APPROXIMATION ALGORITHMS FOR TRAVELING SALESMAN PROBLEMS BASED ON LINEAR PROGRAMMING RELAXATIONS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Hyung Chan An

August 2012

APPROXIMATION ALGORITHMS FOR TRAVELING SALESMAN PROBLEMS BASED ON LINEAR PROGRAMMING RELAXATIONS

Hyung Chan An, Ph.D.
Cornell University 2012

The traveling salesman problem (TSP) is the problem of finding a shortest Hamiltonian circuit or path in a given weighted graph. This problem has been studied in numerous variants, and linear programming has played an important role in the design of approximation algorithms for these problems. In this thesis, we study two versions of the traveling salesman problem and present approximation algorithms for them based on the Held-Karp relaxation.

We first investigate the *s*-*t* path TSP. Hoogeveen showed that the natural variant of Christofides' algorithm is a 5/3-approximation algorithm for this problem; this asymptotically tight bound had remained the best approximation ratio known until now. We surpass this 20-year-old barrier by presenting a deterministic $\frac{1+\sqrt{5}}{2}$-approximation algorithm for the *s*-*t* path TSP for an *arbitrary* metric. The techniques devised in this context can also be applied to other optimization problems including the prize-collecting *s*-*t* path problem and the unit-weight graphical metric *s*-*t* path TSP. The integrality gaps of the LP relaxations for all three problems are studied.

Then we consider the bottleneck asymmetric TSP, where the objective is minimizing the *bottleneck* (or maximum-length) edge cost rather than the total edge cost. We present the first nontrivial approximation algorithm for this problem by giving a novel algorithmic technique to shortcut Eulerian circuits while bounding the lengths of the shortcuts needed. Building on this framework, and the result of Asadpour, Goemans, Mądry, Oveis Gharan, and Saberi, we achieve an $O(\log n/\log\log n)$-approximation algorithm. We also explore the possibility of improvement upon this result through a comparison to the symmetric counterpart of the problem.

BIOGRAPHICAL SKETCH

Hyung-Chan An received his BS in Computer Science and Engineering from Seoul National University in 2006. He started his PhD studies in Computer Science at Cornell University in 2006, and received an MS along the way in 2010. After receiving his PhD, he will be a postdoctoral researcher at École Polytechnique Fédérale de Lausanne, under the supervision of Aleksander Mądry.

아버지, 어머니께
홍찬이에게

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisors David Shmoys and Bobby Kleinberg. I am twice as fortunate as most other students are to have two great advisors at the same time. I am deeply indebted to them for their continuous support, trust, and encouragement, as well as for their wisdom that guided me through my PhD studies. Last few years have been a stage of my life that I had so many things to learn for the first time, both academically and non-academically; and this was made enjoyable by Bobby and David. Thank you.

I would also like to thank David Williamson; in my first semester at Cornell, I was introduced to the field of approximation algorithms through his lectures. I am grateful for his help and advice during my graduate studies and in particular while I was preparing to finish them. I would like to express my appreciation to Mike Todd and Dan Huttenlocher for being on my committee. I thank Mike Todd also for his extensive comments on this thesis.

I was inspired by many people to pursue graduate studies in computer science. I will not be able to thank everyone to each of which I am greatly indebted, but I would like to thank Professors Ha-Jine Kimn and Kunsoo Park for introducing me to theoretical computer science during my early years in school. I thank Professors Myung-Soo Kim and Soonhoi Ha for encouraging me in coming to the United States for the graduate studies.

For more than four years, I have shared the office (and lots of fun) with Lonnie Princehouse, Dustin Tseng, Rachel Lin, and Michael Siegenthaler. It was always a great pleasure to talk to Sungryel Choi while he was still in Ithaca. I am glad that I had a chance to visit Illinois to see Choonghwan Lee before I graduated. Jun Ki Lee, Hyemin Chung, and Sunkyu Han made my stay in Boston delightful. I would like to thank them and other friends back in Korea.

