertices of the cycle, it can re-construct it in a 3 dimensional space such that no two edges cross each other.

This embedding satisfies interesting properties.

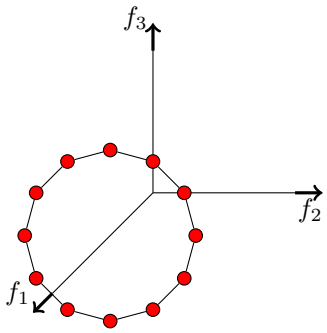

Figure 8.1.1: The Spectral Embedding of a cycle with respect to the first 3 eigenfunctions of the normalized Laplacian matrix.

Average Norm. For $S \subseteq V$, and a mapping $F : V \to \mathbb{R}^k$ we define the ℓ^2 mass of S as follows,

$$\mathcal{M}_F(S) := \sum_{v \in S} w(v) \left\| F(v) \right\|^2 .$$

If F is the spectral embedding, then

$$\mathcal{M}_F(V) = \sum_{v \in V} w(v) \left\| F(v) \right\|^2 = \sum_{v \in V} \sum_{i=1}^{k} w(v) f_i(v)^2 = \sum_{i=1}^{k} \left\| f_i \right\|_w^2 = k. \tag{8.1.2}$$

Therefore, the average norm of the vertices in the spectral embedding is $\sqrt{k/n \cdot w(v)}$.

Lemma 8.1.1. *For any set $S \subseteq V$, and a mapping $F : V \to \mathbb{R}^k$, there is a vertex $v \in S$, such that $\|F(v)\| \geq \sqrt{\frac{\mathcal{M}_F(S)}{n \cdot w(v)}}$.*

Proof. If there is no v satisfying the lemma, then

$$\mathcal{M}_F(S) = \sum_{v \in S} \left\| F(v) \right\|^2 w(v) < \sum_{v \in S} \mathcal{M}_F(S)/n \leq \mathcal{M}_F(S).$$

So we reach a contradiction. □

Isotropy. For a map $F : V \to \mathbb{R}^k$, we say F is isotropic if for any unit vector $\mathbf{x} \in \mathbb{R}^k$,

$$\sum_{v \in V} w(v) \langle \mathbf{x}, F(v) \rangle^2 = 1 . \tag{8.1.3}$$

In the next lemma we show that the spectral embedding is isotropic. This property shows that the mass after projection on a unit vector is 1. Consequently, since by equation (8.1.2) the sum of the norm of all vertices in the spectral embedding is exactly k, each direction in the space contributes

exactly the same amount to the overall ℓ^2 mass. In other words, it is impossible for the ℓ^2 mass of F to "concentrate" along fewer than k directions.

Lemma 8.1.2 (Isotropy)**.** *The spectral embedding $F : V \to \mathbb{R}^k$ is isotropic.*

Proof. The proof simply follows from the fact that the functions $f_1, \ldots, f_k$ are orthonormal in the space $\ell^2(V, w)$. Let $\mathbf{x} \in \mathbb{R}^k$ be a unit vector. Then,

$$\begin{aligned}\sum_{v\in V} w(v)\langle \mathbf{x}, F(v)\rangle^2 = \sum_{v\in V} w(v)\Big(\sum_{i=1}^{k} \mathbf{x}(i) f_i(v)\Big)^2 &= \sum_{1\le i,j\le k} \mathbf{x}(i)\mathbf{x}(j) \sum_{v\in V} w(v) f_i(v) f_j(v) \\ &= \sum_{1\le i,j\le v} \mathbf{x}(i)\mathbf{x}(j)\langle f_i, f_j\rangle_w = \|\mathbf{x}\|^2 = 1.\end{aligned}$$

□

Spreading. One of the important consequences of the isotropy property is the *spreading property*. Roughly speaking, spreading property implies that as $k \to \infty$, the vertices spread all over the space. More formally, in every small ball we have at most $1/k$ fraction of the total mass of the vertices. Let us define the *radial projection distance* function. Redial projection distance is an extended pseudo-metric on V: If $\|F(u)\|, \|F(v)\| > 0$, then

$$d_F(u, v) := \left\| \frac{F(u)}{\|F(u)\|} - \frac{F(v)}{\|F(v)\|} \right\| .$$

Otherwise, if $F(u) = F(v) = 0$, we put $d_F(u, v) := 0$, else $d_F(u, v) := 1$. Observe that a set with small diameter with respect to d_F corresponds to a cone in $\mathbb{R}^k$; see Figure 8.1.2. We do not motivate this distance function here but it is one of the key elements of the proofs in Chapter 10.

Lemma 8.1.3. *For any isotropic map $F : V \to \mathbb{R}^k$, and any set $S \subseteq V$,*

$$\mathcal{M}_F(S) \le \frac{1}{k(1 - \operatorname{diam}(S, d_F)^2)} \mathcal{M}_F(V).$$

Proof. For any vertex $v \in V$, let $\Gamma(v) := F(v)/\|F(v)\|$, and $\Gamma(v) = 0$ if $F(v) = 0$. Fix any $u \in S$ such that $F(u) \neq 0$ (note that if such a vertex does not exist then $\mathcal{M}_F(S) = 0$ and we are done). By Lemma 8.1.2,

$$\begin{aligned}1 = \sum_{v\in V} w(v)\,\langle F(v), \Gamma(u)\rangle^2 &= \sum_{v\in V} w(v)\,\|F(v)\|^2\,\langle \Gamma(v), \Gamma(u)\rangle^2 \\ &= \sum_{v\in V} w(v)\,\|F(v)\|^2\,(\|\Gamma(v)\|^2 + \|\Gamma(u)\|^2 - \|\Gamma(v) - \Gamma(u)\|^2)^2/4 \\ &= \sum_{v\in V} w(v)\|F(v)\|^2 \left(1 - \frac{d_F(u, v)^2}{2}\right)^2 \ge (1 - \operatorname{diam}(S, d_F)^2)\mathcal{M}_F(S)\,.\end{aligned}$$

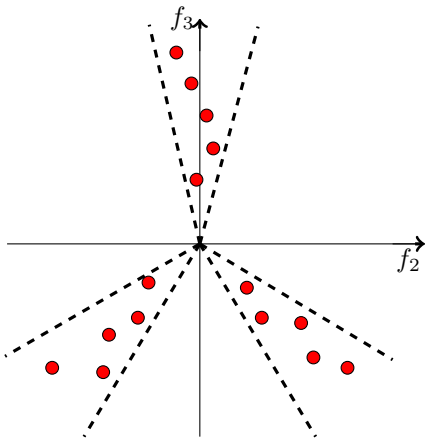

Figure 8.1.2: Each cone represents a subset of vertices with small radial projection distance.

where we used the identity $\|\Gamma(u) - \Gamma(v)\| = d_F(u, v)$. The lemma now follows by equation (8.1.2). □

We say an embedding, $F : V \to \mathbb{R}^k$, is (Δ, η)-*spreading (with respect to G)* if, for all subsets $S \subseteq V$, we have

$$\operatorname{diam}(S, d_F) \leq \Delta \implies \mathcal{M}_F(S) \leq \eta \cdot \mathcal{M}_F(V).$$

In the next corollary we show that for any small constant Δ any isotropic embedding is $(\Delta, O(1/k))$ spreading. It is worth noting that for this lemma we do not need to assume that F is a spectral embedding of G.

Corollary 8.1.4. *For any $\Delta > 0$, any isotropic map $F : V \to \mathbb{R}^k$ is $(\Delta, \frac{1}{k(1-\Delta^2)})$-spreading.*

The following corollary is a simple consequence of Lemma 8.1.3. It shows that any ball of radius $\alpha \|F(u)\|$, with respect to the Euclidean distance, around vertex a u, has at most $O(1/k)$-fraction of the total mass, for any constant $\alpha < 1/2$.

Corollary 8.1.5. *For any isotropic map $F : V \to \mathbb{R}^k$, and any vertex $u \in V$, and $r = \alpha \|F(u)\|$, for $\alpha > 0$,*

$$\mathcal{M}_F(B_F(u, r)) \leq \frac{1}{(1 - 2\alpha^2)^2}.$$

First, we prove the following lemma, which upper bounds the radial projection distance by the Euclidean distance.

Lemma 8.1.6. *For any $F : V \to \mathbb{R}^k$, and for all $u, v \in V$, we have $d_F(u, v) \|F(u)\| \leq 2 \|F(u) - F(v)\|$.*

Proof. For any non-zero functions $\mathbf{x}, \mathbf{y} \in \mathbb{R}^k$, we have

$$\|\mathbf{x}\| \left\| \frac{\mathbf{x}}{\|\mathbf{x}\|} - \frac{\mathbf{y}}{\|\mathbf{y}\|} \right\| = \left\| \mathbf{x} - \frac{\|\mathbf{x}\|}{\|\mathbf{y}\|} \mathbf{y} \right\| \leq \|\mathbf{x} - \mathbf{y}\| + \left\| \mathbf{y} - \frac{\|\mathbf{x}\|}{\|\mathbf{y}\|} \mathbf{y} \right\| \leq 2 \|\mathbf{x} - \mathbf{y}\|.$$

Now, if $F(u) \neq 0$ and $F(v) \neq 0$ we use above inequality. If $F(u) = 0$

$$d_F(u,v) \|F(u)\| = 0 \leq \|F(u) - F(v)\|$$

And, if $F(v) = 0$ and $F(u) \neq 0$, then

$$d_F(u,v) \|F(u)\| = \|F(u)\| \leq \|F(u) - F(v)\|.$$

□

Proof of Corollary 8.1.5. Without loss of generality, assume that $F(u) \neq 0$. For any vertex $v \in B_F(u,r)$,

$$d_F(u,v) \leq \frac{2\|F(u) - F(v)\|}{\|F(u)\|} = \frac{2\alpha \|F(u)\|}{\|F(u)\|} = 2\alpha$$

Thus by proof of Lemma 8.1.3, $\mathcal{M}_F(B(u,r)) \leq 1/(1-2\alpha^2)^2$. □

Energy. The *energy* of a map $F : V \to \mathbb{R}^k$ is defined as follows

$$\mathcal{E}_F := \sum_{(u,v)\in E} w(u,v) \|F(u) - F(v)\|^2.$$

It turns out that if F is the spectral embedding, then we can naturally relate $\mathcal{E}_F$ to the eigenvalues of $\mathcal{L}$.

Lemma 8.1.7. *The spectral embedding $F : V \to \mathbb{R}^k$ satisfies the following,*

$$\mathcal{E}_F = \sum_{i=1}^{k} \lambda_i, \quad \mathcal{R}(F) = \frac{\sum_{i=1}^{k} \lambda_i}{k} \leq \lambda_k.$$

Proof. By equation (7.2.2),

$$\mathcal{E}_F = \sum_{(u,v)\in E} \sum_{i=1}^{k} w(u,v)|f_i(u) - f_i(v)|^2 = \sum_{i=1}^{k} \sum (u,v) \in E|f_i(u) - f_i(v)|^2 = \sum_{i=1}^{k} \mathcal{R}(f_i) = \sum_{i=1}^{k} \lambda_i.$$

where we used the assumption that $\|f_i\|_w = 1$ for all $1 \leq i \leq k$. Consequently,

$$\mathcal{R}(F) = \frac{\mathcal{E}(F)}{\sum_{i=1}^{k} \|f_i\|_w^2} = \frac{\sum_{i=1}^{k} \lambda_i}{k} \leq \lambda_k.$$

□

The next corollary shows that not only spectral embedding has a small energy when $\lambda_k \to 0$, any projection of it have a small energy as well.

Corollary 8.1.8. *Let $F : V \to \mathbb{R}^k$ be the spectral embedding. Then, for any unit vector $\mathbf{x} \in \mathbb{R}^k$, the function $f \in \ell^2(V, w)$ defined as $f(v) := \langle \mathbf{x}, F(v) \rangle$ satisfies*

$$\mathcal{R}(f) \leq \lambda_k, \text{ and } \mathcal{E}_f \leq \lambda_k.$$

Proof. First observe that $f = \sum_{i=1}^k \mathbf{x}(i) f_i$. Therefore, $f \in \text{span}\{f_1, \ldots, f_k\}$. By equation (7.2.3). $\mathcal{R}(f) \leq \lambda_k$. On the other hand, since $\|\mathbf{x}\| = 1$, by Lemma 8.1.2, $\|f\|_w = 1$. Hence, $\mathcal{E}_f = \mathcal{R}(f) \leq \lambda_k$. □

A general theme in many of the proofs in this thesis is that we use geometric arguments to relate graph properties to energy of the spectral embedding, or vice versa. Then, we use Lemma 8.1.7 to relate $\mathcal{E}_F$ to the eigenvalues of $\mathcal{L}$. For example, if one can use graph properties to lower bound $\mathcal{E}_F$, this automatically gives lower bound on the eigenvalues of $\mathcal{L}$.

In addition to above observations, by variational principle we can show that the spectral embedding is an embedding that *minimizes* the energy among all isotropic embeddings. (Note that the embedding that only minimizes the energy is the one that maps every vertex to the same point in $\mathbb{R}^k$.)

Lemma 8.1.9. *For any isotropic map $F : V \to \mathbb{R}^k$,*

$$\mathcal{E}_F \geq \lambda_1 + \lambda_2 + \ldots + \lambda_k.$$

Proof. Let $g_1, \ldots, g_k \in \ell^2(V, w)$ where $g_i(v) := \langle F(v), \mathbf{1}_i \rangle$. We first show that $g_1, \ldots, g_k$ form an orthonormal basis and then we use equation (7.2.3) to prove the lemma. First, for any $1 \leq i \leq j$, by isotropy lemma we have,

$$1 = \sum_{v \in V} w(v) \langle \mathbf{1}_i, F(v) \rangle^2 = \sum_{v \in V} w(v) g_i(v)^2 = \|g_i\|_w^2 . \tag{8.1.4}$$

On the other hand, for any $i \neq j$, and $\mathbf{x} = \mathbf{1}_i/\sqrt{2} = \mathbf{1}_j/\sqrt{2}$.

$$\begin{aligned} 1 = \sum_{v \in V} w(v) \langle \mathbf{x}, F(v) \rangle^2 = \sum_{v \in V} w(v) (g_i(v) + g_j(v))^2/2 &= \|g_i\|_w^2/2 + \|g_j\|_w^2/2 + \langle g_i, g_j \rangle_w/2 \\ &= 1 + \langle g_i, g_j \rangle_w/2. \end{aligned}$$

where the last equality follows by equation (8.1.4). Therefore, $\langle g_i, g_j \rangle = 0$. So, $g_1, \ldots, g_k$ are orthonormal in $\ell^2(V, w)$. Finally, by variational principle,

$$\mathcal{E}_F = \sum_{i=1}^k \mathcal{R}(g_i) \geq \sum_{i=1}^k \lambda_i.$$

□

In the next lemma we develop a simple path argument to lower bound the energy of a mapping F.

Lemma 8.1.10. *Suppose that $w(u,v) \geq 1$ for all edges $(u,v) \in E$. For any map $F : V \to \mathbb{R}^k$ and any path $\mathcal{P} \subseteq V$ in G between vertices $u^*, v^* \in V$,*

$$\mathcal{E}_F(\mathcal{P}) \geq \frac{\|F(u^*) - F(v^*)\|^2}{|\mathcal{P}|}.$$

Proof. Let $\mathcal{P} = (v_0, v_1, v_2, \ldots, v_{l-1}, v_l)$, where $v_0 = u^*$ and $v_l = v^*$. Then by the Cauchy-Schwarz inequality, we have

$$\mathcal{E}_F(\mathcal{P}) \geq \sum_{i=0}^{l-1} \|F(v_i) - F(v_{i+1})\|^2 \geq \frac{1}{l}\left(\sum_{i=0}^{l-1} \|F(v_i) - F(v_{i+1})\|\right)^2 \geq \frac{1}{l}\|F(u^*) - F(v^*)\|^2,$$

where the first inequality uses the assumption that $w(u,v) \geq 1$ for all edges $(u,v) \in E$, the third inequality follows by the triangle inequality in Hilbert space. □

8.2 Beyond log(n) dimensions, Johnson-Lindenstrauss to log(k) dimensional space.

Suppose we have a mapping $F : V \to \mathbb{R}^k$, for some $k \in \mathbb{N}$. Johnson and Lindenstrauss celebrated result [JL84] show that for any $\epsilon > 0$ one can define another mapping $\Gamma : V \to \mathbb{R}^l$, for $l = O(\log(n)\,\mathrm{polylog}(1/\epsilon))$, such that for all $u, v \in V$,

$$(1-\epsilon)\|F(u) - F(v)\|^2 \leq \|\Gamma(u) - \Gamma(v)\|^2 \leq (1+\epsilon)\|F(u) - F(v)\|^2.$$

In other words we can approximately preserve all pairwise distances of vertices of G only in a $\log(n)$ dimensional space. The function Γ can be computed easily by computing l randomly chosen k dimensional Gaussian vectors $\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_l$ and letting

$$\Gamma(v) = \Gamma_{k,l} \circ F(v) = (\langle \boldsymbol{\zeta}_1, F(v)\rangle, \ldots, \langle \boldsymbol{\zeta}_l, F(v)\rangle).$$

This fundamental theorem has many applications in various fields including machine learning, compressed sensing, etc. It is worth noting that if we want to preserve all pairwise distances then we need at least $\Omega(\log n)$ dimensions.

It is known that if we have l different (global) functions of vertex pairwise distances then we can map the vertices to a $\Theta(\log(l))$ dimensional space making sure that the value of all of the functions are preserved (up to constant factors). For example, if we just want to preserve the Rayleigh quotient of F we can map the vertices to a constant dimensional space. Now, suppose F satisfies a global

property, in particular suppose F is (Δ, η)-spreading. What is the smallest number of dimensions necessary to preserve this property (up to constant factors)?

Our goal in this section is to answer this question. We show that we may bypass the $\log(n)$ barrier and only $O(\log(1/\eta))$ dimensions suffice. Note that there is no dependency to k or n. Furthermore, we show that indeed the mapping $\Gamma_{k,l} \circ F$ is $(\Omega(\Delta), O(\eta))$-spreading for $l = O(\log(1/\eta)$, i.e., the same idea of dimension reduction works, but in this case for a significantly smaller number of dimensions.

This observation plays an important role in finding the right dependency to k in higher order Cheeger's inequalities (equation (10.1.4)). We generally believe this new ideas of dimension reduction is one of the fundamental achievements of this thesis and we expect to see several applications or variants of it in the future.

Theorem 8.2.1. *Let $G = (V, E, w)$ be a weighted graph, and $F : V \to \mathbb{R}^k$ be (Δ, η)-spreading for $0 \leq \Delta \leq 1$ and $\eta < 1$. Then for some value*

$$l \lesssim \frac{1 + \log(\frac{1}{\eta}) + \log(41\Delta)}{\Delta^2},$$

the mapping $\Gamma_{k,l}$ as defined in equation (7.5.2) with probability at least $1/2$ satisfies both of the following conditions:

i) $\mathcal{R}_G(\Gamma_{k,l} \circ F) \leq 8 \cdot \mathcal{R}_G(F)$, and

ii) $\Gamma_{k,l} \circ F$ is $(\Delta/4, (1+\Delta)\eta)$-spreading with respect to G.

Proof. Although our proof is delicate the basic idea is this: If F is (Δ, η) spreading but $\Gamma_{k,l} \circ F$ fails to satisfy a this property (up to constant errors), then a $\gg \eta$ fraction of the ℓ^2 mass has to have moved significantly in the dimension reduction step, and such an event is unlikely for a random mapping into $O(\log(1/\eta))$ dimensions.

Let $\delta = \Delta/16$. Choose $l \asymp (1+\log(\frac{1}{\eta})+\log(\frac{1}{\Delta}))/\Delta^2$ large enough such that $2e^{-\delta^2 l/14} \leq \delta^2\eta^3/128$. Let $\Gamma = \Gamma_{k,l}$.

First, observe that equation (7.5.3) combined with Markov's inequality implies that the following holds with probability at least $3/4$,

$$\mathcal{E}_{\Gamma(F)} \leq 4\mathcal{E}_F. \tag{8.2.1}$$

Now define,

$$U := \{v \in V : \|\Gamma(F(v))\|^2 \in [(1-\delta)\|F(v)\|^2, (1+\delta)\|F(v)\|^2],$$

By equation (7.5.4), for each $v \in V$,

$$\mathbb{P}[v \notin U] \leq \delta\eta^3/128. \tag{8.2.2}$$

Next, we bound the amount of ℓ^2 mass that falls outside of U. By Markov's inequality, with probability at least 31/32, we have

$$\mathcal{M}_F(V-U) \leq \frac{\delta\eta^3}{4}\mathcal{M}_F(V). \tag{8.2.3}$$

In particular, with probability at least 31/32, we have

$$\mathcal{M}_{\Gamma(F)}(V) \geq (1-\delta)\mathcal{M}_F(U) \geq (1-2\delta)\mathcal{M}_F(V). \tag{8.2.4}$$

Combining our estimates for (8.2.1) and (8.2.4), we conclude that (i) holds with probability at least 23/32. Thus we can finish by showing that (ii) holds with probability at least 25/32. We first consider property (ii) for subsets of U.

Claim 8.2.2. *With probability at least 7/8, the following holds: equation (8.2.4) implies that, for any subset $S \subseteq U$ with $\operatorname{diam}(S, d_{\Gamma(F)}) \leq \Delta/4$, we have*

$$\mathcal{M}_{\Gamma(F)}(S) \leq (1+6\delta)\eta \cdot \mathcal{M}_{\Gamma(F)}(V)\,.$$

Proof. For every $u, v \in V$, define the event,

$$\mathcal{A}_{u,v} = \left\{d_{\Gamma(F)}(u,v) \in [d_F(u,v)(1-\delta) - 2\delta, d_F(u,v)(1+\delta) + 2\delta]\right\}$$

and let $I_{u,v}$ be the random variable indicating that $\mathcal{A}_{u,v}$ does *not* occur.

We claim that for $u, v \in V$, $\mathcal{A}_{u,v}$ occurs if $u, v \in U$, and

$$\left\|\Gamma\left(\frac{F(u)}{\|F(u)\|} - \frac{F(v)}{\|F(v)\|}\right)\right\| \in [(1-\delta)d_F(u,v), (1+\delta)d_F(u,v)]\,.$$

To see this, observe that,

$$\begin{aligned}
d_{\Gamma(F)}(u,v) &= \left\|\frac{\Gamma(F(u))}{\|\Gamma(F(u))\|} - \frac{\Gamma(F(v))}{\|\Gamma(F(v))\|}\right\| \\
&\geq \left\|\frac{\Gamma(F(u))}{\|F(u)\|} - \frac{\Gamma(F(v))}{\|F(v)\|}\right\| - \left\|\frac{\Gamma(F(u))}{\|\Gamma(F(u))\|} - \frac{\Gamma(F(u))}{\|F(u)\|}\right\| - \left\|\frac{\Gamma(F(v))}{\|\Gamma(F(v))\|} - \frac{\Gamma(F(v))}{\|F(v)\|}\right\| \\
&\geq \left\|\Gamma\left(\frac{F(u)}{\|F(u)\|} - \frac{F(v)}{\|F(v)\|}\right)\right\| - 2\delta \\
&\geq (1-\delta)d_F(u,v) - 2\delta,
\end{aligned}$$

where we have used the fact that Γ is a linear operator. The other direction can be proved similarly.

Therefore, by (7.5.4), (8.2.2), and a union bound, for any $u, v \in V$, $\mathbb{P}[I_{u,v}] \leq 3\delta\eta^3/128$. Let,

$$X := \sum_{u,v \in V} w(u)w(v)\|F(u)\|^2\|F(v)\|^2 I_{u,v}\,.$$

By linearity of expectation,

$$\mathbb{E}[X] \leq \frac{3\delta\eta^2}{128}\mathcal{M}_F(V)^2.$$

Therefore, by Markov's inequality, we conclude that

$$\mathbb{P}\left[X \geq \frac{\delta \cdot \eta^3}{4}\mathcal{M}_F(V)^2\right] \leq \frac{1}{8}\,. \tag{8.2.5}$$

Now suppose there exists a subset $S \subseteq U$ with $\operatorname{diam}(S, d_{\Gamma(F)}) \leq \Delta/4$ and

$$\mathcal{M}_{\Gamma(F)}(S) \geq (1+6\delta)\eta\mathcal{M}_{\Gamma(F)}(V)\,. \tag{8.2.6}$$

Fix a vertex $u \in S$. Since for every $v \in S - B_{d_F}(u, \Delta/2)$, we have $d_F(u,v) \geq \Delta/2$, $d_{\Gamma(F)}(u,v) \leq \Delta/4$, and recalling that $\delta = \Delta/16$, it must be that $I_{u,v} = 1$. On the other hand, we have

$$\begin{aligned}
\mathcal{M}_F(S - B_{d_F}(u,\Delta/2)) &\geq \mathcal{M}_F(S) - \mathcal{M}_F(B_{d_F}(u,\Delta/2)) \\
&\geq (1-\delta)\mathcal{M}_{\Gamma(F)}(S) - \eta\mathcal{M}_F(V) \\
&\overset{(8.2.6)}{\geq} (1-\delta)(1+6\delta)\eta\mathcal{M}_{\Gamma(F)}(V) - \eta\mathcal{M}_F(V) \\
&\overset{(8.2.4)}{\geq} [(1-2\delta)(1-\delta)(1+6\delta) - 1]\eta\mathcal{M}_F(V) \geq \delta\eta \cdot \mathcal{M}_F(V)\,.
\end{aligned}$$

where we have used the fact that $S \subseteq U$ and also $\operatorname{diam}(B_{d_F}(u,\Delta/2)) \leq \Delta$ and the fact that F is (Δ,η)-spreading. In last equation, we have used $\delta \leq \Delta/16 \leq 1/16$.

Thus under our assumption on the existence of S and again using $S \subseteq U$, we have

$$\begin{aligned}
X &\geq \mathcal{M}_F(S) \cdot \mathcal{M}_F(S - B_{d_F}(u,\Delta/2)) \\
&\geq \mathcal{M}_F(S) \cdot \delta\eta \cdot \mathcal{M}_F(V) \\
&\geq (1-\delta)\mathcal{M}_{\Gamma(F)}(S) \cdot \delta\eta \cdot \mathcal{M}_F(V) \\
&\overset{(8.2.6)}{\geq} (1+6\delta)\eta(1-\delta)\mathcal{M}_{\Gamma(F)}(V) \cdot \delta\eta \cdot \mathcal{M}_F(V) \\
&\overset{(8.2.4)}{\geq} \delta(1-\delta)(1-2\delta)(1+6\delta)\eta^2 \cdot \mathcal{M}_F(V)^2 \geq \delta\eta^2 \cdot \mathcal{M}_F(V)^2\,.
\end{aligned}$$

where the last inequality follows from $\delta \leq 1/16$. Combining this with (8.2.5) yields the claim. □

The preceding claim guarantees a spreading property for subsets $S \subseteq U$. Finally, we need to handle points outside U.

Claim 8.2.3. *With probability at least* 15/16*, we have*

$$\mathcal{M}_{\Gamma(F)}(V-U) \leq \delta\eta^3 \cdot \mathcal{M}_F(V).$$

Proof. Let $\mathcal{D}_u$ be the event that $u \notin U$, and let $H_u := \|\Gamma(F(u))\|^2 \cdot \mathbb{I}[\mathcal{D}_u]$. Then,

$$\mathbb{E}\left[\sum_{u\notin U} w(u)\|\Gamma(F(u))\|^2\right] = \sum_{u\in V} w(u)\mathbb{E}[H_u]\,. \tag{8.2.7}$$

Now we can estimate,

$$\frac{\mathbb{E}[H_u]}{\|F(u)\|^2} \leq 2\,\mathbb{P}[\mathcal{D}_u] + \mathbb{P}\left[\frac{\|\Gamma(F(u))\|^2}{\|F(u)\|^2} > 2\right]\cdot\mathbb{E}\left[\frac{\|\Gamma(F(u))\|^2}{\|F(u)\|^2}\,\middle|\,\|\Gamma(F(u))\|^2 > 2\,\|F(u)\|^2\right]. \tag{8.2.8}$$

Using the inequality, valid for all non-negative X,

$$\mathbb{P}[X > r_0]\cdot\mathbb{E}[X \mid X > r_0] \leq \int_{r_0}^{\infty} r\cdot\mathbb{P}[X > r]\,dr\,,$$

we can bound the latter term in (8.2.8) by,

$$\int_2^\infty r\cdot\mathbb{P}\left[\|\Gamma(F(u))\|^2\,dr > r\|F(u)\|^2\right]\,dr \leq \int_2^\infty r\cdot e^{-rl/14}\,dr = \left(\frac{28}{l}+\frac{196}{l^2}\right)e^{-l/7} \leq \frac{\delta\eta^3}{128}\,,$$

where we have used (7.5.5) and the initial choice of l sufficiently large.

It follows from this, (8.2.8), and (8.2.2), that

$$\mathbb{E}[H_u] \leq \frac{3\delta\eta^3}{128}\|F(u)\|^2\,.$$

Therefore, by equation (8.2.7) and Markov's inequality,

$$\mathbb{P}\left[\mathcal{M}_{\Gamma(F)}(V-U) > \delta\eta^3\cdot\mathcal{M}_F(V)\right] \leq \frac{3}{128}\,,$$

completing the proof. □

To conclude the proof of the lemma, we need to verify that (ii) holds with probability at least 25/32. But observe that if (8.2.4) holds, then the conclusion of the preceding claim is,

$$\mathcal{M}_{\Gamma(F)}(V-U) \leq \delta\eta^3\cdot\mathcal{M}_F(V) \leq 2\delta\eta^3\cdot\mathcal{M}_{\Gamma(F)}(V).$$

Combining this with Claim 8.2.2 shows that with probability at least 25/32, $\Gamma\circ F$ is $(\Delta/4, (1+8\delta)\eta)$-spreading, completing the proof. □

8.3 Improved Lower Bounds on Escape Probability

In this section we prove a new technical result on the escape probability of simple random walks. We show that for any $S \subseteq V$ a t-step lazy random walk started at a random (chosen proportional to degree) vertex of S remains entirely in S with probability at least $(1-\phi(S)/2)^t$. Previously, only the lower bound $1 - t\phi(S)/2$ was known (see e.g. [ST13]).

For comparison, when $t = 1/\phi(S)$, the known bound would imply that the walk has probability at least $1/2$ of being entirely contained in S, with no guarantee being available in the case $t = 2/\phi(S)$, while our bound implies that for $t = (\epsilon \log n)/\phi$ the probability of being entirely contained in S is still at least $1/n^{\epsilon}$. As we discuss in Chapter 12 this technical result is our main idea in designing an improved local graph clustering algorithm. As a simple application we in Subsection 8.3.1 use our bound to prove stronger lower bounds on the mixing time of reversible Markov Chains.

Let X_t be the random variable indicating the t^{th} step of the lazy simple random walk on G started from $v \in V$. For $S \subseteq V$, let $\pi_S(.)$ be the following distribution,

$$\pi_S(v) := \begin{cases} w(v)/\mathrm{vol}(S) & \text{if } v \in S, \\ 0 & \text{otherwise.} \end{cases}$$

For $v \in V$, and integer $t > 0$, we write

$$\mathrm{esc}(v, t, S) := \mathbb{P}_{X_0 = v}\left[\cup_{i=0}^{t} X_i \notin S\right]$$

to denote the probability that the random walk started at v leaves S in the first t steps, and $\mathrm{rem}(v, t, S) := 1 - \mathrm{esc}(v, t, S)$ as the probability that the walk stays entirely inside S.

Proposition 8.3.1. *For any set $S \subseteq V$, and integer $t > 0$,*

$$\mathbb{E}_{v \sim \pi_S}\left[\mathrm{rem}(v, t, S)\right] \geq \left(1 - \frac{\phi(S)}{2}\right) \mathbb{E}_{v \sim \pi_S}\left[\mathrm{rem}(v, t-1, S)\right] \geq \ldots \geq \left(1 - \frac{\phi(S)}{2}\right)^t. \tag{8.3.1}$$

Furthermore, there is a subset $S' \subseteq S$, such that $\mathrm{vol}(S') \geq \mathrm{vol}(S)/2$, and for all $v \in S'$

$$\mathrm{rem}(v, t, S) \geq \frac{1}{200}\left(1 - \frac{3\phi(S)}{2}\right)^t. \tag{8.3.2}$$

We remark that the second statement does not follow from a simple application of the Markov inequality to the first statement, as this is the case in [ST13]. Whence, here both of the results incorporate non-trivial spectral arguments. Alternate proofs of above proposition are recently shown by O'Donnell and Witmer [OW12].

Proof. Recall that P is the transition probability operator of the lazy random walk (see Section 7.4).

Let I_S be the identity operator on S, i.e., for any $f : V \to \mathbb{R}$, $I_S f(v) = f(v)$ if $v \in S$ and $I_S f(v) = 0$ otherwise. First, observe that

$$\mathrm{rem}(v, t, S) = \langle (I_S P^T I_S)^t \mathbf{1}_v, \mathbf{1}_S \rangle = \langle \mathbf{1}_v, (I_S P I_S)^t \mathbf{1}_S \rangle.$$

So,

$$\mathbb{E}_{v \sim \pi_S} [\mathrm{rem}(v, t, S)] = \langle \pi_S, (I_S P I_S)^t \mathbf{1}_S \rangle \tag{8.3.3}$$

Therefore, by a simple induction on t, (8.3.1) is equivalent to the following equation,

$$\langle \pi_S, (I_S P I_S)^t \mathbf{1}_S \rangle \geq (1 - \phi(S)/2) \langle \pi_S, (I_S P I_S)^{t-1} \mathbf{1}_S \rangle. \tag{8.3.4}$$

Let $Q := D^{1/2} I_S P I_S D^{-1/2}$, and let $\sqrt{\pi_S} : V \to \mathbb{R}$, where $\sqrt{\pi_S}(v) = \sqrt{\pi_S(v)}$. First we show that (8.3.4) is equivalent to the following equation:

$$\langle \sqrt{\pi_S}, Q^t \sqrt{\pi_S} \rangle \geq \langle \sqrt{\pi_S}, Q \sqrt{\pi_S} \rangle \cdot \langle \sqrt{\pi_S}, Q^{t-1} \sqrt{\pi_S} \rangle. \tag{8.3.5}$$

Then, we use Lemma 8.3.2 to prove the above equation. First observe that by the definition of Q, for any $t > 0$,

$$\langle \pi_S, (I_S P I_S)^t \mathbf{1}_S \rangle = \langle \pi_S, D^{-1/2} Q^t D^{1/2} \mathbf{1}_S \rangle = \langle D^{-1/2} \pi_S, Q^t D^{1/2} \mathbf{1}_S \rangle = \langle \sqrt{\pi_S}, Q^t \sqrt{\pi_S} \rangle \tag{8.3.6}$$

On the other hand,

$$\begin{aligned}
\langle \pi_S, (I_S P I_S) \mathbf{1}_S \rangle &= \left\langle \pi_S, \frac{1}{2}(D^{-1}A + I)\mathbf{1}_S \right\rangle \\
&= \frac{1}{2}\langle \pi_S, D^{-1}A\mathbf{1}_S \rangle + \frac{1}{2}\langle \pi_S, \mathbf{1}_S \rangle \\
&= \frac{1}{2\mathrm{vol}(S)}\langle \mathbf{1}_S, A\mathbf{1}_S \rangle + \frac{1}{2} \\
&= \frac{1}{2\mathrm{vol}(S)} w(S) + \frac{1}{2} \\
&= \frac{1}{2\mathrm{vol}(S)}(\mathrm{vol}(S) - w(S, \overline{S})) + \frac{1}{2} \\
&= 1 - \phi(S)/2.
\end{aligned} \tag{8.3.7}$$

where we used $w(S) = \sum_{u,v \in S} w(u, v)$. Equation (8.3.5) is equivalent to equation (8.3.4) using (8.3.6), (8.3.7). Next we prove equation (8.3.5) using Lemma 8.3.2. First observe that $\sqrt{\pi_S}$ is a norm one vector. On the other hand, by definition

$$Q = D^{1/2} I_S P I_S D^{-1/2} = \frac{1}{2} D^{1/2} I_S (D^{-1}A + I) I_S D^{-1/2} = \frac{1}{2} I_S (D^{-1/2} A D^{-1/2} + I) I_S = I_S (I - \mathcal{L}/2) I_S$$

is a symmetric matrix. On the other hand, for any $f : V \to \mathbb{R}$,

$$\frac{\langle Qf, f\rangle}{\langle f, f\rangle} = \frac{\langle (I - \mathcal{L}/2) f_S, f_S\rangle}{\langle f_S, f_S\rangle} \geq 0$$

where f_S is the restriction of f to S, and in the last equation we used the fact that the largest eigenvalue of $\mathcal{L}$ is at most 2. Thus, by Corollary 7.1.6 Q is positive semidefinite. This completes the proof of (8.3.1).

It remains to prove (8.3.2). We prove it by showing that for any set $T_1 \subseteq S$, of volume $\mathrm{vol}(T_1) \geq \mathrm{vol}(S)/2$, the random walk started at a randomly (proportional to degree) chosen vertex of T_1, remains in T_1 (and S), with probability at least $\frac{1}{200}(1 - 3\phi(S)/2)^t$,

$$\mathbb{E}_{v \sim \pi_{T_1}} \left[\mathrm{rem}(v, t, T_1)\right] = \langle \pi_{T_1}, (I_S P I_S)^t \mathbf{1}_{T_1}\rangle \geq \frac{1}{200}\left(1 - \frac{3\phi(S)}{2}\right)^t. \tag{8.3.8}$$

Therefore, in any such set T_1, there is a vertex that satisfy (8.3.2), so the volume of the set of vertices that satisfy (8.3.2) is at least half of $\mathrm{vol}(S)$.

Using equations (8.3.6) and (8.3.7), (8.3.8) is equivalent to the following equation,

$$\langle \sqrt{\pi_{T_1}}, Q^t \sqrt{\pi_{T_1}}\rangle \geq \frac{1}{200}\left(3\langle \sqrt{\pi_S}, Q\sqrt{\pi_S}\rangle - 2\right)^t. \tag{8.3.9}$$

We prove the above equation using Lemma 8.3.5. Let $T_2 = S - T_1$, and define

$$\begin{aligned} f_{T_1} &:= I_{T_1}\sqrt{\pi_S} = \sqrt{\mathrm{vol}(T_1)\pi_{T_1}/\mathrm{vol}(S)} \\ f_{T_2} &:= I_{T_2}\sqrt{\pi_S} = \sqrt{\mathrm{vol}(T_2)\pi_{T_2}/\mathrm{vol}(S)} \end{aligned} \tag{8.3.10}$$

Since $T_1 \cap T_2 = \emptyset$, $\langle f_{T_1}, f_{T_2}\rangle = 0$, and $\|f_{T_1} + f_{T_2}\| = \left\|\sqrt{\pi_S}\right\| = 1$. Furthermore, since $\mathrm{vol}(T_1) \geq \mathrm{vol}(S)/2 \geq \mathrm{vol}(T_2)$, $\|f_{T_1}\| \geq \|f_{T_2}\|$. Therefore, Q, f_{T_1}, f_{T_2} satisfy the requirements of Lemma 8.3.5. Finally, since

$$\langle \sqrt{\pi_{T_1}}, Q^t \sqrt{\pi_{T_1}}\rangle \geq \langle f_{T_1}, Q^t f_{T_1}\rangle,$$

(8.3.8) follows from Lemma 8.3.5. This completes the proof of Proposition 8.3.1. □

Lemma 8.3.2. *Let $Q \in \mathbb{R}^{V \times V}$ be a symmetric positive semidefinite matrix. Then, for any $f : V \to \mathbb{R}$ with norm $\|f\| = 1$, and integer $t > 0$,*

$$\langle Q^t f, f\rangle \geq \langle Q^{t-1} f, f\rangle \langle Qf, f\rangle \geq \ldots \geq \langle Qf, f\rangle^t.$$

Proof. Since all of the inequalities in lemma's statement follow from the first inequality, we only prove the first inequality. Let $f_1, f_2, \ldots, f_n$ be the set of orthonormal eigenfunctions of Q with the

corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. For any integer $t \geq 1$, by Corollary 7.1.4 and Lemma 7.1.7,

$$\langle Q^t f, f \rangle = \sum_{i=1}^{n} \langle f, f_i \rangle^2 \lambda_i^t. \tag{8.3.11}$$

On the other hand, since $\{f_1, \ldots, f_n\}$ is an orthornormal system, we have

$$\sum_{i=1}^{n} \langle f, f_i \rangle^2 = \|f\|^2 = 1.$$

For any $t > 0$, let $g_t : \mathbb{R} \to \mathbb{R}$ such that $g_t(\lambda) = \lambda^t$; it follows that,

$$\sum_{i} \langle f, f_i \rangle^2 \lambda_i^t = \mathbb{E}_{\Lambda \sim \mathcal{D}} \left[g_t(\Lambda) \right],$$

where $\mathbb{P}_{\Lambda \sim \mathcal{D}} \left[\Lambda = \lambda_i \right] = \langle f, f_i \rangle^2$. Using equation (8.3.11) we may rewrite the lemma's statement as follows,

$$\mathbb{E}_{\mathcal{D}} \left[g_{t-1}(\Lambda) g_1(\Lambda) \right] \geq \mathbb{E}_{\mathcal{D}} \left[g_{t-1}(\Lambda) \right] \mathbb{E}_{\mathcal{D}} \left[g_1(\Lambda) \right]$$

Since Q is positive semidefinite, for all $t > 0$, $g_t(.)$ is increasing in the support of $\mathcal{D}$. The above inequality follows from the Chebyshev's sum inequality, see Fact 8.3.3 below. □

Fact 8.3.3 (Chebyshev's Sum inequality)**.** *Let $a_1 \geq a_2 \geq \ldots \geq a_n$, and $b_1 \geq b_2 \geq \ldots \geq b_n$. Then, for any probability distribution $\mathcal{D}$ defined on $1, 2, \ldots, n$*

$$\mathbb{E}_{i \sim \mathcal{D}} \left[a_i \cdot b_i \right] \geq \mathbb{E}_{i \sim \mathcal{D}} \left[a_i \right] \cdot \mathbb{E}_{i \sim \mathcal{D}} \left[b_i \right].$$

We remark that variants of Lemma 8.3.2 is previously studied in the literature. For example, the following lemma is proved in [MS59, BR65] (see also [Lon66] for various generalizations).

Lemma 8.3.4 ([MS59, BR65])**.** *For any non-negative symmetric matrix $Q \in \mathbb{R}^{V \times V}$, and any non-negative unit norm $f : V \to \mathbb{R}$, and any integer $t > 0$,*

$$\langle Q^t f, f \rangle \geq \langle Qf, f \rangle^t.$$

Lemma 8.3.5. *Let $Q \in \mathbb{R}^{n \times n}$ be a symmetric positive semidefinite matrix such that all eigenvalues are at most 1, and $f, g \in \mathbb{R}^n$ such that $\langle f, g \rangle = 0$, $\|f + g\| = 1$, and $\|f\| \geq \|g\|$. Then, for any integer $t > 0$,*

$$\langle Q^t f, f \rangle \geq \frac{1}{200} \left(3 \langle Q(f+g), (f+g) \rangle - 2 \right)^t$$

Proof. Let $h := f + g$. Since f is orthogonal to g, we have $\|g\|^2 \leq 1/2 \leq \|f\|^2$. Let $f_1, f_2, \ldots, f_n$ be the set of orthonormal eigenfunctions of Q with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Let

$\alpha > 0$ be a constant that will be fixed later in the proof. Define $B := \{i : |\langle f, f_i\rangle| \geq \alpha|\langle g, f_i\rangle|\}$. First observe that,

$$\langle Q^t f, f\rangle = \sum_{i=1}^{n}\langle f, f_i\rangle^2\lambda_i^t \geq \sum_{i\in B}\langle f, f_i\rangle^2\lambda_i^t \geq \frac{1}{(1+1/\alpha)^2}\sum_{i\in B}\langle h, f_i\rangle^2\lambda_i^t, \tag{8.3.12}$$

where the equality follows from equation (8.3.11), the first inequality uses the assumption that Q is positive semidefinite, and the last inequality follows from the definition of B, that is for any $i \in B$, $\langle f, f_i\rangle^2 \geq (\langle h, f_i\rangle/(1+1/\alpha))^2$. Let $s := \sum_{i\in B}\langle h, f_i\rangle^2$. First, we lower bound s by a function of α,

$$s = \sum_{i\in B}\langle h, f_i\rangle^2 = 1 - \sum_{i\notin B}\langle h, f_i\rangle^2 \geq 1 - (1+\alpha)^2\left\|g\right\|^2 \geq \frac{1-\alpha^2-2\alpha}{2}. \tag{8.3.13}$$

The last equation follows by the fact that $\left\|g\right\|^2 \leq 1/2$. On the other hand, since Q is PSD, by Jensen's inequality,

$$\begin{aligned}\frac{1}{s}\sum_{i\in B}\langle h, f_i\rangle^2\lambda_i^t &\geq \left(\frac{1}{s}\sum_{i\in B}\langle h, f_i\rangle^2\lambda_i\right)^t \\ &\geq \left(\frac{\sum_{i=1}^{n}\langle h, f_i\rangle^2\lambda_i - (1-s)}{s}\right)^t \\ &\geq \left(1 - \frac{1-\langle Qh, h\rangle}{(1-\alpha^2-2\alpha)/2}\right)^t,\end{aligned} \tag{8.3.14}$$

where the second inequality follows by the assumptions that $\max_{1\leq i\leq n}\lambda_i \leq 1$, and that $\|h\| = 1$, and the last inequality follows by (8.3.13) and that $\langle Qh, h\rangle \leq 1$. Putting equations (8.3.12) and (8.3.14), and letting $\alpha = 0.154$ we get,

$$\langle Q^t f, f\rangle \geq \frac{1-\alpha^2-2\alpha}{2(1+1/\alpha)^2}\left(1 - \frac{1-\langle Qh, h\rangle}{(1-\alpha^2-2\alpha)/2}\right)^t \geq \frac{1}{200}\left(3\langle Qh, h\rangle - 2\right))^t$$

□

8.3.1 Lower Bounds on Uniform Mixing Time of Random Walks

In this short section we prove lower bounds on the mixing time of reversible Markov Chains. Since any reversible finite state Markov Chain can be realized as a random walk on a weighted undirected graph, for simplicity of notations, we model the Markov Chain as a random walk on a weighted graph G. See Section 7.4 for background on random walks and mixing time.

Recall that the ϵ-*uniform* mixing time of the chain is defined as

$$\tau_\infty(\epsilon) := \min\left\{t : \left|\frac{P^t(u,v)}{\pi(v)} - 1\right| \leq \epsilon, \forall u, v \in V\right\}. \tag{8.3.15}$$

We remark that that the uniform mixing time can be considerably larger than the mixing time in total variation distance. The *bottleneck ratio* uses $\phi(G)$ to provide lower bound on the mixing time of random walks (see e.g. [LPW06, Section 7.2]),. It shows that for any graph G

$$\tau_\infty(1/4) \geq \tau_1(1/4) \geq \frac{1}{4\phi(G)}.$$

Noting equation (7.8.1), there is a $O(\phi(G)\log(\pi_{\min}))$ gap between the upper bound and lower bound of $\tau_\infty(.)$. The $\phi(G)$ gap is because of the square root loss between λ_2 and $\phi(G)$ in Cheeger's inequality and it is unavoidable.

In the next proposition we prove stronger lower bounds on the uniform mixing time of any reversible Markov Chain. Our result shows that if small sets expand, then the $\log(n)$ gap between upper bound and lower bound of $\tau_\infty(.)$ can be recovered.

Proposition 8.3.6. *For any (weighted) graph $G = (V, E)$, any $S \subseteq V$ with $\mathrm{vol}(S) \leq \mathrm{vol}(V)/2$, and $0 < \epsilon < 1$, if $\phi(S) \leq 0.7$, then*

$$\tau(\epsilon) \geq \frac{\log(\mathrm{vol}(V)/2\mathrm{vol}(S))}{2\phi(S)} - 1.$$

Proof. Let $t \geq -\log(2\pi(S))/2\phi(S) - 1$ be an integer. Since the random walk is not necessarily a lazy walk, P is not necessarily a PSD operator, so we cannot directly apply Proposition 8.3.1. Instead we can use Lemma 8.3.4 that does not need a PSD assumption. So, for $Q = D^{1/2} I_S P I_S D^{-1/2}$,

$$\mathbb{E}_{v\sim\pi_S}\left[\mathrm{rem}(v,t,S)\right] = \langle Q^t\sqrt{\pi_S}, \sqrt{\pi_S}\rangle \geq \langle Q\sqrt{\pi_S}, \sqrt{\pi_S}\rangle^t = (1-\phi(S))^t$$

where the first equality follows from equations (8.3.3) and (8.3.6), and the last equality follows by (8.3.7). Thus, there exists a vertex $u \in S$ such that

$$\mathrm{rem}(u,t,S) \geq (1-\phi(S))^t \geq 2\pi(S).$$

where in the last inequality we used the assumption that $\phi(S) \leq 0.7$. Since $\mathbb{P}_{X_0\sim u}[X_t \in S] \geq \mathrm{rem}(u,t,S)$, there is a vertex $v \in S$ such that,

$$\frac{P^t(u,v)}{\pi(v)} \geq \frac{\mathbb{P}_{X_0\sim u}[X_t \in S]}{\pi(S)} \geq \frac{2\pi(S)}{\pi(S)} = 2,$$

where the first inequality uses $\mathbb{P}_{X_0\sim u}[X_t \in S] = \sum_{v\in S} P^t(u,v)$. Therefore, $\frac{|P^t(u,v)-\pi(v)|}{\pi(v)} \geq 1$, and

by equation (8.3.15), for any $\epsilon < 1$, $\tau(\epsilon) \geq t$. □

We remark that the above bound only holds for the uniform mixing time, and it can provide much stronger lower bound than the bottleneck ratio, if $d(S) \ll \mathrm{vol}(V)$.

Chapter 9

Universal Bounds on Laplacian Eigenvalues

In this chapter, we use spectral embedding to provide a unifying framework for bounding all the eigenvalues of graphs. For example, we show that for any finite graph with n vertices and all $k \geq 2$, the kth largest eigenvalue is at most $1 - \Omega(k^3/n^3)$, which extends the only other such result known, which is for $k = 2$ only and is due to [LO81]. This upper bound improves to $1 - \Omega(k^2/n^2)$ if the graph is regular. We generalize these results, and we provide sharp bounds on the spectral measure of various classes of graphs, including vertex-transitive graphs and infinite graphs, in terms of specific graph parameters like the volume growth.

As a consequence, using the entire spectrum, we can provide (improved) upper bounds on the return probabilities and mixing time of random walks with considerably shorter and more direct proofs. Our work introduces spectral embedding as a new tool in analyzing reversible Markov chains. We also provide several algorithmic applications.

The results of this chapter are based on a joint work with Russell Lyons in [LO12]. Many of our results extend naturally to infinite graphs. But, for the sake of accessibility, in this chapter we only work with finite graphs. For reader who are interested in the extension to infinite graphs we refer them to [LO12].

9.1 introduction

There have been a great many papers that upper bound the return probability or the mixing time of random walks. It is known that return probabilities are closely related to the spectrum of the normalized Laplacian. Therefore, once we can control the eigenvalues, we can reproduce, or even improve, bounds on return probabilities (see e.g. Corollary 9.5.5 below). We prove universal lower

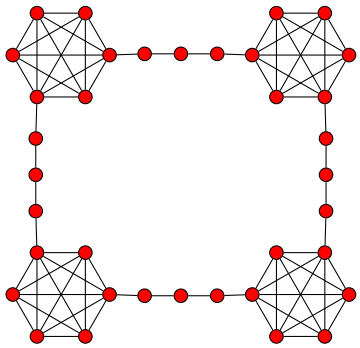

Figure 9.1.1: An example of a graph where $\lambda_k = \Omega(k^3/n^3)$. In this graph each clique has size $\Theta(n/k)$ and cliques are connected by paths of length $\Theta(n/k)$.

bounds on eigenvalues of the normalized Laplacian operator, equivalently, universal upper bounds on the eigenvalues of the random walk matrix. The usual methods for obtaining such bounds involve indirect methods from functional analysis. By contrast, our method is direct, which leads to very short proofs, as well as to improved bounds. By Lemma 8.1.7, all we need to do is to bound from below the energy of an isotropic embedding. We use simple properties of Hilbert spaces, as well as underlying properties of G, to achieve this goal. Our work thus introduces spectral embedding as a new tool in analyzing reversible Markov chains.

Our main contributions are the following results, all of which we believe to be new, as well as the technique used to establish them. The sharpness of these results (up to a constant factor) is discussed briefly here and in more detail in the body of the paper.

Theorem 9.1.1. *For every finite, unweighted, connected graph G, and $k \geq 1$,*

$$\lambda_k \geq \frac{(k-1)^3}{3200n^3}\,.$$

For each k, this is sharp up to a constant factor as shown by the following example: We may assume that $k < n/6$. Let G consist of k cliques of size $\sim 2n/(3k)$ joined in a cycle by paths of length $\sim n/(3k)$ (see Figure 9.1.1 for an illustration). For each $i = 1, \ldots, k$, define f_i to be the function that is 1 on the i-th clique and goes to 0 linearly on each of the paths leaving that clique, reaching 0 at the midpoint. It is straightforward to calculate that $\mathcal{R}(f_i) \sim 27k^3/n^3$. Since f_i's are disjointly supported, by Lemma 7.2.1, $\lambda_k \leq 54k^3/n^3$.

With the additional hypothesis of regularity, above result can improved to $\Omega\big((k-1)^2/n^2\big)$. In fact, only a bound for the ratio of the maximum degree to the minimum degree is needed.

Theorem 9.1.2. *For every unweighted, connected, regular graph G and for $1 \leq k \leq n$, we have*

$$\lambda_k \geq \frac{(k-1)^2}{100n^2}\,.$$

This result is evidently sharp as shown by the example of a cycle, which also shows sharpness of the next result.

Above results can be employed to bound $\tau_\infty(1/4)$ for unweighted (regular) graphs. For example, it implies that $\tau_\infty(1/4) \leq 8n^3$ for any graph G, and $\tau_\infty(1/4) \leq 24n^2$ for regular graphs. These results have been known implicitly in the sense that they can be deduced from known results in probability theory. So, we do not cover the proofs here and we refer the readers to [LO12] for detailed proof.

Finally, the case of transitive graphs is especially interesting and especially well studied, yet, to the best of our knowledge, the following theorem has not been proved in this generality.

An *automorphism* of a graph G is a permutation $\sigma : V \to V$ such that for any edge $(u, v) \in E$, we have $(\sigma(u), \sigma(v)) \in E$. A graph G is *vertex-transitive*, if for every two vertices $u, v \in V$, there is an automorphism $\sigma : V \to V$ such that $\sigma(u) = v$. For example, observe that any vertex-transitive graphs is regular. For a vertex $v \in G$ and $r \geq 0$, let $N(v, r)$ be the number of vertices whose shortest path distance to v is at most r. If G is vertex transitive, then, by definition, $N(u, r) = N(v, r)$ for every two vertices $u, v \in V$. Therefore, we may drop the index v and use $N(r)$. Also for $s \geq 0$, we let

$$N^{-1}(s) := \sup\{r \geq 0 : N(r) \leq s\}.$$

Theorem 9.1.3. *For every connected, unweighted, vertex-transitive graph G of degree w, and any $k \geq 2$,*

$$\frac{8w}{(N^{-1}(4n/(k-1)))^2} \leq \lambda_k$$

As an example, suppose that G is a $n = l \times l$ torus as defined in Subsection 7.3.2. Then, since G is vertex-transitive we can apply the above theorem with $w = 2$, Since $N^{-1}(4n/(k-1)) = \Theta(\sqrt{n/k})$, we get $\lambda_k = \Omega(k/n)$. Noting the spectrum of the torus as proved in Subsection 7.3.2, shows that the theorem is tight up to constant factors,

We provide some algorithmic applications of the above results in Section 9.5. We design a slightly sub-exponential time approximation algorithm to the uniform sparsest cut problem on unweighted graph that beats the [ARV09] $\sqrt{\log(n)}$ approximation factor. Furthermore, building on [Lyo05b], we design a local algorithm to approximate the number of spanning trees of massive graphs. Also, We believe the results and the techniques can be generalized and used in several areas

We prove Theorem 9.1.2 in Section 9.2, Theorem 9.1.1 in Section 9.3, Theorem 9.1.3 in Section 9.4. We describe some algorithmic applications in Section 9.5.

9.1.1 Related Works

There have been many studies bounding from above the eigenvalues of the (normalized) Laplacian (equivalently, bounding the eigenvalues of the (normalized) adjacency matrix from below). For example, Kelner et al. [KLPT11] show that for n-vertex, bounded-degree planar graphs, one has that $\lambda_k = O(k/n)$.

However, to the best of our knowledge, universal lower bounds were known only for the second smallest eigenvalue of the normalized Laplacian. Namely, Landau and Odlyzko [LO81] showed that for every simple connected graph of size n, $\lambda_2 \geq 1/n^3$.

On the other hand, there have also been a great many papers that bound from above the return probabilities of random walks, both on finite and infinite graphs. Such bounds correspond to lower bounds on eigenvalues. In fact, the asymptotics of large-time return probabilities correspond the behavior of the smallest eigenvalues.

Our methods would work as well for the eigenvalues $\tilde{\lambda}_k$ of combinatorial Laplacian L. In this case, [Fri96] has determined the minimum of $\tilde{\lambda}_k$ for each k over all unweighted n-vertex graphs. As noted there, his bound implies that $\tilde{\lambda}_k = \Omega(k^2/n^2)$; this immediately implies that $\lambda_k = \Omega(k^2/n^3)$ by comparison of Rayleigh quotients, but this is not sharp, as indicated by Theorem 9.1.1.

9.1.2 Notations

For a vertex $u \in V$, we use $\mathrm{nei}(u) := \{v : (u,v) \in E\}$ to denote the neighbors of u in G. In this chapter we assume that G is a connected unweighted graph, $w(u,v) = 1$ for all $(u,v) \in E$.

We say an embedding $F : V \to \mathbb{R}^k$ is *non-trivial*, if there is $v \in V$ such that $F(v) \neq \mathbf{0}$. We say F is centered if

$$\sum_{v \in V} w(v) F(v) = 0.$$

In next lemma we show if F is centered, then every ball $B_F(v, \|F(v)\|)$ does not cover at least one vertex of G.

Lemma 9.1.4. *For every non-trivial centered embedding $F\colon V \to \mathcal{H}$ of a finite graph G, and for any $u \in V$,*

$$B_F\big(u, \|F(u)\|\big) \neq V.$$

Proof. If $F(u) = 0$, the statement follows since F is a non-trivial embedding. So assume $\|F(u)\| \neq 0$. For the sake of contradiction, suppose $B_F\big(u, \|F(u)\|\big) = V$. Then for every vertex $v \in V$, we have $\big\langle F(v), F(u)\big\rangle \geq 0$. Since F is centered, we have

$$0 = \sum_{v \in V} w(v) \big\langle F(u), F(v)\big\rangle \geq w(u) \big\langle F(u), F(u)\big\rangle > 0\,,$$

a contradiction. □

For technical reasons in this chapter we slightly change the definition of spectral embedding, and we remove the coordinate corresponding to f_1. So, in all of this chapter we let $F : V \to \mathbb{R}^{k-1}$,

$$F(v) = (f_2(v), \dots, f_k(v)).$$

Recall that $f_1, \ldots, f_k$ form an orthonormal system in $\ell^2(V, w)$, and $\mathcal{R}(f_i) = \lambda_i$.

The consequence of this definition is that spectral embedding is a centered embedding. This is because for any $2 \le i \le k$,

$$\left\langle \sum_{v \in V} F(v)w(v), \mathbf{1}_{i-1} \right\rangle = \langle f_i, \mathbf{1} \rangle_w = 0.$$

Consequently, we can employ Lemma 9.1.4 and argue that balls do not cover all vertices of G.

9.2 Regular Graphs

In this section we prove Theorem 9.1.2. We lower bound the eigenvalues of $\mathcal{L}$, for regular graphs, by $\Omega(k^2/n^2)$. In this section we assume that G is unweighted and w-regular, that is $w(v) = w$ for all $v \in V$.

We next give an overview of the proof. The idea is to choose a vertex u far from the origin in the spectral embedding, i.e., $\|F(u)\|^2 = \Omega\big(k/(nw)\big)$. We consider a ball $B \subset \mathbb{R}^{k-1}$ of radius $\|F(u)\|/2$ centered at $F(u)$. We bound λ_k below by lower bounding the energy of a function $f(v) = \langle F(v), F(u)/\|F(u)\| \rangle$ with $f(u) = \|F(u)\|$ along the shortest path from u to the vertices outside of B. To obtain a good lower bound, we need to show that the length r of this path is $O\big(n/(kw)\big)$. We use the regularity of the graph to show that the shortest path has length $O\big(|B|/w\big)$. Then we use the isotropy property to show that $|B| = O(n/k)$. Together, these give the bound we want on the length of the path. Using the starting value of f, we obtain that $\lambda_k = \Omega\big(F(x)^2/r\big) = \Omega\big(k^2/n^2\big)$, which completes the proof of the theorem.

We now begin the actual proof. Let $F : V \to \mathbb{R}^{k-1}$ be a centered spectral embedding. By Lemma 8.1.1, there is a vertex $x \in V$ such that $\|F(u)\|^2 \ge \frac{k-1}{nw}$. We define $f \in \ell^2(V)$ by

$$\forall v \in V \quad f(v) := \left\langle \frac{F(u)}{\|F(u)\|}, F(v) \right\rangle.$$

In particular, observe that $f(u) = \|F(u)\|$. By Corollary 8.1.8, $\lambda_k \ge \mathcal{R}(f)$, so it suffices to show that $\mathcal{R}(f) = \Omega\big(f(u)^2 kw/n\big) = \Omega\big(k^2/n^2\big)$.

Let

$$B := B_f(u, |f(u)|/2).$$

First, by Lemma 8.1.2,

$$1 = \sum_{v \in V} w \cdot f(v)^2 \ge \sum_{v \in B} w \cdot f(u)^2/4 = |B| w \|F(u)\|^2 /4 \ge \frac{|B|(k-1)}{4n}. \tag{9.2.1}$$

Second, since $f \in \operatorname{span}\{f_2, \ldots, f_k\}$, and F is centered, f is also centered. Thus, by Lemma 9.1.4 we have $B \ne V$. Since $B \ne V$ and G is connected, there is a path from u to a vertex outside of B. Let $\mathcal{P}$ be the shortest path from u to any vertex outside of B, and let r be the length (number of edges)

of this path.

Lemma 9.2.1. *If G is w-regular, then $|B| \geq w \cdot (r-1)/3$.*

Proof. W.l.o.g. we assume that $r \geq 1$. Let $\mathcal{P} = v_0, \ldots, v_r$ where $v_0 = u$ and $v_r \notin B$. Let $S = \{v_0, v_3, v_6, \ldots, v_{3\lfloor (r-2)/3 \rfloor}\}$. Since $\mathcal{P}$ is the shortest path from u to outside of B, each vertex of S is not connected to any other vertex of S, and each pair of vertices of S does not have any common neighbor. Moreover, since $3\lfloor (r-2)/3 \rfloor \leq r-2$, each vertex of S is only connected to the vertices inside B. Since G is w-regular, every vertex of S has $w-2$ neighbors in $B - \mathcal{P}$ that are not adjacent to any other vertices of S. Therefore,

$$|B| \geq r + |S|(w-2) \geq r + \frac{(w-2)\cdot(r-1)}{3} \geq \frac{w\cdot(r-1)}{3}.$$

□

We consider two cases.

i) $|\mathrm{nei}(u) \cap B| \leq w/2$. By Corollary 8.1.8,

$$\lambda_k \geq \mathcal{E}_f \geq \sum_{v \notin B : (v,u) \in E} |f(v) - f(u)|^2 \geq \frac{w}{2}\big(\|F(u)\|/2\big)^2 \geq \frac{(k-1)}{8n}$$

and we are done.

ii) $|\mathrm{nei}(u) \cap B| > w/2$. We first show that $|B| \geq w \cdot r/6$, and then we prove the theorem. If $r = 1$, then since $|\mathrm{nei}(u) \cap B| \geq w/2$, $|B| \geq w/2$. Otherwise, by Lemma 9.2.1 $|B| \geq w(r-1)/3 \geq w \cdot r/6$. Therefore, by Corollary 8.1.8 and Lemma 8.1.10,

$$\lambda_k \geq \mathcal{E}_f \geq \mathcal{E}(P) \geq \frac{1}{r}(\|F(u)\|/2)^2 \geq \frac{w\,\|F(u)\|^2}{24|B|} \geq \frac{(k-1)^2}{96n^2},$$

where the third inequality follows by the fact that the Euclidean distance of endpoints of $\mathcal{P}$ is at least the radius of B which $\|F(u)\|/2$, and the last inequality follows by equation (9.2.1).

This proves Theorem 9.1.2.

9.3 General Graphs

In the preceding section, we proved a bound of $\Omega(k^2/n^2)$ on λ_k, for regular unweighted graphs. In this section, we prove $\Omega(k^3/n^3)$ for all unweighted graphs, and we prove Theorem 9.1.1. This answers a question of [Lyo05b] (see (3.14) there). As we show later this result has applications in estimating the number of spanning trees or approximating sparsest cut problem.

Before getting into the details of the proof of Theorem 9.1.1, we describe a very different approach based on our results on Higher order Cheeger's inequality that provide a worse lower bound on λ_k. By Theorem 10.1.1, for any $k \geq 2$, $\lambda_k \gtrsim \rho(k)^2/\log k$. So, if we prove a lower bound on $\rho(k)$ that provides a lower bound λ_k. Consider k disjoint sets $S_1, \ldots, S_k \subseteq V$. Since these sets are disjoint there is one, say S_1 with only n/k vertices. Since G is unweighted, $w(S_1, \overline{S_1}) \geq 1$, and $\mathrm{vol}(S_1) \leq \binom{|S_1|}{2}$. Therefore,

$$\lambda_k \gtrsim \frac{\rho(k)^2}{\log k} \geq \frac{\phi(S_1)^2}{\log k} = \frac{w(S_1,\overline{S_1})^2}{\log(k)\cdot \mathrm{vol}(S_1)^2} \geq \frac{1}{\log(k)\cdot \binom{n/k}{2}^2} = \Omega\Big(\frac{k^4}{\log(k)\cdot n^4}\Big)$$

We prove Theorem 9.1.1 by showing that $\mathcal{R}(F) = \Omega\big(k^3/n^3\big)$. Our proof is a generalization of the proof of Theorem 9.1.2. Here, instead of just lower bounding the Rayleigh quotient by considering a ball around a single vertex, we take $\Omega(k)$ disjoint balls about $\Omega(k)$ vertices chosen carefully so that their spectral norm is within a constant factor of the average.

This requires us to use the higher-dimensional embedding F, not merely its 1-dimensional projection f.

Let $b := \lceil (k-1)/2 \rceil$. We use Algorithm 11 to choose b disjoint balls based on the spectral embedding of G.

Algorithm 11 Ball-Selection(α)

Let $S_0 \leftarrow V$.
for $i = 1 \to b$ **do**
  Choose a vertex u_i in S_{i-1} that maximizes $\|F(u_i)\| \sqrt{d(u_i)}$.
  Let $S_i \leftarrow S_{i-1} - B_F(u_i, \alpha \|F(u_i)\|)$.
end for
return $B_F(u_1, \alpha \|F(u_1)\|/2), \ldots, B_F(u_b, \alpha \|F(u_b)\|/2)$.

The next lemma shows properties of Ball-Selection that will be used in the proof. In the rest of the proof, we let $\alpha := 1/4$.

Lemma 9.3.1. *The returned balls satisfy*

i) For each $1 \leq i \leq b$, $\|F(u_i)\| \geq \sqrt{k/3n \cdot w(u_i)}$.

ii) For every $1 \leq i < j \leq k/2$,

$$B_F\left(u_i, \frac{\alpha}{2}\|F(u_i)\|\right) \cap B_F\left(u_j, \frac{\alpha}{2}\|F(u_j)\|\right) = \emptyset.$$

Proof. First observe that by property Corollary 8.1.5, for each $1 \leq i \leq k/2$, we have

$$\mathcal{M}_F(B_F(u_i, \alpha\|F(u_i)\|)) \leq \frac{1}{1-4\alpha^2} = 4/3.$$

Since by equation (8.1.2) at the beginning of the algorithm $\mathcal{M}_F(S) = \mathcal{M}_F(V) = k-1$, and by above equation, in each iteration of the for loop $\mathcal{M}_F(S)$ decreases by no more than $4/3$,

$$\mathcal{M}_F(S_{b-1}) \geq k - 1 - (b-1)4/3 \geq (k-1)/3.$$

where we used the definition of b.

Therefore, by Lemma 8.1.1, for every $1 \leq i \leq b$,

$$\|F(u_i)\| \geq \sqrt{\frac{\mathcal{M}_F(S_I)}{n \cdot w(u_i)}} \geq \sqrt{\frac{k-1}{3n \cdot w(u_i)}}.$$

This proves (i). Finally, (ii) simply follows by the fact that each center u_i is only contained in its own ball and none of the rest of $b-1$ balls. □

In the rest of the proof let $B_i := B(u_i, \alpha \|F(u_i)\|/2)$, for all $1 \leq i \leq b$. In the next lemma we prove strong lower bounds on the energy of any ball B_i. Then, we will lower bound the numerator of the Rayleigh quotient of F simply by adding up these lower bounds.

Lemma 9.3.2. *For every* $1 \leq i \leq b$,

$$\mathcal{E}_F(B_i) > \frac{k-1}{200 \cdot n \cdot |B_i|^2}.$$

Proof. We consider two cases. If $w(u_i) \leq |B_i|$, then we lower bound $\mathcal{E}_F(B_i)$ by measuring the energy of the edges of a shortest path from x_i to the outside. Otherwise, we simply lower bound $\mathcal{E}_F(B_i)$ by the stretch of edges of u_i to its neighbors outside of B_i.

i) $w(u_i) \leq |B_i|$. Since F is a centered embedding there is a vertex outside of B_i by Lemma 9.1.4. Let $\mathcal{P}_i$ be the shortest path with respect to the graph distance in G from u_i to any vertex outside of B_i. Since G is connected, $\mathcal{P}_i$ is well defined. Using Lemma 8.1.10, we can lower bound the energy of B_i by

$$\mathcal{E}(B_i) \geq \frac{\alpha^2 \|F(u_i)\|^2}{4|B_i|} \geq \frac{k-1}{200 \cdot n \cdot w(u_i) \cdot |B_i|}, \tag{9.3.1}$$

where the second inequality holds by (i) of Lemma 9.3.1. By the above inequality, if $w(u_i) \leq |B_i|$, then $\mathcal{E}(B_i) > \frac{k-1}{200 \cdot n \cdot |B_i|^2}$, and we are done.

ii) $w(u_i) > |B_i|$. Since G is a simple graph, at least $w(u_i) - |B_i| + 1$ of the neighbors of u_i in G are not contained in B_i. That is, $|\mathrm{nei}(u_i) - B_i| \geq w(u_i) - |B_i| + 1$. We lower bound the energy

of B_i by the energy of the edges between u_i and its neighbors that are not contained in B_i:

$$\mathcal{E}(B_i) \geq \sum_{\substack{(v,u_i)\in E \\ v\notin B_i}} \|F(u_i) - F(v)\|^2 \geq |\mathrm{nei}(u_i) - B_i| \frac{\alpha^2}{4} \|F(u_i)\|^2$$
$$> \big(w(u_i) - |B_i| + 1\big) \frac{k-1}{200 \cdot n \cdot w(u_i)} \geq \frac{k-1}{200 \cdot n \cdot |B_i|}\,.$$

The second inequality uses the radius of the ball B_i, the third inequality follows from (9.3.1).

□

Now we are ready to lower bound $\mathcal{R}(F)$.

Proof of Theorem 9.1.1. By property (ii) of Lemma 9.3.1, the balls are disjoint. Therefore, $\sum_{i=1}^{b} |B_i| \leq n$. Hence, Lemma 8.1.7 yields

$$\lambda_k \geq \mathcal{R}(F) = \frac{\mathcal{E}_F}{k-1} \geq \frac{1}{2(k-1)} \sum_{i=1}^{b} \mathcal{E}(B_i) \geq \frac{1}{k-1} \sum_{i=1}^{b} \frac{k-1}{400 \cdot n \cdot |B_i|^2} \geq \frac{(k-1)^3}{3200 \cdot n^3}$$

where the second inequality follows by the fact that each edge is counted in at most two balls, and the last inequality follows by convexity of the function $s \mapsto 1/s^2$. □

9.4 Vertex-Transitive Graphs

In this section we assume that G is an unweighted vertex-transitive w-regular graph and we prove lower bounds on the eigenvalues of G. Before proving Theorem 9.1.3 we need to prove several properties of vertex-transitive graphs.

We abuse notation and we also use σ to denote an automorphism operator. For a function $f : V \to \mathbb{R}$, $\sigma f(v) = f(\sigma^{-1}(v))$. The next lemma describes several properties of σ as an operator.

Fact 9.4.1. *Any automorphism σ of G satisfies the following*

i) $\mathcal{L}\sigma = \sigma\mathcal{L}$.

ii) σ *is a* unitary *operator, meaning that* $\sigma\sigma^T = \sigma^T\sigma = I$.

Above properties are easy to verify and we don't prove them here.

Lemma 9.4.2. *Let $1 < k < n$ such that $\lambda_k < \lambda_{k+1}$. For any two vertices $u, v \in V$, $\|F(u)\| = \|F(v)\|$. Furthermore, for any automorphism σ,*

$$\|F(u) - F(v)\| = \|F(\sigma(u)) - F(\sigma(v))\|\,.$$

Proof. Let $g_2, \ldots, g_k \in \ell^2(V)$ be orthonormal eigenfunctions of $\mathcal{L}$ corresponding to $\lambda_2, \ldots, \lambda_k$. Let $\mathcal{S}$ be the vector space of any linear combinations of $g_2, \ldots, g_k$. For any automorphism σ of G, and $2 \leq i \leq k$, we have

$$\mathcal{L}g_i = \lambda_i g_i \Rightarrow \sigma\mathcal{L}g_i = \mathcal{L}\sigma g_i = \lambda_i \sigma g_i.$$

So, σg_i is also an eigenfunction of λ_i.

Next, we show that $\sigma g_2, \ldots, \sigma g_k$ also form a basis for $\mathcal{S}$. Since σ is a unitary operator, $\sigma g_2, \ldots, \sigma g_k$ also form an orthonormal system. This is because for any $2 \leq i, j \leq k$, $\langle \sigma g_i, \sigma g_j \rangle = \langle g_i, g_j \rangle$. On the other hand, since $\lambda_{k+1} > \lambda_k$ and $\lambda_2 > \lambda_1$, for all $2 \leq i \leq k$, $\sigma g_i \in \mathcal{S}$. So, $\sigma g_2, \ldots, \sigma g_k$ form another basis for $\mathcal{S}$.

For a function $f \in \ell^2(V)$, let

$$\Pi_{\mathcal{S}}(f) := \sum_{i=2}^{k} \langle f, g_i \rangle g_i,$$

be the projections of f to $\mathcal{S}$. Then, since both $g_2, \ldots, g_k$ and $\sigma g_2, \ldots, \sigma g_k$ form a basis for $\mathcal{S}$,

$$\|\Pi_{\mathcal{S}}(f)\|^2 = \sum_{i=2}^{k} \langle f, g_i \rangle^2 = \sum_{i=2}^{k} \langle \sigma f, \sigma g_i \rangle^2 = \|\Pi_{\mathcal{S}}(\sigma f)\|^2 .$$

Now, fix $u, v \in V$ and choose σ such that $\sigma(u) = \sigma(v)$. Then,

$$\|F(u)\| = \frac{1}{\sqrt{w}} \|\Pi_{\mathcal{S}}(\mathbf{1}_u)\| = \frac{1}{\sqrt{w}} \|\Pi_{\mathcal{S}}(\sigma \mathbf{1}_u)\| = \|F(v)\| .$$

Similarly, for any $u, v \in V$.

$$\|F(u) - F(v)\| = \frac{1}{\sqrt{w}} \|\Pi_{\mathcal{S}}(\mathbf{1}_u - \mathbf{1}_v)\| = \frac{1}{\sqrt{w}} \|\Pi_{\mathcal{S}}(\sigma(\mathbf{1}_u - \mathbf{1}_v))\| = \|F(\sigma(u)) - F(\sigma(v))\| .$$

□

For a vertex u, let $\beta(u) := \max_{v:(u,v)\in E} \|F(u) - F(v)\|^2$. By above lemma $\beta(u) = \beta(v)$ for all u, v. Therefore, we drop u and use β. By the above lemma we can write the Rayleigh quotient of F as follows.

$$\lambda_k \geq \frac{\sum_{(u,v)\in E} \|F(u) - F(v)\|^2}{\sum_{v\in V} w \|F(v)\|^2} \geq \frac{\max_{(u,v)\in E} \|F(u) - F(v)\|^2}{2w \cdot \|F(v)\|^2} = \frac{n \cdot \beta}{2w \cdot (k-1)}. \tag{9.4.1}$$

It remains to lower bound β.

Proof of Theorem 9.1.3. W.l.o.g we assume that $\lambda_k < \lambda_{k+1}$. Our proof strategy is similar to Theorem 9.1.2 Fix a vertex $u \in G$, and let $f : V \to \mathbb{R}$ such that

$$f(v) := \langle F(v), F(u)/\|F(u)\| \rangle.$$

Also, let $B := B_f(u, f(u)/2)$. By equation (9.2.1), $|B| \le 4n/(k-1)$. Let

$$r := \frac{f(u)}{2\sqrt{\beta}} = \frac{\|F(u)\|}{2\sqrt{\beta}} = \sqrt{\frac{k-1}{4n\sqrt{\beta}}},$$

and let B' be the set of vertices where their shortest path distance to u is at most r. We shall show that $B' \subseteq B$. By equation (9.4.1), $r \ge \sqrt{1/(8w\lambda_k)}$. Thus,

$$N\Big(\sqrt{\frac{1}{8w \cdot \lambda_k}}\Big) \le N(r) = |B'| \le |B| \le \frac{4n}{k-1},$$

and we get $8w/N^{-1}(4n/(k-1)) \le \lambda_k^2$ which proves the theorem.

It remains to show that $B' \subseteq B$. First observe that for edge $(v, v') \in V$, we have

$$\sqrt{\beta} \ge \|F(v) - F(v')\| \ge \left|\left\langle F(v) - F(v'), \frac{F(u)}{\|F(u)\|}\right\rangle\right| = |f(v) - f(v')|\,.$$

Since there is a path of length at most r from u to every vertex in B', we have

$$\forall v \in B' \quad |f(v) - f(u)| \le r\sqrt{\beta} = f(u)/2\,.$$

Thus every $v \in B'$ also belongs to B. □

9.5 Applications

Our first application is an approximation algorithm for the uniform sparsest cut problem with a better than $O(\sqrt{\log n})$ approximation factors that run in sub-exponential time.

Theorem 9.5.1. *There exists an algorithm such that for any unweighted graph G and any given $c > 1$, it finds a set of conductance $O(\sqrt{\log(n)/c})\phi(G)$ in time $2^{n \cdot O((c/\log n)^{1/3})}$.*

Proof. We use the following result of Guruswami and Sinop [GS13].

Theorem 9.5.2 (Guruswami and Sinop [GS13])**.** *There exists an algorithm that for any graph G satisfying $\lambda_k \ge 2\phi(G)$, finds a set of conductance $O(\phi(G))$ in time $2^{O(k)}$.*

We consider two cases.

i) $\phi(G) \ge \log(n)/c$. In this case we simply use the Cheeger's inequality and we find a set of conductance $\sqrt{8\phi(G)} = O(\phi(G) \cdot \sqrt{\log(n)/c})$.

ii) $\phi(G) < \log(n)/c$. In this case by Theorem 9.1.1, for $k = O(n \cdot c^{1/3}/\log^{1/3} n)$,

$$\lambda_k = 2\log(n)/c \ge 2\phi(G).$$

Thus, we can employ Theorem 9.5.2 and give a constant factor approximation in time $2^k = 2^{n \cdot O((c/\log n)^{1/3})}$. □

As an example, that for $c = \sqrt{\log(n)}$ we obtain a $O(\log(n)^{1/4})$ approximation algorithm that runs in time $2^{O(n/\log(n)^{1/6})}$

Our second application is a fast local algorithm for approximating the number $\tau(G)$ of spanning trees of a finite massive graph, G. The problem of counting the number of spanning trees of a graph is one of the fundamental problems in graph theory, for which Theorem 2.8.1 (Matrix-Tree Theorem) gives a simple $O(n^3)$-time algorithm. For very large n, however, even this is too slow. For a general graph, $\tau(G)$ can be as large as n^{n-2}, which is its value for a complete graph by Cayley's theorem [Cay89].

A *local* graph algorithm is one that is allowed to look only at the local neighborhood of random samples of vertices of the graph. The notion of graph-parameter estimability involves estimating a graph parameter, such as $\tau(G)$, using a local graph algorithm (see, e.g., [Ele10] or [Lov12, Chapter 22] for a discussion). We prove that $\tau(G)$ is estimable in this sense. In fact, we prove estimability in an even stronger sense. Suppose that we have access to G only through an oracle that supports the following simple operations:

- Select a uniformly random vertex of G.
- For a given vertex $v \in V$, select a uniformly random neighbor of v.
- For a given vertex $v \in V$, return $w(v)$.

The proof of the next corollary presents a local algorithm for approximating the number of spanning trees of G that uses an oracle satisfying the above operations, as well as knowledge of n and $|E|$. For any given $\epsilon > 0$, our algorithm approximates $\frac{1}{n}\log\tau(G)$ within an ϵ-additive error using only $O\big(\operatorname{poly}(\epsilon^{-1}\log n)\big)$ queries.

Corollary 9.5.3. *Let G be a finite, unweighted, connected graph. Given an oracle access to G that satisfies the above operations, together with knowledge of n and $|E|$, there is a randomized algorithm that for any given $\epsilon, \delta > 0$, approximates $\log\tau(G)/n$ within an additive error of ϵ, with probability at least $1-\delta$, by using only $\tilde{O}(\epsilon^{-5} + \epsilon^{-2}\log^2 n)\log(1/\delta)$ many oracle queries.*

We remark that here ϵ can be any number greater than 0. For example letting $\epsilon = n^{0.1}$, we get a $2^{n^{0.9}}$ multiplicative approximation of $\tau(G)$ in time $O(\sqrt{n}\operatorname{polylog} n)$.

We use the following statement of [Lyo05b, Proposition 3.1].

Proposition 9.5.4 (Lyons [Lyo05b])**.** *Suppose that G is an unweighted, connected graph. Then*

$$\log\tau(G) = -\log\big(4|E|\big) + \sum_{v\in V}\log 2w(v) - \sum_{t\geq 1}\frac{1}{t}\Big(\sum_{v\in V}P^t(v,v) - 1\Big).$$

Proof. Write $\det' A$ for the product of the non-zero eigenvalues of a matrix A. As shown by [RS74] (see also [Chu04, Theorem 2]), we may rewrite the Matrix-Tree Theorem as

$$\tau(G) = \frac{\prod_{v\in V} 2w(v)}{\sum_{v\in V} 2w(v)} \det'(I - P)$$

[the proof follows from looking at the coefficient of s in $\det\big(I - P - sI\big) = (\det 2D)^{-1} \det(L - 2sD)$ and using the Matrix-Tree Theorem in its original form with cofactors]. Thus,

$$\log \tau(G) = -\log\big(4|E|\big) + \sum_{v\in V} \log 2w(v) + \log\det'(I - P)\,. \tag{9.5.1}$$

Let $\hat{\lambda}_k$ be the eigenvalues of P with $\hat{\lambda}_1 = 1$. We may rewrite the last term of equation (9.5.1) as

$$\begin{aligned}\log\det'(I - P) &= \sum_{k=2}^{n} \log(1 - \hat{\lambda}_k) = -\sum_{k=2}^{n}\sum_{t\geq 1} \hat{\lambda}_k^t/t \\ &= -\sum_{t\geq 1}\sum_{k=2}^{n} \hat{\lambda}_k^t/t = -\sum_{t\geq 1} \frac{1}{t}(\operatorname{trace} P^t - 1)\,.\end{aligned}$$

where the last equality follows by Lemma 7.1.7. Since $\operatorname{trace} P^t = \sum_{v\in V} P^t(v,v)$, the desired formula now follows from this and (9.5.1). □

The following is an immediate corollary of Theorem 9.1.1.

Corollary 9.5.5. *For any simple, unweighted, connected graph G, and any integer $t > 0$,*

$$\frac{1}{n}\left(\sum_{v\in V} P^t(v,v) - 1\right) < 17 \cdot t^{-1/3}.$$

Proof. First of all, although P is not symmetric, $D^{1/2}PD^{-1/2}$ is symmetric and it has the same eigenvalues as of P. Thus, by Lemma 7.1.7,

$$\sum_{v\in V} P^t(v,v) = \operatorname{trace}(P^t) = \operatorname{trace}(D^{1/2}P^tD^{-1/2}) = \sum_{i=1}^{n} (1 - \lambda_i/2)^t,$$

Since $\lambda_i \le 2$ for all $2 \le i \le n$, by Theorem 9.1.1, we have $0 \le 1 - \lambda_i/2 \le 1 - i^3/6400n^3$. Therefore,

$$\begin{aligned}\sum_{v\in V} P^t(v,v) - 1 &= \sum_{i=1}^{n}(1-\lambda_i/2)^t - 1 \le \sum_{i=1}^{n}\Big(1 - \frac{i^3}{6400n^3}\Big)^t - 1\\ &\le \sum_{i=1}^{n}\exp\Big(-\frac{ti^3}{6400n^3}\Big) - 1 \le \int_{s=0}^{n}\exp\Big(-\frac{ts^3}{6400n^3}\Big)ds\\ &\le \frac{18.57n}{t^{1/3}}\int_{s=0}^{\infty} e^{-s^3}ds \le 17nt^{-1/3}.\end{aligned}$$

□

Proof of Corollary 9.5.3. First, by Proposition 9.5.4,

$$\Big|\log\tau(G) + \log\big(4|E|\big) - \sum_{v\in V}\log 2w(v) + \sum_{1\le t<2r}\frac{1}{t}\Big(\sum_{v\in V}P^t(v,v) - 1\Big)\Big| = \sum_{t\ge 2r}\frac{1}{t}\Big(\sum_{v\in V}P^t(v,v) - 1\Big)$$
$$< n\sum_{t\ge 2r}\frac{17}{t^{4/3}} < n\frac{45}{r^{1/3}}\,.$$

where we used Corollary 9.5.5. Choose $r := \lceil 90\epsilon^{-3}\rceil$, so that $45r^{-1/3} \le \epsilon/2$. Write $s := \sum_{1\le t<2r} 1/t$. Let $W := \frac{1}{n}\sum_v \log 2w(v)$ and $Y := \sum_v \frac{1}{n}\sum_{t=1}^{2r-1} P^t(v,v)/(st)$. Then, by above inequality,

$$\Big|\frac{\tau(G)}{n} - \frac{\log(4|E|)}{n} + W - sY + \frac{s}{n}\Big| \le \epsilon/2\,.$$

Therefore, we just have to approximate $W - sY$ within an additive error of $\epsilon/2$. The details of the algorithm are described below.