Last but not least, I am thankful for my parents' unconditional love. Thank you. Hongchan, thank you, and I look forward to seeing your thesis.

My graduate studies have been supported in part by NSF under grants no. CCF-1017688, CCF-0729102, CCR-0635121, DMS-0732196, and CCF-0832782, a Graduate School Fellowship, and the Korea Foundation for Advanced Studies. I would like to express my appreciation to them. Part of this research was conducted while I was a visiting student at CSAIL, MIT.

Preliminary versions of some of the results in this thesis have appeared in two papers presented each at APPROX '10 and STOC '12, coauthored with Robert Kleinberg and David Shmoys.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The traveling salesman problem (TSP) is the problem of finding a shortest Hamiltonian circuit in a given weighted graph. One of the most celebrated problems in combinatorial optimization, the problem has inspired the evolution of numerous algorithmic methodologies for solving optimization problems [41]. Dantzig, Fulkerson, and Johnson [18], through finding an optimal solution to an instance consisting of 49 cities in the United States and proving its optimality, presented the concept of the cutting-plane method. Branch-and-bound using a Lagrangian relaxation as a lower bound, proposed by Held and Karp [31], was employed to solve the problem.

As is widely known, the traveling salesman problem is NP-hard [36], and hence its theoretical investigation has mainly been focused on approximation algorithms. The worst-case analysis of approximation algorithms aims to guarantee the worst-case ratio of the cost of the algorithm's output to that of the optimal solution; we call a polynomial-time algorithm whose output cost is no more than ρ times the optimum a *ρ-approximation algorithm*. This worst-case analysis is performed using various techniques including combinatorial, polyhedral, or stochastic methods; research on the traveling salesman problem has contributed to the development of these techniques as well [15, 54, 46, 43, 44, 52].

The traveling salesman problem does not admit an approximation algorithm with a polynomial approximation ratio unless P=NP. However, for the metric TSP, where the cost function is given as a (pseudo)metric on the vertices, the folklore algorithm gives 2-approximation. The metric TSP is a particularly interesting case since many cost functions of practical interest satisfy the triangle inequality, and even when the cost function does not satisfy the triangle inequality, if it is allowed to visit a vertex multiple times and weights are nonnegative, the problem reduces to the metric TSP. In 1976, Christofides [15] gave a 3/2-approximation algorithm for the metric TSP, and various special cases and variants of the metric TSP have been studied since. Yet, despite the historical and practical importance of the problem, no algorithms with better approximation ratio have been discovered for this problem. On the other hand, the complexity-theoretic inapproximability bound currently known, assuming P≠NP, is only 185/184, as was recently improved by Lampis [39] over 220/219 due to Papadimitriou and Vempala [48].

For the asymmetric TSP, where the distance from vertex a to b can be different from b to a, there is an even larger gap between the known upper and lower bounds on its approximability. The complexity-theoretic lower bound due to Papadimitrou and Vempala [48] is 117/116 in this case; however, since Frieze, Galbiati, and Maffoli [22] gave the first $O(\log n)$-approximation algorithm, it was only recently that this was improved to $O(\log n / \log \log n)$, the current best known, by Asadpour, Goemans, Mądry, Oveis Gharan, and Saberi [8].

In solving combinatorial optimization problems, linear programming (LP) is not only important for the approaches through integer programming such as the cutting-plane method and branch-and-bound, but it also has proven to be a very useful tool in the theoretical analysis of approximation algorithms. LPs can be solved in polynomial time to yield a good bound on the optimal solution value, and the LP solution itself often roughly reflects the structure of the true optimal solution; the LP rounding approach finds an answer to a hard optimization problem by first solving an LP relaxation of the problem and then "rounding" the fractional solution to an integral feasible solution. Some algorithms work without actually solving an LP: primal-dual algorithms directly produce an integral solution with an accompanying near-optimality certificate in the form of a feasible LP dual solution.