Algorithm 12 Approximate Counting of Spanning Trees

Input: $\epsilon > 0$.
Let $r \leftarrow \lceil 90\epsilon^{-3}\rceil$ and $s \leftarrow \sum_{1\le t<2r} 1/t$.
$N \leftarrow \lceil 64\log(1/\delta)s^2/\epsilon^2\rceil$.
for $i = 1 \to N$ **do**
  Let v be a randomly chosen vertex of G.
  Sample $1 \le t < 2r$ with probability $1/st$.
  Run a t-step lazy simple random walk from v, and let $Y_i \leftarrow \mathbb{I}[X_t = v]$.
end for
Sample $\lceil 256\log(1/\delta)(\log n)^2/\epsilon^2\rceil$ random vertices of G, and let $\tilde{W}$ be the average of the logarithm of twice the degree of sampled vertices. **return** $-n^{-1}\log(4|E|) + \tilde{W} - s(Y_1 + \ldots + Y_N)/N + s/n$.

We start by describing how to approximate Y within an $\epsilon/4s$ error (hence, to approximate sY within an $\epsilon/4$ error). We use a Monte Carlo sampling method. Let $X_0, X_1, \ldots, X_t$ represent a t-step

lazy simple random walk started from a vertex of G. Then

$$\begin{aligned} Y &= \sum_{v\in V} \frac{1}{nst} \sum_{t=1}^{2r-1} \mathbb{P}\left[X_t = v \mid X_0 = v\right] \\ &= \sum_{v\in V} \sum_{1\le t<2r} \sum_{\substack{v_1,\ldots,v_t\in V \\ v_t=v}} \frac{1}{nst} \mathbb{P}\left[X_1 = v_1, \ldots, X_t = v_t \mid X_0 = v\right] . \end{aligned}$$

Consider a random walk starting at a random vertex and lasting a random length of time. Namely, let $\mathcal{D}$ be the distribution on walks of lengths in $[1, 2r)$ where

$$\mathbb{P}_{\mathcal{D}}\left[(v_0, v_1, \ldots, v_t)\right] = \frac{1}{nst} \mathbb{P}\left[X_1 = v_1, \ldots, X_t = v_t \mid X_0 = v_0\right].$$

Then $Y = \mathbb{P}_{\mathcal{D}}\left[\{(v_0, v_1, \ldots, v_t) : v_t = v_0\}\right]$. First we describe how to sample from $\mathcal{D}$, then show how to approximate Y. First we sample a random vertex v of G, then we select a random $1 \le t < 2r$ with probability $1/st$ (note that $\sum_{1\le t<2r} 1/st = 1$ by definition of s). Finally, we choose a t-step random walk started from y and compute $\mathbb{I}\left[X_t = v\right]$. See the details in Algorithm 12.

We approximate Y by sampling $N := \left\lceil 64 \log(1/\delta) s^2/\epsilon^2 \right\rceil$ independent elements of $\mathcal{D}$ and computing their average. Let $Y_i := \mathbb{I}\left[X_t = v\right]$ be the i^{th} sample of $\mathcal{D}$. By definition, $Y_i \in [0, 1]$ and $\mathbb{E}\left[Y_i\right] = Y$. Since $Y_1, \ldots, Y_N$ are independent, Hoeffding's inequality gives

$$\mathbb{P}\left[\left|\frac{Y_1 + \ldots + Y_N}{N} - Y\right| \ge \frac{\epsilon}{4s}\right] \le 2\exp\left(-\frac{\epsilon^2 N}{16 s^2}\right) \le \delta/2 .$$

Therefore, with probability at least $1-\delta/2$, we have that $s(Y_1+\ldots+Y_N)/N$ approximates sY within an error of $\epsilon/4$. It remains to approximate W within error $\epsilon/4$ and with probability at least $1-\delta/2$. That can be done easily by sampling $O(\epsilon^{-2} \log \delta^{-1} \log^2 n)$ independent uniform random vertices of G and taking the average of the logarithm of twice their degrees, $\tilde{W}$ (see step 8 of Algorithm 12). Since $\log 2w(y) \le 2 \log n$ for all $y \in V$, again by Hoeffding's inequality we have

$$\mathbb{P}\left[|\tilde{W} - W| \ge \frac{\epsilon}{4}\right] \le \delta/2 .$$

Therefore, by the union bound the algorithm succeeds with probability at least $1 - \delta$.

It remains to compute the number of oracle accesses. We used $O(\epsilon^{-2} \log \delta^{-1} \log^2 n)$ accesses to approximate W. On the other hand, we can compute each Y_i with at most $2r = O(\epsilon^{-3})$ oracle accesses. Therefore, we can approximate Y with at most

$$2Nr = O\big(\epsilon^{-5} \log \delta^{-1} s^2\big) = O\big(\epsilon^{-5} \log^2 \epsilon^{-1} \log \delta^{-1}\big) = \tilde{O}\big(\epsilon^{-5} \log \delta^{-1}\big)$$

many queries. □

We note that knowing $|E|$ is not really necessary for this algorithm, since it contributes a term of size $O(\log(n)/n)$, which will be much less than ϵ in any reasonable example where one might use this algorithm.

Chapter 10

Higher Order Cheeger's Inequality

In this chapter we study higher-order analogs of Theorem 7.8.1. and develop new multi-way spectral partitioning algorithms. The results of this chapter are based on a joint work with James Lee and Luca Trevisan in [LOT12]. Since the appearance of this paper our results and techniques have been used in several recent papers (see e.g. [Tan12, Mic13, LO12, TY12]).

We will extend and generalize several of results in this chapter later in Chapter 11 and Chapter 13.

10.1 Introduction

For any $k \in \mathbb{N}$, the *k-way expansion constant,* is defined as

$$\rho(k) := \min_{\text{disjoint } S_1,S_2,\ldots,S_k} \max_{1\leq i\leq k} \phi(S_i), \tag{10.1.1}$$

where the minimum is over all collections of k non-empty, disjoint subsets $S_1, S_2, \ldots, S_k \subseteq V$. For example, observe that for any graph G, $\rho(2) = \phi(G)$. As another example, observe that $\rho(k) = 0$ if and only if G has at least k connected components. One can also consider other norms of $\phi(S_1), \ldots, \phi(S_k)$ as the quality of a k-way partitioning. For example, we can define $\rho_1(k)$ as the average conductance,

$$\rho_1(k) := \min_{\text{disjoint } S_1,\ldots,S_k} \frac{\phi(S_1) + \ldots + \phi(S_k)}{k}.$$

In this sense $\rho(k)$ can be considered as the infinity norm and any upper bound on $\rho(k)$ provides an upper bound on all other norms.

In Section 7.2 we proved that for any graph G $\lambda_k = 0$ if and only if G has k connected components. By above argument, we obtain $\rho_G(k) = 0 \iff \lambda_k = 0$. Recall that Cheeger's inequality provides a robust version of this fact for $k = 2$. Our main theorem implies that for any graph G, $\rho(k) \approx 0 \iff \lambda_k \approx 0$. In this sense, it can be seen as a generalization of Cheeger's inequality.

Theorem 10.1.1. *For every graph G, and every $k \in \mathbb{N}$, we have*

$$\frac{\lambda_k}{2} \leq \rho_G(k) \leq O(k^2)\sqrt{\lambda_k}\,. \tag{10.1.2}$$

Furthermore, there is a polynomial time algorithm that construct the k disjoint sets, $S_1, \ldots, S_k$ such that $\max_{1\leq i\leq k} \phi(S_i) \leq O(k^2)\sqrt{\lambda_k}$.

This resolves a conjecture of Miclo [Mic08]; see also [DJM12], where some special cases are considered. We prove several variants of this theorem in Subsection 10.2.4 and Theorem 10.4.5. We remark that from Theorem 10.1.1, it is easy to find a *partition* of the vertex set into k non-empty pieces such that every piece in the partition has expansion $O(k^3)\sqrt{\lambda_k}$ (see Theorem 10.2.7 for more details).

The left side of equation (10.1.2) is very easy to prove and follows from Claim 7.8.2. So, we only need to prove the right side of equation (10.1.2). Our proof is algorithmic and leads to new algorithms for k-way spectral partitioning. Let $F : V \to \mathbb{R}^k$ be any isotropic mapping of G (as defined in equation (8.1.3)). Our proof shows that,

$$\rho(k) \leq O(k^2)\sqrt{\mathcal{R}(F)}. \tag{10.1.3}$$

Note that to obtain equation (10.1.2) from above equation, we only need to use the fact that the spectral embedding is isotropic (see Lemma 8.1.2), and that its Rayleigh quotient is λ_k, (see Lemma 8.1.7). In other words, our theorems are robust in the sense that they extend to any isotropic embedding.

Our algorithm provides a theoretical justification for clustering algorithms that use the bottom k eigenvectors of the Laplacian[1] to embed the vertices into $\mathbb{R}^k$, and then apply geometric considerations to the embedding. See [VM03, Lux07] for a survey of such approaches. As a particular example, consider the work of Jordan, Ng and Weiss [NJW02] which applies a k-means clustering algorithm to the embedding in order to achieve a k-way partitioning. Our proof of Theorem 10.1.1 employs a similar algorithm, where the k-means step is replaced by a random geometric partitioning. It remains an interesting open problem whether k-means itself can be analyzed in this setting.

Tightness of Higher Order Cheeger's inequality

Observe that although there is a dependency to k in the RHS of equation (10.1.2), similar to the original Cheeger's inequality both sides of the inequality are independent of the size of the graph. As we will show in Section 10.5 a poly-logarithmic dependency to k is necessary in the RHS of equation (10.1.2). However, it is still an open question if the dependency to k can be improved to a poly-logarithmic function. For all we know, a polynomial dependency to k is necessary in the RHS of

[1] Equivalently, algorithms that use the top k eigenvectors of the adjacency matrix.

equation (10.1.3), and because of that our current proof techniques cannot improve the dependency to k in the RHS of equation (10.1.2) to a poly-logarithmic function. In particular, let G be a union of a cycle and $k-2$ isolated vertices. Then, $\lambda_1 = \lambda_2 \ldots = \lambda_{k-1} = 0$ and $\lambda_k = \Theta(1/n^2)$, while, $\rho(k) = \Theta(1/n)$. Consequently, if $F : V \to \mathbb{R}^k$ is the spectral embedding as defined in equation (8.1.1),

$$\rho(k) = \Theta(\sqrt{\lambda_1 + \ldots + \lambda_k}) = O(\sqrt{k\mathcal{R}(F)}).$$

Example 10.1.2 (Cycle)**.** *Let G be a cycle of length n. By our discussion in Subsection 7.3.1, for $k = o(n)$, $\lambda_k = \Theta(k^2/n^2)$. On the other hand, if we choose k disjoint paths of length $\approx n/k$ we obtain that $\rho(k) = \Theta(n/k)$. Therefore, $\phi(k) = \Theta(\sqrt{\lambda_k})$. This shows that we must have $\sqrt{\lambda_k}$ in the RHS of equation (10.1.2).*

10.1.1 Finding many sets and small-set expansion

If one is interested in finding slightly fewer sets, our approach performs significantly better.

Theorem 10.1.3. *For every graph G, and every $k \in \mathbb{N}$, we have*

$$\rho_G(k) \leq O(\sqrt{\lambda_{2k} \log k}). \tag{10.1.4}$$

If G is planar then, the bound improves to,

$$\rho_G(k) \leq O(\sqrt{\lambda_{2k}}), \tag{10.1.5}$$

and if G excludes K_h as a minor, then $\rho_G(k) \leq O(h^2\sqrt{\lambda_{2k}})$.

We remark that the bound in equation (10.1.4) holds with $2k$ replaced by $(1+\delta)k$ for any $\delta > 0$, but where the leading constant now becomes δ^{-3}; see Corollary 10.4.2. Louis, Raghavendra, Tetali and Vempala [LRTV12] have independently proved a somewhat weaker version of the bound in equation (10.1.4), using rather different techniques. Specifically, they show that there exists an absolute constant $C > 1$ such that $\rho_G(k) \leq O(\sqrt{\lambda_{Ck} \log k})$.

Theorem 10.1.3 has applications to the small-set expansion problem in graphs (see Subsection 7.7.3 for background).

Arora, Barak and Steurer [ABS10] prove the bound,

$$\phi^c(k^{1/100}) \leq O(\sqrt{\lambda_k \log_k n}),$$

Note that for $k = n^\varepsilon$ and $\epsilon \in (0,1)$, one achieves an upper bound of $O(\sqrt{\lambda_k})$, and this small loss in the expansion constant is crucial for applications to approximating small-set expansion. This was

recently improved further [OT12, OW12] by showing that for every $\alpha > 0$,

$$\phi^c(k^{1-\alpha}) \leq O(\sqrt{(\lambda_k/\alpha)\log_k n})\,.$$

These bounds work fairly well for large values of k, but give less satisfactory results when k is smaller.

Louis, Raghavendra, Tetali and Vempala [LRTV11] proved that

$$\phi^c(\sqrt{k}) \leq O(\sqrt{\lambda_k \log k}),$$

and conjectured that $\sqrt{k}$ could be replaced by k. Since for any $1 \leq k$, $\phi^c(k) \leq \rho(k)$, Theorem 10.1.3 immediately yields,

$$\phi^c(k/2) \leq O(\sqrt{\lambda_k \log k}) \tag{10.1.6}$$

resolving their conjecture up to a factor of 2 (and actually, as discussed earlier, up to a factor of $1+\delta$ for every $\delta > 0$).

Moreover, (10.1.6) is quantitatively optimal for the noisy hypercube graphs (see Section 10.5), yielding an optimal connection between the kth Laplacian eigenvalue and expansion of sets of size $\approx n/k$.

It is interesting to note that in [KLPT11], it is shown that for n-vertex, bounded-degree planar graphs, one has $\lambda_k = O(k/n)$. Thus the spectral algorithm guaranteeing (10.1.5) partitions such a planar graph into k disjoint pieces, each of expansion $O(\sqrt{k/n})$. This is tight, up to a constant factor, as one can easily see for an $\sqrt{n} \times \sqrt{n}$ planar grid, in which case the set of size $\approx n/k$ with minimal expansion is a $\sqrt{n/k} \times \sqrt{n/k}$ subgrid.

Large gaps in the spectrum.

We recall that in the practice of spectral clustering, it is often observed that the correct number of clusters is indicated by a large gap between adjacent eigenvalues, i.e., if $\lambda_{k+1} \gg \lambda_k$, then one expects the input graph can be more easily partitioned into k pieces than $k+1$. In Section 10.3, we prove a result supporting this phenomenon.

Theorem 10.1.4. *There is a constant $C > 0$ such that for every graph G and $k \in \mathbb{N}$, the following holds. If $\lambda_{4k} \geq C(\log k)^2 \lambda_{2k}$, then*

$$\rho_G(k) \leq O(\sqrt{\lambda_{2k}})\,.$$

The key point is that the implicit constant in the upper bound is independent of k, unlike the bound (10.1.4).

10.1.2 Proof Techniques

We now present an overview of the proofs of our main theorems, as well as explain our general approach to multi-way spectral partitioning. By Lemma 7.8.3 for any non-negative function $f : V \to \mathcal{R}$, it is possible to find a subset $S \subseteq \{v \in V : f(v) \neq 0\}$ such that that $\phi_G(S) \leq \sqrt{2\mathcal{R}(f)}$. Thus, in order to find k disjoint, non-expanding subsets $S_1, S_2, \ldots, S_k \subseteq V$, it suffices to find k disjointly supported functions $h_1, h_2, \ldots, h_k : V \to \mathbb{R}$ such that $\mathcal{R}(h_i)$ is small for each $i = 1, 2, \ldots, k$.

In fact, in the same paper that Miclo conjectured the validity of Theorem 10.1.1, he conjectured that finding such a family $\{h_i\}$ should be possible [Mic08, DJM12]. We resolve this conjecture and prove the following theorem in Subsection 10.2.4.

Theorem 10.1.5. *For any graph $G = (V, E)$ and any $1 \leq k \leq n$, there exist disjointly supported functions $h_1, h_2, \ldots, h_k : V \to \mathbb{R}$ such that for each $i = 1, 2, \ldots, k$, we have*

$$\mathcal{R}(h_i) \leq O(k^6)\, \lambda_k\,.$$

Observe Lemma 7.2.3 can be seen as a special case of the above theorem for the case of $k = 2$. Let F be the spectral embedding with respect to the first k eigenfunctions of $\mathcal{L}$ as defined in equation (8.1.1). Since by Lemma 8.1.7 $\mathcal{R}(F) \leq \lambda_k$, our goal is to "localize" F on k disjoint regions to produce disjointly supported functions $h_1, h_2, \ldots, h_k : V \to \mathbb{R}$, each with small Rayleigh quotient compared to $\mathcal{R}(F)$. In order to ensure that $\mathcal{R}(h_i)$ is small for each i, we must ensure that each region captures a large fraction of the ℓ^2 mass of F, and that our localization process is sufficiently smooth.

Isotropy and spreading. The first problem we face is that, in order to find k disjoint regions each with large ℓ^2 mass, it should be that the ℓ^2 mass of F is sufficiently well-spread. This directly follows from Lemma 8.1.2 and Corollary 8.1.4. As we mentioned earlier, this is the only property of the spectral embedding that we use in the proof. We show that if F is isotropic, then there are k disjointly supported functions $h_1, \ldots, h_k$ such that $\mathcal{R}(h_i) \leq O(k^6)\mathcal{R}(F)$.

A natural approach would be to find k regions $R_1, \ldots, R_k$ and define,

$$h_i(v) = \begin{cases} \|F(v)\| & \text{if } v \in R_i \\ 0 & \text{otherwise.} \end{cases}$$

Unfortunately, this sharp cutoff could make $\mathcal{E}_{h_i}$ significantly larger than $\mathcal{E}_F$, henceforth, $\mathcal{R}(h_i)$ is much larger than $\mathcal{R}(F)$. For example, suppose that G has only one edge (u, v), and $\|F(u) - F(v)\|^2 \ll \|F(v)\|^2$. Then $\mathcal{E}_F = \|F(u) - F(v)\|^2$. Now, if $v \in R_1$ and $u \notin R_1$, then

$$\mathcal{E}_{h_1} \geq \|F(v)\|^2 \gg \|F(u) - F(v)\|^2 = \mathcal{E}_F.$$

Thus we must pursue a smoother approach for localizing F. Let $d(u,v) = \|F(u) - F(v)\|$. Suppose we choose $S_1, \ldots, S_k$ such that they are *well separated*, i.e., for all $1 \leq i < j \leq k$, $d(S_i, S_j) \geq \Omega_k(1)$. Now, we may define,

$$h_i(v) = \|F(v)\| \cdot \max\left\{0, 1 - \frac{d(v, S_i)}{\epsilon}\right\}.$$

where $d(u,v)$ is just the Euclidean distance between $F(u)$ and $F(v)$, and $\epsilon = O_k(1)$. Unfortunately, this doesn't work either. Suppose (u,v) is the only edge of G for which $F(u) \neq F(v)$, $\|F(v)\|^2 = \text{poly}(n)$. If $v \in S_1$ and $d(u, S_1) = d(u,v) = \epsilon$, then

$$\mathcal{E}_{h_1} = \|F(v)\|^2 \gg \|F(u) - F(v)\|^2 = \mathcal{E}_F.$$

To resolve this problem instead of the Euclidean distance we use the radial projection distance in the above definition.

The radial projection distance. We would like to think of two vertices $u, v \in V$ as close if their Euclidean distance $\|F(u) - F(v)\|$ is small compared to their norms $\|F(u)\|, \|F(v)\|$. Recall that the radial projection distance is defined as follows, for all $u, v \in V$,

$$d_F(u,v) = \left\| \frac{F(u)}{\|F(u)\|} - \frac{F(v)}{\|F(v)\|} \right\|.$$

and if $F(u) = 0 (\text{resp.} F(v) = 0)$ we use 0 instead of $F(u)/\|F(u)\|$ (resp. $F(v)/\|F(v)\|$. Observe that this distance function has several nice properties:

i) For all $u, v \in V$, $d_F(u,v) \leq 2$.

ii) Radial projection distance is a Euclidean distance. If we define $\Gamma(v) = F(v)/\|F(v)\|$ when $F(v) \neq 0$ and $\Gamma(v) = 0$ otherwise, then $d_F(.,.)$ is just a Euclidean distance metric on the vectors $\{\Gamma(v)\}_{v \in V}$.

Now, similar to above, suppose we have $S_1, \ldots, S_k$ that are well separated with respect to $d_F(.,.)$, and let

$$h_i(v) = \|F(v)\| \cdot \max\left\{0, 1 - \frac{d_F(v, S_i)}{\epsilon}\right\}.$$

for some $\epsilon = \Omega_k(1)$. Suppose we have we have an edge (u,v) such that $\|F(u) - F(v)\| \ll \|F(u)\|$, $v \in S_1$. Then,

$$\mathcal{E}_h(\{(u,v)\}) \lesssim \|F(v)\|^2 \cdot \frac{d_F(u,v)^2}{\epsilon^2} \approx \frac{\|F(u) - F(v)\|^2}{\epsilon^2} = \mathcal{E}_F(\{(u,v)\})/\epsilon^2.$$

In Lemma 10.2.1 we show that by a similar argument we can upper bound $\mathcal{R}(h_i)$ as a function of $\mathcal{R}(F)$.

Our goal now becomes to find *dense, well separated regions* in d_F, i.e., $S_1, \ldots, S_k \subseteq V$ such that each of which contains a large fraction of the ℓ^2 mass of F and such that their pairwise distance in d_F is $\Omega_k(1)$.

Finding dense well-separated regions: Random space partitions. In order to find many separated regions, we rely on the theory of random partitions discussed in Section 7.9. We use padded-partitioning of Euclidean metrics discussed in Theorem 7.9.2. Since $d_F(.,.)$ is a Euclidean metric, we can partition our set of points randomly into pieces of diameter at most $1/2$ so that the expected fraction of ℓ^2 mass which is close to the boundary of the partition is small i.e., $o(1)$. Then, we take unions of the interiors of the pieces to find dense well-separated sets. Since by the spreading property of d_F Lemma 8.1.3. no set in the partition can contain a large fraction of the ℓ^2 mass, we can merge these set making sure that each of the merged regions contain $\Omega(1/k)$ fraction of the total mass. The details are described in Subsection 10.2.2 Subsection 10.2.3. We use these separated sets as the supports of our family $\{h_i\}$, allowing us to complete the proof of Theorem 10.1.5.

As we described in Section 7.9, the notion of "close to the boundary" depends on the dimension k, and thus the smoothness of our maps $\{h_i\}$ will degrade as the dimension grows. For many families of graphs, however, we can appeal to special properties of their intrinsic geometry.

Exploiting the intrinsic geometry. It is well-known that the shortest-path metric on a planar graph has many nice properties, but d_F is, in general, not a shortest-path geometry. Thus it is initially unclear how one might prove a bound like (10.1.5) using our approach. The answer is to combine information from the spectral embedding with the intrinsic geometry of the graph.

We define $\hat{d}_F$ as the shortest-path pseudometric on G, where the length of an edge $(u, v) \in E$ is precisely $d_F(u, v)$. In Section 10.2 we show that it is possible to do the partitioning in the metric $\hat{d}_F$, and thus for planar graphs (and other generalizations), we are able to achieve dimension-independent bounds in Theorem 10.1.3.

This technique also addresses a common shortcoming of spectral methods: The spectral embedding can lose auxiliary information about the input data that could help with clustering. Our "hybrid" technique for planar graphs suggests that such information (in this case, planarity) can be fruitfully combined with the spectral computations.

Dimension reduction. In order to obtain the tight bound in equation (10.1.4) for general graphs, we have to improve the quantitative parameters of our construction significantly. The main loss in our preceding construction comes from the ambient dimension k.

Thus our first step is to apply dimension-reduction techniques: We randomly project our points from $\mathbb{R}^k$ into $\mathbb{R}^{O(\log k)}$ and we use Theorem 8.2.1 to argue that with a constant probability the Rayleigh quotient and the spreading property are preserved under the dimension reduction, and that is all we need for the proof.

A new multi-way Cheeger inequality. Dimension reduction only yields a loss of $O(\log k)$ in

(10.1.4). In order to get the bound down to $\sqrt{\log k}$, we have to abandon our goal of localizing eigenfunctions. In Section 10.4, we give a new multi-way Cheeger rounding algorithm that combines random partitions of the radial projection distance d_F, and random thresholding based on $\|F(\cdot)\|$ (as in Cheeger's inequality). By analyzing these two processes simultaneously, we are able to achieve the optimal loss.

10.1.3 A general algorithm

Given a graph $G = (V, E)$ and any embedding $F : V \to \mathbb{R}^k$ (in particular, the spectral embedding (8.1.1)), our approach yields a general algorithmic paradigm for finding many non-expanding sets. For some $r \in \mathbb{N}$, do the following:

1. **(Radial decomposition)**

   Find disjoint subsets $S_1, S_2, \ldots, S_r \subseteq V$ using the values $\{F(v)/\|F(v)\| : v \in V\}$.

2. **(Cheeger sweep)**

   For each $i = 1, 2, \ldots, r$,

   Sort the vertices $S_i = \{v_1, v_2, \ldots, v_{n_i}\}$ so that

   $$\|F(v_1)\| \geq \|F(v_2)\| \geq \cdots \geq \|F(v_{n_i})\|.$$

   Output the set with smallest conductance among the $n_i - 1$ sets of the form,

   $$\{v_1, v_2, \ldots, v_j\}$$

   for $1 \leq j \leq n_i - 1$.

As discussed in the preceding section, each of our main theorems is proved using an instantiation of this schema. For instance, the proof of Theorem 10.1.1 partitions uses the radial projection distance d_F. The proof of equation (10.1.5) uses the induced shortest-path metric $\hat{d}_F$. And the proof of equation (10.1.4) uses $d_{F'}$ where $F' : V \to \mathbb{R}^{O(\log k)}$ is obtained from random projection. The details of the scheme for equation (10.1.4) is provided in Section 10.6. A practical algorithm might use r-means to cluster according to the radial projection distance (c.f. Algorithm 2). We will provide some justification for using r-means in Section 10.3.

We remark that partitioning the normalized vectors as in step (i) is used in the approach of [NJW02], but not in some other methods of spectral partitioning (see [SM00, VM03] for alternatives). Interestingly, while we were working on this problem we were not aware of this fact. Coincidently, we reached to the idea of using normalized vectors and in particular the radial projection distance function when we were trying find the right way to localize the spectral embedding F. Our analysis suggests a theoretical justification for partitioning using the normalized vectors.

10.2 Localizing eigenfunctions

Given a mapping $F : V \to \mathbb{R}^k$, in the present section, we show how to find disjointly supported functions $h_1, h_2, \ldots, h_k : V \to \mathbb{R}$ with $\mathcal{R}(h_i) \leq k^{O(1)}\mathcal{R}(F)$, where $k \in \mathbb{N}$. In order to find many disjointly supported functions from a geometric representation $F : V \to \mathbb{R}^h$, it should be that the ℓ^2 mass of F is not too concentrated. As we discussed earlier, all we need to assume is that F satisfies the spreading property.

10.2.1 Smooth localization

Given a map $F : V \to \mathbb{R}^l$ and a subset $S \subseteq V$, we now show how to construct a function supported on a small-neighborhood S, which retains the ℓ^2 mass of F on S, and which doesn't stretch edges by too much.

Recall that $\hat{d}$ is the shortest path metric defined on a metric $d(.,.)$. We prove several of our statements for $\hat{d}_F$ to get the tightest results, but we recommend the readers to first understand the whole arguments for the actual radial projection distance, d_F.

Lemma 10.2.1 (Localization)**.** *Let $F : V \to \mathbb{R}^l$, and let $d : V \times V \to \mathbb{R}_+$ be a distance function such that for any $u, v \in V$,*

$$\|F(u)\| \cdot d(u,v) \leq \alpha \cdot \|F(u) - F(v)\| .$$

Then, for every subset $S \subseteq V$ and number $\epsilon > 0$, there exists a mapping $h : V \to \mathbb{R}$ which satisfies the following three properties:

1. *$h(v) = \|F(v)\|$ for $v \in S$,*
2. *$\mathrm{supp}(h) \subseteq B_{\hat{d}}(S, \epsilon)$, and*
3. *if $\{u, v\} \in E$, then $|h(u) - h(v)| \leq (1 + \frac{\alpha}{\epsilon})\|F(u) - F(v)\|$.*

Proof. First, define

$$\psi(v) := \max\left(0, 1 - \frac{\hat{d}(v,S)}{\epsilon}\right) .$$

Since $\hat{d}(v,S)$ is a metric, $\psi(.)$ is $(1/\epsilon)$-Lipschitz with respect to $\hat{d}$, so since $\hat{d}$ and d agree on edges, we have for every $(u,v) \in E$,

$$|\psi(u) - \psi(v)| \leq \frac{1}{\epsilon} \cdot d(u,v) . \tag{10.2.1}$$

Finally, set $h(v) := \psi(v) \cdot \|F(v)\|$.

Properties (i) and (ii) are immediate from the definition, thus we turn to property (iii). Fix

$\{u, v\} \in E$. We have,

$$\begin{aligned} |h(u) - h(v)| &= \left|\psi(u)\,\|F(u)\| - \psi(v)\,\|F(v)\|\,\right| \\ &\leq |\psi(v)| \cdot \|F(u) - F(v)\| + \|F(u)\| \cdot |\psi(u) - \psi(v)|\,. \end{aligned}$$

Since $\psi \leq 1$, the first term is at most $\|F(u) - F(v)\|$. Now, using equation (10.2.1), we have

$$\|F(u)\| \cdot |\psi(u) - \psi(v)| \leq \frac{1}{\epsilon} \cdot \|F(u)\| \cdot d(u, v) \leq \frac{\alpha}{\epsilon} \cdot \|F(u) - F(v)\|,$$

completing the proof of (iii). □

Note that by Lemma 8.1.6, d_F satisfies the assumption of above lemma. The preceding construction reduces the problem of finding disjointly supported set functions to finding separated regions in $(V, \hat{d}_F)$, each of which contains a large fraction of the ℓ^2 mass of F.

Lemma 10.2.2. *Let $F : V \to \mathbb{R}^l$ be given, and let $d : V \times V \to \mathbb{R}_+$ be a distance function such that for any $u, v \in V$,*

$$\|F(u)\| \cdot d(u, v) \leq \alpha \cdot \|F(u) - F(v)\|\,.$$

Suppose that for some $\epsilon, \delta > 0$ and $r \in \mathbb{N}$, there exist r disjoint subsets $T_1, T_2, \ldots, T_r \subseteq V$ such that $\hat{d}(T_i, T_j) \geq 2\epsilon$ for $i \neq j$, and for every $i = 1, 2, \ldots, r$, we have $\mathcal{M}_F(T_i) \geq \delta \cdot \mathcal{M}_F(V)$. Then, there exist disjointly supported functions $h_1, h_2, \ldots, h_r : V \to \mathbb{R}$ such that for $i = 1, 2, \ldots, r$, we have

$$\mathcal{R}(h_i) \leq \frac{2}{\delta(r - i + 1)} \left(1 + \frac{\alpha}{\epsilon}\right)^2 \mathcal{R}(F)\,. \tag{10.2.2}$$

Proof. For each $i \in [r]$, let $h_i : V \to \mathbb{R}$ be the result of applying Lemma 10.2.1 to the domain T_i. Since $\hat{d}(T_i, T_j) \geq \beta$ for $i \neq j$, property (ii) of Lemma 10.2.1 ensures that the functions $\{h_i\}_{i=1}^r$ are disjointly supported.

Additionally property (i) implies that for each $i \in [r]$,

$$\mathcal{M}_{h_i}(V) \geq \mathcal{M}_F(T_i) \geq \delta \cdot \mathcal{M}_F(V).$$

and by property (iii) of Lemma 10.2.1, and since the supports are disjoint,

$$\sum_{i=1}^{r} \mathcal{E}_{h_i} = \sum_{u \sim v} \sum_{i=1}^{r} w(u, v) \|h_i(u) - h_i(v)\|^2 \leq 2 \left(1 + \frac{\alpha}{\epsilon}\right)^2 \sum_{u \sim v} w(u, v) \|F(u) - F(v)\|^2 \leq 2 \left(1 + \frac{\alpha}{\epsilon}\right)^2 \mathcal{E}_F\,.$$

Now we reorder the maps so that $\mathcal{E}_{h_1} \leq \mathcal{E}_{h_2} \leq \ldots \leq \mathcal{E}_{h_r}$, then for each $1 \leq i \leq r$, we have

$$\mathcal{E}_{h_i} \leq \frac{1}{r - i + 1} \cdot \sum_{j=i}^{r} \mathcal{E}_{h_j}.$$

Therefore, the preceding three inequalities imply (10.2.2). □

10.2.2 Merging

From Lemma 10.2.2, to find many disjointly supported functions with small Rayleigh quotient, it suffices to partition $(V, \hat{d}_F)$ into well separated regions, each of which contains a large fraction of the ℓ^2 mass of F. Given a random partition of the vertices using a padded random partition $\mathcal{P}$, in the next lemma we show that we can choose a random sample $P \sim \mathcal{P}$ and merge the sets in P to obtain r dense well-seperated sets for $k/2 \le r \le k$.

Lemma 10.2.3. *Let $r, k \in \mathbb{N}$ be given with $k/2 \le r \le k$, and suppose that the map $F : V \to \mathbb{R}^l$ is $(\Delta, \frac{1}{k} + \frac{k-r+1}{8kr})$-spreading for some $\Delta > 0$. Suppose additionally there is a random partition $\mathcal{P}$ with the properties that*

1. *For every $S \in \mathcal{P}$, $\operatorname{diam}(S, d_F) \le \Delta$, and*
2. *For every $v \in V$, $\mathbb{P}\left[B_{\hat{d}_F}(v, \Delta/\alpha) \subseteq \mathcal{P}(v)\right] \ge 1 - \frac{k-r+1}{4r}$.*

Then there exist r disjoint subsets $T_1, T_2, \ldots, T_r \subseteq V$ such that for each $i \ne j$, we have $\hat{d}_F(T_i, T_j) \ge 2\Delta/\alpha$, and for every $i = 1, 2, \ldots, k$,

$$\mathcal{M}_F(T_i) \ge \frac{1}{2k}\mathcal{M}_F(V).$$

Proof. For a subset $S \subseteq V$, define

$$\tilde{S} := \{v \in S : B_{\hat{d}_F}(v, \Delta/\alpha) \subseteq S\}.$$

By linearity of expectation, there exists a partition P such that for every $S \in P$, $\operatorname{diam}(S, d_F) \le \Delta$, and also

$$\sum_{S \in P} \mathcal{M}_F(\tilde{S}) \ge \left(1 - \frac{k-r+1}{4r}\right)\mathcal{M}_F(V). \tag{10.2.3}$$

Furthermore, by the spreading property of F, we have, for each $S \in P$,

$$\mathcal{M}_F(S) \le \frac{1}{k}\left(1 + \frac{k-r+1}{8r}\right)\mathcal{M}_F(V).$$

Therefore, we may take disjoint unions of the sets $\{\tilde{S} : S \in P\}$ to form at least r disjoint sets $T_1, T_2, \ldots, T_r$ with the property that for every $i = 1, 2, \ldots, r$, we have

$$\mathcal{M}_F(T_i) \ge \frac{1}{2k}\mathcal{M}_F(V)$$

because the first $r-1$ pieces will have total mass at most

$$\frac{r-1}{k}\left(1+\frac{k-r+1}{8r}\right)\mathcal{M}_F(V) \leq \left(1-\frac{k-r+1}{4r}-\frac{1}{2k}\right)\mathcal{M}_F(V),$$

for all $r \in [k/2, k]$, leaving at least $\frac{\mathcal{M}_F(V)}{2k}$ mass left over from (10.2.3). □

We mention a representative corollary that follows from the conjunction of Lemmas 8.1.6, 10.2.2 and 10.2.3.

Corollary 10.2.4. *Let $k \in \mathbb{N}$ and $\delta \in (0,1)$ be given. Suppose the map $F : V \to \mathbb{R}^l$ is $(\Delta, \frac{1}{k}+\frac{\delta}{48k})$-spreading for some $\Delta \leq 1$, and there is a random partition $\mathcal{P}$ with the properties that*

1. *For every $S \in \mathcal{P}$, $\operatorname{diam}(S, d_F) \leq \Delta$, and*
2. *For every $v \in V$, $\mathbb{P}\left[B_{\hat{d}_F}(v, \Delta/\alpha) \subseteq \mathcal{P}(v)\right] \geq 1 - \frac{\delta}{24}$.*

Then there are at least $r \geq \lceil (1-\delta)k \rceil$ disjointly supported functions $h_1, h_2, \ldots, h_r : V \to \mathbb{R}$ such that

$$\mathcal{R}(h_i) \lesssim \frac{\alpha^2}{\delta\Delta^2}\mathcal{R}(F).$$

Proof. In this case, we set $r = \lceil (1-\delta/2)k \rceil$ in our application of Lemma 10.2.3. After extracting at least $\lceil (1-\delta/2)k \rceil$ sets, we apply Lemma 10.2.2, but only take the first $r' = \lceil (1-\delta)k \rceil$ functions $h_1, h_2, \ldots, h_{r'}$. □

Note, in particular, that we can apply the preceding corollary with $\delta = \frac{1}{2k}$ to obtain $r = k$.

10.2.3 Random Partitioning

We now present some theorems applying our machinery to isotropic mappings of G to prove Theorem 10.1.5. In the proof we use several of tools in random partitioning of metric spaces discussed in Section 7.9.

Theorem 10.2.5. *For any $\delta \in (0,1)$, and any weighted graph $G = (V, E, w)$, there exist there exists $r \geq \lceil (1-\delta)k \rceil$ disjointly supported functions $h_1, h_2, \ldots, h_r : V \to \mathbb{R}$ such that*

$$\mathcal{R}_G(h_i) \lesssim \frac{k^2}{\delta^4}\lambda_k. \tag{10.2.4}$$

where λ_k is the kth smallest eigenvalue of $\mathcal{L}_G$. If G excludes K_c as a minor, then the bound improves to

$$\mathcal{R}_G(h_i) \lesssim \frac{c^4}{\delta^4}\lambda_k, \tag{10.2.5}$$

and if G has genus at most $g \geq 1$, then one gets

$$\mathcal{R}_G(h_i) \lesssim \frac{\log^2(g+1)}{\delta^4}\lambda_k\,. \tag{10.2.6}$$

Proof. Let $F : V \to \mathbb{R}^k$ be the spectral embedding of G with respect to the first k eigenfunctions of $\mathcal{L}$ as defined in equation (8.1.1). Choose $\Delta \asymp \sqrt{\delta}$ so that $(1-\Delta^2)^{-1} \leq 1 + \frac{\delta}{48}$. In this case, Lemma 8.1.2 and Lemma 8.1.3 imply that F is $(\Delta, \frac{1}{k} + \frac{\delta}{48k})$-spreading. Now, for general graphs, since d_F is Euclidean, we can use Theorem 7.9.2 applied to d_F to achieve $\alpha \asymp k/\delta$ in the assumptions of Corollary 10.2.4. Observe that $\hat{d}_F \geq d_F$, so that $B_{\hat{d}_F}(v, \Delta/\alpha) \subseteq B_{d_F}(v, \Delta/\alpha)$, meaning that we can satisfy both conditions (i) and (ii), verifying (10.2.4).

For (10.2.5) and (10.2.6), we use Theorems 7.9.4 and 7.9.5, respectively, applied to the shortest-path metric $\hat{d}_F$. Again, since $\hat{d}_F \geq d_F$, we have that $\operatorname{diam}(S, \hat{d}_F) \leq \Delta$ implies $\operatorname{diam}(S, d_F) \leq \Delta$, so conditions (i) and (ii) are satisfied with $\alpha \asymp c^2/\delta$ and $\alpha \asymp \log(g+1)/\delta$, respectively. □

Observe that the above proof extends to any isotropic mapping F where λ_k is replaced by $\mathcal{R}(F)$.

Next, we use Theorem 8.2.1 to give an alternate bound of $O(\delta^{-7}\log^2 k)\cdot\lambda_k$ for (10.2.4), which is better for moderate values of δ. We use the essentially the same proof, but first we perform a dimension reduction of the spectral embedding to a $O(\delta^{-2}\log k)$-dimensional space, and then we use random partitioning in a much lower dimensional space.

Theorem 10.2.6. *For any weighted graph $G = (V, E, w)$ and $\delta > 0$ the following holds. For every $k \in \mathbb{N}$, there exist $r \geq \lceil (1-\delta)k \rceil$ disjointly supported functions $h_1, h_2, \ldots, h_r : V \to \mathbb{R}$ such that*

$$\mathcal{R}(h_i) \lesssim \delta^{-9}\log^2(k)\lambda_k\,. \tag{10.2.7}$$

where λ_k is the k-th smallest eigenvalue of $\mathcal{L}_G$.

Proof. Let $F : V \to \mathbb{R}^k$ be the spectral embedding as defined in equation (8.1.1). We may clearly assume that $\delta \geq \frac{1}{2k}$. Choose $\Delta \asymp \delta$ so that $(1-16\Delta^2)^{-1}(1+4\Delta) \leq 1 + \frac{\delta}{48}$. By Lemma 8.1.2 and Lemma 8.1.3 F is $(4\Delta, (k(1-16\Delta^2))^{-1})$-spreading. In this case, for some choice of

$$l \lesssim \frac{1 + \log(k) + \log(41\Delta)}{\Delta^2} \lesssim \frac{O(\log k)}{\delta^2}\,,$$

with probability at least $1/2$, $\Gamma_{k,l}$ satisfies the conclusions of Theorem 8.2.1. Assume that $\Gamma : \mathbb{R}^k \to \mathbb{R}^l$ is some map satisfying these conclusions.

Then combining (ii) from Theorem 8.2.1 with Lemma 8.1.3, we see that $\Gamma \circ F : V \to \mathbb{R}^l$ is $(\Delta, \frac{1}{k} + \frac{\delta}{48k})$-spreading. Now we finish as in the proof of Theorem 10.2.5, using the fact that $l = O(\delta^{-2}\log k)$. □

10.2.4 Higher-order Cheeger inequalities

Finally, we can use theorems in preceding subsection in conjunction with Lemma 7.8.3 to produce many non-expanding sets, thus obtaining higher order Cheeger inequalities.

Theorem 10.2.7. *(Non-expanding k-partition) For any weighted graph $G = (V, E, w)$, there exists a partition $V = S_1 \cup S_2 \cup \cdots \cup S_k$ such that*

$$\phi_G(S_i) \lesssim k^4 \sqrt{\lambda_k}\,.$$

where λ_k is the kth smallest eigenvalue of $\mathcal{L}_G$. If G excludes K_c as a minor, then the bound improves to

$$\phi_G(S_i) \lesssim c^2 k^3 \sqrt{\lambda_k}\,,$$

and if G has genus at most $g \geq 1$, then one gets

$$\phi_G(S_i) \lesssim \log(g+1) k^3 \sqrt{\lambda_k}\,.$$

Proof. First apply Theorem 10.2.5 with $\delta = \frac{1}{2k}$ to find disjointly supported functions $h_1, h_2, \ldots, h_k : V \to \mathbb{R}$ satisfying (10.2.4). Now apply Lemma 7.8.3 to find sets $S_1, S_2, \ldots, S_k$ with $S_i \subseteq \mathrm{supp}(h_i)$ and $\phi_G(S_i) \leq \sqrt{2\mathcal{R}_G(h_i)}$ for each $i = 1, 2, \ldots, k$.

Now reorder the sets so that $w(S_1) \leq w(S_2) \leq \cdots \leq w(S_k)$, and replace S_k with the larger set $S_k' = V - (S_1 \cup S_2 \cup \cdots \cup S_{k-1})$ so that $V = S_1 \cup S_2 \cup \cdots \cup S_{k-1} \cup S_k'$ forms a partition. One can now easily check that

$$\phi_G(S_k') = \frac{w(S_k', \overline{S_k'})}{\mathrm{vol}(S_k')} \leq \frac{\sum_{i=1}^{k-1} w(S_i, \overline{S_i})}{\mathrm{vol}(S_k')} \leq k \cdot \max_{1 \leq i \leq k} \phi_G(S_i) \lesssim k^4 \sqrt{\lambda_k}\,.$$

A similar argument yields the other two bounds. □

Using Theorem Theorem 10.2.5 in conjunction with Lemma 7.8.3 again yields the following.

Theorem 10.2.8. *For every $\delta \in (0, 1)$ and any weighted graph $G = (V, E, w)$, there exist $r \geq \lceil (1-\delta)k \rceil$ disjoint sets $S_1, S_2, \ldots, S_r \subseteq V$ such that,*

$$\phi_G(S_i) \lesssim \frac{k}{\delta^2} \sqrt{\lambda_k}\,. \tag{10.2.8}$$

where λ_k is the kth smallest eigenvalue of $\mathcal{L}_G$. If G excludes K_c as a minor, then the bound improves to

$$\phi_G(S_i) \lesssim \frac{c^2}{\delta^2} \sqrt{\lambda_k}\,,$$

and if G has genus at most $g \geq 1$, then one gets

$$\phi_G(S_i) \lesssim \frac{\log(g+1)}{\delta^2}\sqrt{\lambda_k}\,.$$

We remark that the bound (10.2.8) will be improved, in various ways, in Section 10.4.

10.3 Gaps in the Spectrum

We now show that if there are significant gaps in the spectrum of G, one can obtain a higher-order Cheeger inequality with no dependence on k.

Theorem 10.3.1. *There is a constant $c > 0$ such that for any $0 < \delta < 1/3$ such that δk is an integer, if $c \cdot \frac{\log^2(k)}{\delta^9}\lambda_k < \lambda_{(1+\delta)k}$ then there are $(1-3\delta)k$ disjointly supported functions $f_1, \ldots, f_{(1-3\delta)k}$ such that $\mathcal{R}(f_i) \lesssim O(\lambda_k/\delta^3)$.*

Proof. We start similar to Theorem 10.2.6. Let $F : V \to \mathbb{R}^k$ be the spectral embedding as defined in equation (8.1.1). We may clearly assume that $\delta \geq 1/k$. Choose $\Delta \asymp \delta$ so that $(1-16\Delta^2)^{-1}(1+4\Delta) \leq 1 + \frac{\delta}{16}$. By Lemma 8.1.2 and Lemma 8.1.3 F is $(4\Delta, (k(1-16\Delta^2))^{-1})$-spreading. For some choice of

$$l \lesssim \frac{1 + \log(k) + \log(41\Delta)}{\Delta^2} \lesssim \frac{O(\log k)}{\delta^2}\,,$$

with probability at least $1/2$, $\Gamma_{k,l}$ satisfies the conclusions of Theorem 8.2.1. So, assume $\Gamma : \mathbb{R}^k \to \mathbb{R}^l$ is some map satisfying these conclusions, and let $F' = F \circ \Gamma$. Combining (ii) from Theorem 8.2.1 with Lemma 8.1.3, F' is (Δ, η)-spreading where $\eta = \frac{1}{k} + \frac{\delta}{16k}$ and $\mathcal{R}(F') \leq 8\mathcal{R}(F)$.

Since $d_{F'}$ is Euclidean, we can use Theorem 7.9.2 for $d_{F'}$ and $\alpha \asymp l/\delta$ to achieve a $(\Delta/4, \alpha, 1 - \delta/16)$ random partitioning $\mathcal{P}$. Recall that for a set $S \subseteq V$, $\tilde{S} := \{v \in S : B_{\hat{d}_F}(v, \Delta/\alpha) \subseteq S\}$. By linearity of expectation there is a partition P such that for any $S \in P$, $\operatorname{diam}(S, d_{F'}) \leq \Delta$ and

$$\sum_{S \in P} \mathcal{M}_{F'}(\tilde{S}) \geq (1 - \delta/16)\mathcal{M}_{F'}(V). \tag{10.3.1}$$

Let us order the sets of P with respect to $\mathcal{M}_{F'}(\tilde{S})$, i.e., let $\mathcal{M}_{F'}\tilde{S}_1) \leq \mathcal{M}_{F'}(\tilde{S}_2) \leq \ldots$. We consider two cases

Case 1: $\mathcal{M}_{F'}(\tilde{S}_{(1-2\delta)k}) > \eta \cdot \mathcal{M}_{F'}(V)/2$**.** First, we show that for any $1 \leq i, j \leq (1-2\delta)k$ such that $i \neq j$,

$$B_{d_{F'}}(S_i, \Delta/4) \cap B_{d_{F'}}(S_j, \Delta/4) = \emptyset.$$

Otherwise, let $S = B_{d_{F'}}(S_i, \Delta/4) \cup B_{d_{F'}}(S_j, \Delta/4)$. Then, $\operatorname{diam}(S, d_{F'}) \leq \Delta$ but $\mathcal{M}_{F'}(S) > \eta \cdot \mathcal{M}_F(V)$ which is a contradiction with the fact that F' is (Δ, η) spreading.

Now by applying Lemma 10.2.2 to $S_1, \ldots, S_{(1-2\delta)k}$ for $\alpha = 2$, $\epsilon = \Delta/4$ and $\delta = \eta$ we obtain $k(1-3\delta)$ disjointly supported functions $f_1, \ldots, f_{(1-3\delta)k}$ such that

$$\mathcal{R}(f_i) \lesssim \frac{1}{\eta \cdot \delta k \cdot \Delta^2} \mathcal{R}(F') = O(\lambda_k/\delta^3).$$

So, we are done.

Case 2: $\mathcal{M}_{F'}(\tilde{S}_{(1-2\delta)k}) \leq \eta \cdot \mathcal{M}_{F'}(V)/2$**.** We will reach to a contradiction with the fact that $\log^2(k)\lambda_k/\delta^9 \lesssim \lambda_{(1+\delta)k}$. Since F' is a (Δ, η) spreading, for any $i \leq (1-2\delta)k$,

$$\mathcal{M}_{F'}(\tilde{S}_i) \leq \eta \cdot \mathcal{M}_{F'}(V).$$

Therefore we can take disjoint unions of the sets $\{\tilde{S} : S \in P\}$ to form at least r disjoint sets $T_1, \ldots, T_r$ for $r = \lceil (1+3\delta/2)k \rceil$ such that

$$\mathcal{M}_{F'}(T_i) \geq \frac{1}{4k} \mathcal{M}_{F'}(V).$$

This is because the first $r-1$ piece will have total mass at most

$$\begin{aligned} \sum_{i=1}^{r-1} \mathcal{M}_{F'}(\tilde{S}_i) &\leq (1-2\delta)k \cdot \eta \cdot \mathcal{M}_{F'}(V) + \frac{7\delta k}{2} \max\left\{\frac{1}{2k}, \frac{\eta}{2}\right\} \mathcal{M}_{F'}(V) \\ &\leq (1-\delta/4)(1+\delta/16)\mathcal{M}_{F'}(V) \end{aligned}$$

leaving at least $\frac{\delta k}{8} \mathcal{M}_{F'}(V) \geq \mathcal{M}_{F'}(V)/4k$ mass left over from equation (10.3.1). Applying Lemma 10.2.2 to $T_1, \ldots, T_r$ for $\epsilon = \Delta/\alpha$, $\delta = 1/4k$, we obtain $(1+\delta)k$ disjoint supported functions $f_1, \ldots, f_{(1+\delta)k}$ such that

$$\mathcal{R}(f_i) \lesssim \frac{1}{\frac{1}{4k} \cdot \frac{\delta k}{2} \cdot \frac{\Delta^2}{\alpha^2}} \mathcal{R}(F') = \frac{c' \cdot \log^2(k)\lambda_k}{\delta^9}.$$

for some constant $c' > 0$. By Lemma 7.2.1 we get $\lambda_{(1+\delta)k} \leq c \cdot \log^2(k)\lambda_k/\delta^9$ for $c = 2c'$. But this condtriacts with the assumption of the theorem. So this case does not happen. □

Using the above theorem in conjunction with Lemma 7.8.3 yields the following.

Corollary 10.3.2. *There is a constant $c > 0$ such that for any $0 < \delta < 1/3$ such that δk is an integer, if $c \cdot \frac{\log^2(k)}{\delta^9} \lambda_k \leq \lambda_{(1+\delta)k}$, then there are $(1-3\delta)k$ disjoint sets $S_1, \ldots, S_{(1-3\delta)k}$ each of conductance $\phi(S_i) \leq O(\sqrt{\lambda_k/\delta^3})$.*

Let us conclude this section by describing the consequences of above results to the spectral clustering algorithms. The proof of the above theorem confirms the folklore believes that, in a spectral clustering algorithm, the number of clusters, is best chosen based on the largest gap in the

spectrum of the underlying graph. Additionally, the proof of the above theorem also provides a justification for the use of the k-means heuristic in spectral clustering algorithms (see Algorithm 2). Observe that in Case 1 (the only possible case under the assumptions of the theorem), the support of each of the functions $f_1, \ldots, f_{(1-3\delta)k}$ is *a ball* of radius Δ with respect to the metric $d_{F'}(.,.)$. In other words, in this case the vertices are concentrated in $\Theta(k)$ balls of small radius after the dimension reduction. It seems plausible that the k-means heuristic could successfully locate a good partition of the vertices in such a scenario.

10.4 A New Multiway Cheeger Inequality

A main result of this section is the following theorem.

Theorem 10.4.1. *For $k \in \{1, 2, \ldots, n\}$ and $\delta \in (0, 1)$, let $F : V \to \mathbb{R}^k$ be an isotropic mapping. Suppose that Then there exist $r \geq \lceil (1-\delta)k \rceil$ disjoint sets $S_1, S_2, \ldots, S_r \subseteq V$ such that for all $1 \leq i \leq k$,*

$$\phi(S_i) \leq \frac{1}{\delta^3} \sqrt{\mathcal{R}(F) \cdot \log(k)}\,.$$

Corollary 10.4.2. *For any weighted graph $G = (V, E, w)$, $k \in \{1, 2, \ldots, n\}$, and $\delta \in (0, 1)$, there exist $r \geq \lceil (1-\delta)k \rceil$ disjoint sets $S_1, S_2, \ldots, S_r \subseteq V$ with*

$$\phi_G(S_i) \lesssim \frac{1}{\delta^3} \sqrt{\lambda_k \log k}\,,$$

where λ_k is the kth smallest eigenvalue of $\mathcal{L}_G$.

Note that Theorem 10.2.6 combined with Lemma 7.8.3 is still not strong enough to prove Theorem 10.4.1. To do that, we need to combine Theorem 8.2.1 with a strong Cheeger inequality for Lipschitz partitions.

For $F : V \to \mathbb{R}^l$, let $t \in \{0, \max_v \|F(v)\|^2\}$ be chosen uniformly at random, and for any subset $S \subseteq V$, define

$$\hat{S} = \{v \in S : \|F(v)\|^2 \geq t\}\,.$$

Let $\mathbb{E}_t[.]$ be the expectation over random choice of t.

Lemma 10.4.3. *For every $\Delta > 0$, there exists a partition $V = S_1 \cup S_2 \cup \cdots \cup S_m$ such that for every $i \in [m]$, $\operatorname{diam}(S_i, d_F) \leq \Delta$, and*

$$\frac{\mathbb{E}_t\left[w(\hat{S}_1, \overline{\hat{S}_1}) + w(\hat{S}_2, \overline{\hat{S}_2}) + \cdots + w(\hat{S}_m, \overline{\hat{S}_m})\right]}{\mathbb{E}_t\left[\operatorname{vol}(\hat{S}_1) + \cdots + \operatorname{vol}(\hat{S}_m)\right]} \lesssim \frac{\sqrt{l}}{\Delta} \sqrt{\mathcal{R}(F)}\,. \tag{10.4.1}$$

Proof. Since the statement of the lemma is homogeneous in F, we may assume that $\max_v \|F(v)\| = 1$.

By Theorem 7.9.1, there exists a Δ-bounded random partition $\mathcal{P}$ satisfying, for every $u, v \in V$,

$$\mathbb{P}\left[\mathcal{P}(u) \neq \mathcal{P}(v)\right] \lesssim \frac{\sqrt{l}}{\Delta} \cdot d_F(u,v)\,. \tag{10.4.2}$$

Let $\mathcal{P} = S_1 \cup S_2 \cup \cdots \cup S_m$, where we recall that m is a random number and $S_1, \ldots, S_m$ are random sets.

First, observe that, $\mathbb{E}_t\left[\mathrm{vol}(\hat{S}_i)\right] = \sum_{v \in S_i} w(v)\|F(v)\|^2 = \mathcal{M}_F(S_i)$, thus,

$$\mathbb{E}_t\left[\mathrm{vol}(\hat{S}_1) + \cdots + \mathrm{vol}(\hat{S}_m)\right] = \mathcal{M}_F(V)\,. \tag{10.4.3}$$

Next, if $\{u,v\} \in E$ with $\|F(u)\|^2 \leq \|F(v)\|^2$, then we have

$$\begin{aligned}
&\mathbb{P}_{\mathcal{P},t}\left[\{u,v\} \in E(\hat{S}_1, \overline{\hat{S}_1}) \cup \cdots \cup E(\hat{S}_m, \overline{\hat{S}_m})\right] \\
&\quad\leq \mathbb{P}\left[\mathcal{P}(u) \neq \mathcal{P}(v)\right] \cdot \mathbb{P}_t\left[\|F(u)\|^2 \geq t \text{ or } \|F(v)\|^2 \geq t \mid \mathcal{P}(u) \neq \mathcal{P}(v)\right] \\
&\qquad + \mathbb{P}_t\left[t \in [\|F(u)\|^2, \|F(v)\|^2] \,\middle|\, \mathcal{P}(u) = \mathcal{P}(v)\right] \\
&\quad\lesssim \frac{\sqrt{l}}{\Delta} \cdot d_F(u,v)\left(\|F(u)\|^2 + \|F(v)\|^2\right) + \|F(v)\|^2 - \|F(u)\|^2 \\
&\quad\leq \left(\|F(u)\| + \|F(v)\|\right)\left(\frac{\sqrt{l}}{\Delta} \cdot d_F(u,v)(\|F(u)\| + \|F(v)\|) + \|F(v)\| - \|F(u)\|\right) \\
&\quad\leq \frac{5\sqrt{l}}{\Delta}\left(\|F(u)\| + \|F(v)\|\right)\|F(u) - F(v)\|,
\end{aligned}$$

where we use $\mathbb{P}_{\mathcal{P}}[.]$ to denote probability over the random choice of $\mathcal{P}$, and in the final line we have used Lemma 8.1.6.

Thus, we can use Cauchy-Schwarz to write,

$$\begin{aligned}
\mathbb{E}_{\mathcal{P},t}\left[w(\hat{S}_1, \overline{\hat{S}_1}) + \cdots + w(\hat{S}_m, \overline{\hat{S}_m})\right] &\lesssim \frac{\sqrt{l}}{\Delta} \sum_{\{u,v\} \in E} w(u,v)(\|F(u)\| + \|F(v)\|)\|F(u) - F(v)\| \\
&\leq \frac{\sqrt{l}}{\Delta} \sqrt{\sum_{\{u,v\} \in E} w(u,v)(\|F(u)\| + \|F(v)\|)^2} \\
&\qquad \cdot \sqrt{\sum_{\{u,v\} \in E} w(u,v)\|F(u) - F(v)\|^2} \\
&\leq \frac{\sqrt{l}}{\Delta}\sqrt{2\mathcal{M}_F(V)} \cdot \sqrt{\mathcal{E}_F}\,.
\end{aligned}$$

Combining this with (10.4.3) yields,

$$\frac{\mathbb{E}_{\mathcal{P}}\left[\mathbb{E}_t\left[w(\hat{S}_1,\overline{\hat{S}_1})+\cdots+w(\hat{S}_m,\overline{\hat{S}_m})\right]\right]}{\mathbb{E}_t\left[\mathrm{vol}(\hat{S}_1)+\cdots+\mathrm{vol}(\hat{S}_m)\right]} \lesssim \frac{\sqrt{l}}{\Delta}\cdot\sqrt{\mathcal{R}(F)}\,,$$

In particular, by Markov inequality a random partition P satisfy the above inequality with probability $1/2$. □

We can use the preceding theorem to find many non-expanding sets, assuming that $F : V \to \mathbb{R}^l$ has sufficiently good spreading properties.

Lemma 10.4.4. *Let $k \in \mathbb{N}$ and $\delta \in (0,1)$ be given. If the map $F : V \to \mathbb{R}^l$ is $(\Delta, \frac{1}{k}+\frac{\delta}{4k})$-spreading, then there exist $r \geq \lceil(1-\delta)k\rceil$ disjoint sets $T_1^*, T_2^*, \ldots, T_r^*$, such that*

$$\phi_G(T_i^*) \lesssim \frac{\sqrt{l}}{\delta\Delta}\sqrt{\mathcal{R}_G(F)}\,.$$

Proof. Let $V = S_1 \cup S_2 \cup \cdots \cup S_m$ be the partition guaranteed by applying Lemma 10.4.3 to the mapping $F : V \to \mathbb{R}^l$. Since F is $(\Delta, \frac{1}{k}+\frac{\delta}{4k})$-spreading and each S_i satisfies $\mathrm{diam}(S_i, d_F) \leq \Delta$, we can form $r' \geq \lceil(1-\delta/2)k\rceil$ sets $T_1, T_2, \ldots, T_{r'}$ by taking disjoint unions of the sets $\{S_i\}$ so that for each $i = 1, 2, \ldots, r'$, we have

$$\frac{\mathcal{M}_F(V)}{2k} \leq \mathcal{M}_F(T_i) \leq \frac{\mathcal{M}_F(V)}{k}\left(1+\frac{\delta}{4}\right).$$

In particular, $\mathbb{E}_t\left[\mathrm{vol}(\hat{T}_i)\right] = \mathcal{M}_F(T_i) \in [\frac{1}{2}\frac{\mathcal{E}}{k}, (1+\frac{\delta}{4})\frac{\mathcal{E}}{k}]$.

Order the sets so that $\mathbb{E}_t\left[w(\hat{T}_i,\overline{\hat{T}_i})\right] \leq \mathbb{E}_t\left[w(\hat{T}_{i+1},\overline{\hat{T}_{i+1}})\right]$ for $i = 1, 2, \ldots, r'-1$, and let $r = \lceil(1-\delta)k\rceil$. Then from (10.4.1), it must be that each $i = 1, 2, \ldots, r$ satisfies

$$\begin{aligned}\mathbb{E}_t\left[w(\hat{T}_i,\overline{\hat{T}_i})\right] &\lesssim \frac{1}{\delta k}\mathbb{E}_t\left[\sum_{j=1}^m w(\hat{S}_j,\overline{\hat{S}_j})\right]\\ &\lesssim \frac{\sqrt{l}}{\delta k\cdot\Delta}\cdot\sqrt{\mathcal{R}(F)}\cdot\mathbb{E}_t\left[\sum_{j=1}^m \mathrm{vol}(\hat{S}_j)\right] \lesssim \frac{\sqrt{l}}{\delta k\cdot\Delta}\cdot\sqrt{\mathcal{R}(F)}\cdot\mathcal{E}\,.\end{aligned}$$

But $\mathbb{E}_t\left[\mathrm{vol}(\hat{T}_i)\right] \asymp \mathcal{M}_F(V)/k$ for each $i = 1, 2, \ldots, r$, showing that

$$\frac{\mathbb{E}_t\left[w(\hat{T}_i,\overline{\hat{T}_i})\right]}{\mathbb{E}_t\left[\mathrm{vol}(\hat{T}_i)\right]} \lesssim \frac{\sqrt{l}}{\delta\Delta}\cdot\sqrt{\mathcal{R}(F)}\,.$$

Therefore, for each $1 \le i \le r$, there is a set T_i^* such that $\phi(T_i^*) \le \sqrt{l \cdot \mathcal{R}(F)}/\delta\Delta$. □

We can already use this to improve (10.2.8) in Theorem 10.2.8.

Theorem 10.4.5. *For every $\delta \in (0,1)$ and any weighted graph $G = (V, E, w)$, there exist $r \ge \lceil (1-\delta)k \rceil$ disjoint, non-empty sets $S_1, S_2, \ldots, S_r \subseteq V$ such that,*

$$\phi_G(S_i) \lesssim \frac{\sqrt{k}}{\delta^{3/2}} \sqrt{\lambda_k}\,. \tag{10.4.4}$$

where λ_k is the kth smallest eigenvalue of $\mathcal{L}_G$.

Proof. Let $\Delta \asymp \sqrt{\delta}$ be such that $(1-\Delta^2)^{-1} \le 1 + \frac{\delta}{4}$. If we take $F : V \to \mathbb{R}^k$ to be the spectral embedding coming from the first k eigenfunctions of $\mathcal{L}$, then Lemma 8.1.2 and Lemma 8.1.3 implies that F is $(\Delta, \frac{1}{k} + \frac{\delta}{4k})$-spreading. Now apply Lemma 10.4.4. □

Observe that setting $\delta = \frac{1}{2k}$ in the preceding theorem yields Theorem 10.1.1.

And now we can complete the proof of Theorem 10.4.1.

Proof of Theorem 10.4.1. Choose $\Delta \asymp \delta$ so that $(1 - 16\Delta^2)^{-1}(1 + 4\Delta) \le 1 + \frac{\delta}{4}$. In this case, for some choice of

$$l \lesssim \frac{1 + \log(k) + \log(41\Delta)}{\Delta^2} \lesssim \frac{O(\log k)}{\delta^2}\,,$$

with probability at least $1/2$, $\Gamma_{k,l}$ satisfies the conclusions of Theorem 8.2.1. Assume that $\Gamma : \mathbb{R}^k \to \mathbb{R}^l$ is some map satisfying these conclusions.

Then combining the conclusions of Theorem 8.2.1 with Lemma 8.1.3, we see that $F^* := \Gamma \circ F$ is $(\Delta, \frac{1}{k} + \frac{\delta}{4k})$-spreading, takes values in $\mathbb{R}^l$, and satisfies $\mathcal{R}(F^*) \le 8 \cdot \mathcal{R}(F)$. Now applying Lemma 10.4.4 yields the desired result. □

10.5 Noisy hypercubes

In the present section, we review examples for which equation (10.1.4) is tight. They also show that a poly-logarithmic dependency to k is necessary in the RHS of equation (10.1.2).

For $k \in \mathbb{N}$ and $\epsilon \in (0,1)$ let $H_{k,\epsilon} = (V, E)$ be the "noisy hypercube" graph, where $V = \{0,1\}^k$, and for any $u, v \in V$ there is an edge of weight $w(u,v) = \epsilon^{\|u-v\|_1}$. We put $n = |V| = 2^k$.

Theorem 10.5.1. *For any $1 \le C < k$ and $k \in \mathbb{N}$, and $S \subseteq V$ with $|S| \le Cn/k$, we have*

$$\phi_{H_{k,\epsilon}}(S) \gtrsim \sqrt{\lambda_k \log(k/C)}\,,$$

where $\epsilon = \frac{\log(2)}{\log(k/C)}$.

Proof. Let $H = H_{k,\epsilon}$. Since H is a Cayley graph on abelian group $(\mathbb{Z}/2\mathbb{Z})^k$, the eigenfunctions of $\mathcal{L}_H$ are $\chi_{\mathbf{r}}$ for any vector $S \subseteq [k]$ (see Section 7.3), where

$$\chi_S(v) = (-1)^{\sum_{i\in S} v(i)}.$$

First, the weighted degree of every vertex is

$$w(u) = \sum_{v\in V} \epsilon^{\|u-v\|_1} = (1+\epsilon)^k .$$

For any $S \subseteq [k]$, let λ_S be the eigenvalue of $\mathcal{L}_H$ corresponding to χ_S. By Theorem 7.3.3

$$\lambda_S = 1 - \frac{1}{(1+\epsilon)^k} \sum_{T\subseteq[k]} \chi_T(S) = \frac{1}{(1+\epsilon)^k} \sum_{T\subseteq[k]} \epsilon^T (-1)^{S\cap T} = 1 - \left(\frac{1-\epsilon}{1+\epsilon}\right)^{|S|}. \tag{10.5.1}$$

Since there are k sets of size 1, $\lambda_k \le 2\epsilon$. We will now show that for $S \subseteq V$ such that $|S| \le Cn/k$, one has $\phi_H(S) \ge \frac{1}{2}$, completing the proof of the theorem.

To bound $\phi_H(\cdot)$, we need to recall some Fourier analysis. For $f, g\colon \{0,1\}^k \to \mathbb{R}$ define the inner product:

$$\langle f, g\rangle_{L^2(V)} := \frac{1}{n} \sum_{v\in\{0,1\}^k} f(v)g(v).$$

The function χ_S form an orthonormal basis with respect to the above inner product. Therefore, any function $f : \{0,1\}^k \to \mathbb{R}$ has a unique representation as $f = \sum_{S\subseteq[k]} \widehat{f}(S)\chi_S$, where $\widehat{f}(S) := \langle f, \chi_S\rangle_{L^2(V)}$.

For $\eta \in [0,1]$, the Bonami-Beckner operator T_η is defined as

$$T_\eta f := \sum_{S\subseteq[k]} \eta^{|S|} \widehat{f}(S)\chi_S.$$

The Bonami-Beckner inequality [Bon70, Bec75] states that

$$\sum_{S\subseteq[k]} \eta^{|S|} \widehat{f}(S)^2 = \|T_{\sqrt{\eta}} f\|_2^2 \le \|f\|_{1+\eta}^2 = \left\{ \frac{1}{n} \sum_{v\in\{0,1\}^k} f(v)^{1+\eta} \right\}^{\frac{2}{1+\eta}}. \tag{10.5.2}$$

Let A be the normalized adjacency matrix of H, i.e. $A(u,v) = \frac{\epsilon^{|u\oplus v|}}{(1+\epsilon)^k}$. By (10.5.1), χ_S is an eigenfunction of A with corresponding eigenvalue $(\frac{1-\epsilon}{1+\epsilon})^{|S|}$, i.e.

$$A\chi_S = \left(\frac{1-\epsilon}{1+\epsilon}\right)^{|S|} \chi_S.$$

For $S \subseteq V(H)$, let $\mathbf{1}_S$ be the indicator function of S. Therefore,

$$\langle \mathbf{1}_S, A\mathbf{1}_S \rangle_{L^2(V)} = \sum_{T \subseteq [k]} \widehat{\mathbf{1}}_S(T)^2 \left(\frac{1-\epsilon}{1+\epsilon}\right)^{|T|} \leq \|\mathbf{1}_S\|^2_{\frac{2}{1+\epsilon}} = \left(\frac{|S|}{n}\right)^{1+\epsilon},$$

where the inequality follows from (10.5.2).

Now, observe that for any $S \subseteq V$, we have

$$w(S, \overline{S}) = \mathrm{vol}(S) - w(S) = \mathrm{vol}(S) - (1+\epsilon)^k n \langle \mathbf{1}_S, A\mathbf{1}_S \rangle_{L^2(V)}$$

where we have written $w(S) = \sum_{u,v \in S} w(u,v)$.

Hence, for any subset $S \subseteq V$ of size $|S| \leq Cn/k$, we have

$$\phi_H(S) = \frac{w(S,\overline{S})}{w(S)} = \frac{|S| - n\langle \mathbf{1}_S, A\mathbf{1}_S \rangle_{L^2(V)}}{|S|} \geq 1 - \left(\frac{|S|}{n}\right)^{\epsilon} \geq 1 - (k/C)^{-\epsilon} \geq \frac{1}{2}$$

where the last inequality follows by the choice of $\epsilon = \log(2)/\log(k/C)$. □

Remark 10.5.2. *The preceding theorem shows that even if we only want to find a set S of size $n/\sqrt{k}$, then for values of $k \leq O(\log n)$, we can still only achieve a bound of the form $\phi_H(S) \lesssim \sqrt{\lambda_k \log k}$. The state of affairs for $k \gg \log n$ is a fascinating open question.*

10.6 Conclusion

In Subsection 10.1.3, we gave a generic outline of our spectral partitioning algorithm. We remark that our instantiations of this algorithm are simple to describe. As an example, suppose we are given a weighted graph $G = (V, E, w)$ and want to find k disjoint sets, each of expansion $O(\sqrt{\lambda_{2k} \log k})$ (recall Theorem 10.1.3). We specify a complete randomized algorithm.

One starts with the spectral embedding $F : V \to \mathbb{R}^k$, given by $F(v) = (f_1(v), f_2(v), \ldots, f_{2k}(v))$, where $f_1, f_2, \ldots, f_{2k}$ is the $\ell^2(V,w)$-orthogonal system comprised of the first $2k$ eigenfunctions of $\mathcal{L}$. Then, for some $l = O(\log k)$, we perform random projection into $\mathbb{R}^l$. Let $\Gamma_{2k,l} : \mathbb{R}^{2k} \to \mathbb{R}^l$ be the random linear map given by

$$\Gamma_{2k,l}(\mathbf{x}) = \frac{1}{\sqrt{l}} \left(\langle \boldsymbol{\zeta}_1, \mathbf{x} \rangle, \ldots, \langle \boldsymbol{\zeta}_l, \mathbf{x} \rangle\right),$$

where $\{\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_l\}$ are i.i.d. $2k$ dimensional Gaussian vectors. We now have an embedding $F^* := \Gamma_{2k,l} \circ F : V \to \mathbb{R}^l$.

Next, for some $R = \Theta(1)$, we perform the random space partitioning algorithm from [CCG+98]. Let $\mathcal{B}$ denotes the closed Euclidean unit ball in $\mathbb{R}^l$. Consider $V \subseteq \mathcal{B}$ by identifying each vertex with

its image under the map $v \mapsto F^*(v)/\|F^*(v)\|$. If $\{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$ is an i.i.d. sequence of points in $\mathcal{B}$ (chosen according to the Lebesgue measure), then we form a partition of V into the sets

$$V = \bigcup_{i=1}^{\infty} \Big[V \cap B(\mathbf{x}_i, R) - (B(\mathbf{x}_1, R) \cup \cdots \cup B(\mathbf{x}_{i-1}, R)) \Big]$$

Here, $B(\mathbf{x}, R)$ represents the closed Euclidean ball of radius R about $\mathbf{x}$, and it is easy to see that this induces a partition of V in a finite number of steps with probability one. Let $V = S_1 \cup S_2 \cup \cdots \cup S_m$ be this partition.

We sort the partition $\{S_1, S_2, \ldots, S_m\}$ in decreasing order according to $\mathcal{M}_{F^*}(S_i)$. Let $k' = \lceil \frac{3}{2}k \rceil$. Then for each $i = k'+1, k'+2, \ldots, m$, we iteratively set $S_a := S_a \cup S_i$ where

$$a = \operatorname{argmin}\{\mathcal{M}_{F^*}(S_j) : j \leq k\}.$$

(Intuitively, we form k' sets from our total of $m \geq k'$ sets by balancing the $\mathcal{M}_{F^*}(\cdot)$-value among them.) At the end, we are left with a partition $V = S_1 \cup S_2 \cup \cdots \cup S_{k'}$ of V into $k' \geq 3k/2$ sets.

To complete the algorithm, for each $i = 1, 2, \ldots, k'$, we choose a value t such that

$$\hat{S}_i = \{v \in S_i : \|F^*(v)\|^2 \geq t\}$$

has the least expansion. We then output k of the sets $\hat{S}_1, \hat{S}_2, \ldots, \hat{S}_{k'}$ that have the smallest expansion.

Chapter 11

Improved Cheeger's Inequality

In this chapter we improve the Cheeger's inequality (Theorem 7.8.1) using higher eigenvalues of normalized Laplacian matrix, $\mathcal{L}$. Our result shows that the spectral partitioning algorithm (Algorithm 10) is a constant factor approximation algorithm for finding a set with smallest conductance if λ_k is a constant for some constant k. This provides some theoretical justification to the empirical performance of spectral partitioning algorithm in image segmentation and clustering problems. We extend the analysis to other graph partitioning problems, including multi-way partition, balanced separator and provide improved approximation algorithms.

The results of this chapter are based on a joint work with Tsz Chiu Kwok, Lap Chi Lau, Yin Tat Lee and Luca Trevisan [KLL+13].

11.1 Introduction

Spectral partitioning algorithm provides a constant factor approximation algorithm to the sparsest cut problem when λ_2 is an absolute constant. But the approximation factor can be $\Omega(n)$ for general graphs when λ is very small (see Example 7.8.7 for some examples). Nonetheless, this algorithm has been quite successful in practice in image segmentation [SM00, TM06] or community detection [LLM10]. Our main goal in this chapter is to justify this success and provide a tighter analysis of Cheeger's inequality and in particular the spectral partitioning algorithm.

In Subsection 7.8.2 we showed that both sides of Cheeger's inequality are tight (up to constant factors). The left side is tight when G is a hypercube, and the right side is tight when G is a cycle. Our main result of this chapter is an improved variant of the RHS of Cheeger's inequality using higher eigenvalues of the normalized Laplacian matrix.

Theorem 11.1.1. *For every undirected graph G and any $k \geq 2$, it holds that*

$$\phi(G) = O(k)\frac{\lambda_2}{\sqrt{\lambda_k}}.$$

Furthermore, this guarantee is achieved by the spectral partitioning algorithm, Algorithm 10 when given an eigenfunction of λ_2.

This improves Cheeger's inequality, as it shows that λ_2 is a better approximation of $\phi(G)$ when there is a large gap between λ_2 and λ_k for any $k \geq 3$. It also shows that the spectral partitioning algorithm is a $O(k/\sqrt{\lambda_k})$-approximation algorithm to the sparsest cut problem, even though it does not employ any information about higher eigenvalues or higher eigenfunctions. This was quite surprising for us! In particular, it shows that spectral partitioning algorithm provides a constant factor approximation to $\phi(G)$ when λ_k is a constant for some constant k, i.e., when G is a low threshold rank graph. Next we provide an explicit example where the above theorem provides a significantly tighter analysis of the spectral partitioning algorithm.

Example 11.1.2 (Generalized Barbell graph)**.** *Let G consist of k cliques of size $\sim n/k$ joined in a cycle (see Figure 11.1.1 for an illustration). Observe that $\phi(G) = \Theta(k/n)$. Now, for $i = 1, 2$ define f_i to be the function that is 1 on the $ik/4$-th clique and goes to 0 linearly on the neighboring cliques. That is f_i is equal to $1 - 4j/k$ on the clique $ik/4 \pm j$, for $1 \leq j \leq k/4$. It is straightforward that $\mathcal{R}(f_i) = \Theta(k^2/n^2)$ for $i = 1, 2$. Thus, $\lambda_2 = \Theta(k^2/n^2)$.*

Above calculations show that by Theorem 7.8.1 spectral partitioning gives an $O(\sqrt{\lambda_2}) = O(n/k)$ approximation algorithm to $\phi(G)$. But since $\lambda_{2k} \geq 1/\log(k)$, our theorem shows spectral partitioning has a significantly better performance. First observe that in any $k + 1$ disjoint sets $S_1, \ldots, S_{k+1}$, there is one, say S_1, that has at most half of the vertices of each of the cliques. Thus, $\phi(S_1) = \Omega(1)$, and consequently, $\rho(k + 1) = \Omega(1)$. By equation (10.1.4) we get $\lambda_{2k} = \Omega(1/\log(k))$. Now, by Theorem 11.1.1 spectral partitioning algorithm provides an $O(k \log(k))$ approximation to $\phi(G)$.

The bound in Theorem 11.1.1 is tight up to a constant factor for any $k \geq 2$. Let G be a cycle of length n. As we discussed in Subsection 7.3.1 for $k \leq n/100$, $\lambda_k = \Theta(k^2/n^2)$. Since $\phi(G) = \Theta(1/n)$, we get $\phi(G) = \Theta(k\lambda_2/\sqrt{\lambda_k})$ for any $2 \leq k \leq n/100$.

Noting our result, one may try to even improve the left side of Cheeger's inequality using the higher eigenvalues of $\mathcal{L}$. This is a very interesting open problem and we leave it open for future works.

11.1.1 Improvements on Applications of Cheeger's Inequality

Once we improve the Cheeger's inequality we can apply the same machinery to extensions and generalizations of this inequality. Roughly speaking, when we have inequalities with the terms $\sqrt{\lambda_k}$,

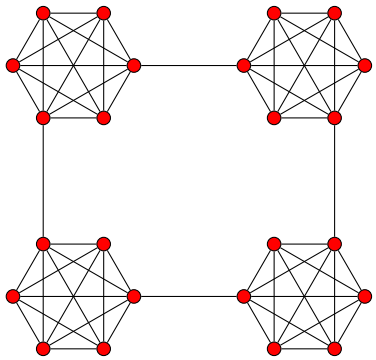

Figure 11.1.1: An example a graph where Theorem 11.1.1 provide significantly tighter analysis of spectral partitioning algorithm compared to Theorem 7.8.1. In this graph each clique has size n/k and cliques placed around a cycle. $\phi(G) = \Theta(n/k)$, $\lambda_2 = \Theta(k^2/n^2)$, and $\lambda_{2k} = \Omega(1/\log(k))$.

we may to use our machinery to replace this with $\lambda_k/\sqrt{\lambda_l}$ for some $l > k$ at the cost of losing a factor of poly(l/k).

Our first applications are improvements of many of our results proved in Chapter 10.

Corollary 11.1.3. *For any graph $G = (V, E, w)$ and any $l > k \geq 2$,*

(i)

$$\rho(k) \leq O(lk^6)\frac{\lambda_k}{\sqrt{\lambda_l}}.$$

(ii) For any $\delta \in (0, 1)$,

$$\rho(k/2) \leq O\left(\frac{l\log^2 k}{k}\right)\frac{\lambda_k}{\sqrt{\lambda_l}}.$$

(iii) If G excludes K_c as a minor,

$$\rho(k/2) \leq O\left(\frac{c^4 l}{k}\right)\frac{\lambda_k}{\sqrt{\lambda_l}}.$$

Part (i) shows that λ_k is a better approximation of $\rho(k)$ when there is a large gap between λ_k and λ_l for any $l > k$. Part (ii) implies that $\rho(k/2) \leq O(\lambda_k \log^2 k/\sqrt{\lambda_{2k}})$. The factor 2 can be replaced by $(1 - \delta)$ where the leading constant in the RHS is δ^{10}. Similarly part (iii) implies that $\rho(k/2) \leq O(\lambda_k/\sqrt{\lambda_{2k}})$ for planar graphs.

Furthermore, our proof shows that the spectral algorithms discussed in Chapter 10 achieve the above performance. For instance, if λ_l is a constant for a constant $l > k$, there is a constant factor approximation algorithm for the k-way partitioning problem.

Our second application of Theorem 11.1.1 is the minimum bisection problem. In the minimum bisection problem, the objective is to find a set S with minimum conductance among the sets with $|V|/2$ vertices. While it is very nontrivial to find a sparse cut with exactly $|V|/2$ vertices [FK02, R08],

it is well known that a simple recursive spectral algorithm can find a *balanced separator* S with $\phi(S) = O(\sqrt{\epsilon})$ with $|S| = \Omega(|V|)$, where ϵ denotes the conductance of the minimum bisection (e.g. [KVV04]). We use Theorem 11.1.1 to generalize the recursive spectral algorithm to obtain a better approximation guarantee when λ_k is large for a small k.

Theorem 11.1.4. *Let*

$$\epsilon := \min_{\mathrm{vol}(S)=\mathrm{vol}(V)/2} \phi(S).$$

There is a polynomial time algorithm that finds a set S such that $\mathrm{vol}(V)/5 \leq |S| \leq 4\mathrm{vol}(V)/5$ *and* $\phi(S) \leq O(k\epsilon/\lambda_k)$.

We remark that although a bisection is defined as a cut with equal number of vertices in the both sides, above corollary finds a cut with (approximately) equal volume in both sides of the cut. This is a limitation of spectral algorithms. Nonetheless, the applications are very similar (e.g. we can use above theorem in divide and conquer algorithms to partition the graph into small pieces with few edges in between).

11.1.2 More Related Work

Recently, several approximation algorithms are designed for various classes of problems on low threshold rank graphs (see Subsection 7.7.2). Recall that the subspace enumeration algorithm of [ABS10] provides an $O(1/\lambda_k)$ approximation algorithm for the sparsest cut problem in time $n^{O(k)}$. The algorithm simply searches for a sparse cut in the $(k-1)$-dimensional eigen-space corresponding to $\lambda_1, \ldots, \lambda_{k-1}$. It is worth noting that for $k = 3$ the subspace enumeration algorithm is exactly the same as the spectral partitioning algorithm. Nonetheless, the result in [ABS10] is incomparable to Theorem 11.1.1 since it does not upper bound $\phi(G)$ by a function of λ_2 and λ_3. Although algorithms designed specifically for low threshold rank graphs [ABS10, BRS11, GS11] provide much better approximation guarantees for large values of k compare to Theorem 11.1.1, our results show that simple spectral algorithms provide nontrivial performance guarantees.

Another direction to show that spectral algorithms work well is to analyze their performance in a family of graphs with a fixed combinatorial property. Spielman and Teng [ST96] showed that $\lambda_2 = O(1/n)$ for a bounded degree planar graph, they use this to show that a spectral algorithm can find a separator of size $O(\sqrt{n})$ in such graphs. This result is extended to bounded genus graphs by Kelner [Kel04] and to fixed minor free graphs by Biswal, Lee and Rao [BLR08]. This is further extended to higher eigenvalues by Kelner, Lee, Price and Teng [KLPT11]: $\lambda_k = O(k/n)$ for planar graphs, bounded genus graphs, and fixed minor free graphs when the maximum degree is bounded. Recall that combining these results with Theorem 10.2.8 implies that $\rho(k) = O(\sqrt{k/n})$ for bounded degree planar graphs. Note that these results give mathematical bounds on the conductances of the resulting partitions, but they do not imply that the approximation guarantee of Cheeger's inequality

could be improved for these graphs, neither does our result as these graphs have slowly growing spectrums.

11.1.3 Proof Overview

We start by describing an informal intuition of the proof of Theorem 11.1.1 for $k = 3$, and then we describe how this intuition can be generalized. Let f be a non-constant function such that $\mathcal{R}(f) \approx \lambda_2$.

Suppose λ_2 is small and λ_3 is large. Then the higher order Cheeger's inequality implies that there is a partitioning of the graph into two sets of small conductance, but in every partitioning into at least three sets, there is a set of large conductance. So, we expect the graph to have a sparse cut of which the two parts are expanders; see Chapter 13 for quantitative statements. Since $\mathcal{R}(f)$ is small and f is orthogonal to the constant function, we expect that the vertices in the same expander have similar values in f and the average values of the two expanders are far apart. Hence, f is similar to a step function with two steps representing a cut, and we expect that $\mathcal{R}(f) \approx \phi(G)$ in this case (see Theorem 10.3.1 for high dimensional variant of this observation). Therefore, roughly speaking, $\lambda_3 \gg \lambda_2$ implies $\lambda_2 \approx \phi(G)$.