The subtour elimination LP relaxation, or the Held-Karp relaxation, is a standard LP relaxation to the (variants of) TSP [18, 31], and has been successfully used by many algorithms [12, 25, 8, 2, 46, 43, 44]. In the LP-based design of an approximation algorithm, one important measure of the strength of a particular relaxation is its integrality gap, i.e., the worst-case ratio between the integral and fractional optimal values; however, there exists a significant gap between currently known lower and upper bounds on the integrality gap of the Held-Karp relaxation: the best upper bound known of 3/2 is constructively proven by the analysis of Christofides' algorithm due to Wolsey [54]; yet, the best lower bound known is 4/3 (see Figure 1.1(a) due to Goemans [24]).

In this thesis, we investigate two important versions of the traveling salesman problem that are closely related to improving our understanding of the metric TSP and Christofides' algorithm. In particular, we present approximation algorithms for the s-t path TSP, the bottleneck asymmetric TSP, and other related problems; these algorithms are based on various versions of the Held-Karp relaxation and provide provable bounds on their integrality gaps.

***s*-*t* path TSP.** After 35 years, Christofides' 3/2-approximation algorithm [15] still provides the best performance guarantee known for the metric traveling salesman problem (TSP), and improving upon this bound is a fundamental open question in combinatorial optimization. For the path variant of the metric TSP in which the aim is to find a shortest Hamiltonian path between given endpoints

s and t, Hoogeveen [33] showed that the natural variant of Christofides' algorithm yields an approximation ratio of 5/3 that is asymptotically tight, and this has been the best approximation algorithm known for this s-t path variant for the past 20 years. Recently, there has been progress for the special case of metrics derived as shortest paths in unit-weight (undirected) graphs: Oveis Gharan, Saberi, and Singh [46] gave a $(3/2 - \epsilon_0)$-approximation algorithm for the TSP, where ϵ_0 is an absolute positive constant that is very close to zero; we show in Subsection 2.3.3 of this thesis that their method can be extended to yield an analogous result of a $(5/3 - \epsilon_1)$-approximation algorithm for the s-t path TSP in the same special case, for a very small constant ϵ_1. Mömke and Svensson [43] gave a 1.4605-approximation algorithm for this special case of the TSP, as well as a 1.5858-approximation algorithm for the s-t path TSP in the same case (where the results of Mömke & Svensson [43] and Subsection 2.3.3 were obtained independently). Subsequent to our results, Sebő and Vygen [52] recently announced improved ratios of 7/5 and 3/2 for each variant. We note the techniques devised in these results for the unit-weight graphical metric case proved useful in both path and ordinary (circuit) variants. The main result of Chapter 2 is to provide the first improvement for the *general* metric case of the s-t path TSP: more specifically, we give a deterministic $\left(\frac{1+\sqrt{5}}{2}\right)$-approximation algorithm for the metric s-t path TSP for an arbitrary metric, breaking the 5/3 barrier. It remains an open question whether these techniques can be extended to yield a comparable improvement (over the 3/2 barrier) for the general-metric ordinary (circuit) TSP.

Our analysis gives the first constant upper bound on the integrality gap of the Held-Karp relaxation analogously defined for the path problem as well. Even though Hoogeveen [33] shows the natural variant of Christofides' algorithm is a 5/3-approximation algorithm, the analysis compares the output solu-

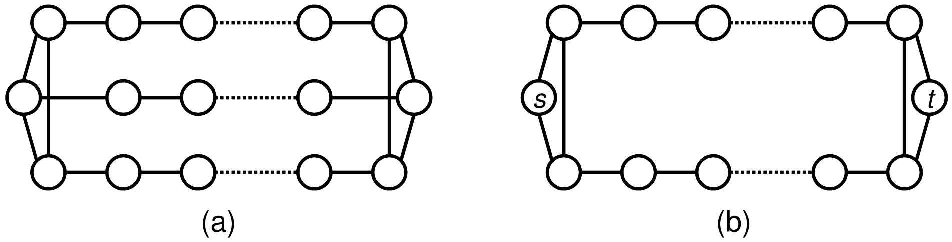

Figure 1.1: Examples establishing the integrality gap lower bounds for the circuit- and path-variant Held-Karp relaxations.

tion value to the optimal (integral) solution; therefore it is unclear whether the algorithm yields an integrality gap bound for the Held-Karp relaxation formulated for the path problem. The analysis of the present algorithm, in contrast, reveals an upper bound of $\frac{1+\sqrt{5}}{2}$ on its integrality gap, matching the approximation ratio. (Subsequent to Hoogeveen, several papers [3, 29, 9] present alternate algorithms and analyses of tight 5/3-approximation algorithms; in particular, with hindsight, it would not be hard to yield a weaker 5/3 integrality gap upper bound from some of these ideas.) We also show an alternative LP-based analysis of Christofides' algorithm proves an upper bound of 5/3. We observe that the family of graphs in Figure 1.1(b) establishes the integrality gap lower bound of 3/2 under the unit-weight graphical metric. Note that this lower bound is strictly greater than the known upper bound of $(3/2 - \epsilon_0)$ on the integrality gap of the circuit-variant Held-Karp relaxation under the unit-weight graphical metric [46]; this suggests that the lack of a performance guarantee known for the *s*-*t* path TSP matching the 3/2 for other TSP variants has a true structural cause. We will also demonstrate how our techniques can be applied to other problems, such as the prize-collecting *s*-*t* path problem and the unit-weight graphical metric *s*-*t* path TSP, to obtain better approximation ratios and better LP integrality gap upper bounds than the current best known.

Bottleneck asymmetric TSP. In Chapter 3, we study the bottleneck asymmetric TSP; that is, in contrast to the variant of traveling salesman problem most commonly studied, the objective is to minimize the maximum edge cost in the tour, rather than the sum of the edge costs. Furthermore, while the edge costs still satisfy the triangle inequality, we do not require that they be symmetric. We present the first nontrivial approximation algorithm for the bottleneck asymmetric traveling salesman problem, by giving an $O(\log n / \log\log n)$-approximation algorithm. At the heart of our result is a new algorithmic technique for converting Eulerian circuits into tours while introducing "shortcuts" that are of bounded length.

For any optimization problem defined in terms of pairwise distances between nodes, it is natural to consider both the symmetric case and the asymmetric one, as well as the min-sum variant and the bottleneck one. For the bottleneck symmetric TSP, Lau [40], and Parker & Rardin [49], building on structural results of Fleischner [21], give a 2-approximation algorithm, and based on the metric in which all costs are either 1 or 2, it is easy to show that, for any $\rho < 2$, the existence of a ρ-approximation algorithm implies that P=NP. This cross-section of results is mirrored in other optimization settings. For example, for the min-sum symmetric k-median problem in which k points are chosen as "medians" and each point is assigned to its nearest median, Arya, Garg, Khandekar, Meyerson, Munagala, and Pandit [7] give a ρ-approximation algorithm for each $\rho > 3$, whereas Jain, Mahdian, Markakis, Saberi and Vazirani prove hardness results for $\rho < 1+2/e$ [35]. In contrast, for the bottleneck symmetric version, the k-center problem, Hochbaum and Shmoys [32] gave a 2-approximation algorithm, whereas Hsu and Nemhauser [34] showed the NP-hardness of a performance guarantee of $\rho < 2$. For the asymmetric k-center, a matching upper and lower bound of $\Theta(\log^* n)$ for the best performance guarantee was shown by Panigrahy & Vishwanathan [47] and Chuzhoy, Guha, Halperin, Khanna, Kortsarz, Krauthgamer & Naor [17], respectively. In contrast, for the asymmetric k-median problem, a bicriterion result which allowed a constant factor increase in cost with a logarithmic increase in the number of medians was shown by Lin and Vitter [42], and a hardness tradeoff matching this (up to constant factors) was proved by Archer [5].