Conversely, Theorem 11.1.1 shows that if $\lambda_2 \approx \phi^2(G)$, then $\lambda_3 \approx \lambda_2$. One way to prove that $\lambda_2 \approx \lambda_3$ is to find a function f' of Rayleigh quotient close to λ_2 such that f' is orthogonal to both f and the constant function. For example, if G is a cycle, then $\lambda_2 = \Theta(1/n^2)$, $\phi(G) = \Theta(1/n)$, and f (up to normalizing factors) could represent the cosine function. In this case we may define f' to be the sine function (see Subsection 7.3.1 for a discussion on eigenvalues/eigenfunctions of a cycle). Unfortunately, finding such a function f' in general is not as straightforward. Instead, we find three *disjointly supported* functions f_1, f_2, f_3 of Rayleigh quotient close to λ_2 and by Lemma 7.2.1 this would upper bound λ_3 by $2\max\{\mathcal{R}(f_1), \mathcal{R}(f_2), \mathcal{R}(f_3)\}$. For the cycle example, if f is the cosine function, we may construct f_1, f_2, f_3 simply by first dividing the support of f into three disjoint intervals and then constructing each f_i by defining a smooth localization of f (similar to Lemma 10.2.1) in one of those intervals. To ensure that $\max\{\mathcal{R}(f_1), \mathcal{R}(f_2), \mathcal{R}(f_3)\} \approx \lambda_2$ we need to show that f is a "smooth" function, whose values change continuously. We make this rigorous by showing that if $\lambda_2 \approx \phi(G)^2$, then the function f must be smooth. Therefore, we can construct three disjointly supported functions based on f and show that $\lambda_2 \approx \lambda_3$.

In [KLL+13] we present two proofs of Theorem 11.1.1 where the first proof generalizes the first observation. In this thesis we only include the proof based on the second observation. Say $\mathcal{R}(f) \approx \phi(G)^2$. We partition the support of f into disjoint intervals of the form $[2^{-i}, 2^{-(i+1)}]$, and we show that the vertices are distributed almost uniformly in most of these intervals in the sense that if we divide $[2^{-i}, 2^{-(i+1)}]$ into $\Theta(k)$ equal length subintervals, then we expect to see the same amount of mass in the subintervals. This shows that f is a smooth function. We then argue that $\lambda_k \lesssim k^2\lambda_2$, by constructing k disjointly supported functions, $f_1, \ldots, f_k$, each of Rayleigh quotient

$O(k^2)\mathcal{R}(f)$.

To construct $f_1, \ldots, f_k$ we use the machinery that we developed in Section 10.2. We use the smoothness of f to find k well separated regions each covering $\Omega_k(1)$ fraction of the mass of f. Then, we use Lemma 10.2.1 to localize these regions and construct $f_1, \ldots, f_k$. The main difference with the proofs in Chapter 10 is that we use a different metric, *normalized euclidean metric*.

11.2 Proof of the Main Result

Throughout this section we assume that $f \in \ell^2(V, w)$ is a non-negative function of norm $\|f\|_w^2 = 1$ such that $\text{vol}(\text{supp}(f)) \leq \text{vol}(V)/2$. Such a function can be obtained using Lemma 7.2.2 (or Lemma 7.2.3 if we are given an eigenfunction of $\mathcal{L}$). Recall that for a threshold $t \in \mathbb{R}$, a threshold set of f is defined as $S_f(t) := \{v : f(v) \geq t\}$. We let

$$\phi(f) := \min_{t>0} \phi(S_f(t)).$$

be the conductance of the best threshold set of the function f.

Instead of directly proving Theorem 11.1.1 we prove a stronger version, as it will be used later to prove Corollary 11.1.3. In particular, instead of directly upper bounding λ_k, we construct k disjointly supported functions with small Rayleigh quotients.

Theorem 11.2.1. *For any non-negative function $f \in \ell^2(V, w)$ such that $\text{vol}(\text{supp}(f)) \leq \text{vol}(V)/2$, at least one of the following holds*

i) $\phi(f) \leq O(k)\mathcal{R}(f)$;

ii) There exist k disjointly supported functions $f_1, f_2, \ldots, f_k$ such that for all $1 \leq i \leq k$, $\text{supp}(f_i) \subseteq \text{supp}(f)$ and

$$\mathcal{R}(f_i) \leq O(k^2)\mathcal{R}(f)/\delta.$$

where and $\delta \asymp \phi^2(f)/\mathcal{R}(f)$. Furthermore, the support of each f_i is an interval $[a_i, b_i]$ such that $|a_i - b_i| = \Theta(1/k)a_i$.

Observe that Theorem 11.1.1 can be obtained from above theorem and an application of Lemma 7.2.1. We will show that if $\mathcal{R}(f) = \Theta(\phi(G)^2)$ (i.e., $\delta = \Theta(1)$), then f is a smooth function of the vertices, in the sense that in any interval of the form $[t, 2t]$ we expect the vertices to be embedded in equidistance positions. It is instructive to verify this for the second eigenvector of the cycle.

11.2.1 Normalized Euclidean Distance Function

In this section we introduce the *normalized euclidean* distance function $d^{\diamond}(.,.)$. For a function $F : V \to \mathbb{R}$, and $u, v \in V$,

$$d_F^{\diamond}(u,v) = \frac{\|F(u) - F(v)\|}{\max\{\|F(u)\|, \|F(v)\|\}}.$$

if $\max\{\|F(u)\|, \|F(v)\|\} > 0$, and $d_F^{\diamond}(u,v) = 0$ otherwise. In the next lemma we show that $d_F^{\diamond}(.,.)$ indeed defines a metric.

Lemma 11.2.2. *For any $F : V \to \mathbb{R}^l$, and $v_1, v_2, v_3 \in V$ and $f : V \to \mathbb{R}$,*

$$d_F^{\diamond}(v,u) \leq d_F^{\diamond}(v,v') + d_F^{\diamond}(u,v').$$

Proof. If $\max\{\|F(v)\|, \|F(u)\|\} = \max\{\|F(v)\|, \|F(u)\|, \|F(v')\|\}$, then

$$\max\{\|F(v)\|, \|F(u)\|\} \cdot (d_F^{\diamond}(v,v') + d_F^{\diamond}(u,v')) \geq \|F(v) - F(v')\| + \|F(u) - F(v')\| \geq \|F(v) - F(u)\|,$$

and we are done. Otherwise, it must be that $\|F(v')\| = \max\{\|F(v)\|, \|F(u)\|, \|F(v')\|\}$.

The rest of the proof is based on a proof by S.B. in `math.stackexchange.com`. Without loss of generality assume that $F(v), F(u) \neq 0$. For a vertex $v \in V$, let $\Gamma(v) = F(v)/\|F(v)\|^2$. First observe that for any $u, v \in V$,

$$\|\Gamma(u) - \Gamma(v)\| = \frac{\|F(u) - F(v)\|}{\|F(v)\| \cdot \|F(v)\|}. \tag{11.2.1}$$

This is because,

$$\|\Gamma(u) - \Gamma(v)\|^2 = \frac{1}{\|F(u)\|^2} + \frac{1}{\|F(v)\|^2} - \frac{\langle F(u), F(v)\rangle}{\|F(u)\| \cdot \|F(v)\|} = \frac{\|F(u) - F(v)\|^2}{\|F(u)\|^2 \cdot \|F(v)\|^2}$$

Therefore,

$$\begin{aligned}
\max\{\|F(v)\|, \|F(u)\|\} \quad \cdot \quad & (\|F(v) - F(v')\| + \|F(u) - F(v')\|) \\
\geq \quad & \|F(v) - F(v')\| \cdot \|F(u)\| + \|F(u) - F(v')\| \cdot \|F(v)\| \\
= \quad & \|F(u)\| \cdot \|F(v)\| \cdot \|F(v')\| \cdot (\|\Gamma(v) - \Gamma(v')\| + \|\Gamma(u) - \Gamma(v')\|) \\
\geq \quad & \|F(u)\| \cdot \|F(v)\| \cdot \|F(v')\| \cdot \|\Gamma(v) - \Gamma(u)\| \\
= \quad & \|F(v')\| \cdot \|F(u) - F(v)\|.
\end{aligned}$$

where the second equation follows by (11.2.1) and the third equation follows by the triangle inequality. □

Fact 11.2.3. *For any $F : V \to \mathbb{R}^l$, and $u, v \in V$,*

$$\|F(u)\| \cdot d_F^{\diamond}(u, v)| \leq \|F(u) - F(v)\| .$$

Our overall idea to construct $f_1, \ldots, f_k$ is very similar to what we did in Section 10.2. We find well-separated regions with respect to f such that each of them contains $\Omega_k(1)$ fraction of the total mass.

The following corollary follows from Lemma 10.2.2, Fact 11.2.3, and that $\mathcal{M}_f(V) = \|f\|_w = 1$.

Corollary 11.2.4. *Let $S_1, S_2, \ldots, S_{2k} \subseteq V$ such that $\mathcal{M}_f(S_i) \geq \alpha$ and $d_f^{\diamond}(S_i, S_j) \geq 2\epsilon$ for all $1 \leq i \leq j \leq 2k$. Then, there are k disjointly supported functions $f_1, \ldots, f_k$, each supported on the ϵ-neighborhood of one of the regions such that*

$$\forall\ 1 \leq i \leq k,\ \mathcal{R}(f_i) \leq \frac{2\mathcal{R}(f)}{k \cdot \alpha}(1 + 1/\epsilon)^2.$$

11.2.2 Construction of Dense Well Separated Regions

For $t_1, t_2 \in \mathbb{R}$, we define the interval $[t_1, t_2]$,

$$[t_1, t_2] := \{x \in \mathbb{R} : \min\{t_1, t_2\} < x \leq \max\{t_1, t_2\}\}.$$

In this chapter all intervals are defined to be closed on the larger value and open on the smaller value. For an interval $I = [t_1, t_2] \subseteq \mathbb{R}$, we use $\text{len}(I) := |t_1 - t_2|$ to denote the length of I. For a function $f \in \mathbb{R}^V$, we abuse the notation and use $I_f := \{v : f(v) \in I\}$ to denote the vertices within I.

Let $0 < \alpha < 1$ be a constant that will be fixed later in the proof. For $i \in \mathbb{Z}$, we define the interval $I(i) := [\alpha^i, \alpha^{i+1}]$. Observe that these intervals partition the vertices with positive value in f. We partition each interval I_i into $12k$ subintervals of equal length,

$$I(i, j) := \left[\alpha^i \left(1 - \frac{j(1-\alpha)}{12k}\right), \alpha^i \left(1 - \frac{(j+1)(1-\alpha)}{12k}\right)\right],$$

for $0 \leq j < 12k$. Observe that for all i, j,

$$\text{len}(I(i, j)) = \frac{\alpha^i(1-\alpha)}{12k}. \tag{11.2.2}$$

We say a subinterval $I(i, j)$ is *heavy*, if

$$\mathcal{M}(I_f(i, j)) \geq \frac{\delta \cdot \mathcal{M}(I_f(i-1))}{k},$$

where $c > 0$ is a constant that will be fixed later in the proof; we say it is *light* otherwise. We also say

an interval $I(i)$ is *balanced* if the number of heavy sub-intervals is at least $6k$. We use $\mathcal{B}$ to denote the set of balanced intervals. Intuitively, an interval $I(i)$ is *balanced* if the vertices are distributed uniformly inside that interval.

Next we describe our proof strategy. Using Corollary 11.2.4 to prove the theorem it is sufficient to find $2k$ sets each covering $\Omega(\delta/k)$ mass of f such that their distance is $\Omega(1/k)$. Each of our $2k$ sets will be a union of vertices in *heavy* subintervals. Our construction is simple: from each balanced interval we choose $2k$ *separated* heavy subintervals and include each of them in one of the sets. In order to promise that the sets are well separated, once we include a heavy subinterval, say $I_f(i,j)$, in a set S we leave the two neighboring subintervals $I(i,j-1)$ and $I(i,j+1)$ unassigned, so as to separate S from the rest of the sets. In particular, for all $1 \le a \le 2k$ and all $I(i) \in \mathcal{B}$, we include the $(3a-1)$-th heavy subinterval of $I(i)$ in S_a. Note that if an interval $I(i)$ is balanced, then it has $6k$ heavy subintervals and we can include one heavy subinterval in each of the $2k$ sets. Consequently, for $1 \le a < b \le 2k$,

$$d_f^\diamond(S_a, S_b) \ge \max_{i\in\mathbb{Z}} \max_{0\le j<12k} \frac{\mathrm{len}(I(i,j))}{\alpha^2} = \frac{(1-\alpha)}{12k}.$$

where the last equality follows by equation (11.2.2), It remains to prove that these $2k$ regions are dense. Let

$$\Delta := \sum_{I(i)\in\mathcal{B}} \mathcal{M}(I(i-1))$$

be the summation of the mass of the preceding interval of balanced intervals. Then, since each heavy subinterval $I(i,j)$ has a mass of $\delta \cdot \mathcal{M}(I(i-1))/k$, by the above construction all regions are $\delta\cdot\Delta/k$-dense. Hence, the following proposition follows from Corollary 11.2.4.

Proposition 11.2.5. *There are k disjoint supported functions $f_1, \ldots, f_k$ such that for all $1 \le i \le k$,* $\mathrm{supp}(f_i) \subseteq \mathrm{supp}(f)$ *and*

$$\forall\ 1 \le i \le k,\ \mathcal{R}(f_i) \le \frac{1250k^2\mathcal{R}(f)}{(1-\alpha)^2\cdot\delta\cdot\Delta}.$$

11.2.3 Lower Bounding the Energy

Our approach to lower bound Δ is by lower bounding the energy of edges inside each light subinterval. Then, we use this to upper bound $\mathcal{M}(I(i-1))$ for when $I(i)$ is not balanced.

We define the energy of f *restricted* to an interval I as follows:

$$\mathcal{E}_f(I) := \sum_{\{u,v\}\in E} w(u,v)\cdot \mathrm{len}(I \cap [f(u), f(v)])^2.$$

When the function f is clear from the context we drop the subscripts from the above definitions.

The next fact shows that by restricting the energy of f to disjoint intervals we may only decrease the energy.

Fact 11.2.6. *For any set of disjoint intervals $I_1, \ldots, I_m$, we have*

$$\mathcal{E}_f \geq \sum_{i=1}^{m} \mathcal{E}_f(I_i).$$

Proof.

$$\mathcal{E}_f = \sum_{u \sim v} w(u,v)|f(u) - f(v)|^2 \geq \sum_{u \sim v} \sum_{i=1}^{m} w(u,v) \operatorname{len}(I_i \cap [f(u), f(v)])^2 = \sum_{i=1}^{m} \mathcal{E}_f(I_i).$$

□

Proposition 11.2.7. *If $I(i)$ is not balanced, then*

$$\mathcal{E}_f(I_i) \geq \frac{\alpha^6 \cdot (1-\alpha)^2 \cdot \phi(f)^2 \cdot \mathcal{M}_f(I(i-1))}{24(\alpha^4 \cdot k \cdot \phi(f) + \delta)}.$$

The following is the key lemma is an extension of Lemma 8.1.10 allows us to lower bound the energy of a function f as a function of $\phi(f)$. It shows that a long interval with small volume must have a significant contribution to the energy of f.

Lemma 11.2.8. *For any subinterval $I = I(i,j)$,*

$$\mathcal{E}_f(I) \geq \frac{\phi(f)^2 \cdot \operatorname{vol}(I(i-1))^2 \cdot \operatorname{len}(I)^2}{\phi(f) \cdot \operatorname{vol}(I(i-1)) + \operatorname{vol}(I)}$$

Proof. By the definition of $\phi(f)$, for any $t \in I$. the total weight of the edges going out of any threshold set $S_f(t)$ is at least $\phi(f) \cdot \operatorname{vol}(I(i-1))$. Therefore, by summing over these threshold sets, we have

$$\sum_{\{u,v\} \in E} w(u,v) \operatorname{len}(I \cap [f(u), f(v)]) \geq \operatorname{len}(I) \cdot \phi(f) \cdot \operatorname{vol}(I(i-1)).$$

Let $E_1 := \{\{u,v\} : \operatorname{len}(I \cap [f(u), f(v)]) > 0\}$ be the set of edges with nonempty intersection with the interval I. Let $\beta \in (0,1)$ be a parameter to be fixed later. Let $E_2 \subseteq E_1$ be the set of edges of E_1 that are not adjacent to any of the vertices in I. If $w(E_2) \geq \beta w(E_1)$, then

$$\mathcal{E}_f(I) \geq w(E_2) \cdot \operatorname{len}(I)^2 \geq \beta \cdot w(E_1) \cdot \operatorname{len}(I)^2 \geq \beta \cdot \phi(f) \cdot \operatorname{vol}(I(i-1)) \cdot \operatorname{len}(I)^2.$$

Otherwise, $\operatorname{vol}(I) \geq (1-\beta) w(E_1)$. Therefore, by a variant of Cauchy Schwarz inequality in

Fact 11.2.9, we have

$$\mathcal{E}_f(I) = \sum_{\{u,v\}\in E_1} w(u,v)(\text{len}(I\cap[f(u),f(v)]))^2 \geq \frac{\left(\sum_{\{u,v\}\in E_1} w(u,v)\,\text{len}(I\cap[f(u),f(v)])\right)^2}{w(E_1)} \geq \frac{(1-\beta)\,\text{len}(I)^2\cdot\phi(f)^2\cdot\text{vol}(I(i-1))^2}{\text{vol}(I)}.$$

Choosing $\beta = (\phi(f)\cdot\text{vol}(I(i-1)))/(\phi(f)\cdot\text{vol}(I(i-1))+\text{vol}(I))$ such that the above two terms are equal gives the lemma. □

Fact 11.2.9 (A variant of Cauchy-Schwarz inequality)**.** *For any* $a_1,\ldots,a_m,b_1,\ldots,b_m\geq 0$,

$$\sum_{i=1}^m \frac{a_i^2}{b_i} \geq \frac{\left(\sum_{i=1}^m a_i\right)^2}{\sum_{i=1}^m b_i}.$$

We note that above lemma can be used to give a new proof of Cheeger's inequality with a weaker constant. Using the above lemma we can lower bound the energy of a light subinterval $I(i,j)$ in terms of $\mathcal{M}(I(i-1))$.

Corollary 11.2.10. *For any light subinterval* $I = I_{i,j}$,

$$\mathcal{E}_f(I) \geq \frac{\alpha^6\cdot(1-\alpha)^2\cdot\phi(f)^2\cdot\mathcal{M}_f(I(i-1))}{144k\cdot(\alpha^4\cdot k\cdot\phi(f)+\delta)}.$$

Proof. First, observe that

$$\text{vol}(I) \leq \sum_{v\in I} w(v)\frac{f^2(v)}{\alpha^{2i+2}} = \frac{\mathcal{M}_f(I)}{\alpha^{2i+2}} \leq \frac{\delta\cdot\mathcal{M}_f(I(i-1))}{k\cdot\alpha^{2i+2}} \leq \frac{\delta\cdot\text{vol}(I(i-1))}{k\cdot\alpha^4}, \tag{11.2.3}$$

where we use the assumption that I is a light subinterval in the second last inequality, and that $f(v)\leq\alpha^{2i-2}$ for $v\in I(i-1)$ in the last inequality. By Lemma 11.2.8,

$$\mathcal{E}_f(I) \geq \frac{\phi(f)^2\cdot\text{vol}(I(i-1))^2\cdot\text{len}(I)^2}{\phi(f)\cdot\text{vol}(I(i-1))+\text{vol}(I)} \geq \frac{k\alpha^4\phi(f)^2\cdot\text{vol}(I(i-1))\cdot\text{len}(I)^2}{k\cdot\alpha^4\phi(f)+\delta} \geq \frac{\alpha^6\cdot(1-\alpha)^2\cdot\phi(f)^2\cdot\mathcal{M}_(I(i-1))}{144k\cdot(\alpha^4\cdot k\cdot\phi(f)+\delta)},$$

where the first inequality holds by equation (11.2.3), and the last inequality holds by equation (11.2.2). □

Above corollary directly implies Proposition 11.2.7. If $I(i)$ is not balanced it has at least $6k$ light interval. Since the subintervals of $I(i)$ are disjoint by Fact 11.2.6, Proposition 11.2.7 follows from above corollary.

11.2.4 Improved Cheeger's Inequality

Now we are ready to prove Theorem 11.2.1.

Proof of Theorem 11.2.1. First we show that $\Delta \geq 1/2$, unless (i) holds, and then we use Proposition 11.2.5 to prove the theorem. If $\phi(f) \leq \sqrt{10^9} \cdot k \cdot \mathcal{R}(f)$, then (i) holds and we are done. So, assume that

$$\frac{10^9 \cdot k^2 \cdot \mathcal{R}^2(f)}{\phi^2(f)} \leq 1, \tag{11.2.4}$$

and we prove (ii). Since $\|f\|_w^2 = 1$, by Proposition 11.2.7,

$$\mathcal{R}(f) = \mathcal{E}_f \geq \sum_{I_i \notin \mathcal{B}} \mathcal{E}(I(i)) \geq \sum_{I(i) \notin \mathcal{B}} \frac{\alpha^6 \cdot (1-\alpha)^2 \cdot \phi(f)^2 \cdot \mathcal{M}(I(i-1))}{24(\alpha^4 \cdot k \cdot \phi(f) + \delta)}.$$

Set $\alpha = 1/2$ and $\delta = \frac{\alpha^6(1-\alpha)^2 \cdot \phi^2(f)}{96 \cdot \mathcal{R}(f)}$. If $\alpha^4 \cdot k \cdot \phi(f) \geq \delta$, then we get

$$\sum_{I(i) \notin \mathcal{B}} \mathcal{M}(I(i-1)) \leq \frac{48 \cdot k \cdot \mathcal{R}(f)}{\alpha^2 \cdot (1-\alpha)^2 \cdot \phi(f)} \leq \frac{1}{2},$$

where the last inequality follows from (11.2.4). Otherwise,

$$\sum_{I(i) \notin \mathcal{B}} \mathcal{M}(I(i-1)) \leq \frac{48\delta \cdot \mathcal{R}(f)}{\alpha^6 \cdot (1-\alpha)^2 \cdot \phi^2(f)} \leq \frac{1}{2},$$

where the last inequality follows from the definition of c and δ. Since $\mathcal{M}(V) = \|f\|_w^2 = 1$, it follows from the above equations that $\Delta \geq \frac{1}{2}$. Therefore, by Proposition 11.2.5, we get k disjointly supported functions $f_1, \ldots, f_k$ such that

$$\mathcal{R}(f_i) \leq \frac{1250 \cdot k^2 \cdot \mathcal{R}(f)}{(1-\alpha)^2 \cdot \delta \cdot \Delta} \leq \frac{10^9 \cdot k^2 \cdot \mathcal{R}(f)^2}{\phi(f)^2}.$$

Although each function f_i is defined on a region which is a union of many heavy subintervals, we can simply restrict it to only one of those subintervals guaranteeing that $\mathcal{R}(f_i)$ only decreases. Therefore each f_i is defined on an interval $[a_i, b_i]$ where by (11.2.2), $|a_i - b_i| = \Theta(1/k)a_i$. This proves (ii). □

We remark that the constant in the above argument can be significantly improved using very different proof techniques (see [KLL+13] for more details).

11.3 Extensions and Connections

In this section, we extend our technique to other graph partitioning problems, including multiway partitioning (Subsection 11.3.1), balanced separator (Subsection 11.3.2). We also remark that our

techniques can be applied to maximum cut, and to the Remanian manifolds to obtain improved quantitate bounds (we refer the interested readers [KLL+13]).

11.3.1 Spectral Multiway Partitioning

In this subsection, we use Theorem 11.1.1 and results in Chapter 10 to prove Corollary 11.1.3.

We start by part (i). First of all, by Theorem 10.1.5, there are k non-negative disjointly supported functions $f_1, \ldots, f_k : V \to \mathbb{R}$ such that for each $1 \le i \le k$ we have $\mathcal{R}(f_i) \le O(k^6)\lambda_k$. Let $S_{f_i}(t_{\mathsf{opt}})$ be the best threshold set of f_i. We consider two cases.

$\mathrm{vol}(\mathrm{supp}(f_i)) \le \mathrm{vol}(V)/2$ for all $1 \le i \le k$: Let $S_i := S_{f_i}(t_{\mathsf{opt}})$. Then, for each function f_i, by Theorem 11.2.1,

$$\phi(S_i) = \phi(f_i) \le O(l)\frac{\mathcal{R}(f_i)}{\sqrt{\lambda_l}} \le O(lk^6)\frac{\lambda_k}{\sqrt{\lambda_l}}.$$

Furthermore, since $S_i \subseteq \mathrm{supp}(f_i)$ and $f_1, \ldots, f_k$ are disjointly supported, $S_1, \ldots, S_k$ are disjoint. Hence,

$$\rho(k) = \max_{1\le i\le k} \phi(S_i) \le O(lk^6)\frac{\lambda_k}{\sqrt{\lambda_l}},$$

and we are done.

There exists a function, say f_k, with $\mathrm{vol}(\mathrm{supp}(f_k)) > \mathrm{vol}(V)/2$: Let $S_i = S_{f_i}(t_{\mathsf{opt}})$ for $1 \le i \le k-1$, and $S_k := V - S_1 - \ldots - S_{k-1}$. Similar to the above, the sets $S_1, \ldots, S_{k-1}$ are disjoint, and $\phi(S_i) \le O(lk^6\lambda_k/\sqrt{\lambda_l})$ for all $1 \le i \le k-1$. Observe that

$$\phi(S_k) = \frac{w(S_1, S_k) + \ldots + w(S_{k-1}, S_k)}{\mathrm{vol}(V) - \mathrm{vol}(S_k)} \le \frac{\sum_{i=1}^{k-1} w(E(S_i, \overline{S_i}))}{\sum_{i=1}^{k-1} \mathrm{vol}(S_i)} \le O(lk^6)\frac{\lambda_k}{\sqrt{\lambda_l}},$$

where the first equality uses $\mathrm{vol}(S_k) \ge \mathrm{vol}(V)/2$. Hence, $\rho(k) \le O(lk^6)\lambda_k/\sqrt{\lambda_l}$.

This completes the proof of (i) of Corollary 11.1.3.

To prove (ii) we use Theorem 10.2.6. It follows from (i) that without loss of generality we can assume that $\delta > 10/k$. Let $\delta' := \delta/2$. Then, by Theorem 10.2.6, there exist $r \ge (1-\delta')k$ non-negative disjointly supported functions $f_1, \ldots, f_r$ such that $\mathcal{R}(f_i) \le O(\delta^{-9}\log^2 k)\lambda_k$ and $\mathrm{vol}(\mathrm{supp}(f_i)) \le \mathrm{vol}(V)/2$. For each $1 \le i \le r$, let $S_i := S_{f_i}(t_{\mathsf{opt}})$. Similar to the argument in part (i), since $S_i \subseteq \mathrm{supp}(f_i)$, the sets $S_1, \ldots, S_r$ are disjoint. Without loss of generality assume that $\phi(S_1) \le \phi(S_2) \le \ldots \phi(S_r)$. Since $S_1, \ldots, S_{(1-\delta)k}$ are disjoint,

$$\rho((1-\delta)k) \le \phi(S_{(1-\delta)k+1}) \le \ldots \le \phi(S_r). \tag{11.3.1}$$

Let $m := l/(\delta' k) = 2l/(\delta k)$. If $\phi(f_i) \le O(m)\mathcal{R}(f_i)$ for some $(1-\delta)k < i \le r$, then we get

$$\rho((1-\delta)k) \le \phi(S_i) = \phi(f_i) \le O(m)\mathcal{R}(f_i) \le O\left(\frac{l\log^2 k}{\delta^9 k}\right)\lambda_k,$$

and we are done. Otherwise, by Theorem 11.2.1, for each $(1-\delta)k < i \le r$, there exist m disjointly supported functions $h_{i,1}, \ldots h_{i,m}$ such that for all $1 \le j \le m$, $\mathrm{supp}(h_{i,j}) \subseteq \mathrm{supp}(f_i)$ and

$$\mathcal{R}(h_{i,j}) \le O(m^2)\frac{\mathcal{R}(f_i)^2}{\phi(f_i)^2} \le O\left(\frac{l^2}{\delta^2 k^2}\right)\frac{O(\delta^{-18}\log^4 k)\lambda_k^2}{\rho^2((1-\delta)k)} = O\left(\frac{l^2\log^4 k}{\delta^{20}k^2}\right)\frac{\lambda_k^2}{\rho^2((1-\delta)k)} \tag{11.3.2}$$

where the second inequality follows from (11.3.1). Since $f_{(1-\delta)k+1}, \ldots, f_r$ are disjointly supported, all functions $h_{i,j}$ are disjointly supported as well. Therefore, since $l = m(\delta' k) \le m(r-(1-\delta)k)$, by Lemma 7.2.1,

$$\lambda_l \le 2 \max_{\substack{(1-\delta)k<i\le r \\ 1\le j\le m}} \mathcal{R}(h_{i,j}) \le O\left(\frac{l^2\log^4 k}{\delta^{20}k^2}\right)\frac{\lambda_k^2}{\rho^2((1-\delta)k)},$$

where the second inequality follows from (11.3.2). This completes the proof of (ii) of Corollary 11.1.3.

Part (iii) can be proved in a very similar way to part (ii) using Theorem 10.2.5 We follow the same proof steps as in part (ii) except that we upper bound $\mathcal{R}(f_i)$ by $O(h^4\delta^{-4})\lambda_k$. This completes the proof of Corollary 11.1.3.

In the remaining part of this section we describe some examples.

Example 11.3.1. *First we show that there exists a graph where $\rho(k) \ge \Omega(l-k+1)\lambda_k/\sqrt{\lambda_l}$. Let G be a union of $k-2$ isolated vertices and a cycle of length n. Then, $\rho(k) = \Theta(1/n)$, $\lambda_k = \Theta(1/n^2)$ and for $l > k$, $\lambda_l = \Theta((l-k+1)^2/n^2)$. Therefore,*

$$\rho(k) \ge \Omega(l-k+1)\frac{\lambda_k}{\sqrt{\lambda_l}}$$

This shows that for $l \gg k$, the dependency on l in the right hand side of part (i) of Corollary 11.1.3 is necessary.

Example 11.3.2. *In the second example we show that there exists a graph where $\rho(k/2) \ge \Omega(l/k)\lambda_k/\sqrt{\lambda_l}$. Let G be a cycle of length n. Then, $\rho(k/2) = \Theta(k/n)$, $\lambda_k = \Theta(k^2/n^2)$ and $\lambda_l = \Theta(l^2/n^2)$. Therefore,*

$$\rho(k/2) \ge \Omega(l/k)\frac{\lambda_k}{\sqrt{\lambda_l}}.$$

This shows that part (iii) of Corollary 11.1.3 is tight (up to constant factors).

11.3.2 Balanced Separator

In this section we give a simple polynomial time algorithm with approximation factor $O(k/\lambda_k)$ for the balanced separator problem.

We will prove Theorem 11.1.4 by repeated applications of Theorem 11.2.1. Our algorithm is similar to the standard algorithm for finding a balanced separator by applying Cheeger's inequality repeatedly. We inductively remove a subset of vertices of the remaining graph such that the union of the removed vertices is a non-expanding set in G, until the set of removed vertices has at least a quarter of the total volume. Besides that, there is an additional step that removes a subset of vertices such that the conductance of the union of the removed vertices does not increase. The details are described in Algorithm 13.

Algorithm 13 A Spectral Algorithm for Balanced Separator

$U \leftarrow V$.
while $\mathrm{vol}(U) > \frac{4}{5}\mathrm{vol}(V)$ **do**
  Let $H = (U, E(U))$ be the induced subgraph of G on U, and λ'_2 be the second smallest eigenvalue of $\mathcal{L}_H$.
  Let $f \in \ell^2(U, w)$ be a non-negative function such that $\mathrm{vol}(\mathrm{supp}(f)) \leq \mathrm{vol}(H)/2$, and $\mathcal{R}_H(f) \leq \lambda'_2$.
  if $\phi_H(f) \leq O(k)\mathcal{R}_H(f)/\sqrt{\lambda_k}$ **then**
    $U \leftarrow U - U_f(t_{\mathsf{opt}})$.
  else
    Let $f_1, \ldots, f_k$ be k disjointly supported functions such that $\mathrm{supp}(f_i) \subseteq \mathrm{supp}(f)$ and

$$\phi_H(f) \leq O(k)\frac{\mathcal{R}_H(f)}{\sqrt{\max_{1\leq i\leq k}\mathcal{R}_H(f_i)}},$$

as defined in Theorem 11.2.1.
    Find a threshold set $S = S_{f_i}(t)$ for $1 \leq i \leq k$, and $t > 0$ such that

$$w(S, U - S) \leq w(S, V - U).$$

    $U \leftarrow U - S$.
  end if
end while
return $\overline{U}$.

Let U be the set of vertices remained after a number of steps of the induction, where initially $U = V$. We will maintain the invariant that $\phi_G(\overline{U}) \leq O(k\epsilon/\lambda_k)$. Suppose $\mathrm{vol}(U) > \frac{4}{5}\mathrm{vol}(V)$. Let $H = (U, E(U))$ be the induced subgraph of G on U, and $0 = \lambda'_1 \leq \lambda'_2 \leq \ldots$ be the eigenvalues of $\mathcal{L}_H$. First, observe that $\lambda'_2 = O(\epsilon)$ as the following lemma shows.

Lemma 11.3.3. *For any set $U \subseteq V$ with $\mathrm{vol}(U) \geq \frac{4}{5}\mathrm{vol}(V)$, let $H(U, E(U))$ be the induced subgraph of G on U. Then the second smallest eigenvalue λ'_2 of $\mathcal{L}_H$ is at most 10ϵ.*

Proof. Let $(T, \overline{T})$ be the optimum bisection, and let $T' := U \cap T$. Since $\mathrm{vol}(U) \geq \frac{4}{5}\mathrm{vol}(V)$, and

$\mathrm{vol}(T) = \mathrm{vol}(V)/2$, we have

$$\mathrm{vol}_H(T') \geq \mathrm{vol}_G(T) - 2\mathrm{vol}_G(\overline{U}) \geq \mathrm{vol}(V)/2 - 2\mathrm{vol}(V)/5 = \mathrm{vol}(V)/10 = \mathrm{vol}(T)/5.$$

Furthermore, since $E(T', U - T') \subseteq E(T, \overline{T})$, we have

$$\phi_H(T') = \frac{w(T', U - T')}{\mathrm{vol}_H(T')} \leq \frac{w(T, \overline{T})}{\mathrm{vol}_G(T)/5} \leq 5\phi(T) = 5\epsilon.$$

Therefore, by Claim 7.8.2, we have $\lambda_2' \leq 10\epsilon$. □

To prove Theorem 11.1.4, it is sufficient to find a set $S \subseteq U$ with $\mathrm{vol}_H(S) \leq \frac{1}{2}\mathrm{vol}_H(U)$ and conductance $\phi_H(S) \leq O(k\lambda_2'/\lambda_k) = O(k\epsilon/\lambda_k)$, because

$$\phi_G(\overline{U} \cup S) \leq \frac{w(\overline{U}, U) + w(S, \overline{S})}{\mathrm{vol}_G(\overline{U}) + \mathrm{vol}_H(S)} \leq \max(\phi_G(\overline{U}), \phi_H(S))) \leq O(k\epsilon/\lambda_k),$$

and so we can recurse until $\frac{1}{5}\mathrm{vol}(V) \leq \mathrm{vol}(\overline{U} \cup S) \leq \frac{4}{5}\mathrm{vol}(V)$. Let $f \in \ell^2(U, w)$ be a non-negative function such that $\mathrm{vol}_H(\mathrm{supp}(f)) \leq \frac{1}{2}\mathrm{vol}_H(U)$ and $\mathcal{R}_H(f) \leq \lambda_2'$, as defined in Lemma 7.2.3. If $\phi_H(f) \leq O(k\lambda_2'/\lambda_k)$, then we are done by induction. Otherwise, we will find a set S such that $\mathrm{vol}_H(S) \leq \frac{1}{2}\mathrm{vol}_H(U)$ and $w(E(S, U - S)) \leq w(E(S, \overline{U}))$. This implies that we can simply remove S from U without increasing the conductance of the union of the removed vertices, i.e., $\phi_G(S \cup \overline{U}) \leq \phi_G(\overline{U})$. This is because the numerator (total weight of the cut edges) does not increase while the denominator (volume of the set) may only increase.

It remains to find a set S with either of the above properties. We can assume that $\phi_H(f) \not\leq O(k)\mathcal{R}_H(f)$ as otherwise we are done. By Theorem 11.2.1, there are k disjointly supported functions $f_1, \ldots, f_k \in \ell^2(U, w)$ such that $\mathrm{supp}(f_i) \subseteq \mathrm{supp}(f)$ and

$$\phi_H(f) \leq O(k)\frac{\lambda_2'}{\sqrt{\max \mathcal{R}_H(f_i)}}.$$

We extend $f_i \in \ell^2(U, w)$ to $f_i \in \ell^2(V, w)$ by defining $f_i(v) = 0$ for $v \in V - U$. We will prove that either $\phi_H(f) \leq O(k\lambda_2'/\lambda_k)$, or there is a threshold set $S = S_{f_i}(t)$ for some $1 \leq i \leq k$ and $t > 0$ such that $w(S, U - S) \leq w(S, \overline{U})$. As $f_1, \ldots, f_k$ can be computed in polynomial time, this will complete the proof of Theorem 11.1.4.

Suppose that for every f_i and any threshold set $S = S_{f_i}(t)$ we have $w(S, \overline{U}) \leq w(S, U - S)$. Then, by Lemma 11.3.4 that we will prove below, $\mathcal{R}_H(f_i) \geq \Omega(\mathcal{R}_G^2(f_i))$ for every $1 \leq i \leq k$. This implies that

$$\phi_H(f) \leq O(k)\frac{\lambda_2'}{\sqrt{\max_{1\leq i\leq k} \mathcal{R}_H(f_i)}} \leq O(k)\frac{\lambda_2'}{\sqrt{\max_{1\leq i\leq k} \mathcal{R}_G^2(f_i)}} \leq O(k)\frac{\lambda_2'}{\lambda_k},$$

where the last inequality follows by Lemma 7.2.1 and the fact that $f_1, \ldots, f_k$ are disjointly supported.

Lemma 11.3.4. *For any set $U \subseteq V$, let $H(U, E(U))$ be the induced subgraph of G on U, and $f \in \ell^2(V, w)$ be a non-negative function such that $f(v) = 0$ for any $v \in V - U$. Suppose that for any threshold set $S_f(t)$, we have*

$$w(S_f(t), \overline{U}) \leq w(S_f(t), U - S_f(t)),$$

then

$$\sqrt{8\mathcal{R}_H(f)} \geq \mathcal{R}_G(f).$$

Proof. Without loss of generality we assume $w(v) \geq 1$ for all $v \in V$. Since both sides of the inequality are homogeneous in f, we may assume that $\max_v f(v) \leq 1$. Furthermore, we can assume that $\sum_v w(v) f^2(v) = 1$. Observe that, since $w_H(v) \leq w_G(v)$ for all $v \in U$,

$$\sum_{v \in U} w_H(v) f^2(v) \leq \sum_{v \in U} w_G(v) f^2(v) = \sum_v w_G(v) f^2(v) = 1. \tag{11.3.3}$$

Since $f(v) = 0$ for any $v \in \overline{U}$, $S_f(t) \subseteq U$ for any $t > 0$. Therefore, by Lemma 7.8.3,

$$\mathbb{E}\left[w(S_f(\sqrt{t}), U - S_f(\sqrt{t}))\right] \leq \sqrt{2\mathcal{R}_H(f)}. \tag{11.3.4}$$

where we used the fact that $\|f\|_w = 1$.

On the other hand, since $w(S_f(t), \overline{U}) \leq w(S_f(t), U - S_f(t))$ for any t,

$$\begin{aligned} \mathbb{E}\left[w(S_f(\sqrt{t}), U - S_f(\sqrt{t}))\right] &\geq \frac{1}{2}\mathbb{E}\left[w(S_f(\sqrt{t}), V - S_f(\sqrt{t}))\right] \\ &= \frac{1}{2} \sum_{\{u,v\} \in E} w(u,v) |f^2(u) - f^2(v)| \\ &\geq \frac{1}{2} \sum_{\{u,v\} \in E} w(u,v) |f(u) - f(v)|^2 = \frac{1}{2}\mathcal{R}_G(f). \end{aligned} \tag{11.3.5}$$

where the last inequality follows by the fact that $f(v) \geq 0$ for all $v \in V$, and the last equality follows by the normalization $\sum_v w(v) f^2(v) = 1$. Putting together (11.3.4) and (11.3.5) proves the lemma. □

Chapter 12

Almost Optimal Local Graph Clustering

In this chapter we provide a local variant of the Spectral Partitioning algorithm (Algorithm 10). We design an algorithm with the same guarantee (up to a constant factor) as the Cheeger's inequality, that runs in time slightly super linear in the size of the output. This is the first sublinear (in the size of the input) time algorithm with almost the same guarantee as the Cheeger's inequality. As a byproduct of our results, we prove a bicriteria approximation algorithm for the conductance profile of any graph. Let $\phi(k)$ be the conductance profile of a graph as defined in equation (7.7.1). There is a polynomial time algorithm that, for any $k, \epsilon > 0$, finds a set S of volume $\mathrm{vol}(S) \leq O(k^{1+\epsilon})$, and conductance $\phi(S) \leq O(\sqrt{\phi(k)/\epsilon})$. Our proof techniques also provide a simpler proof of the structural result of Arora, Barak, Steurer [ABS10], that can be applied to irregular graphs.

The results of this chapter are based on a joint work with Luca Trevisan [OT12].

12.1 Introduction

Often, one is interested in applying a sparsest cut approximation algorithm *iteratively*, that is, first find an approximate sparsest cut in the graph, and then recurse on one or both of the subgraphs induced by the set found by the algorithm and by its complement. Such iteration might be used to find a *balanced* sparse cut if one exists (c.f. [OSV12]), or to find a good *clustering* of the graph, an approach that lead to approximate clusterings with good worst-case guarantees, as shown by Kannan, Vempala and Vetta [KVV04]. Even though each application of the spectral partitioning algorithm runs in nearly linear time, iterated applications of the algorithm can result in a quadratic running time.

Spielman and Teng [ST04], and subsequently [ACL06, AP09] studied local graph partitioning

algorithms that find a set S of approximately minimal conductance in time nearly linear in the size of the output set S. Note that the running time can be sub linear in the size of the input graph if the algorithm finds a small output set S. When iterated, such an algorithm finds a balanced sparse cut in nearly linear time in the size of the graph, and can be used to find a good clustering in nearly linear time as well.

Another advantage of such "local" algorithms is that if there are both large and small sets of near-optimal conductance, the algorithm is more likely to find the smaller sets. Thus, such algorithms can be used to approximate the small set expansion problem and the unique games conjecture (see Subsection 7.7.3 for background). Finding small, low-conductance, sets is also interesting in clustering applications. In a social network, for example, a low-conductance set of users in the "friendship" graph represents a "community" of users who are significantly more likely to be friends with other members of the community than with non-members, and discovering such communities has several applications. While large communities might correspond to large-scale, known, factors, such as the fact that American users are more likely to have other Americans as friends, or that people are more likely to have friends around their age, small communities generally contain more interesting and substantial information. Leskovec et al. [LLDM09, LLM10] observed that in large networks the sets which mostly resemble communities are of size around only 100, while larger communities gradually "blend into" the expander-like core of the network and thus become less "community-like". There is also an experimental evidence that a significant fraction of vertices in networks belong to small communities [LPP11, LP11].

12.1.1 Almost Optimal Local Graph Clustering

A *local graph clustering* algorithm, is a *local* graph algorithm that finds a non-expanding set in the local neighborhood of a given vertex v, in time proportional to the size of the output set. The *work/volume ratio* of such an algorithm, which is the ratio of the computational time of the algorithm in a single run, and the volume of the output set, may depend only poly logarithmically to the size of the graph.