In considering these comparative results, there is a mixed message as to whether a bottleneck problem is easier or harder to approximate than its min-sum counterpart. We examine the challenges that are unique to the bottleneck problems, and show that the bottleneck asymmetric TSP reduces to a problem with a purely combinatorial statement yet retaining a close relation to the directed Held-Karp relaxation; this suggests that the problem could be an easier-to-approach stepping stone to the ordinary (min-sum) asymmetric TSP. Our result can also be combined with the result of Oveis Gharan and Saberi [45] to yield an $O(1)$-approximation algorithm for the bottleneck asymmetric TSP when the support of the Held-Karp solution has a bounded orientable genus.

CHAPTER 2

IMPROVING CHRISTOFIDES' ALGORITHM FOR THE *s*-*t* PATH TSP

We present a $\frac{1+\sqrt{5}}{2}$-approximation algorithm for the s-t path TSP in this chapter.

The present algorithm is based on the Held-Karp relaxation defined for the path problem. A feasible solution to the path-variant Held-Karp relaxation is in the spanning tree polytope; thus, given a feasible Held-Karp solution, there exists a probability distribution over spanning trees whose marginal edge probabilities are given by the Held-Karp solution. Our algorithm first computes an optimal solution to the Held-Karp relaxation, and samples a spanning tree from a probability distribution whose marginal is given by the Held-Karp solution. Then it augments this tree with a minimum T-join, where T is the set of vertices with "wrong" parity of degree, to obtain an Eulerian path visiting every vertex; this Eulerian path can be shortcut into an s-t Hamiltonian path of no greater cost. Our analysis of this algorithm shows that the expected cost of the Eulerian path is at most $\frac{1+\sqrt{5}}{2}$ times the Held-Karp optimum; the analysis relies only on the marginal probabilities, and therefore holds for *any* arbitrary distribution with the given marginals. Our algorithm is similar in its basic outline to the algorithm of Oveis Gharan et al. [46], although that result both relies on a specific means for probabilistically generating spanning trees and adds complications in the algorithm design. We note that the flexibility of our probabilistic choice enables a simple derandomization: a feasible Held-Karp solution can be efficiently decomposed into a convex combination of polynomially many spanning trees (see Grötschel, Lovász, and Schrijver [28]) and trying every spanning tree in this convex combination yields a simple deterministic algorithm. We also note that

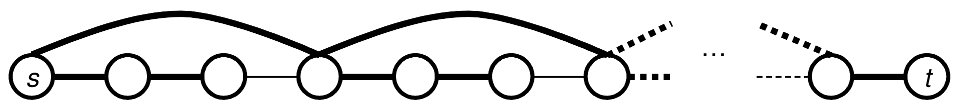

Figure 2.1: Example showing 5/3 is asymptotically tight (Hoogeveen [33]): a minimum spanning tree is marked with thick edges.

our algorithm differs from Christofides' in only one crucial respect: rather than taking a single tree and augmenting it with a T-join, we try out polynomially many trees and then take the one whose augmentation yields the lowest-cost path. The example in Figure 2.1 due to Hoogeveen [33] shows that this simple modification of the original algorithm is crucial to achieving the improved approximation ratio: if one only tries augmenting the minimum spanning tree, the approximation ratio remains no better than 5/3.

As the expected cost of the sampled spanning tree is equal to the Held-Karp optimum, the rest of the analysis focuses on bounding the cost of the minimum T-join by providing a low-cost *fractional T-join dominator* that serves as an upper bound on the cost of the minimum T-join. First we show that the Held-Karp solution and the spanning tree, while being costly fractional T-join dominators themselves, are complementary: a certain linear combination of them is a fractional T-join dominator whose expected cost is no greater than 2/3 times the Held-Karp optimum, thereby recovering the same 5/3 performance guarantee provided by Hoogeveen's analysis of Christofides' algorithm. Based on this beginning analysis, we present progressively better ways of constructing a low-cost fractional T-join dominator. In all of these approaches, we perturb the coefficients of the tree and the Held-Karp solution to reduce the cost of their linear combination at the expense of potentially violating some constraints of the fractional T-join dominator linear program, and then we add a low-cost correction to repair the violated constraints. To construct this correction vector and to bound its cost, we show that the only potentially violated constraints correspond to *narrow cuts* having a layered structure, as illustrated in Figure 2.2. The layered structure allows us to choose disjoint sets of representative edges for each cut and to correct the violated constraints using a sum of vectors each supported on the representative edge set of the corresponding narrow cut. We show that this idea leads to a slight improvement upon 5/3, using the fact that the representative edge sets, while being mutually disjoint, occupy a large portion of each cut and that each narrow cut constraint has only a small probability of being violated. After that, we present a tighter analysis with a similar construction. Finally, pushing the performance guarantee towards the golden ratio requires relaxing the disjointness of the representatives to a notion of "fractional disjointness". We define this relaxed disjointness, construct the requisite fractionally disjoint vectors via the analysis of an auxiliary flow network, and prove the performance guarantee of $\frac{1+\sqrt{5}}{2}$. We note that neither the fractional T-join dominator nor the narrow cuts are actually computed by the algorithm; these progressive analyses all analyze the same single algorithm while different fractional T-join dominators are considered in each analysis. That is, it might be possible to obtain a better performance guarantee for the same algorithm by providing a better construction of a fractional T-join dominator. The narrow cuts are purely for the purpose of analysis in Section 2.2 and never determined by the algorithm; however, their algorithmic use is explored in Section 2.3.

Section 2.3 demonstrates how the present results can be applied to other problems to obtain better approximation algorithms than the current best known. We first consider the metric prize-collecting s-t path problem. In a prize-collecting problem, we are given "prize" values defined on vertices, and the objective function becomes the sum of the "regular" solution cost and the total "missed" prize of the vertices that are not included in the solution. For example, the prize-collecting s-t path problem finds a (not necessarily spanning) s-t path that minimizes the sum of the path cost and the total prize of the vertices not on the path. Chaudhuri, Godfrey, Rao, and Talwar [13] give a primal-dual 2-approximation algorithm for this problem. Prize-collecting TSP, the circuit version of this problem, has been introduced in Balas [10]; Bienstock, Goemans, Simchi-Levi, and Williamson [12] give an LP-rounding 2.5-approximation algorithm, and Goemans & Williamson [27] show a primal-dual 2-approximation algorithm. For both problems, Archer, Bateni, Hajiaghayi, and Karloff [6] give improvement on approximation ratios: using the path-variant Christofides' algorithm as a black box, Archer et al. give a 241/121-approximation algorithm for the prize-collecting s-t path problem; a 97/49-approximation algorithm is given for the prize-collecting TSP, using Christofides' algorithm as a black box again. For the prize-collecting (circuit) TSP, Goemans [25] combines Bienstock et al. [12] and Goemans & Williamson [27] to obtain a 1.9146-approximation algorithm, the current best known.

As the analysis of Archer et al. [6] treats Christofides' algorithm as a black box, replacing this with the present algorithm readily gives an improvement over the best approximation ratio known. Furthermore, we will show that, since the present analysis produces the performance guarantee in terms of the Held-Karp optimum, it enables an LP-rounding approach analogous to Bienstock et al. [12] utilizing the parsimonious property due to Goemans & Bertsimas [26], Goemans [23], and Bertsimas & Teo [11]. This further leads to an extension of Goemans' analysis [25], yielding a 1.9535-approximation algorithm for the prize-collecting s-t path problem; the same upper bound is established on the integrality gap of the LP relaxation used.