The problem first studied in the remarkable work of Spielman and Teng [ST04]. Spielman and Teng design an algorithm `Nibble` such that for any set $A \subseteq V$, if the initial vertex, v, is sampled randomly according to the degree of vertices in A, with a constant probability, `Nibble` finds a set of conductance $O(\phi^{1/2}(A)\log^{3/2} n)$, with a work/volume ratio of $O(\phi^{-2}(A)\,\mathrm{polylog}(n))$, `Nibble` finds the desired set by looking at the *threshold sets* of the probability distribution of a t-step random walk started at v. To achieve the desired computational time they keep the support of the probability distribution small by removing a small portion of the probability mass at each step.

Andersen, Chung and Lang [ACL06], used the approximate PageRank vector rather than approximate random walk distribution, and they managed to improve the conductance of the output set to $O(\sqrt{\phi(A)\log n})$, and the work/volume ratio to $O(\phi^{-1}(A)\,\mathrm{polylog}\, n)$. More recently, Andersen and

Peres [AP09], use the *volume biased evolving set process* developed in [DF90, MP03], and improved the work/volume ratio to $O(\phi^{-1/2}(A)\operatorname{polylog} n)$, while achieving the same guarantee as [ACL06] on the conductance of the output set.

It has been a long standing open problem to design a local variant of the Cheeger's inequalities: that is to provide a *sublinear* time algorithm with an approximation guarantee that does not depend on the size of G, assuming that the size of the optimum set is sufficiently smaller than n, and a randomly chosen vertex of the optimum set is given. In this work we answer this question, and we prove the following theorem:

Theorem 12.1.1. $\texttt{ParESP}(v, k, \phi, \epsilon)$ *takes as input a starting vetex* $v \in V$*, a target conductance* $\phi \in (0, 1)$*, a target size* k*, and* $0 < \epsilon < 1$*, and outputs a set of vertices. For a given run of the algorithm let* S *be the output and let* W *be the computational complexity of the algorithm. Then, both* S *and* W *depend on the randomness of the algorithm.*

1. *The work per volume ratio* $W/\operatorname{vol}(S)$ *satisfies,*

$$\mathbb{E}\left[W/\operatorname{vol}(S)\right] = O(k^{\epsilon}\phi^{-1/2}\log^2 n).$$

2. *If* $A \subseteq V$ *is a set of vertices that satisfy* $\phi(A) \leq \phi$*, and* $\operatorname{vol}(A) \leq k$*, then there is a subset* $A' \subseteq A$ *with volume at least* $\operatorname{vol}(A)/2$*, such that if* $v \in A'$*, then with a constant probability* S *satisfies,*

   a) $\phi(S) = O(\sqrt{\phi/\epsilon})$,

   b) $\operatorname{vol}(S) = O(k^{1+\epsilon})$.

We remark that unlike the previous local graph clustering algorithms, the running time of the algorithm is slightly super linear in the size of the optimum. Nonetheless, by choosing $\epsilon = \Theta(1/\log k)$, one can reproduce the best previous local graph clustering algorithm of Andersen, Peres [AP09].

12.1.2 Approximating the Conductance Profile

As a byproduct of the above result we give an approximation algorithm for the conductance profile of G (see Subsection 7.7.3 for background on conductance profile and small set expansion problem). We prove k independent approximation of $\phi(k)$ as a function of $\phi(k^{1-\epsilon})$, without any dependency in the size of the graph; specifically, we prove the following corollary:

Corollary 12.1.2. *There is a polynomial time algorithm that takes as input a target conductance* ϕ*, and* $0 < \epsilon < 1$*, and outputs a set* S*, s.t. if* $\phi(A) \leq \phi$*, for some* $A \subseteq V$*, then* $\operatorname{vol}(S) = O(\operatorname{vol}(A)^{1+\epsilon})$*, and* $\phi(S) = O(\sqrt{\phi/\epsilon})$.

This is the first approximation algorithm for the small set expansion problems where the conductance of the output is only a function of the optimum conductance, and does not depend on the

size of the graph. The approximation factor of the previous algorithms [RST10, BFK+11] depend poly-logarithmically on n/k.

Our corollary indicates that the hard instance of the small set expansion problem are those where $\phi(m^{1-\Omega(1)}) \approx 1$, while $\phi(\delta m) \leq \phi$. In other words, if in an instance of SSE problem $\phi(m^{1-\Omega(1)})$ is bounded away from 1, then using our algorithm we can prove $\phi(\delta m) \leq \phi$ for any constant $\delta > 0$.

Independent of our work, Kwok and Lau [KL12] have obtained a somewhat different proof of Corollary 12.1.2.

12.1.3 Approximating Balanced Separator

One application of our local partitioning algorithm is a fast algorithm for finding balanced cuts. Spielman and Teng showed how to find a balanced cut in nearly linear time by repeatedly removing small sets from a graph using local partitioning [ST04]. Applying their technique with our algorithm yields an algorithm with the following properties. The algorithm has complexity $m^{1+O(\epsilon)}\phi^{-1/2}\operatorname{polylog}(n)$, and it outputs a set of vertices whose conductance is $O(\sqrt{\phi/\epsilon})$ and whose volume at least half that of any set with conductance at most ϕ and volume at most $m^{1-\epsilon}/c$, where ϕ, ϵ are inputs to the algorithm and $c > 0$ is an absolute constant.

Orecchia, Sachdeva and Vishnoi [OSV12] very recently designed almost linear time algorithm $O(m\operatorname{polylog}(n))$ that gives a $\sqrt{\phi}$ approximation to the balanced separator problem. Their algorithm is the current fastest algorithm that provides a nontrivial approximation guarantee for the balanced cut problem (note that compare to our algorithm, the running time of [OSV12] does not depend on ϕ). There are also several algorithms that provide stronger approximation guarantees with a slower running time (see e.g. Arora et al. [AHK10] and Sherman [She09]). These algorithms produce cuts with conductance $O(\phi\operatorname{polylog}(n))$, and their computational complexity is dominated by the cost of solving $O(n^{o(1)})$ many single-commodity flow problems.

12.1.4 Techniques

Our main technical result is a new upper bound on escape probability of simple random walks that we covered in Section 8.3. We showed for $S \subseteq V$, a t-step *lazy* random walk started at a random vertex of S remains entirely in S with probability at least $(1 - \phi(S)/2)^t$. Previously, only the lower bound $1 - t\phi(S)/2$ was known, and the analysis of other local clustering algorithms implicitly or explicitly depended on such a bound.

For comparison, when $t = 1/\phi(S)$, the known bound would imply that the walk has probability at least $1/2$ of being entirely contained in S, with no guarantee being available in the case $t = 2/\phi(S)$, while our bound implies that for $t = (\epsilon \ln n)/\phi$ the probability of being entirely contained in S is still at least $1/n^{\epsilon}$. Roughly speaking, the $\Omega(\log n)$ factor that we gain in the length of walks that we can study corresponds to our improvement in the conductance bound, while the $1/n^{\epsilon}$ factor that we lose in the probability corresponds to the factor that we lose in the size of the output set.

Our local algorithm uses the volume biased evolving set process. This process starts with a vertex v of the graph (or possibly any $S_0 \subset V$), and then produces a sequence of sets $S_1, S_2, \ldots, S_\tau$ for any stopping time τ. Anderson and Peres [AP09] show that with constant probability at least one set S_t is such that $\phi(S_t) \lesssim \sqrt{\log \mathrm{vol}(S_\tau)/\tau}$. If one can show that up to some time T the process constructs sets all of volume at most k, then we get a set of volume at most k and conductance at most $O(\sqrt{\log(k)/T})$. Andersen and Peres were able to show that if the graph has a set A of conductance ϕ, then the process is likely to construct sets all of volume at most $2\mathrm{vol}(A)$ for at least $T = \Omega(1/\phi)$ steps, if started from a random element of the set A, leading to their $O(\sqrt{\phi \log n})$ guarantee. We show that for any chosen $\epsilon < 1/2$, the process will construct sets of volume at most $O(\mathrm{vol}(A)^{1+\epsilon})$ for $T \gtrsim \epsilon \cdot \log(\mathrm{vol}(A))/\phi$ steps, with probability at least $1/\mathrm{vol}(A)^\epsilon$. This is enough to guarantee that, at least with probability $1/\mathrm{vol}(A)^\epsilon$, the process constructs at least one set of conductance $O(\sqrt{\phi/\epsilon})$. To obtain this conclusion, we also need to strengthen the first part of the analysis of Andersen and Peres: we need to show that the process with probability at least $1 - 1/\mathrm{vol}(A)^{\Omega(1)}$ of constructing a set with low conductance in the first τ steps, because we need to take a union bound with the event that t is large, for which probability we only have a $\mathrm{vol}(A)^{-\Omega(1)}$ lower bound. Finally, to achieve a constant probability of success, we run $\mathrm{vol}(A)^\epsilon$ copies of the evolving set process simultaneously, and stop as soon as one of the copies finds a small non-expanding set.

12.2 Background

12.2.1 Evolving Set Process

The *evolving set process* (ESP) is a Markov chain on subsets of the vertex set V. Given the current state S, the next state S_1 is chosen by the following rule: pick a threshold Z uniformly at random from the interval $[0, 1]$, and let

$$S_1 = \{u : \sum_{v \in S} P(u, v) \geq Z\}. \tag{12.2.1}$$

Notice that $\emptyset$ and V are absorbing states for the process. Given a starting state $S_0 \subseteq V$, we write $\mathbf{P}_{S_0}[\cdot] := \mathbf{P}[\cdot \mid S_0]$ to denote the probability measure for the ESP Markov chain started from S_0. Similarly, we write $\mathbf{E}_{S_0}[\cdot]$ for the expectation. For a singleton set, we use the shorthand $\mathbf{P}_v[\cdot] = \mathbf{P}_{\{v\}}[\cdot]$. We define the transition kernel $\mathbf{K}(S, S') = \mathbf{P}_S[S_1 = S']$.

Morris and Peres [MP03] used the evolving set process to prove upper bounds on the strong stationary time of markov chains based on the conductance profile. They proved the following propositions to relate the conductance of a set in the ESP to the change in volume in the next step. The first proposition strengthens the fact that the sequence $(\mathrm{vol}(S_t))_{t \geq 0}$ is a martingale.

Proposition 12.2.1 (Morris and Peres [MP03])**.** *Let Z be the uniform random variable used to*

generate S_1 from S in the ESP. Then,

$$\mathbf{E}_S\left[\mathrm{vol}(S_1) \mid Z \le \frac{1}{2}\right] = \mathrm{vol}(S) + \partial(S) = \mathrm{vol}(S)(1+\phi(S)).$$

$$\mathbf{E}_S\left[\mathrm{vol}(S_1) \mid Z > \frac{1}{2}\right] = \mathrm{vol}(S) - \partial(S) = \mathrm{vol}(S)(1-\phi(S)).$$

Proposition 12.2.2 (Morris and Peres [MP03])**.** *The* growth gauge $\psi(S)$ *of a set S is defined by the following equation:*

$$1 - \psi(S) := \mathbf{E}_S\left[\sqrt{\frac{\mathrm{vol}(S_1)}{\mathrm{vol}(S)}}\right].$$

For any set $S \subseteq V$, the growth gauge and conductance satisfy $\psi(S) \ge \phi(S)^2/8$.

Proof. Using Proposition 12.2.1, for any set $S \subseteq V$, we have

$$\begin{aligned}
\mathbf{E}_S\left[\sqrt{\frac{\mathrm{vol}(S_1)}{\mathrm{vol}(S)}}\right] &= \frac{1}{2}\mathbf{E}_S\left[\sqrt{\frac{\mathrm{vol}(S_1)}{\mathrm{vol}(S)}} \mid Z \le \frac{1}{2}\right] + \frac{1}{2}\mathbf{E}_S\left[\sqrt{\frac{\mathrm{vol}(S_1)}{\mathrm{vol}(S)}} \mid Z > \frac{1}{2}\right] \\
&\le \frac{1}{2}\sqrt{\mathbf{E}_S\left[\frac{\mathrm{vol}(S_1)}{\mathrm{vol}(S)} \mid Z \le \frac{1}{2}\right]} + \frac{1}{2}\sqrt{\mathbf{E}_S\left[\frac{\mathrm{vol}(S_1)}{\mathrm{vol}(S)} \mid Z > \frac{1}{2}\right]} \\
&= \frac{1}{2}\sqrt{1+\phi(S)} + \frac{1}{2}\sqrt{1-\phi(S)} \le 1 - \phi^2(S)/8,
\end{aligned}$$

where the first inequality follows by the Jensen's inequality and the last inequality follows from the taylor expansion of the square root function. □

12.2.2 The Volume-Biased Evolving Set Process

The **volume-biased evolving set process** (volume-biased ESP) is a Markov chain on subsets of V with the following transition kernel:

$$\widehat{\mathbf{K}}(S, S') = \frac{\mathrm{vol}(S')}{\mathrm{vol}(S)}\mathbf{K}(S, S'), \tag{12.2.2}$$

where $\mathbf{K}(S, S')$ is the transition kernel for the ESP. We remark that $\widehat{\mathbf{K}}$ is the *Doob h-transform* of $\mathbf{K}$ with respect to vol (see chapter 17 of [LPW06]), and that the volume-biased ESP is equivalent to the ESP conditioned to absorb in the state V. Given a starting state S_0, we write $\widehat{\mathbf{P}}_{S_0}[\cdot] := \widehat{\mathbf{P}}[\cdot \mid S_0]$ for the probability measure of the Markov chain. Similarly, we write $\widehat{\mathbf{E}}_{S_0}[\cdot]$ for the expectation.

The following proposition relates the volume-biased ESP and the ESP. This is a standard consequence of the Doob h-transform, but we include a proof for completeness.

Proposition 12.2.3. *For any function f and any starting set $S_0 \neq \emptyset$,*

$$\widehat{\mathbf{E}}_{S_0}\left[f(S_0,\ldots,S_n)\right] = \mathbf{E}_{S_0}\left[\frac{\mathrm{vol}(S_n)}{\mathrm{vol}(S_0)} f(S_0,\ldots,S_n)\right]. \tag{12.2.3}$$

Proof. Assume that $S_0 \neq \emptyset$. Let $\mathcal{C}$ be the collection of sample paths $(S_0,\ldots,S_t)$ such that $\widehat{\mathbf{P}}_{S_0}[S_1,\ldots,S_t] > 0$. If $(S_0,\ldots,S_t) \in \mathcal{C}$, then $\mathrm{vol}(S_i) > 0$ for all $i \in [0,t]$, so

$$\widehat{\mathbf{P}}_{S_0}[S_1,\ldots,S_t] = \prod_{i=0}^{t-1} \frac{\mathrm{vol}(S_{i+1})}{\mathrm{vol}(S_i)} \mathbf{P}_{S_i}[S_{i+1}] = \frac{\mathrm{vol}(S_t)}{\mathrm{vol}(S_0)} \mathbf{P}_{S_0}[S_1,\ldots,S_t].$$

Therefore,

$$\begin{aligned}
\widehat{\mathbf{E}}_{S_0}\left[f(S_0,\ldots,S_t)\right] &= \sum_{(S_0,\ldots,S_t)\in\mathcal{C}} f(S_0,\ldots,S_t)\widehat{\mathbf{P}}_{S_0}[S_1,\ldots,S_t] \\
&= \sum_{(S_0,\ldots,S_t)\in\mathcal{C}} f(S_0,\ldots,S_t)\frac{\mathrm{vol}(S_t)}{\mathrm{vol}(S_0)}\mathbf{P}_{S_0}[S_1,\ldots,S_t] \\
&= \mathbf{E}_{S_0}\left[\frac{\mathrm{vol}(S_t)}{\mathrm{vol}(S_0)} f(S_0,\ldots,S_t)\right].
\end{aligned}$$

□

Andersen and Peres used the volume biased ESP as a local graph clustering algorithm [AP09]. They show that for any non-expanding set A, if we run the volume biased ESP from a randomly chosen vertex of A, with a constant probability, there is a set in the sample path of expansion $O(\sqrt{\phi(A)\log n})$, and volume at most $2\mathrm{vol}(A)$. As a part of their proof, they designed an efficient simulation of the volume biased ESP, called `GenerateSample`. They prove the following theorem,

Theorem 12.2.4 (Andersen, Peres [AP09, Theorems 3,4])**.** *There is an algorithm,* `GenerateSample`*, that simulates the volume biased ESP such that for any vertex $v \in V$, any sample path $(S_0 = \{v\},\ldots,S_\tau)$, is generated with probability $\widehat{\mathbf{P}}_v[S_0,\ldots,S_\tau]$. Furthermore, for a stopping time τ that is bounded above by T, let $W(\tau)$ be the time complexity of* `GenerateSample` *if it is run up to time τ. Then, the expected work per volume ratio of the algorithm is*

$$\widehat{\mathbf{E}}_v\left[\frac{W(\tau)}{\mathrm{vol}(S_\tau)}\right] = O(T^{1/2}\log^{3/2}\mathrm{vol}(V)).$$

Here, we do not include the proof of above theorem and we refer the reader to [AP09].

12.2.3 The Diaconis-Fill Coupling

Diaconis-Fill [DF90] introduced the following coupling between the random walk process and the volume-biased ESP. Let (X_t, S_t) be a Markov chain, where X_t is a vertex and $S_t \subseteq V$ is a subset of

vertices. Let $\mathbf{P}^*[\cdot]$ be the probability measure for the Markov chain. Given a starting vertex v, let $X_0 = v$ and $S_0 = \{v\}$, and let $\mathbf{P}^*_v[\cdot] = \mathbf{P}^*[\cdot \mid X_0 = v, S_0 = \{v\}]$. Given the current state (X_t, S_t), the transition probabilities are defined as follows.

$$\begin{aligned}\mathbf{P}^*[X_{t+1} = v' \mid X_t = v, S_t = S] &= P(v, v'),\\ \mathbf{P}^*[S_{t+1} = S' \mid S_t = S, X_{t+1} = v'] &= \frac{\mathbf{K}(S, S')\mathbb{I}[v' \in S']}{\mathbf{P}[v' \in S_{t+1} \mid S_t = S]}.\end{aligned}$$

In words, we first select X_{t+1} according to the random walk transition kernel, then select S_{t+1} according to the ESP transition kernel restricted to sets that contain X_{t+1}. We define the transition kernel $\mathbf{K}^*((v, S), (v', S')) = \mathbf{P}^*[X_1 = v', S_1 = S' \mid X_0 = v, S_0 = S]$.

The following proposition shows that $\mathbf{P}^*[\cdot]$ is a coupling between the random walk process and the volume-biased ESP, and furthermore the distribution of X_t conditioned on $(S_0, \ldots, S_t)$ is the stationary distribution restricted to S_t. A proof of Proposition 12.2.5 is given in of [LPW06, Chapter 17].

Proposition 12.2.5 (Diaconis and Fill [DF90])**.** *Let (X_t, S_t) be a Markov chain started from $(v, \{v\})$ with the transition kernel $\mathbf{K}^*$.*

1. *The sequence (X_t) is a Markov chain started from v with the transition kernel $P(.,.)$.*
2. *The sequence (S_t) is a Markov chain started from $\{v\}$ with transition kernel $\widehat{\mathbf{K}}$.*
3. *For any vertex u and time $t \geq 0$,*

$$\mathbf{P}^*_v[X_t = u \mid S_1, \ldots, S_t] = \mathbb{I}[u \in S_t]\frac{w(u)}{\mathrm{vol}(S_t)}.$$

12.3 Main Proof

In this section, we show how to find sets with small conductance by generating sample paths from the volume-biased ESP. The following theorem is the main theorem of this section from which we will prove Theorem 12.1.1.

Theorem 12.3.1. *Let $A \subset V$ be a set of vertices of volume $\mathrm{vol}(A) \leq k$, and conductance $\phi(A) \leq \phi$. For any $\epsilon \in (0, 1)$, fix $T = \epsilon \log k/3\phi$. There is a constant $c > 0$, and a subset $A' \subseteq A$ of volume $\mathrm{vol}(A') \geq \mathrm{vol}(A)/2$ for which the following holds. For any $v \in A'$, with probability at least $ck^{-\epsilon}/8$, a sample path $(S_1, S_2, \ldots, S_T)$ of the volume biased ESP started from $S_0 = \{v\}$ satisfies the following,*

i) For some $0 \leq t \leq T$, $\phi(S_t) \leq \Phi_\epsilon(\phi)$, where $\Phi_\epsilon(\phi) := \sqrt{100(1 - \log c)\phi/\epsilon}$,

ii) For all $0 \leq i \leq T$, $\mathrm{vol}(S_i \cap A) \geq ck^{-\epsilon}\mathrm{vol}(S_i)/2$, and henceforth,

iii) For all $0 \leq i \leq T$, $\mathrm{vol}(S_i) \leq \mathcal{K}_\epsilon(k)$, where $\mathcal{K}_\epsilon(k) := 2k^{1+\epsilon}/c$.

The proof of Theorem 12.3.1 is at the end of this section, after we present two necessary lemmas.

Let $(S_0, S_1, \ldots, S_\tau)$ be a sample path of the volume biased ESP, for a stopping time τ, Andersen and Peres show that with a constant probability the conductance of at least one of the sets in the sample path is at most $O(\sqrt{\frac{1}{\tau}\log \mathrm{vol}(S_\tau)})$,

Lemma 12.3.2 ([AP09, Corollary 1])**.** *For any starting set S_0, and any stopping time τ, and $\alpha > 0$,*

$$\widehat{\mathbf{P}}_{S_0}\left[\sum_{i=1}^{\tau}\phi^2(S_i) \le 4\alpha \ln\frac{\mathrm{vol}(S_\tau)}{\mathrm{vol}(S_0)}\right] \ge 1 - \frac{1}{\alpha}.$$

Here, we strengthen the above result, and we show the event occurs with significantly higher probability. In particular, we show the with probability at least $1 - 1/\alpha$, the conductance of at least one of the sets in the sample path is at most $O(\sqrt{\frac{1}{\tau}\log(\alpha \cdot \mathrm{vol}(S_\tau))})$.

Lemma 12.3.3. *For any starting set $S_0 \subseteq V$ and any stopping time , and $\alpha > 1$,*

$$\widehat{\mathbf{P}}_{S_0}\left[\sum_{i=1}^{\tau}\phi^2(S_i) \le 8\Big(\log\alpha + \log\frac{\mathrm{vol}(S_\tau)}{\mathrm{vol}(S_0)}\Big)\right] \ge 1 - \frac{1}{\alpha}.$$

Proof. We use the same martingale argument as in [AP09, Lemma 1], we include the full proof for the sake of completeness. We define a martingale M_t using the rate of change at each step $1-\psi(S_t)$, and then we use the optional sampling theorem to lower bound the growth in the size of the set at the stopping time τ. We define

$$M_t := F_t\frac{\sqrt{\mathrm{vol}(S_0)}}{\sqrt{\mathrm{vol}(S_t)}}, \quad \text{where } F_t := \prod_{j=0}^{t-1}(1-\psi(S_j))^{-1}, \text{ and } F_0 := 1, \tag{12.3.1}$$

First, we verify that (M_t) is a martingale in the volume-biased ESP:

$$\begin{aligned}
\widehat{\mathbf{E}}\left[M_t \middle| S_0, \ldots, S_{t-1}\right] &= \sqrt{\mathrm{vol}(S_0)}F_t\widehat{\mathbf{E}}\left[\frac{1}{\sqrt{\mathrm{vol}(S_t)}}\,\middle|\, S_{t-1}\right] \\
&= F_t\frac{\sqrt{\mathrm{vol}(S_0)}}{\sqrt{\mathrm{vol}(S_{t-1})}}\widehat{\mathbf{E}}_{S_{t-1}}\left[\frac{\sqrt{\mathrm{vol}(S_{t-1})}}{\sqrt{\mathrm{vol}(S_t)}}\right] \\
&= F_t\frac{\sqrt{\mathrm{vol}(S_0)}}{\sqrt{\mathrm{vol}(S_{t-1})}}\mathbf{E}_{S_{t-1}}\left[\frac{\sqrt{\mathrm{vol}(S_t)}}{\sqrt{\mathrm{vol}(S_{t-1})}}\right] \\
&= F_t\frac{\sqrt{\mathrm{vol}(S_0)}}{\sqrt{\mathrm{vol}(S_{t-1})}}(1-\psi(S_{t-1})) \\
&= F_{t-1}\frac{\sqrt{\mathrm{vol}(S_0)}}{\sqrt{\mathrm{vol}(S_{t-1})}} = M_{t-1}.
\end{aligned}$$

Let τ be a stopping time for the volume-biased ESP. By the optional stopping theorem for nonnegative martingales (see [Wil91]), we have

$$\widehat{\mathbf{E}}\left[M_\tau\right] \leq M_0 = 1.$$

Then by applying Jensen's inequality, and Markov's Inequality to the above equation we have

$$\widehat{\mathbf{P}}\left[\log M_\tau \leq \log \alpha\right] \geq 1 - \frac{1}{\alpha} \tag{12.3.2}$$

By the definition of M_τ,

$$\begin{aligned}
\log M_\tau = \log F_\tau + \frac{1}{2}\log\frac{\mathrm{vol}(S_0)}{\mathrm{vol}(S_\tau)} &= \log\prod_{i=0}^{\tau-1}\frac{1}{1-\psi(S_i)} - \frac{1}{2}\log\frac{\mathrm{vol}(S_\tau)}{\mathrm{vol}(S_0)} \\
&\geq \sum_{i=0}^{\tau-1}\psi(S_i) - \frac{1}{2}\log\frac{\mathrm{vol}(S_\tau)}{\mathrm{vol}(S_0)} \\
&\geq \frac{1}{8}\sum_{i=0}^{\tau-1}\phi^2(S_i) - \frac{1}{2}\log\frac{\mathrm{vol}(S_\tau)}{\mathrm{vol}(S_0)},
\end{aligned}$$

where the first inequality follows by the fact that $1/(1-\psi(S_i)) \geq e^{\psi(S_i)}$, and the last inequality follows by Proposition 12.2.2. The lemma follows from putting (12.3.2) and above equation together. □

The previous lemma shows that for any $k, \phi > 0$, if we can run the volume biased evolving set process for $T \approx \epsilon \log k/\phi$ steps without observing a set larger than $k^{O(1)}$, then, with probability $1 - 1/k$, one of the sets in the sample path must have a conductance of $O(\sqrt{\phi/\epsilon})$, which is what we are looking for.

The above lemma is quite strong and it may lead to disproving the SSE conjecture. We conjecture that if G has a set A of conductance ϕ, then there is a vertex $v \in A$ such that all of the sets in a sample path of the volume biased ESP started from $\{v\}$ and ran for $T = \Omega(\log(\mathrm{vol}(A))/\phi)$ steps have size at most $O(\mathrm{vol}(A))$. with probability $1/\mathrm{vol}(A)^{O(1)}$.

Conjecture 12.3.4. *There are universal constants $\epsilon_0 > 0, c > 1$ such that for any set $A \subseteq V$, there is a vertex $v \in A$ such that with probability at least $\mathrm{vol}(A)^{-c}$ a sample path $S_0, S_1, \ldots, S_T$ of the volume biased ESP started from $\{v\}$ satisfies*

$$\max_{1 \leq i \leq T} \mathrm{vol}(S_i) \leq O(\mathrm{vol}(A)),$$

where $T = \epsilon_0 \cdot \log(\mathrm{vol}(A))/\phi(A)$.

It is easy to see that the above conjecture combined with Lemma 12.3.3 refutes Conjecture 7.7.2.

In the rest of this section we use the following lemma of [AP09, Lemma 2] together with Proposition 8.3.1 to prove a weaker version of the above conjecture: we only upper bound the volume of the sets in a sample path of the ESP process with $\mathrm{vol}(A)^{O(1)}$.

Lemma 12.3.5. *[AP09, Lemma 2] For any set $A \subseteq V$, vertex $v \in A$, and integer $T \geq 0$, the following holds for all $\beta > 0$,*

$$\widehat{\mathbf{P}}_v \left[\max_{t \leq T} \frac{\mathrm{vol}(S_t - A)}{\mathrm{vol}(S_t)} > \beta \cdot \mathrm{esc}(v, T, A) \right] < \frac{1}{\beta}.$$

Proof. We use the Diaconis-Fill coupling, Proposition 12.2.5, between the volume-biased ESP Markov chain (S_t) and the random walk Markov chain (X_t). Recall that for any $t \geq 0$,

$$\mathbf{P}^* \left[X_t = u \mid S_0, \ldots, S_t \right] = \frac{w(u)}{\mathrm{vol}(S_t)} \mathbb{I} \left[u \in S_t \right].$$

Fix a value $\alpha \in [0, 1]$ and let τ be the first time t when $\mathrm{vol}(S_t - A) > \alpha \cdot \mathrm{vol}(S_t)$, or let $\tau = \infty$ if this does not occur. Consider the probability that $X_\tau \notin A$, conditioned on S_τ:

$$\mathbf{P}^* \left[X_\tau \notin A \mid S_\tau = S \right] = \sum_{u \in S - A} \frac{w(u)}{\mathrm{vol}(S)} = \frac{\mathrm{vol}(S - A)}{\mathrm{vol}(S)}.$$

By the definition of τ, we have $\mathbf{P}^* \left[X_\tau \notin A \mid \tau \leq T \right] > \alpha$, so

$$\begin{aligned} \mathrm{esc}(v, T, A) &= \mathbf{P}^* \left[\cup_{i=0}^{T} (X_i \notin A) \right] \\ &\geq \mathbf{P}^* \left[X_\tau \notin A \wedge \tau \leq T \right] \\ &= \mathbf{P}^* \left[X_\tau \notin A \mid \tau \leq T \right] \mathbf{P}^* \left[\tau \leq T \right] \\ &> \alpha \cdot \mathbf{P}^* \left[\tau \leq T \right]. \end{aligned}$$

Therefore,

$$\widehat{\mathbf{P}}_v \left[\max_{t \leq T} \frac{\mathrm{vol}(S_t - A)}{\mathrm{vol}(S_t)} > \alpha \right] = \mathbf{P}^* \left[\tau \leq T \right] < \frac{\mathrm{esc}(v, T, A)}{\alpha}.$$

The lemma follows by taking $\alpha = \beta \, \mathrm{esc}(v, T, A)$. □

We now combine the results of this section to prove Theorem 12.3.1. The proof simply follows from a simple application of the union bound.

Proof of Theorem 12.3.1. If $\phi \geq 1/2$ we simply return v. Otherwise assume $\phi < 1/2$. First of all, we let A' be the set of vertices $v \in A$ such that

$$\mathrm{rem}(v, T, A) \geq c \left(1 - \frac{3\phi(A)}{2} \right)^T.$$

By Proposition 8.3.1, there exists a constant $c > 0$ such that $\mathrm{vol}(A') \geq \mathrm{vol}(A)/2$. In the rest of the proof let v be a vertex in A'. We have,

$$\mathrm{esc}(v, T, A) \leq 1 - c\left(1 - \frac{3\phi(A)}{2}\right)^{T} \leq 1 - c\left(1 - \frac{3\phi}{2}\right)^{\frac{\epsilon \log k}{3\phi}} \leq 1 - ck^{-\epsilon},$$

where we used $\phi \leq 1/2$. Now, let $\beta := 1 + ck^{-\epsilon}/2$. By Lemma 12.3.5, we have

$$\widehat{\mathbf{P}}_v\left[\max_{t \leq T} \frac{\mathrm{vol}(S_t - A)}{\mathrm{vol}(S_t)} \leq \beta\, \mathrm{esc}(v, T, A) \leq 1 - \frac{ck^{-\epsilon}}{2}\right] \geq 1 - \frac{1}{\beta} \geq \frac{ck^{-\epsilon}}{4}$$

Since for any $S \subset V$, $\mathrm{vol}(S - A) + \mathrm{vol}(S \cap A) = \mathrm{vol}(S)$, we have

$$\widehat{\mathbf{P}}_v\left[\min_{t \leq T} \frac{\mathrm{vol}(S_t \cap A)}{\mathrm{vol}(S_t)} \geq \frac{ck^{-\epsilon}}{2}\right] \geq \frac{ck^{-\epsilon}}{4}$$

On the other hand, let $\alpha := k$. By Lemma 12.3.3, with probability $1 - 1/k$, for some $t \in [0, T]$,

$$\phi^2(S_t) \leq \frac{1}{T}\sum_{i=0}^{T} \phi^2(S_i) \leq \frac{8(\log k + \log \mathrm{vol}(S_T))}{T}.$$

Therefore, since $\epsilon < 1$, by the union bound we have

$$\widehat{\mathbf{P}}_v\left[\min_{t \leq T} \frac{\mathrm{vol}(S_t \cap A)}{\mathrm{vol}(S_t)} \geq \frac{ck^{-\epsilon}}{2} \bigwedge \exists\, t :\ \phi(S_t) \leq \sqrt{\frac{8(\log k + \log \mathrm{vol}(S_T))}{T}}\right] \geq \frac{ck^{-\epsilon}}{8}$$

Finally, since for any set $S \subseteq V$, $\mathrm{vol}(S \cap A) \leq \mathrm{vol}(A) \leq k$, in the above event, $\mathrm{vol}(S_T) \leq \frac{2k^{1+\epsilon}}{c}$. Therefore,

$$\phi(S_t) \leq \sqrt{\frac{8(\log k + \log(2k^{1+\epsilon}/c))}{T}} \leq \sqrt{\frac{100(1 - \log c)\phi}{\epsilon}},$$

which completes the proof. □

To prove Theorem 12.1.1, we can simply run k^{ϵ} copies of the volume biased ESP in parallel. Using the previous lemma with a constant probability at least one of the copies finds a non-expanding set. Moreover, we may bound the time complexity of the algorithm using Theorem 12.2.4. The details of the algorithm is described in Algorithm 14.

Algorithm 14 `ParESP`(v, k, ϕ, ϵ)

$T \leftarrow \epsilon \log k / 6\phi$.
Run $k^{\epsilon/2}$ independent copies of the volume biased ESP, using the simulator `GenerateSample`, starting from $\{v\}$, in parallel. Stop each copy as soon as the length of its sample path reaches T. As soon as any of the copies finds a set S, of volume $\mathrm{vol}(S) \leq \mathcal{K}_{\epsilon/2}(k)$, and conductance $\phi(S) \leq \Phi_{\epsilon/2}(\phi)$, stop the algorithm and return S.

Now we are ready to prove Theorem 12.1.1.

Proof of Theorem 12.1.1. Let A' be as defined in Theorem 12.3.1. First of all, for any $v \in A'$, by Theorem 12.3.1, each copy finds a set of volume $\mathcal{K}_{\epsilon/2}(k)$, and conductance $\Phi_{\epsilon/2}(\phi)$ with probability $\Omega(k^{-\epsilon/2})$; in this case the algorithm will output a set satisfying theorem's statement. But, since $k^{\epsilon/2}$ copies are executed independently, at least one of them will succeed with a constant probability. This proves the correctness of the algorithm.

It remains to compute the time complexity. Let, $l := k^{\epsilon/2}$ be the number of copies, and $W_1, \ldots, W_l$ be random variables indicating the work done by each of the copies in a single run of `ParESP`, thus $\sum_i W_i$ is the time complexity of the algorithm. Note that it is possible that $W_i < W_j$ for some $i \neq j$, since the i-th copy may stop without finding any small non-expanding sets. Let S_{out} be the output of the algorithm. If the algorithm returns the output of the i^{th} copy we define $I_i = 1/\text{vol}(S_{out})$ and we let $I_i = 0$ otherwise. Also, let $I := \sum I_i$; note that if the algorithm returns the empty set, then $I = 0$. We write $\widehat{\mathbf{P}}_v^l[.]$ to denote the probability measure of the l independent volume biased ESP all started from $S_0 = \{v\}$, and $\widehat{\mathbf{E}}_v^l[.]$ for the expectation. To prove the theorem it is sufficient to show

$$\widehat{\mathbf{E}}_v^l\left[I\sum_{i=1}^{l} W_i\right] = O(k^{\epsilon}\phi^{-1/2}\log^2 n).$$

By linearity of expectation, it is sufficient to show that for all $1 \leq i \leq k$,

$$\widehat{\mathbf{E}}_v^l\left[I_i\sum_{j=1}^{l} W_j\right] = O(k^{\epsilon/2}\phi^{-1/2}\log^2 n),$$

By symmetry of the copies, it is sufficient to show the above equation only for $i = 1$. Furthermore, since conditioned on $I_1 \neq 0$, $W_1 = \max_i W_i$, we just need to show,

$$\widehat{\mathbf{E}}_v^l[I_1 W_1] = O(\phi^{-1/2}\log^2 n),$$

Let τ be a stopping time for the first copy which indicates the first time t where $\text{vol}(S_t) \leq \mathcal{K}_{\epsilon/2}(k)$ and $\phi(S_t) \leq \Phi_{\epsilon/2}(\phi)$, S_τ be the corresponding set at time τ, and $W_1(\tau)$ be the amount of work done by time τ in the first copy. Note that we always have $W_1 \leq W_1(\tau)$ because the first copy may be stopped since one of the other copies succeeded. Since $I_1 \leq 1/\text{vol}(S_\tau)$ with probability 1, for any element of the joint probability space we have $I_1 W_1 \leq W_1(\tau)/\text{vol}(S_\tau)$. Therefore,

$$\widehat{\mathbf{E}}_v^l[I_1 W_1] \leq \widehat{\mathbf{E}}_v^l\left[\frac{W_1(\tau)}{\text{vol}(S_\tau)}\right] = \widehat{\mathbf{E}}_v\left[\frac{W(\tau)}{\text{vol}(S_\tau)}\right] = O(T^{1/2}\log^{3/2} n) = O(\phi^{-1/2}\log^2 n),$$

where the second to last equation follows from Theorem 12.2.4. □

Chapter 13

Partitioning into Expanders

There is a basic fact in algebraic graph theory that $\lambda_k > 0$ if and only if G has at most $k-1$ connected components. We prove a robust version of this fact. If $\lambda_k > 0$, then for some $1 \leq l \leq k-1$, V can be *partitioned* into l sets $P_1, \ldots, P_l$ such that each P_i is a low-conductance set in G and induces a high conductance induced subgraph. In particular, $\phi(P_i) \lesssim l^3\sqrt{\lambda_l}$ and $\phi(G[P_i]) \gtrsim \lambda_k/k^2$.

Our main technical lemma shows that if $(1+\epsilon)\rho(k) < \rho(k+1)$, then V can be partitioned into k sets $P_1, \ldots, P_k$ such that for each $1 \leq i \leq k$, $\phi(G[P_i]) \gtrsim \epsilon \cdot \rho(k+1)/k$ and $\phi(P_i) \leq k \cdot \rho(k)$. This significantly improves a recent result of Tanaka [Tan12] who assumed an exponential (in k) gap between $\rho(k)$ and $\rho(k+1)$.

The results of this chapter are based on a joint work with Luca Trevisan [OT13].

13.1 Introduction

In the preceding chapters we study several algorithms for finding sets of small conductance in G. As we described in the introduction a set of small conductance can represent a cluster of graph G. For example, Shi and Malik [SM00] used a partitioning of G to sets of small conductance and obtained

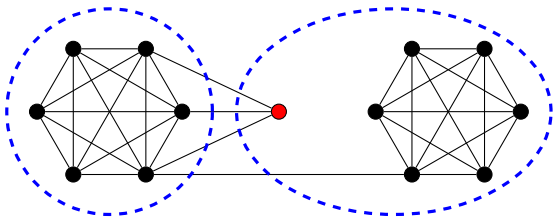

Figure 13.1.1: In this example although both sets in the 2-partitioning are of small conductance, in a natural clustering the red vertex (middle vertex) will be merged with the left cluster

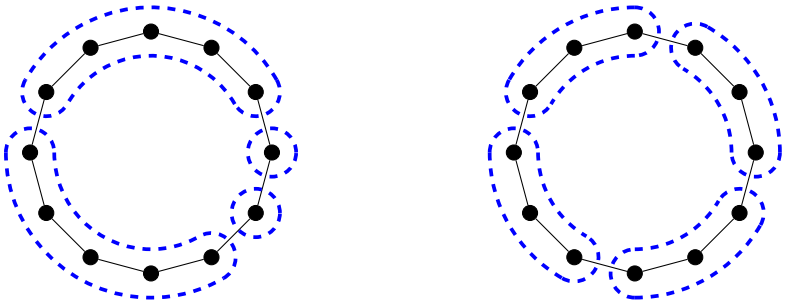

Figure 13.1.2: Two 4-partitioning of the cycle graph. In both of the partitionings the number of edges between the clusters are exactly 4, and the inside conductance of all components is at least 1/2 in both cases. But, the right clustering is a more natural clustering of cycle.

high quality solutions in image segmentation applications. It turns out that for many graphs just the fact that a set S has a small conductance is not enough to argue that it is a good cluster; this is because although $\phi(S)$ is small, S can be loosely-connected or even disconnected inside (see Figure 13.1.1).

Kannan, Vempala and Vetta [KVV04] proposed a bicriteria measure, where they measure the quality of a k-clustering based on the *inside* conductance of sets and the number of edges between the clusters. For $P \subseteq V$ let $\phi(G[P])$ be the *inside conductance* of P, i.e., the conductance of the induced subgraph of G on the vertices of P. Kannan et al. [KVV04] suggested that a k-partitioning into $P_1, \ldots, P_k$ is good if $\phi(G[P_i])$ is large, and $\sum_{i \neq j} w(P_i, P_j)$ is small. It turns out that an approximate solution for this objective function can be very different than the "correct" k-partitioning. Consider a 4-partitioning of a cycle as we illustrate in Figure 13.1.2. Although the inside conductance of every set in the left partitioning is within a factor 2 of the right partitioning, the left partitioning does not provide the "correct" 4-partitioning of a cycle.

In this chapter we propose a third objective which uses both of the inside/outside conductance of the clusters. Roughly speaking, $S \subseteq V$ represents a good cluster when $\phi(S)$ is small, but $\phi(G[S])$ is large. In other words, although S doesn't expand in G, the induced subgraph $G[S]$ is an expander.

Definition 13.1.1. *We say k disjoint subsets $A_1, \ldots, A_k$ of V are a (ϕ_{in}, ϕ_{out})-clustering, if for all $1 \leq i \leq k$,*

$$\phi(G[A_i]) \geq \phi_{in} \quad \text{and} \quad \phi_G(A_i) \leq \phi_{out}.$$

To the best of our knowledge, the only theoretical result that guarantees a $(\phi_{\text{in}}, \phi_{\text{out}})$ partitioning of G is a recent result of Tanaka [Tan12]. Tanaka [Tan12] proved that if there is a large enough gap between $\rho(k)$ and $\rho(k+1)$ then G has a k-partitioning that is a $(\exp(k)\rho(k), \rho(k+1)/\exp(k))$-clustering (see equation (10.1.1) for the definition of $\rho(k)$).

Theorem 13.1.2 (Tanaka [Tan12]). *If $\rho_G(k+1) > 3^{k+1}\rho_G(k)$ for some k, then G has a k-partitioning that is a $(\rho(k+1)/3^{k+1}, 3^k\rho(k))$-clustering.*

Unfortunately, Tanaka requires a very large gap (exponential in k) between $\rho(k)$ and $\rho(k+1)$. Furthermore, the above result is not algorithmic, in the sense that he needs to find the optimum sparsest cut of G or its induced subgraphs to construct the k-partitioning.

13.1.1 Related Works

Kannan, Vempala and Vetta in [KVV04] designed an approximation algorithm to find a partitioning of a graph that cuts very few edges and each set in the partitioning has a large inside conductance. Comparing to Definition 13.1.1 instead of minimizing $\phi(A_i)$ for each set A_i they minimize $\sum_i \phi(A_i)$. Very recently, Zhu, Lattanzi and Mirrokni [ZLM13] designed a *local algorithm* to find a set S such that $\phi(S)$ is small and $\phi(G[S])$ is large assuming that such a set exists. Both of these results do not argue about the existence of a partitioning with large inside conductance. Furthermore, unlike Cheeger type inequalities the quality of approximation factor of these algorithms depends on the size of the input graph (or the size of the cluster S).

13.1.2 Our Contributions

Partitioning into Expanders There is a basic fact in algebraic graph theory that for any graph G and any $k \geq 2$, $\lambda_k > 0$ if and only if G has at most $k-1$ connected components. It is a natural question to ask for a robust version of this fact. Our main existential theorem provides a robust version of this fact.

Theorem 13.1.3. *For any $k \geq 2$ if $\lambda_k > 0$, then for some $1 \leq l \leq k-1$ there is a l-partitioning of V into sets $P_1, \ldots, P_l$ that is a $(\Omega(\rho(k)/k^2), O(l\rho(l))) = (\Omega(\lambda_k/k^2), O(l^3)\sqrt{\lambda_l})$ clustering.*

The above theorem can be seen as a generalization of Theorem 10.1.1.

Algorithmic Results The above result is not algorithmic but with some loss in the parameters we can make them algorithmic.

Theorem 13.1.4 (Algorithmic Theorem). *There is a simple local search algorithm that for any $k \geq 1$ if $\lambda_k > 0$ finds a l-partitioning of V into sets $P_1, \ldots, P_l$ that is a $(\Omega(\lambda_k^2/k^4), O(k^6\sqrt{\lambda_{k-1}})$ where $1 \leq l < k$. If G is unweighted the algorithm runs in a polynomial time in the size of G.*

The details of the above algorithm are described in Algorithm 17. We remark that the algorithm does not use any SDP or LP relaxation of the problem. It only uses the Spectral Partitioning algorithm as a subroutine. Furthermore, unlike the spectral clustering algorithms studied in [NJW02, LOT12], our algorithm does not use multiple eigenfunctions of the normalized laplacian matrix. It rather iteratively refines a partitioning of G by adding non-expanding sets that induce an expander.

Suppose that there is a large gap between λ_k and λ_{k+1}. Then, the above theorem (together with 13.1.9) implies that there is a k partitioning of V such that inside conductance of each set is significantly larger than its outside conductance in G. Furthermore, such a partitioning can be found in polynomial time. This partitioning may represent one of the best k-clusterings of the graph G.

If instead of the Spectral Partitioning algorithm we use the $O(\sqrt{\log n})$-approximation algorithm for $\phi(G)$ developed in [ARV09] the same proof implies that $P_1, \ldots, P_l$ are a

$$\Big(\Omega\Big(\frac{\lambda_k}{k^2 \cdot \sqrt{\log(n)}}\Big), k^3\sqrt{\lambda_{k-1}}\Big)$$

clustering.

To the best of our knowledge, the above theorem provides the first polynomial time algorithm that establishes a Cheeger-type inequality for the inside/outside conductance of sets in a k-way partitioning.

Main Technical Result The main technical result of this paper is the following theorem. We show that even if there is a very small gap between $\rho(k)$ and $\rho(k+1)$ we can guarantee the existence of a $(\Omega_k(\rho(k+1)), O_k(\rho(k)))$-clustering.

Theorem 13.1.5 (Existential Theorem)**.** *If $\rho_G(k+1) > (1+\epsilon)\rho_G(k)$ for some $0 < \epsilon < 1$, then*

i) There exists k disjoint subsets of V that are a $(\epsilon \cdot \rho(k+1)/7, \rho(k))$-clustering.

ii) There exists a k-partitioning of V that is a $(\epsilon \cdot \rho(k+1)/(14k), k\rho(k))$-clustering.

The importance of the above theorem is that the gap is even independent of k and it can be made arbitrarily close to 0. Compared to Theorem 13.1.2, we require a very small gap between $\rho(k)$ and $\rho(k+1)$ and the quality of our k-partitioning has a linear loss in terms of k. We show tightness of above theorem in Subsection 13.1.3.

Using the above theorem it is easy to prove Theorem 13.1.3.

Proof of Theorem 13.1.3. Assume $\lambda_k > 0$ for some $k \geq 2$. By Theorem 10.1.1 we can assume $\rho(k) \geq \lambda_k/2 > 0$. Since $\rho(1) = 0$ we have $(1+1/k)\rho(l) < \rho(l+1)$ at least for one index $1 \leq l < k$. Let l be the largest index such that $(1+1/k)\rho(l) < \rho(l+1)$; it follows that

$$\rho(k) \leq (1+1/k)^{k-l-1}\rho(l+1) \leq e \cdot \rho(l+1). \tag{13.1.1}$$

Therefore, by part (ii) of Theorem 13.1.5 there is a l-partitioning of V into sets $P_1, \ldots, P_l$ such that for all $1 \leq i \leq l$,

$$\phi(G[P_i]) \geq \frac{\rho(l+1)}{14k \cdot l} \geq \frac{\rho(k)}{40k^2} \geq \frac{\lambda_k}{80k^2}, \text{and}$$

$$\phi(P_i) \leq l\rho(l) \leq O(l^3)\sqrt{\lambda_l}.$$

where we used (13.1.1) and Theorem 10.1.1. The following corollary follows. □

Building on Theorem 10.1.1 we can also prove the existence of a good k-partitioning of G if there is a large enough gap between λ_k and λ_{k+1}.

Corollary 13.1.6. *There is a universal constant $c > 0$, such that for any graph G if $\lambda_{k+1} \geq c \cdot k^2\sqrt{\lambda_k}$, then there exists a k-partitioning of G that is a $(\Omega(\lambda_{k+1}/k), O(k^3\sqrt{\lambda_k}))$-clustering.*

13.1.3 Tightness of Existential Theorem

In this part we provide several examples showing the tightness of Theorem 13.1.5. In the first example we show that if there is no gap between $\rho(k)$ and $\rho(k+1)$ then G cannot be partitioned into expanders.

Example 13.1.7. *In the first example we construct a graph such that there is no gap between $\rho(k)$ and $\rho(k+1)$ and we show that in any k-partitioning there is a set P such that $\phi(G[P]) \ll \rho(k+1)$. Suppose G is a star. Then, for any $k \geq 2$, $\rho(k) = 1$. But, among any k disjoint subsets of G there is a set P with $\phi(G[P]) = 0$. Therefore, for any $k \geq 2$, there is a set P with $\phi(G[P]) \ll \rho(k+1)$.*

In the next example we show that a *linear* loss in k is necessary in the quality of our k-partitioning in part (ii) of Theorem 13.1.5.

Example 13.1.8. *In this example we construct a graph such that in any k-partitioning there is a set P with $\phi(P) \geq \Omega(k \cdot \rho(k))$. Furthermore, in any k partitioning where the conductance of every set is $O_k(\rho(k))$, there is a set P such that $\phi(G[P]) \leq O(\rho(k+1)/k)$.*

Let G be a union of $k+1$ cliques $C_0, C_1, \ldots, C_k$ each with $\approx n/(k+1)$ vertices where $n \gg k$. Also, for any $1 \leq i \leq k$, include an edge between C_0 and C_i. In this graph $\rho(k) = \Theta(k^2/n^2)$ by choosing the k disjoint sets $C_1, \ldots, C_k$. Furthermore, $\rho(k+1) = \Theta(k \cdot \rho(k))$.

Now consider a k partitioning of G. First of all if there is a set P in the partitioning that contains a proper subset of one the cliques, i.e., $\emptyset \subset (P \cap C_i) \subset C_i$ for some i, then $\phi(P) \geq \Omega_k(1/n) = \Omega_k(n \cdot \rho(k))$. Otherwise, every clique is mapped to one of the sets in the partitioning. Now, let P be the set containing C_0 (P may contain at most one other clique). It follows that $\phi(P) = \Omega(k \cdot \rho(k))$.

Now, suppose we have a partitioning of G into k sets such that the conductance of each set is $O_k(\rho(k))$. By the arguments in above paragraph none of the sets in the partitioning can have a proper subset of one cliques. Since we have $k+1$ cliques there is a set P that contains exactly two cliques C_i, C_j, for $i \neq j$. It follows that $\phi(G[P]) \leq O(\rho(k)/k)$.

13.1.4 Notations

For a function $f : V \to \mathbb{R}$ let

$$\mathcal{R}(f) := \frac{\sum_{(u,v)\in E} |f(u) - f(v)|^2}{\sum_{v\in V} w(v)f(v)^2}$$

The support of f is the set of vertices with non-zero value in f,

$$\text{supp}(f) := \{v \in V : f(v) \neq 0\}.$$

We say two functions $f, g : V \to \mathbb{R}$ are disjointly supported if $\text{supp}(f) \cap \text{supp}(g) = \emptyset$.

For $S \subseteq P \subseteq V$ we use $\phi_{G[P]}(S)$ to denote the conductance of S in the induced subgraph $G[P]$. For $S, T \subseteq V$ we use

$$w(S \to T) := \sum_{u\in S, v\in T-S} w(u, v).$$

We remark that in the above definition S and T are not necessarily disjoint, so $w(S \to T)$ is not necessarily the same as $w(T \to S)$.

For $S \subseteq B_i \subseteq V$ we define

$$\varphi(S, B_i) := \frac{w(S \to B_i)}{\frac{\text{vol}(B_i - S)}{\text{vol}(B_i)} \cdot w(S \to V - B_i)}$$

Let us motivate the above definition. Suppose $B_i \subseteq V$ such that $\phi_G(B_i)$ is very small but $\phi(G[B_i])$ is very large. Then, any $S \subseteq B_i$ such that $\text{vol}(S) \leq \text{vol}(B_i)/2$ satisfy the following properties.

- Since $\phi_{G[B_i]}(S)$ is large, a large fraction of edges adjacent to vertices of S must leave this set.
- Since $\phi_G(B_i)$ is small, a small fraction of edges adjacent to S may leave B_i.

Putting above properties together we obtain that $w(S \to B_i) \gtrsim w(S \to V - B_i)$, thus $\varphi(S, B_i)$ is a constant. As we describe in the next section the converse of this argument is a crucial part of our proof. In particular, if for any $S \subseteq B_i$, $\varphi(S, B_i)$ is large, then B_i has large inside conductance, and it can be used as the "backbone" of our k-partitioning.

13.1.5 Overview of the Proof

We prove Theorem 13.1.5 in two steps. Let $A_1, \ldots, A_k$ be any k disjoint sets such that $\phi(A_i) \leq (1 + \epsilon)\rho(k + 1)$. In the first step we find $B_1, \ldots, B_k$ such that for $1 \leq i \leq k$, $\phi(B_i) \leq \phi(A_i)$ with the crucial property that any subset of B_i has at least a constant fraction of its outgoing edges inside B_i. We then use $B_1, \ldots, B_k$ as the "backbone" of our k-partitioning. We merge the remaining vertices with $B_1, \ldots, B_k$ to obtain $P_1, \ldots, P_k$ making sure that for each $S \subseteq P_i - B_i$ at least $1/k$ fraction of the outgoing edges of S go to P_i (i.e., $w(S \to P_i) \geq w(S \to V)/k$).

We show that if $2\max_{1\leq i\leq k}\phi(A_i) < \rho(k+1)$ then we can construct $B_1,\ldots,B_k$ such that every $S\subseteq B_i$ satisfies $\varphi(S,B_i)\geq\Omega(1)$ (see Lemma 13.2.1). For example, if $\mathrm{vol}(S)\leq\mathrm{vol}(B_i)/2$, we obtain that

$$w(S\to B_i - S)\gtrsim w(S\to V).$$

This property shows that each B_i has an inside conductance of $\Omega(\rho(k+1))$ (see Lemma 13.2.3). In addition, it implies that any superset of B_i, $P_i\supseteq B_i$, has an inside conductance $\phi(G[P_i])\gtrsim\alpha\cdot\rho(k+1)$ as long as for any $S\subseteq P_i - B_i$, $w(S\to B_i)\geq\alpha\cdot w(S\to V)$ (see Lemma 13.2.6). By latter observation we just need to merge the vertices in $V - B_1 - \ldots - B_k$ with $B_1,\ldots,B_k$ and obtain a k-partitioning $P_1,\ldots,P_k$ such that for any $S\subseteq P_i - B_i$, $w(S\to P_I)\geq w(S\to V)/k$.

13.1.6 Background on Higher Order Cheeger's Inequality

In this short section we use the machinery developed in Chapter 10 to show that for any partitioning of V into $l<k$ sets $P_1,\ldots,P_l$ the minimum inside conductance of P_i's is $\mathrm{poly}(k)\sqrt{\lambda_k}$.

Lemma 13.1.9. *There is a universal constant $c_0>$ such that for any $k\geq 2$ and any partitioning of V into l sets $P_1,\ldots,P_l$ of V where $l\leq k-1$, we have*

$$\min_{1\leq i\leq l}\lambda_2(G[P_i])\leq 2c_0k^6\lambda_k.$$

where $\lambda_2(G[P_i])$ is the second eigenvalue of the normalized laplacian matrix of the induced graph $G[P_i]$.

Proof. Let $f_1,\ldots,f_k$ be the first k eigenfunctions of $\mathcal{L}$ corresponding to $\lambda_1,\ldots,\lambda_k$. By definition $\mathcal{R}(f_i)=\lambda_i$.

By Theorem 10.1.5 there are k disjointly supported functions $g_1,\ldots,g_k$ such that $\mathcal{R}(g_i)\leq c_0k^6\lambda_k$. For any $1\leq j\leq l$ let $g_{i,j}$ be the restriction of g_i to the induced subgraph $G[P_i]$. It follows that

$$\mathcal{R}(g_i)\geq\frac{\sum_{j=1}^{l}\sum_{(u,v)\in E(P_j)}|g_i(v)-g_i(u)|^2}{\sum_{j=1}^{l}\sum_{v\in P_j}g_i(v)^2}\geq\min_{1\leq j\leq l}\frac{\sum_{(u,v)\in E(P_j)}|g_i(u)-g_i(v)|^2}{\sum_{v\in P_j}g_i(v)^2}=\min_{1\leq j\leq l}\mathcal{R}(g_{i,j}). \tag{13.1.2}$$

For each $1\leq i\leq l$ let $j(i):=\mathrm{argmin}_{1\leq j\leq l}\mathcal{R}(g_{i,j})$. Since $l<k$, by the pigeon hole principle, there are two indices $1\leq i_1<i_2\leq k$ such that $j(i_1)=j(i_2)=j^*$ for some $1\leq j^*\leq l$. Since $g_1,\ldots,g_k$ are disjointly supported, by Lemma 7.2.1

$$\lambda_2(G[P_{j^*}])\leq 2\max\{\mathcal{R}(g_{i_1,j^*}),\mathcal{R}(g_{i_2,j^*})\}\leq 2\max\{\mathcal{R}(g_{i_1}),\mathcal{R}(g_{i_2})\}\leq 2c_0k^6\lambda_k.$$

where the second inequality follows by (13.1.2). □

The above lemma is used in the proof of Theorem 13.1.4.

13.2 Proof of Existential Theorem

In this section we prove Theorem 13.1.5. Let $A_1, \ldots, A_k$ are k disjoint sets such that $\phi(A_i) \leq \rho(k)$ for all $1 \leq i \leq k$. In the first lemma we construct k disjoint sets $B_1, \ldots, B_k$ such that their conductance in G is only better than $A_1, \ldots, A_k$ with the additional property that $\varphi(S, B_i) \geq \epsilon/3$ for any $S \subseteq B_i$.

Lemma 13.2.1. *Let $A_1, \ldots, A_k$ be k disjoint sets s.t. $(1+\epsilon)\phi(A_i) \leq \rho(k+1)$ for $0 < \epsilon < 1$. For any $1 \leq i \leq k$, there exist a set $B_i \subseteq A_i$ such that the following holds:*

1. *$\phi(B_i) \leq \phi(A_i)$.*
2. *For any $S \subseteq B_i$, $\varphi(S, B_i) \geq \epsilon/3$.*

Proof. For each $1 \leq i \leq k$ we run Algorithm 15 to construct B_i from A_i. Note that although the algorithm is constructive, it may not run in polynomial time. The reason is that we don't know any (constant factor approximation) algorithm for $\min_{S \subseteq B_i} \varphi(S, B_i)$.

Algorithm 15 Construction of $B_1, \ldots, B_k$ from $A_1, \ldots, A_k$

```
B_i = A_i.
loop
    if ∃S ⊂ B_i such that φ(S, B_i) ≤ ε/3 then,
        Update B_i to either of S or B_i − S with the smallest conductance in G.
    else
        return B_i.
    end if
end loop
```

First, observe that the algorithm always terminates after at most $|A_i|$ iterations of the loop since $|B_i|$ decreases in each iteration. The output of the algorithm always satisfies conclusion 2 of the lemma. So, we only need to bound the conductance of the output set. We show that throughout the algorithm we always have

$$\phi(B_i) \leq \phi(A_i). \tag{13.2.1}$$

In fact, we prove something stronger. That is, the conductance of B_i never increases in the entire run of the algorithm. We prove this by induction. At the beginning $B_i = A_i$, so (13.2.1) obviously holds. It remains to prove the inductive step.

Let $S \subseteq B_i$ such that $\varphi(S, B_i) \leq \epsilon/3$. Among the $k+1$ disjoint sets $\{A_1, \ldots, A_{i-1}, S, T, A_{i+1}, A_k\}$ there is one of conductance $\rho_G(k+1)$. So,

$$\max\{\phi(S), \phi(B_i - S)\} \geq \rho_G(k+1) \geq (1+\epsilon)\phi(A_i) \geq (1+\epsilon)\phi(B_i).$$

The inductive step follows from the following lemma.

Lemma 13.2.2. *For any set $B_i \subseteq V$ and $S \subset B_i$, if $\varphi(S, B_i) \leq \epsilon/3$ and*

$$\max\{\phi(S), \phi(B_i - S)\} \geq (1+\epsilon)\phi(B_i), \tag{13.2.2}$$

then $\min\{\phi(S), \phi(B_i - S)\} \leq \phi(B_i)$.

Proof. Let $T = B_i - S$. Since $\varphi(S, B_i) \leq \epsilon/3$,

$$w(S \to T) \leq \frac{\epsilon}{3} \cdot \frac{\mathrm{vol}(T)}{\mathrm{vol}(B_i)} \cdot w(S \to V - B_i) \leq \frac{\epsilon}{3} \cdot w(S \to V - B_i). \tag{13.2.3}$$

We consider two cases depending on whether $\phi(S) \geq (1+\epsilon)\phi(B_i)$.

Case 1: $\phi(S) \geq (1+\epsilon)\phi(B_i)$**.** First, by (13.2.3).

$$(1+\epsilon)\phi(B_i) \leq \phi(S) = \frac{w(S \to T) + w(S \to V - B_i)}{\mathrm{vol}(S)} \leq \frac{(1+\epsilon/3)w(S \to V - B_i)}{\mathrm{vol}(S)} \tag{13.2.4}$$

Therefore,

$$\begin{aligned}
\phi(T) &= \frac{w(B_i \to V) - w(S \to V - B_i) + w(S \to T)}{\mathrm{vol}(T)} \\
&\leq \frac{w(B_i \to V) - (1-\epsilon/3)w(S \to V - B_i)}{\mathrm{vol}(T)} \\
&\leq \frac{\phi(B_i)(\mathrm{vol}(B_i) - \mathrm{vol}(S)(1+\epsilon/2)(1-\epsilon/3))}{\mathrm{vol}(T)} \\
&\leq \frac{\phi(B_i)\mathrm{vol}(T)}{\mathrm{vol}(T)} = \phi(B_i).
\end{aligned}$$

where the first inequality follows by (13.2.3) and the second inequality follows by (13.2.4) and that $\epsilon \leq 1$.

Case 2: $\phi(T) \geq (1+\epsilon)\phi(B_i)$**.** First,

$$(1+\epsilon)\phi(B_i) \leq \phi(T) = \frac{w(S \to T) + w(T \to V - B_i)}{\mathrm{vol}(T)} \tag{13.2.5}$$

Therefore,

$$\begin{aligned}\phi(S) &= \frac{w(B_i \to V) - w(T \to V - B_i) + w(S \to T)}{\mathrm{vol}(S)} \\ &\leq \frac{w(B_i \to V) - (1+\epsilon)\phi(B_i)\mathrm{vol}(T) + 2w(S \to T)}{\mathrm{vol}(S)} \\ &\leq \frac{\phi(B_i)(\mathrm{vol}(B_i) - (1+\epsilon)\mathrm{vol}(T)) + \frac{2\epsilon}{3} \cdot \mathrm{vol}(T) \cdot \phi(B_i)}{\mathrm{vol}(S)} \\ &\leq \frac{\phi(B_i)\mathrm{vol}(S)}{\mathrm{vol}(S)} = \phi(B_i).\end{aligned}$$

where the first inequality follows by (13.2.5), the second inequality follows by (13.2.3) and that $w(S \to V - B_i) \leq w(B_i \to V - B_i)$. So we get $\phi(S) \leq \phi(B_i)$.

This completes the proof of Lemma 13.2.2. □

This completes the proof of Lemma 13.2.1. □

Note that sets that we construct in the above lemma do not necessarily define a partitioning of G. In the next lemma we show that the sets $B_1, \ldots, B_k$ that are constructed above have large inside conductance.

Lemma 13.2.3. *Let $B_i \subseteq V$, and $S \subseteq B_i$ such that $\mathrm{vol}(S) \leq \mathrm{vol}(B_i)/2$. If $\varphi(S, B_i), \varphi(B_i - S, B_i) \geq \epsilon/3$ for $\epsilon \leq 1$, then*

$$\phi_{G[B_i]}(S) \geq \frac{w(S \to B_i)}{\mathrm{vol}(S)} \geq \frac{\epsilon}{7} \cdot \max\{\phi(S), \phi(B_i - S)\},$$

Proof. Let $T = B_i - S$. First, we lower bound $\phi_{G[B_i]}(S)$ by $\epsilon \cdot \phi(S)/7$. Since $\varphi(S, B_i) \geq \epsilon/3$,

$$\frac{w(S \to B_i)}{\mathrm{vol}(S)} = \frac{\varphi(S, B_i) \cdot \frac{\mathrm{vol}(T)}{\mathrm{vol}(B_i)} \cdot w(S \to V - B_i)}{\mathrm{vol}(S)} \geq \frac{\epsilon \cdot w(S \to V - B_i)}{6\mathrm{vol}(S)}$$

where the first inequality follows by the assumption $\mathrm{vol}(S) \leq \mathrm{vol}(B_i)/2$. Summing up both sides of the above inequality with $\frac{\epsilon w(S \to B_i)}{6\mathrm{vol}(S)}$ and dividing by $1 + \epsilon/6$ we obtain

$$\frac{w(S \to B_i)}{\mathrm{vol}(S)} \geq \frac{\epsilon/6}{(1+\epsilon/6} \cdot \frac{\cdot w(S \to V)}{\mathrm{vol}(S)} \geq \frac{\epsilon \cdot \phi(S)}{7}.$$

where we used $\epsilon \leq 1$. It remains to $\phi_{G[B_i]}(S)$ by $\epsilon \cdot \phi(B_i - S)/7$. Since $\varphi(T, B_i) \geq \epsilon/3$,

$$\begin{aligned}\frac{w(S \to B_i)}{\mathrm{vol}(S)} = \frac{w(T \to B_i)}{\mathrm{vol}(S)} &= \frac{\varphi(T, B_i) \cdot w(T \to V - B_i)}{\mathrm{vol}(B_i)} \\ &\geq \frac{\epsilon}{3} \cdot \frac{w(T \to V - B_i)}{\mathrm{vol}(B_i)} \\ &\geq \frac{\epsilon}{6} \cdot \frac{w(T \to V - B_i)}{\mathrm{vol}(T)}\end{aligned}$$

where the last inequality follows by the assumption $\mathrm{vol}(S) \leq \mathrm{vol}(B_i)/2$. Summing up both sides of the above inequality with $\frac{\epsilon \cdot w(S \to B_i)}{6\mathrm{vol}(S)}$ we obtain,

$$(1 + \epsilon/6)\frac{w(S \to B_i)}{\mathrm{vol}(S)} \geq \frac{\epsilon}{6} \cdot \frac{w(T \to V)}{\mathrm{vol}(T)} \geq \frac{\epsilon \cdot \phi(T)}{6}.$$

where we used the assumption $\mathrm{vol}(S) \leq \mathrm{vol}(B_i)/2$. The lemma follows using the fact that $\epsilon \leq 1$. □

Let $B_1, \ldots, B_k$ be the sets constructed in Lemma 13.2.1. Then, for each B_i and $S \subseteq B_i$ since $\phi(B_j) < \rho(k+1)$ for all $1 \leq j \leq k$, we get

$$\max(\phi(S), \phi(T)) \geq \rho(k+1).$$

Therefore, by the above lemma, for all $1 \leq i \leq k$,

$$\phi(G[B_i]) \geq \epsilon \cdot \rho(k+1)/7, \text{ and } \phi(B_i) \leq \max_{1 \leq i \leq k} \rho(A_i) \leq \rho(k).$$

This completes the proof of part (i) of Theorem 13.1.5.

It remains to prove part (ii). To prove part (ii) we have to turn $B_1, \ldots, B_k$ into a k-partitioning. We run the following algorithm to merge the vertices that are not included in $B_1, \ldots, B_k$. Again, although this algorithm is constructive, it may not run in polynomial time. The main difficulty is in finding a set $S \subset P_i - B_i$ such that $w(S \to P_i) < w(S \to P_j)$, if such a set exists.

Algorithm 16 Construction of $P_1, \ldots, P_k$ based on the $B_1, \ldots, B_k$

Let $P_i = B_i$ for all $1 \leq i \leq k-1$, and $P_k = V - B_1 - B_2 - \ldots - B_{k-1}$ (note that $B_k \subseteq P_k$).
while there is $i \neq j$ and $S \subset P_i - B_i$, such that $w(S \to P_i) < w(S \to P_j)$, **do**
  Update $P_i = P_i - S$, and merge S with $\mathrm{argmax}_{P_j} w(S \to P_j)$.
end while

First, observe that above algorithm always terminates in a finite number of steps. This is because in each iteration of the loop the weight of edges between the sets decreases. That is,

$$\sum_{1 \leq i < j \leq k} w(P_i \to P_j)$$

decreases. The above algorithm has two important properties which are the key ideas of the proof.

Fact 13.2.4. *The output of the above algorithm satisfies the following.*

1. *For all* $1 \leq i \leq k$, $B_i \subseteq P_i$.
2. *For any* $1 \leq i \leq k$, *and any* $S \subseteq P_i - B_i$, *we have*

$$w(S \to P_i) \geq w(S \to V)/k.$$

Next, we use the above properties to show that the resulting sets $P_1, \ldots, P_k$ are non-expanding in G

Lemma 13.2.5. *Let* $B_i \subseteq P_i \subseteq V$ *such that* $w(P_i - B_i \to B_i) \geq w(P_i - B_i \to V)/k$. *Then*

$$\phi(P_i) \leq k\phi(B_i).$$

Proof. Let $S = P_i - B_i$. Therefore,

$$\begin{aligned} \phi(P_i) = \frac{w(P_i \to V)}{\mathrm{vol}(P_i)} &\leq \frac{w(B_i \to V) + w(S \to V - P_i) - w(S \to B_i)}{\mathrm{vol}(B_i)} \\ &\leq \phi(B_i) + \frac{(k-1)w(B_i \to S)}{\mathrm{vol}(B_i)} \leq k\phi(B_i). \end{aligned}$$

The second inequality uses conclusion 2 of Fact 13.2.4. □

It remains to lower-bound the inside conductance of each P_i. This is proved in the next lemma. For a $S \subseteq P_i$ we use the following notations in the next lemma (see Figure 13.2.3 for an illustration).

$$\begin{aligned} S_B &:= B_i \cap S, \qquad & \overline{S}_B &:= B_i \cap \overline{S}, \\ S_P &:= S - B_i, \qquad & \overline{S}_P &:= \overline{S} - B_i. \end{aligned}$$

Lemma 13.2.6. *Let* $B_i \subseteq P_i \subseteq V$ *and let* $S \subseteq B_i$ *such that* $\mathrm{vol}(S_B) \leq \mathrm{vol}(B_i)/2$. *Let* $\rho \leq \phi(S_P)$ *and* $\rho \leq \max\{\phi(S_B), \phi(\overline{S}_B))\}$ *and* $0 < \epsilon < 1$. *If the following conditions hold then* $\phi(S) \geq \epsilon \cdot \rho/14k$.

1) If $S_P \neq \emptyset$, *then* $w(S_P \to P_i) \geq w(S_P \to V)/k$,

2) If $S_B \neq \emptyset$ *and* $S_B \neq B_i$, *then* $\varphi(S_B, B_i) \geq \epsilon/3$ *and* $\varphi(\overline{S}_B, B_i) \geq \epsilon/3$.

Proof. We consider 2 cases.

Case 1: $\mathrm{vol}(S_B) \geq \mathrm{vol}(S_P)$: Because of assumption (2) and $\mathrm{vol}(S_B) \leq \mathrm{vol}(B_i)/2$ we can apply Lemma 13.2.3, and we obtain

$$\phi_{G[P_i]}(S) \geq \frac{w(S \to P_i)}{\mathrm{vol}(S)} \geq \frac{w(S_B \to B_i)}{2\mathrm{vol}(S_B)} \geq \frac{\epsilon \cdot \max\{\phi(S_B), \phi(\overline{S}_B)\}}{14\mathrm{vol}(S_B)} \geq \frac{\epsilon \cdot \rho}{14}.$$

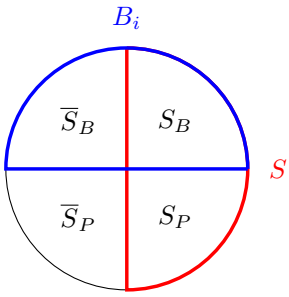

Figure 13.2.3: The circle represents P_i, the top (blue) semi-circle represents B_i and the right (red) semi-circle represents the set S.

Case 2: $\text{vol}(S_P) \geq \text{vol}(S_B)$:

$$\begin{aligned}
\phi_{G[P_i]}(S) \geq \frac{w(S \to P_i)}{\text{vol}(S)} &\geq \frac{w(S_P \to P_i - S) + w(S_B \to B_i)}{2\text{vol}(S_P)} \\
&\geq \frac{w(S_P \to P_i - S) + \epsilon \cdot w(S_B \to S_P)/6}{2\text{vol}(S_P)} \\
&\geq \frac{\epsilon \cdot w(S_P \to P_i)}{12\text{vol}(S_P)} \\
&\geq \frac{\epsilon \cdot w(S_P \to V)}{12k\text{vol}(S_P)} \\
&\geq \epsilon \cdot \phi(S_P)/12k \geq \epsilon \cdot \rho_G(k+1)/12k.
\end{aligned}$$

where the third inequality follows by the assumption that $\varphi(S_B, B_i) \geq \epsilon/3$ and $\text{vol}(S_B) \leq \text{vol}(B_i)/2$, and the fifth inequality follows by assumption (1). □

Let $B_1, \ldots, B_k$ be the sets constructed in Lemma 13.2.1 and $P_1, \ldots, P_k$ the sets constructed in Algorithm 16, First, observe that we can let $\rho = \rho(k+1)$. This is because among the $k+1$ disjoint sets $\{B_1, \ldots, B_{i-1}, S_B, \overline{S}_B, B_{i+1}, B_k\}$ there is a set of conductance $\rho(k+1)$. Similarly, among the sets $\{B_1, B_2, \ldots, B_k, S_P\}$ there is a set of conductance $\rho(k+1)$. Since for all $1 \leq i \leq k$, $\phi(B_i) < \rho(k+1)$, we $\max\{\phi(S_B, \overline{S}_B\} \geq \rho(k+1)$ and $\phi(P_S) \geq \rho(k+1)$. Therefore, by the above lemma,

$$\phi(G[P_i]) = \min_{S \subset P_i} \max\{\phi_{G[P_i]}(S), \phi_{G[P_i]}(P_i - S)\} \geq \epsilon \cdot \rho(k+1)/14k.$$

This completes the proof of part (ii) of Theorem 13.1.5.

13.3 Proof of Algorithmic Theorem

In this section we prove Theorem 13.1.4. Let

$$\rho^* := \min\{\lambda_k/10, 30c_0k^5\sqrt{\lambda_{k-1}}\}. \tag{13.3.1}$$

where c_0 is the constant defined in **??**. We use the notation $\phi_{\text{in}} := \lambda_k/140k^2$ and $\phi_{\text{out}} := 90c_0 \cdot k^6\sqrt{\lambda_{k-1}}$.

The idea of the algorithm is simple: we start with one partitioning of G, $P_1 = B_1 = V$. Each time we try to find a set S of small conductance in one P_i. Then, either we can use S to introduce a new set B_{l+1} of small conductance, i.e., $\phi(B_{l+1}) \le 4\rho^*$, or we can improve the current l-partitioning by refining B_i to one of its subsets (similar to Algorithm 15) or by moving parts of P_i to the other sets P_j (similar to Algorithm 16).

The details of our polynomial time algorithm are described in Algorithm 17. Our algorithm is a simple local search designed based on Algorithm 15 and Algorithm 16.

Algorithm 17 A polynomial time algorithm for partitioning G into k expanders

Input: $k > 1$ such that $\lambda_k > 0$.
Output: A $(\phi_{\text{in}}^2/4, \phi_{\text{out}})$ l-partitioning of G for some $1 \le l < k$.

1: Let $l = 1$, $P_1 = B_1 = V$.
2: **while** $\exists\ 1 \le i \le l$ such that $w(P_i - B_i \to B_i) < w(P_i - B_i \to P_j)$ for $j \ne i$, or Spectral Partitioning finds $S \subseteq P_i$ s.t. $\phi_{G[P_i]}(S), \phi_{G[P_i]}(P_i - S) < \phi_{\text{in}}$ **do**
3: Assume (after renaming) $\text{vol}(S \cap B_i) \le \text{vol}(B_i)/2$.
4: Let $S_B = S \cap B_i, \overline{S}_B = B_i \cap \overline{S}, S_P = (P_i - B_i) \cap S$ and $\overline{S}_P = (P_i - B_i) \cap \overline{S}$ (see Figure 13.2.3).
5: **if** $\max\{\phi(S_B), \phi(\overline{S}_B)\} \le (1 + 1/k)^{l+1}\rho^*$ **then**
6: Let $B_i = S_B, P_{l+1} = B_{l+1} = \overline{S}_B$ and $P_i = P_i - \overline{S}_B$. Set $l \leftarrow l + 1$ and **goto** step 2.
7: **end if**
8: **if** $\max\{\varphi(S_B, B_i), \varphi(\overline{S}_B, B_i)\} \le 1/3k$, **then**
9: Update B_i to either of S_B or $\overline{S}_B$ with the smallest conductance, and **goto** step 2.
10: **end if**
11: **if** $\phi(S_P) \le (1 + 1/k)^{l+1}\rho^*$ **then**
12: Let $P_{l+1} = B_{l+1} = S_P, P_i = P_i - S_P$. Set $l \leftarrow l + 1$ and **goto** step 2.
13: **end if**
14: **if** $w(P_i - B_i \to B_i) < w(P_i - B_i \to B_j)$ for $j \ne i$, **then**
15: Update $P_j = P_j \cup (P_i - B_i)$, and let $P_i = B_i$ and **goto** step 2.
16: **end if**
17: **if** $w(S_P \to P_i) < w(S_P \to P_j)$ for $j \ne i$, **then**
18: Update $P_i = P_i - S_P$ and merge S_P with $\text{argmax}_{P_j} w(S_P \to P_j)$.
19: **end if**
20: **end while**
return $P_1, \ldots, P_k$.

Observe that in the entire run of the algorithm $B_1, \ldots, B_l$ are always disjoint, $B_i \subseteq P_i$ and $P_1, \ldots, P_l$ form an l-partitioning of V. We prove Algorithm 17 by a sequence of steps.

Claim 13.3.1. *Throughout the algorithm we always have*

$$\max_{1\le i\le l} \phi(B_i) \le \rho^*(1+1/k)^l.$$

Proof. We prove the claim inductively. By definition, at the beginning $\phi(B_1) = 0$. In each iteration of the algorithm, $B_1, \ldots, B_l$ only change in steps 6,9 and 12. It is straightforward that by executing either of steps 6 and 12 we satisfy induction claim, i.e., we obtain $l+1$ sets $B_1, \ldots, B_{l+1}$ such that for all $1 \le i \le l+1$,

$$\phi(B_i) \le \rho^*(1+1/k)^{l+1}.$$

On the other hand, if step 9 is executed, then the condition of 5 is not satisfied, i.e.,

$$\max\{\phi(S_B), \phi(\overline{S}_B)\} > (1+1/k)^{l+1}\rho^* \ge (1+1/k)\phi(B_i).$$

where the last inequality follows by the induction hypothesis. Since $\min\{\varphi(S_B, B_i), \varphi(\overline{S}_B, B_i)\} \le 1/3k$ for $\epsilon = 1/k$ by Lemma 13.2.2 we get

$$\min\{\phi(S_B), \phi(\overline{S}_B)\} \le \phi(B_i) \le (1+1/k)^l\rho^*,$$

which completes the proof. □

Claim 13.3.2. *In the entire run of the algorithm we have $l < k$.*

Proof. The follows from the previous claim. If $l = k$, then by previous claim we have disjoint sets $B_1, \ldots, B_k$ such that

$$\max_{1\le i\le k} B_i \le \rho^*(1+1/k)^k \le e\cdot\rho^* \le e\lambda_k/10 < \lambda_k/2.$$

where we used (13.3.1). But, the above inequality implies $\rho(k) < \lambda_k/2$ which contradicts Theorem 10.1.1. □

Claim 13.3.3. *If the algorithm terminates, then it returns a l-partitioning of V that is a $(\phi_{in}^2/4, \phi_{out})$-clustering.*

Proof. Suppose the algorithm terminates with sets $B_1, \ldots, B_l$ and $P_1, \ldots, P_l$. Since by the loop condition, for each $1 \le i \le k$,

$$w(P_i - B_i \to B_i) \ge w(P_i - B_i \to V)/k,$$

by Lemma 13.2.5,

$$\phi(P_i) \le l\phi(B_i) \le l\cdot e\cdot\rho^* \le 90c_0\cdot k^6\sqrt{\lambda_{k-1}}.$$

where the second inequality follows by Claim 13.3.1, and the last inequality follows by Claim 13.3.2 and (13.3.1).

On the other hand, by the condition of the loop and the performance of Spectral Partitioning algorithm as described in Theorem 7.8.1, for each $1 \leq i \leq k$,

$$\phi(G[P_i]) \geq \phi_{\text{in}}^2/4 = \Omega(\lambda_k^2/k^4).$$

□

It remains to show that the algorithm indeed terminates. First, we show that in each iteration of the loop at least one of the conditions are satisfied.

Claim 13.3.4. *In each iteration of the loop at least one of the conditions hold.*

Proof. We use Lemma 13.2.6 to show that if none of the conditions in the loop are satisfied then $\phi(S) \geq \phi_{\text{in}}$ which is a contradiction. So, for the sake of contradiction assume in an iteration of the loop none of the conditions hold.

First, since conditions of 8 and 17 do not hold, for $\epsilon = 1/k$ assumptions (1) and (2) of Lemma 13.2.6 are satisfied. Furthermore, since condition of steps 5 and 11 do not hold

$$\max\{\phi(S_B, \overline{S}_B)\} = \max\{\phi(B_1), \ldots, \phi(B_{i-1}), \phi(S_B), \phi(\overline{S}_B), \phi(B_{i+1}, \ldots, \phi(B_l)\} \geq \max\{\rho^*, \rho(l+1)\}.$$
$$\phi(P_S) = \max\{\phi(B_1), \ldots, , \ldots, \phi(B_l), \phi(P_S)\} \geq \max\{\rho^*, \rho(l+1)\}.$$

where we used Claim 13.3.1. So, for $\rho = \rho^*$ and $\epsilon = 1/k$ by Lemma 13.2.6 we get

$$\phi(S) \geq \frac{\epsilon \cdot \rho}{14k} = \frac{\max\{\rho^*, \rho(l+1)\}}{14k^2}. \tag{13.3.2}$$

Now, if $l = k-1$, then by Theorem 10.1.1 we get

$$\phi(S) \geq \frac{\rho(k)}{14k^2} \geq \frac{\lambda_k}{28k^2} \geq \phi_{\text{in}},$$

which is a contradiction and we are done. Otherwise, we must have $l < k-1$. Then by Lemma 13.1.9,

$$\phi(S) \leq \min_{1 \leq i \leq l} \sqrt{2\lambda_2(G[P_i])} \leq \sqrt{4c_0 k^6 \lambda_{k-1}}, \tag{13.3.3}$$

where the first inequality follows by the Cheeger's inequality (Theorem 7.8.1), Putting (13.3.2) and (13.3.3) together we have

$$\rho^* \leq 14k^2 \sqrt{4c_0 k^6 \lambda_{k-1}}.$$

But, by definition of ρ^* in equation (13.3.1)), we must have $\rho^* = \lambda_k/10$. Therefore, by (13.3.2),

$$\phi(S) \geq \frac{\lambda_k}{140k^2} = \phi_{\text{in}},$$

which is a contradiction, and we are done. □

It remains to show that the algorithm actually terminates and if G is unweighted it terminates in polynomial time.

Claim 13.3.5. *For any graph G the algorithm terminates in finite number of iterations of the loop. Furthermore, if G is unweighted, the algorithm terminates after at most $O(kn \cdot |E|)$ iterations of the loop.*

Proof. In each iteration of the loop at least one of conditions in lines 5,8,11,14 and 17 are satisfied. By Claim 13.3.2, Lines 5 and 11 can be satisfied at most $k-1$ times. Line 8 can be satisfied at most n times (this is because each time the size of one B_i decreases by at least one vertex). Furthermore, for a fixed $B_1, \ldots, B_k$, 14,17 may hold only finite number of iterations, because each time the total weight of the edges between $P_1, \ldots, P_k$ decreases. In particular, if G is unweighted, the latter can happen at most $O(|E|)$ times. So, for undirected graphs the algorithm terminates after at most $O(kn \cdot |E|)$ iterations of the loop. □

This completes the proof of Theorem 13.1.4.

13.4 Concluding Remarks

We propose a new model for measuring the quality of k-partitionings of graphs which involves both the inside and the outside conductance of the sets in the partitioning. We believe that this is often an accurate model of the quality of solutions in practical applications. Furthermore, the simple local search Algorithm 17 can be used as a pruning step in any graph clustering algorithm.

Chapter 14

Open Problems

We conclude this thesis with a few open problems and future works. We think each of the Conjectures 5.3.2, 6.1.2 and 12.3.4 are very influential and each of them is a groundbreaking achievement in the field of theoretical CS.

A proof of Conjecture 5.3.2 shows that the integrality gap of the Held-Karp LP relaxation for ATSP (2.4.2) is constant. It also proves Conjecture 5.3.6 which has been open for decades. An algorithmic proof of Conjecture 5.3.2 provides a constant factor approximation algorithm for Asymmetric TSP. Recently, there has been some progress on a close variant of Conjecture 5.3.2 where it is shown $O(1/k)$ "spectrally" thin trees exist in graphs where the effective resistance between the endpoints of each edge is at most $1/k$ [MSS13]. We refer an interested to a recent work of Harvey and Olver [HO13, Section 4.3] for more information.

A proof of Conjecture 6.1.2 would finally break the 3/2 approximation algorithm of Christofides [Chr76]. Apart from that, such a proof shows that maximum entropy rounding by sampling method can be considered as a promising Heuristic in many applications of the traveling salesman problem.

A proof of Conjecture 12.3.4 would refute the Small-set expansion conjecture. There is a consensus that an algorithm for the small set expansion problem can be adopted to the unique games problem and refute this conjecture as well. Such a result can dramatically change our understanding of approximation algorithms, since it shows that the best approximation algorithm that we know for several of the most important optimization problems including maximum cut, minimum vertex cover, and maximum constraint satisfiability problems are not necessarily optimal.

There are several other open problems that we suggest for future works. We describe a list of them below.

1. It is easy to design a polynomial-time dynamic programming algorithm for Asymmetric TSP on graphs with bounded treewidth. Nevertheless, we still do not know if every k-edge connected graph with bounded treewidth has an $O(1)/k$-thin tree. Interestingly, Conjecture 5.3.6 is

also open for bounded treewidth graphs. An answer to any of these questions may lead to a constant factor approximation algorithm for ATSP in families of graphs with excluded minors.

2. In Chapter 10 we proved a higher order Cheeger's inequality and we argued that a dependency of $\Omega(\sqrt{\log k})$ is necessary in the right hand side of (10.1.2). It is an interesting open problem to find the right dependency to k in the RHS of (10.1.2) In particular, whetber for any graph G and any $k \geq 2$, $\rho_G(k) \leq \text{polylog}(k)\sqrt{\lambda_k}$.

3. Our higher order Cheeger's inequality, Theorem 10.1.1 and Theorem 10.3.1 provide a rigorous justification for the spectral clustering algorithm and in particular the use of kmeans Heuristic in the last step of this algorithm, see Algorithm 2. Our algorithm that we described in Section 10.6 suggests some natural questions. First, does dimension reduction always help to improve the quality of clusterings in practice? For instance, if one runs the k-means algorithm on the randomly projected points, does it yield better results? In Figure 1.2.2 we show that at least in some examples this idea helps. Another interesting question is whether, at least in certain circumstances, the quality of the k-means clustering can be rigorously analyzed when used in place of our random geometric partitioning.

4. In Theorem 13.1.5 we significantly improve Theorem 13.1.2 of Tanaka [Tan12] and we show that even if there is a small gap between $\rho(k)$ and $\rho(k+1)$, for some $k \geq 1$, then the graph admits a k-partitioning that is a $(\text{poly}(k)\rho(k+1), \text{poly}(k)\rho(k))$-clustering. Unfortunately, to carry-out this result to the domain of eigenvalues we need to look for a significantly larger gap between λ_k, λ_{k+1} (see Corollary 13.1.6). It is an interesting open problem if such a partitioning of G exists under only a constant gap between λ_k, λ_{k+1}. There has been a long line of works on the sparsest cut problem and partitioning of a graph into sets of small outside conductance [LR99, LLR95, ARV09, ALN08, BFK+11] but none of these works study the inside conductance of the sets in the partitioning. We think it is a fascinating open problem to study efficient algorithms based on linear programming or semidefinite programming relaxations that provide a bicriteria approximation to the $(\phi_{\text{in}}, \phi_{\text{out}})$-clustering problem.

Bibliography

[ABCC07] David L. Applegate, Robert E. Bixby, Vasek Chvatal, and William J. Cook. *The Traveling Salesman Problem: A Computational Study (Princeton Series in Applied Mathematics)*. Princeton University Press, Princeton, NJ, USA, 2007. 3, 30

[ABS10] Sanjeev Arora, Boaz Barak, and David Steurer. Subexponential algorithms for unique games and related problems. In *FOCS*, pages 563–572, Washington, DC, USA, 2010. IEEE Computer Society. 13, 177, 178, 223, 247, 261

[ACL06] Reid Andersen, Fan R. K. Chung, and Kevin J. Lang. Local graph partitioning using pagerank vectors. In *FOCS*, pages 475–486, 2006. 261, 262, 263

[AG11] Sanjeev Arora and Rong Ge. New tools for graph coloring. In *APPROX-RANDOM*, pages 1–12, 2011. 177

[AGG11] Nishita Aggarwal, Naveen Garg, and Swati Gupta. A 4/3-approximation for tsp on cubic 3-edge-connected graphs. CoRR abs/1101.5586, 2011. 25

[AGK+98] Sanjeev Arora, Michelangelo Grigni, David Karger, Philip Klein, and Andrzej Woloszyn. A polynomial-time approximation scheme for weighted planar graph tsp. In *SODA*, pages 33–41, 1998. 26

[AGM+10] Arash Asadpour, Michel X. Goemans, Aleksander Madry, Shayan Oveis Gharan, and Amin Saberi. An o(log n/ log log n)-approximation algorithm for the asymmetric traveling salesman problem. In *SODA*, pages 379–389, 2010. 6, 7, 32, 70, 103

[AHK10] Sanjeev Arora, Elad Hazan, and Satyen Kale. O(sqrt(log(n)) approximation to sparsest cut in $\tilde{o}(n^2)$ time. *SIAM J. Comput.*, 39(5):1748–1771, 2010. 264

[AK97] Noga Alon and Nabil Kahale. A spectral technique for coloring random 3-colorable graphs. *SIAM Journal on Computing*, 26:1733–1748, 1997. 3, 13

[AKS12] Hyung-Chan An, Robert Kleinberg, and David B. Shmoys. Improving christofides' algorithm for the s-t path tsp. In *STOC*, pages 875–886, 2012. 64

[AKY95] Charles J. Alpert, Andrew B. Kahng, and So-zen Yao. Spectral partitioning: The more eigenvectors, the better. In *DAC*, pages 195–200, 1995. 11, 15

[Ald90] D J Aldous. A random walk construction of uniform spanning trees and uniform labelled trees. *SIAM Journal on Discrete Mathematics*, 3(4):450–465, 1990. 49

[ALN08] Sanjeev Arora, James R. Lee, and Assaf Naor. Euclidean distortion and the Sparsest Cut. *J. Amer. Math. Soc.*, 21(1):1–21, 2008. 113, 176, 292

[Alo86] N Alon. Eigenvalues and expanders. *Combinatorica*, 6:83–96, January 1986. 3, 13, 14, 179

[AM85] N. Alon and V. Milman. Isoperimetric inequalities for graphs, and superconcentrators. *Journal of Combinatorial Theory, Series B*, 38(1):73–88, feb 1985. 3, 13, 14, 19, 20, 176, 179

[AP09] Reid Andersen and Yuval Peres. Finding sparse cuts locally using evolving sets. In *STOC*, pages 235–244, 2009. 18, 261, 263, 265, 267, 269, 271

[Aro96] Sanjeev Arora. Polynomial time approximation schemes for euclidean tsp and other geometric problems. In *FOCS*, pages 2–11, 1996. 26

[ARV09] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings and graph partitioning. *J. ACM*, 56:5:1–5:37, April 2009. 3, 18, 176, 207, 277, 292

[AS07] Arash Asadpour and Amin Saberi. An approximation algorithm for max-min fair allocation of indivisible goods. In *STOC*, pages 114–121, 2007. 66

[AS09] A. Asadpour and A. Saberi. Maximum entropy selection: a randomized rounding method, 2009. 66

[Bar98] Yair Bartal. On approximating arbitrary metrics by tree metrics. In *STOC*, pages 183–193, 1998. 185

[BBL09] Julius Borcea, Petter Branden, and Thomas M. Liggett. Negative dependence and the geometry of polynomials. *Journal of American Mathematical Society*, 22:521–567, 2009. 6, 54, 55, 56, 57, 58, 59

[BC11] Sylvia Boyd and Robert Carr. Finding low cost tsp and 2-matching solutions using certain half-integer subtour vertices. *Discrete Optimization*, 8(4):525–539, 2011. 4

[Bec75] William Beckner. Inequalities in Fourier analysis. *Ann. of Math. (2)*, 102(1):159–182, 1975. 241

[BEM10] S. Boyd and P. Elliott-Magwood. Structure of the extreme points of the subtour elimination polytope of the stsp. In *Combinatorial Optimization and Discrete Algorithms*, volume B23, pages 33–47, Research Institute for Mathematical Sciences, Kyoto University, Kyoto, Japan, 2010. 4

[Ben95] András A. Benczúr. A representation of cuts within 6/5 times the edge connectivity with applications. In *FOCS*, pages 92–102, 1995. 38, 44

[Ben97] Andras A. Benczúr. *Cut structures and randomized algorithms in edge-connectivity problems*. PhD thesis, MIT, 1997. 35, 39, 40, 44, 45, 46

[BFK+11] Nikhil Bansal, Uriel Feige, Robert Krauthgamer, Konstantin Makarychev, Viswanath Nagarajan, Joseph Naor, and Roy Schwartz. Min-max graph partitioning and small set expansion. In *FOCS*, pages 17–26. IEEE, 2011. 178, 264, 292

[BG08] András A. Benczúr and Michel X. Goemans. Deformable polygon representation and near-mincuts. *Building Bridges: Between Mathematics and Computer Science, M. Groetschel and G.O.H. Katona, Eds., Bolyai Society Mathematical Studies*, 19:103–135, 2008. 39, 44, 46, 47

[BH03] Matthew Brand and Kun Huang. A unifying theorem for spectral embedding and clustering, 2003. 11

[Big81] N. L. Biggs. T. p. kirkman, mathematician. *The Bulletin of the London Mathematical Society*, 13(2):97–120, 1981. 4

[BJ03] Francis R. Bach and Michael I. Jordan. Learning spectral clustering. In *NIPS*, 2003. 11

[Blä02] M. Bläser. A new approximation algorithm for the asymmetric tsp with triangle inequality. In *SODA*, pages 638–645, 2002. 4, 25

[BLR08] Punyashloka Biswal, James R. Lee, and Satish Rao. Eigenvalue bounds, spectral partitioning, and metrical deformations via flows. *Journal of The ACM*, abs/0808.0:751–760, 2008. 179, 247

[Bol02] B. Bollobas. *Modern Graph Theory*. Springer, 2002. 48

[Bon70] Aline Bonami. Étude des coefficients de Fourier des fonctions de $L^p(G)$. *Ann. Inst. Fourier (Grenoble)*, 20(fasc. 2):335–402 (1971), 1970. 241

[BP91] S. C. Boyd and William R. Pulleyblank. Optimizing over the subtour polytope of the travelling salesman problem. *Math. Program.*, 49:163–187, 1991. 4, 31

[BP93] R Burton and R Pemantle. Local characteristics, entropy and limit theorems for spanning trees and domino tilings via transfer-impedances. *Ann. Probab.*, 21(math.PR/0404048):1329–1371, 1993. 53

[BR65] G. R. Blakley and Prabir Roy. A holder type inequality for symmetric matrices with nonnegative entries. *Proceedings of the American Mathematical Society*, 16(6):1244–1245, Dec 1965. 201

[Brä07] Petter Brändén. Polynomials with the half-plane property and matroid theory. *Advances in Mathematics*, 216(1):302–320, 2007. 55

[Bro89] A. Broder. Generating random spanning trees. In *FOCS*, pages 442–447, 1989. 49

[BRS11] Boaz Barak, Prasad Raghavendra, and David Steurer. Rounding semidefinite programming hierarchies via global correlation. In *FOCS*, pages 472–481, 2011. 177, 247

[BS93] Stephen T. Barnard and Horst D. Simon. A fast multilevel implementation of recursive spectral bisection for partitioning unstructured problems. In *Parallel Processing for Scientific Computing*, pages 711–718, 1993. 14

[BS00] Hans-Joachim Böckenhauer and Sebastian Seibert. Improved lower bounds on the approximability of the traveling salesman problem. *RAIRO - Theoretical Informatics and Applications*, 34(3):213–255, 2000. 25, 26

[BSvdSS11] Sylvia Boyd, Rene Sitters, Suzanne van der Ster, and Leen Stougie. Tsp on cubic and subcubic graphs. In *IPCO*, pages 65–77, 2011. 25

[BV06] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2006. 27

[Cay89] Arthur Cayley. A theorem on trees. *Quart. J. Math.*, 23:376–378, 1889. 66, 73, 216

[CCG+98] M. Charikar, C. Chekuri, A. Goel, S. Guha, and S. Plotkin. Approximating a finite metric by a small number of tree metrics. In *Proceedings of the 39th Annual IEEE Symposium on Foundations of Computer Science*, 1998. 185, 242

[CCPV07] Gruia Calinescu, Chandra Chekuri, Martin Pál, and Jan Vondrák. Maximizing a submodular set function subject to a matroid constraint (extended abstract). In *IPCO*, pages 182–196, 2007. 75

[CGK06] Moses Charikar, Michel X. Goemans, and Howard Karloff. On the integrality ratio for the asymmetric traveling salesman problem. *MOR*, 31(2):245–252, 2006. 31, 109

[CGR05] S. Chawla, A. Gupta, and H. Räcke. An improved approximation to sparsest cut. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms*, Vancouver, 2005. ACM. 176

[Che70] J. Cheeger. A lower bound for the smallest eigenvalue of the laplacian. *Problems in Analysis (R. C. Gunning Ed.)*, pages 195–199, 1970. 179

[Chr76] Nicos Christofides. Worst case analysis of a new heuristic for the traveling salesman problem. Report 388, Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, PA, 1976. 4, 6, 25, 26, 30, 64, 119, 291

[Chu89] F. R. K. Chung. Diameters and eigenvalues. *J. Amer. Math. Soc.*, 2:187–196, 1989. 20

[Chu96] F. R. K. Chung. Laplacians of graphs and Cheeger's inequalities. In *Combinatorics, Paul ErdGos is eighty, Vol. 2 (Keszthely, 1993)*, volume 2 of *Bolyai Soc. Math. Stud.*, pages 157–172. János Bolyai Math. Soc., Budapest, 1996. 180

[Chu04] Fan Chung. Discrete isoperimetric inequalities. In *Surveys in Differential Geometry IX*, pages 53–82. International Press, 2004. 217

[CKK+05] Shuchi Chawla, Robert Krauthgamer, Ravi Kumar, Yuval Rabani, and D. Sivakumar. On the hardness of approximating multicut and sparsest-cut. In *CCC*, pages 144–153, 2005. 176

[CMN96] Charles J. Colbourn, Wendy J. Myrvold, and Eugene Neufeld. Two algorithms for unranking arborescences. *Journal of Algorithms*, 20(2):268–281, 1996. 49

[CSZ94] Pak K. Chan, Martine D. F. Schlag, and Jason Y. Zien. Spectral k-way ratio-cut partitioning and clustering. *IEEE Trans. on CAD of Integrated Circuits and Systems*, 13(9):1088–1096, 1994. 11

[CV00] Robert D. Carr and Santosh Vempala. Towards a 4/3 approximation for the asymmetric traveling salesman problem. In *SODA*, pages 116–125, 2000. 4, 31, 111, 118

[CVZ10] Chandra Chekuri, Jan Vondrák, and Rico Zenklusen. Dependent randomized rounding via exchange properties of combinatorial structures. In *FOCS*, pages 575–584, 2010. 75

[Dar64] J. N. Darroch. On the distribution of the number of successes in independent trials. *Ann. Math. Stat.*, 36:1317–1321, 1964. 59

[DF90] Persi Diaconis and James A. Fill. Strong stationary times via a new form of duality. *Ann. Probab.*, 18(4):1483–1522, 1990. 263, 267, 268

[DFJ54] G Dantzig, R Fulkerson, and S Johnson. Solution of a large-scale traveling-salesman problem. *Operations Research*, 2:393–410, 1954. 4, 29

[DFJ59] G.B. Dantzig, D.R. Fulkerson, and S. Johnson. On a linear programming combinatorial approach to the traveling salesman problem. *OR*, 7:58–66, 1959. 4

[DHM07] Erik D. Demaine, MohammadTaghi Hajiaghayi, and Bojan Mohar. Approximation algorithms via contraction decomposition. In *SODA*, pages 278–287, 2007. 26

[DJM12] Amir Daneshgar, Ramin Javadi, and Laurent Miclo. On nodal domains and higher-order Cheeger inequalities of finite reversible Markov processes. *Stochastic Processes and their Applications*, 122(4):1748–1776, 2012. 222, 225

[DKL76] E.A. Dinits, A.V. Karzanov, and M.V. Lomonosov. On the structure of a family of minimal weighted cuts in graphs. *Studies in Discrete Mathematics (in Russian), ed. A.A. Fridman, 290-306, Nauka (Moskva)*, 1976. 37, 38

[DM05] Luca Donetti and Miguel A. Muñoz. Improved spectral algorithm for the detection of network communities. In *Modeling cooperative behavior in the social sciences*, pages 104–107, 2005. 11

[Edm70] Jack Edmonds. Submodular functions, matroids and certain polyhedra. In *Combinatorial Structures and Their Applications*, pages 69–87, New York, NY, USA, 1970. Gordon and Breach. 31

[EJ73] Jack Edmonds and Ellis L. Johnson. Matching, euler tours and the chinese postman. *Mathematical Programming*, 5(1):88–124, 1973. 32

[Ele10] Gábor Elek. Parameter testing in bounded degree graphs of subexponential growth. *Random Structures Algorithms*, 37(2):248–270, 2010. 216

[Eng99] Lars Engebretsen. An explicit lower bound for tsp with distances one and two. In *STACS*, pages 373–382. Springer-Verlag, 1999. 25

[ES13] Jeff Erickson and Anastasios Sidiropoulos. A near-optimal approximation algorithm for asymmetric tsp on embedded graphs. CoRR abs/1304.1810, 2013. 112

[Fel68] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 1 of *3rd edition*. John Wiley & Sons, Inc., 1968. 168

[FF09] Tamás Fleiner and András Frank. A quick proof for the cactus representation of mincuts. Technical Report QP-2009-03, Egerváry Research Group, Budapest, 2009. `www.cs.elte.hu/egres`. 37

[FGM82] A. M. Frieze, G. Galbiati, and F. Maffioli. On the worst-case performance of some algorithms for the asymmetric traveling salesman problem. *Networks*, 12:23–39, 1982. 4, 7, 25, 31

[FK02] U. Feige and R. Krauthgamer. A polylogarithmic approximation of the minimum bisection. *SIAM Journal on Computing*, 31(3):1090–1118, 2002. 246

[FM92] Tomás Feder and Milena Mihail. Balanced matroids. In *Proceedings of the twenty-fourth annual ACM symposium on Theory of Computing*, pages 26–38, New York, NY, USA, 1992. ACM. 57

[FMMM09] Jon Feldman, Aranyak Mehta, Vahab Mirrokni, and S. Muthukrishnan. Online stochastic matching: Beating 1-1/e. In *FOCS*, pages 117–126, 2009. 7, 8, 9, 10

[For09] Lance Fortnow. The status of the p versus np problem. *Commun. ACM*, 52(9):78–86, September 2009. 2

[Fri96] Joel Friedman. Minimum higher eigenvalues of Laplacians on graphs. *Duke Math. J.*, 83(1):1–18, 1996. 208

[FS07] U. Feige and M. Singh. Improved approximation ratios for traveling salesman tours and path sin directed graphs. In *APPROX*, pages 104–118, 2007. 4, 25

[FT03] J. Fakcharoenphol and K. Talwar. An improved decomposition theorem for graphs excluding a fixed minor. In *Proceedings of 6th Workshop on Approximation, Randomization, and Combinatorial Optimization*, volume 2764 of *Lecture Notes in Computer Science*, pages 36–46. Springer, 2003. 185

[GBS08] Arpita Ghosh, Stephen Boyd, and Amin Saberi. Minimizing effective resistance of a graph. *SIAM Rev.*, 50(1):37–66, 2008. 51

[GHGG99] Luis A. Goddyn, Pavol Hell, Anna Galluccio, and Anna Galluccio. High girth graphs avoiding a minor are nearly bipartite. *J. Comb. Theory (B)*, 83:1–14, 1999. 118

[GKL03] Anupam Gupta, Robert Krauthgamer, and James R. Lee. Bounded geometries, fractals, and low-distortion embeddings. In *44th Symposium on Foundations of Computer Science*, pages 534–543, 2003. 185

[GKP95] M. Grigni, E. Koutsoupias, and C. Papadimitriou. An approximation scheme for planar graph tsp. In *FOCS '95: Proceedings of the 36th Annual Symposium on Foundations of Computer Science*, page 640, Washington, DC, USA, 1995. IEEE Computer Society. 26

[GL96] G. H. Golub and C. F. Van Loan. *Matrix Computations*. 3rd. Edition. The Johns Hopkins University Press, Baltimore, MD,, 1996. 51

[Gle75] Leon Jay Gleser. On the distribution of the number of successes in independent trials. *Ann. Probab.*, 3(1):182–188, 1975. 59

[GLS93] Martin Grötschel, László Lovász, and Alexander Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, second edition, 1993. 8, 9, 27

[GLS05] David Gamarnik, Moshe Lewenstein, and Maxim Sviridenko. An improved upper bound for the tsp in cubic 3-edge-connected graphs. *Oper. Res. Lett.*, 33(5):467–474, September 2005. 4

[God04] Luis A. Goddyn. Some open problems i like. available at *http://www.math.sfu.ca/ goddyn/Problems/problems.html*, 2004. 116

[Goe95] Michel X. Goemans. Worst-case comparison of valid inequalities for the tsp. *MATH. PROG*, 69:335–349, 1995. 4

[Goe06] Michel X. Goemans. Minimum bounded degree spanning trees. In *FOCS*, pages 273–282, 2006. 31

[GS11] Venkatesan Guruswami and Ali Kemal Sinop. Lasserre hierarchy, higher eigenvalues, and approximation schemes for graph partitioning and quadratic integer programming with psd objectives. In *FOCS*, pages 482–491, 2011. 177, 247

[GS12] Venkatesan Guruswami and Ali Kemal Sinop. Faster sdp hierarchy solvers for local rounding algorithms. In *FOCS*, 2012. 177

[GS13] Venkatesan Guruswami and Ali Kemal Sinop. Approximating non-uniform sparsest cut via generalized spectra. In *SODA*, pages 295–305, 2013. 177, 215

[GT12] Shayan Oveis Gharan and Luca Trevisan. A universal upper bound on graph diameter based on laplacian eigenvalues. CoRR abs/1212.2701, 2012. 19

[Gué83] A. Guénoche. Random spanning tree. *J. Algorithms*, 4(3):214–220, 1983. 49

[GW95] Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6):1115–1145, November 1995. 3

[HK70] M. Held and R.M. Karp. The traveling salesman problem and minimum spanning trees. *Operations Research*, 18:1138–1162, 1970. 29, 120

[HL92] Bruce Hendrickson and Robert Leland. An improved spectral graph partitioning algorithm for mapping parallel computations. *Siam Journal on Scientific Computing*, 16, 1992. 14

[HLP52] G. H. Hardy, J. E. Littlewood, and G. Polya. *Inequalities*. Cambridge Univ. Press, 1952. 59

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006. 14, 161, 179

[HO13] Nicholas J. A. Harvey and Neil Olver. Pipage rounding, pessimistic estimators and matrix concentration. CORR abs/1307.2274, 2013. 291

[Hoe56] W. Hoeffding. On the distribution of the number of successes in independent trials. *Ann. Math. Statist.*, 27:713–721, 1956. 59, 60

[Jae84] F. Jaeger. On circular flows in graphs, finite and infinite sets. *Colloquia Mathematica Societatis Janos Bolyai 37*, pages 391–402, 1984. 118

[Jai01] Kamal Jain. A factor 2 approximation algorithm for the generalized steiner network problem. *Combinatorica*, 21:39–60, 2001. 3

[JL84] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conference in modern analysis and probability (New Haven, Conn., 1982)*, pages 189–206. Amer. Math. Soc., Providence, RI, 1984. 16, 193

[JM85] S Jimbo and A Maruoka. Expanders obtained from affine transformations. In *STOC*, pages 88–97, New York, NY, USA, 1985. ACM. 14, 179

[JSV04] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *J. ACM*, 51(4):671–697, July 2004. 13, 14, 70, 179

[Kal12] Gil Kalai, 2012. Personal Communication. 20

[Kar72] Richard M. Karp. Reducibility among combinatorial problems. In *Complexity of Computer Computations*, pages 85–103, 1972. 4

[Kar77] Richard M. Karp. Probabilistic analysis of partitioning algorithms for the traveling-salesman in the plane. *MOR*, 2:209–224, 1977. 4

[Kar79] Richard M. Karp. A patching algorithm for the nonsymmetric traveling-salesman problem. *SIAM J. Comput.*, 8(4):561–573, 1979. 4

[Kar93] David R. Karger. Global min-cuts in rnc, and other ramifications of a simple min-out algorithm. In *SODA*, pages 21–30, Philadelphia, PA, USA, 1993. Society for Industrial and Applied Mathematics. 47

[Kar00] David R. Karger. Minimum cuts in near-linear time. *J. ACM*, 47(1):46–76, January 2000. 30, 48

[Kel04] Jonathan A. Kelner. Spectral partitioning, eigenvalue bounds, and circle packings for graphs of bounded genus. In *STOC*, pages 455–464. Press, 2004. 13, 179, 247

[KL12] Tsz Chiu Kwok and Lap Chi Lau. Finding small sparse cuts by random walk. In *APPROX-RANDOM*, pages 615–626, 2012. 264

[Kle05] Philip N. Klein. A linear-time approximation scheme for planar weighted tsp. In *FOCS*, pages 647–657, 2005. 26

[KLL+13] Tsz Chiu Kwok, Lap Chi Lau, Yin Tat Lee, Shayan Oveis Gharan, and Luca Trevisan. Improved cheeger's inequality: analysis of spectral partitioning algorithms through higher order spectral gap. In *STOC*, pages 11–20, 2013. 17, 160, 244, 248, 255, 256

[KLPT11] J. Kelner, James R. Lee, G. Price, and S.-H. Teng. Metric uniformization and spectral bounds for graphs. *Geom. Funct. Anal.*, 21(5):1117–1143, 2011. 179, 207, 224, 247

[KLSS05] H. Kaplan, M. Lewenstein, N. Shafrir, and M. Sviridenko. Approximation algorithms for asymmetric TSP by decomposing directed regular multigraphs. *J. ACM*, pages 602–626, 2005. 4, 25

[KM09] Jonathan Kelner and Aleksander Madry. Faster generation of random spanning trees. In *FOCS*, 2009. 49

[KMP10] Ioannis Koutis, Gary L. Miller, and Richard Peng. Approaching optimality for solving sdd linear systems. In *FOCS*, pages 235–244, 2010. 170, 171

[KMP11] Ioannis Koutis, Gary L. Miller, and Richard Peng. A nearly-m log n time solver for sdd linear systems. In *FOCS*, pages 590–598, 2011. 170, 171

[Kol11] Alexandra Kolla. Spectral algorithms for unique games. *Comput. Complex.*, 20(2):177–206, June 2011. 177

[KOSZ13] Jonathan A. Kelner, Lorenzo Orecchia, Aaron Sidford, and Zeyuan Allen Zhu. A simple, combinatorial algorithm for solving sdd systems in nearly-linear time. In *STOC*, pages 911–920, 2013. 170, 171

[KPR93] Philip N. Klein, Serge A. Plotkin, and Satish Rao. Excluded minors, network decomposition, and multicommodity flow. In *Proceedings of the 25th Annual ACM Symposium on Theory of Computing*, pages 682–690, 1993. 185

[KT07] Alexandra Kolla and Madhur Tulsiani. Playing unique games using graph spectra. manuscript, 2007. 177

[Kul90] V. G. Kulkarni. Generating random combinatorial objects. *J. Algorithms*, 11(2):185–207, 1990. 49

[KVV04] Ravi Kannan, Santosh Vempala, and Adrian Vetta. On clusterings: Good, bad and spectral. *J. ACM*, 51(3):497–515, May 2004. 14, 173, 179, 247, 261, 275, 276

[Lam12] Michael Lampis. Improved inapproximability for tsp. In *APPROX-RANDOM*, Lecture Notes in Computer Science, pages 243–253. Springer, 2012. 4, 25, 26

[Lee12] James R. Lee, 2012. http://tcsmath.wordpress.com/2012/04/18/gabber-galil-analysis-of-margulis-expanders/. 14, 179

[LK99] László Lovász and Ravi Kannan. Faster mixing via average conductance. In *STOC*, pages 282–287. ACM, 1999. 177

[LLDM09] Jure Leskovec, Kevin J. Lang, Anirban Dasgupta, and Michael W. Mahoney. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*, 6(1):29–123, 2009. 262

[LLM10] Jure Leskovec, Kevin J. Lang, and Michael Mahoney. Empirical comparison of algorithms for network community detection. In *WWW*, pages 631–640, New York, NY, USA, 2010. ACM. 244, 262

[LLR95] Nathan Linial, Eran London, and Yuri Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15:577–591, 1995. 176, 292

[LN05] James R. Lee and Assaf Naor. Extending Lipschitz functions via random metric partitions. *Invent. Math.*, 160(1):59–95, 2005. 185

[LO81] H.J. Landau and A.M. Odlyzko. Bounds for eigenvalues of certain stochastic matrices. *Linear Algebra Appl.*, 38:5–15, 1981. 205, 208

[LO12] Russell Lyons and Shayan Oveis Gharan. Sharp bounds on random walk eigenvalues via spectral embedding. submitted, 2012. 18, 205, 207, 221

[Lon66] David London. Inequalities in quadratic forms. *Duke Math. J.*, 33(3):511–522, 1966. 201

[LOS12] Bundit Laekhanukit, Shayan Oveis Gharan, and Mohit Singh. A rounding by sampling approach to the minimum size k-arc connected subgraph problem. In *ICALP (1)*, pages 606–616, 2012. 6, 64

[LOT12] James R. Lee, Shayan Oveis Gharan, and Luca Trevisan. Multi-way spectral partitioning and higher-order cheeger inequalities. In *STOC*, pages 1117–1130, 2012. 15, 221, 276

[Lov12] László Lovász. *Large Networks and Graph Limits*. Amer. Math. Soc., Providence, RI, 2012. 216

[LP11] Angsheng Li and Pan Peng. Community structures in classical network models. *Internet Mathematics*, 7(2):81–106, 2011. 262

[LP13] Russell Lyons and Yuval Peres. *Probability on Trees and Networks*. Cambridge University Press, 2013. To appear. 48, 52, 53, 68

[LPP11] Angsheng Li, Yicheng Pan, and Pan Peng. Testing conductance in general graphs. *Electronic Colloquium on Computational Complexity (ECCC)*, 18:101, 2011. 262

[LPW06] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, 2006. 166, 203, 266, 268

[LR99] Tom Leighton and Satish Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *Journal of The ACM*, 46:787–832, November 1999. 175, 176, 185, 292

[LRS11] Lap-Chi Lau, R. Ravi, and Mohit Singh. *Iterative Methods in Combinatorial Optimization*. Cambridge University Press, New York, NY, USA, 1st edition, 2011. 3

[LRTV11] Anand Louis, Prasad Raghavendra, Prasad Tetali, and Santosh Vempala. Algorithmic extensions of Cheeger's inequality to higher eigenvalues and partitions. In *APPROX-RANDOM*, pages 315–326, 2011. 224

[LRTV12] Anand Louis, Prasad Raghavendra, Prasad Tetali, and Santosh Vempala. Many sparse cuts via higher eigenvalues. In *STOC*, 2012. 223

[LS93] Nathan Linial and Michael Saks. Low diameter graph decompositions. *Combinatorica*, 13(4):441–454, 1993. 185

[LS10] J. R. Lee and A. Sidiropoulos. Genus and the geometry of the cut graph. In *Proceedings of the 21st ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 193–201, 2010. 186

[LT11] Michel Ledoux and Michel Talagrand. *Probability in Banach spaces*. Classics in Mathematics. Springer-Verlag, Berlin, 2011. Isoperimetry and processes, Reprint of the 1991 edition. 168

[Lux07] Ulrike Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007. 13, 222

[Lyo03] Russell Lyons. Determinantal probability measures. *Publications Mathématiques de l'IHÉS*, 98:167–212, 2003. 54

[Lyo05a] Russell Lyons. Asymptotic enumeration of spanning trees. *Combin. Probab. Comput.*, 14(4):491–522, 2005. 18

[Lyo05b] Russell Lyons. Asymptotic enumeration of spanning trees. *Combin. Probab. Comput.*, 14(4):491–522, 2005. 207, 210, 216

[Mat02] J. Matoušek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002. 168

[McS01] Frank McSherry. Spectral partitioning of random graphs. In *FOCS*, pages 529–537, 2001. 13

[Mic08] Laurent Miclo. On eigenfunctions of Markov processes on trees. *Probability Theory and Related Fields*, 142(3-4):561–594, 2008. 222, 225

[Mic13] Laurent Miclo. On hyperboundedness and spectrum of markov operators. hal-00777146, 2013. 221

[Mih89] M. Mihail. Conductance and convergence of markov chains-a combinatorial treatment of expanders. In *FOCS*, volume 0, pages 526–531, Los Alamitos, CA, USA, October 1989. IEEE. 180

[Mit99] Joseph S. B. Mitchell. Guillotine subdivisions approximate polygonal subdivisions: A simple polynomial-time approximation scheme for geometric tsp, k-mst, and related problems. *SIAM J. Comput.*, 28(4):1298–1309, 1999. 26

[MMP90] Clyde L. Monma, Beth Spellman Munson, and William R. Pulleyblank. Minimum-weight two-connected spanning networks. *Math. Program.*, 46:153–171, 1990. 123

[MOS11] Vahideh H. Manshadi, Shayan Oveis Gharan, and Amin Saberi. Online stochastic matching: Online actions based on offline statistics. In *SODA*, pages 1285–1294, 2011. 6, 7

[MP03] Ben Morris and Yuval Peres. Evolving sets and mixin. In *STOC*, pages 279–286, 2003. 18, 263, 265, 266

[MR95] Rajeev Motwani and Prabhakar Raghavan. *Randomized algorithms*. Cambridge University Press, New York, NY, USA, 1995. 110

[MS59] H. P. Mulholland and C. A. B. Smith. An inequality arising in genetical theory. *American Mathematical Monthly*, 66:673–683, 1959. 201

[MS90] David W. Matula and Farhad Shahrokhi. Sparsest cuts and bottlenecks in graphs. *Discrete Applied Mathematics*, 27:113–123, May 1990. 176

[MS11] Tobias Moemke and Ola Svensson. Approximating graphic tsp by matchings. In *FOCS*, pages 560–569, 2011. 25, 119

[MSS13] Adam Marcus, Daniel A Spielman, and Nikhil Srivastava. Interlacing families ii: Mixed characteristic polynomials and the kadison-singer problem. In *FOCS*, 2013. 291

[Muc12] M Mucha. $\frac{13}{9}$-approximation for graphic tsp. In *STACS*, pages 30–41, 2012. 119

[Nem05] Arkadi Nemirovski. Lectures on modern convex optimization, 2005. 66

[NJW02] Andrew Ng, Michael Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *NIPS*, 2002. 11, 187, 222, 228, 276

[NNI94] Hiroshi Nagamochi, Kazuhiro Nishimura, and Toshihide Ibaraki. Computing all small cuts in an undirected network. *SIAM Journal on Discrete Mathematics*, 10:469–481, 1994. 46

[NW61] C. St. J. A. Nash-Williams. Edge disjoint spanning trees of finite graphs. *J. London Math. Soc.*, 36:445–45, 1961. 47

[OS11] Shayan Oveis Gharan and Amin Saberi. The asymmetric traveling salesman problem on graphs with bounded genus. In *SODA*, pages 967–975, 2011. 7, 112

[OSS11] Shayan Oveis Gharan, Amin Saberi, and Mohit Singh. A randomized rounding approach to the traveling salesman problem. In *FOCS*, pages 550–559, 2011. 6, 119

[OSV12] Lorenzo Orecchia, Sushant Sachdeva, and Nisheeth K. Vishnoi. Approximating the exponential, the lanczos method and an $\tilde{O}(m)$-time spectral algorithm for balanced separator. In *STOC*, 2012. 261, 264

[OT12] Shayan Oveis Gharan and Luca Trevisan. Approximating the expansion profile and almost optimal local graph clustering. In *FOCS*, 2012. 17, 177, 224, 261

[OT13] Shayan Oveis Gharan and Luca Trevisan. Partitioning into expanders. In *SODA*, 2013. 18, 274

[OW12] Ryan O'Donnell and David Witmer. Improved small-set expansion from higher eigenvalues. *CoRR*, abs/1204.4688, 2012. 198, 224

[Pap77] Christos H. Papadimitriou. The euclidean traveling salesman problem is np-complete. *Theor. Comput. Sci.*, 4(3):237–244, 1977. 26

[Pit97] Jim Pitman. Probabilistic bounds on the coefficients of polynomials with only real zeros. *J. Comb. Theory Ser. A*, 77:279–303, February 1997. 59

[PS97] Alessandro Panconesi and Aravind Srinivasan. Randomized distributed edge coloring via an extension of the Chernoff-Hoeffding bounds. *SIAM J. Comput.*, 26:350–368, 1997. 53

[PV06] Christos H. Papadimitriou and Santosh Vempala. On the approximability of the traveling salesman problem. *Combinatorica*, 26(1):101–120, 2006. 25, 26

[RÖ8] Harald Räcke. Optimal hierarchical decompositions for congestion minimization in networks. In *STOC*, pages 255–264, 2008. 246

[Rei05] Omer Reingold. Undirected st-connectivity in log-space. In *STOC*, pages 376–385, 2005. 13

[RS74] Fritz Runge and Horst Sachs. Berechnung der Anzahl der Gerüste von Graphen und Hypergraphen mittels deren Spektren. In *Balkan Mathematical Congress*, volume 4, pages 529–536. Beogradski Izdavačko-Grafički Zavod, Belgrade, 1974. 217

[RS10] Prasad Raghavendra and David Steurer. Graph expansion and the unique games conjecture. In *STOC*, pages 755–764. ACM, 2010. 177, 178

[RST10] Prasad Raghavendra, David Steurer, and Prasad Tetali. Approximations for the isoperimetric and spectral profile of graphs and related parameters. In *STOC*, pages 631–640. ACM, 2010. 178, 264

[RT87] Prabhakar Raghavan and Clark D. Tompson. Randomized rounding: a technique for provably good algorithms and algorithmic proofs. *Combinatorica*, 7(4):365–374, December 1987. 3, 63

[RVW00] Omer Reingold, Salil P. Vadhan, and Avi Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders and extractors. In *FOCS*, pages 3–13, 2000. 13

[SC10] Huawei Shen and Xueqi Cheng. Spectral methods for the detection of network community structure: a comparative analysis. CoRR abs/1010.4098, 2010. 11

[Sch03] Alexander Schrijver. *Combinatorial Optimization*, volume 2 of *Algorithms and Combinatorics*. Springer, 2003. 104, 105, 106

[Sey95] P. D. Seymour. Nowhere zero flows. In R. L. Graham, M. Grötschel, and L. Lovász, editors, *Handbook of combinatorics*, volume 1. MIT Press, Cambridge, MA, USA, 1995. 118

[SG76] Sartaj Sahni and Teofilo Gonzalez. P-complete approximation problems. *J. ACM*, 23(3):555–565, July 1976. 25

[She09] Jonah Sherman. Breaking the multicommodity flow barrier for $O(\sqrt{\log n})$-approximations to sparsest cut. In *FOCS*, pages 363–372, 2009. 264

[Sim91] H. D. Simon. Partitioning of unstructured problems for parallel processing. *Computing Systems in Engineering*, 2:135–148, 1991. 14

[SJ89] Alistair Sinclair and Mark Jerrum. Approximate counting, uniform generation and rapidly mixing markov chains. *Inf. Comput.*, 82(1):93–133, July 1989. 13, 14, 179, 180

[SM00] Jianbo Shi and Jitendra Malik. Normalized Cuts and Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000. 11, 14, 17, 173, 179, 187, 228, 244, 274

[Spi96] Daniel A. Spielman. Linear-time encodable and decodable error-correcting codes. *IEEE Transactions on Information Theory*, 42:388–397, 1996. 13

[SS11] Daniel A. Spielman and Nikhil Srivastava. Graph sparsification by effective resistances. *SIAM J. Comput.*, 40(6):1913–1926, 2011. 50

[ST96] Daniel A. Spielman and Shang-Hua Teng. Spectral partitioning works: Planar graphs and finite element meshes. In *FOCS*, pages 96–105, 1996. 13, 14, 179, 247

[ST04] Daniel A. Spielman and Shang-Hua Teng. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In *STOC*, pages 81–90, 2004. 170, 171, 261, 262, 264

[ST13] Daniel A. Spielman and Shang-Hua Teng. A local clustering algorithm for massive graphs and its application to nearly linear time graph partitioning. *SIAM J. Comput.*, 42(1):1–26, 2013. 198

[SV] Sushant Sachdeva and Nisheeth Vishnoi. Personal Communication. 172

[SV12] András Sebö and Jens Vygen. Shorter tours by nicer ears:. CoRR abs/1201.1870, 2012. 119

[SV13] Mohit Singh and Nisheeth K. Vishnoi. Entropy, optimization and counting. CoRR abs/1304.8108, 2013. 69, 70, 75

[SW90] D. B. Shmoys and D. P. Williamson. Analyzing the held-karp tsp bound: a monotonicity property with application. *Inf. Process. Lett.*, 35(6):281–285, September 1990. 4

[SWvZ12] Frans Schalekamp, David P. Williamson, and Anke van Zuylen. A proof of the boyd-carr conjecture. In *SODA*, pages 1477–1486, 2012. 4, 31

[Tan12] Mamoru Tanaka. Higher eigenvalues and partitions of a graph. Preprint, `http://arxiv.org/abs/1112.3434`, 2012. 221, 274, 275, 276, 292

[TM06] David A. Tolliver and Gary L. Miller. Graph partitioning by spectral rounding: Applications in image segmentation and clustering. In *CVPR*, pages 1053–1060. IEEE Computer Society, 2006. 11, 173, 244

[Tre09] Luca Trevisan. Max cut and the smallest eigenvalue. In *STOC*, pages 263–272, 2009. 3, 13

[Tre13] Luca Trevisan. The cheeger inequality in manifolds. http://lucatrevisan.wordpress.com/2013/03/20/the-cheeger-inequality-in-manifolds/, 2013. 179

[TY12] Suguru Tamaki and Yuichi Yoshida. Approximation guarantees for the minimum linear arrangement problem by higher eigenvalues. In *APPROX-RANDOM*, pages 313–324, 2012. 221

[Vis13] N.K. Vishnoi. *Lx = B - Laplacian Solvers and Their Algorithmic Applications*. Now Publishers, 2013. 170, 171

[VM03] Deepak Verma and Marina Meila. Comparison of spectral clustering methods. Technical Report UW-CSE-03-05-01, Department of Computer Science, University of Washington, March 2003. 222, 228

[Wan93] Y. H. Wang. On the number of successes in independent trials. *Statistica Sinica*, 3:295–312, 1993. 59

[Whi32] Hassler Whitney. Non-separable and planar graphs. *Trans. Amer. Math. Soc*, 34(2):339–362, 1932. 113, 116

[Wil91] David Williams. *Probability with Martingales (Cambridge Mathematical Textbooks)*. Cambridge University Press, February 1991. 270

[Wil96] David Bruce Wilson. Generating random spanning trees more quickly than the cover time. In *Proceedings of the Twenty-eighth Annual ACM Symposium on the Theory of Computing*, pages 296–303. ACM, 1996. 49

[Wol80] Laurence A. Wolsey. Heuristic analysis, linear programming and branch and bound. In *Combinatorial Optimization II*, volume 13 of *Mathematical Programming Studies*, pages 121–134. Springer Berlin Heidelberg, 1980. 4, 30, 34, 123

[WS05] S. White and P. Smyth. A spectral clustering approach to finding communities in graphs. *SIAM International Conference on Data Mining*, 2005. 11

[YGS02] Stella X. Yu, Ralph Gross, and Jianbo Shi. Concurrent object recognition and segmentation by graph partitioning. In *NIPS*, pages 1383–1390. MIT Press, 2002. 11

[YS03] Stella X. Yu and Jianbo Shi. Multiclass spectral clustering. In *ICCV*, pages 313–319, 2003. 11

[Zha02] Cun Quan Zhang. Circular flows of nearly eulerian graphs and vertex-splitting. *J. Graph Theory*, 40(3):147–161, 2002. 118

[ZLM13] Zeyuan A. Zhu, Silvio Lattanzi, and Vahab Mirrokni. A local algorithm for finding well-connected clusters. In *ICML*, 2013. 276