Secondly, we study the *unit-weight* graphical metric *s*-*t* path TSP to present a 1.5780-approximation algorithm. As discussed above, there has been progress for this special case in both the ordinary (circuit) TSP and the *s*-*t* path TSP. Recently, Mucha [44] gave an improved analysis of Mömke & Svensson's algorithm [43] to prove the performance guarantee of 13/9 for the circuit case and $19/12 + \epsilon$ for the path case, for any $\epsilon > 0$. We observe that the critical case of this analysis is when the Held-Karp optimum is small, and we show how to obtain an algorithm that yields a better performance guarantee on this critical case, based on the main results of this chapter. In particular, we devise an algorithm that works on narrow cuts, to be run in parallel with the present algorithm; this illustrates that the narrow cuts are a useful algorithmic tool as well, not only an analytic tool. Our algorithm establishes an upper bound on the integrality gap of the path-variant Held-Karp relaxation under the unit-weight graphical metric, which does not match the performance guarantee but is smaller than $\frac{1+\sqrt{5}}{2}$. We also present an alternative analysis of the path-variant Christofides' algorithm that yields a slight improvement over Christofides' for the unit-weight graphical metric case. Subsequent to these results, Sebő and Vygen [52] recently announced a 3/2-approximation algorithm for the unit-weight graphical metric *s*-*t* path TSP.

2.1 Preliminaries

In this section, we introduce some definitions and notation to be used throughout this chapter.

Let $G = (V, E)$ be the input complete graph with metric cost function $c : E \to \mathbb{R}_+$. *Endpoints* $s, t \in V$ are given as a part of the input; we call the other vertices *internal points*.

For $A, B \subset V$ such that $A \cap B = \emptyset$, $E(A, B)$ denotes the set of edges between A and B: i.e., $E(A, B) = \{\{u, v\} \in E \mid u \in A, v \in B\}$. Let $E(A)$ denote the set of edges within A: $E(A) := \{\{u, v\} \in E \mid u, v \in A\}$.

For nonempty $U \subsetneq V$, let $(U, \bar{U})$ denote the cut defined by U, and $\delta(U)$ be the edge set in the cut: $\delta(U) = E(U, \bar{U})$. $(U, \bar{U})$ is called an *s-t cut* if $|U \cap \{s, t\}| = 1$; we call $(U, \bar{U})$ *nonseparating* otherwise.

For $x, c \in \mathbb{R}^E$ and $F \subset E$, $x(F)$ is a shorthand for $\sum_{f \in F} x_f$; $c(x)$ is $\sum_{e \in E} c_e x_e$. The incidence vector $\chi_F \in \mathbb{R}^E$ of $F \subset E$ is a $(0, 1)$-vector defined as follows:

$$(\chi_F)_e := \begin{cases} 1 & \text{if } e \in F, \\ 0 & \text{otherwise.} \end{cases}$$

For two vectors $a, b \in \mathbb{R}^I$, let $a * b \in \mathbb{R}^I$ denote the vector defined by $(a * b)_i := a_i b_i$.

Definition 1 ([31])**.** *The* path-variant Held-Karp relaxation *is defined as follows:*

$$\begin{aligned} \text{minimize} \quad & c(x) & & \\ \text{subject to} \quad & x(\delta(S)) \geq 1, & & \forall S \subsetneq V, |\{s, t\} \cap S| = 1; \\ & x(\delta(S)) \geq 2, & & \forall S \subsetneq V, |\{s, t\} \cap S| \neq 1, S \neq \emptyset; \\ & x(\delta(\{s\})) = x(\delta(\{t\})) = 1; & & \\ & x(\delta(\{v\})) = 2, & & \forall v \in V \setminus \{s, t\}; \\ & x \geq 0. & & \end{aligned} \tag{2.1}$$

This linear program can be solved in polynomial time via the ellipsoid method using a min-cut algorithm to solve the separation problem [28]. The following observation gives an equivalent formulation of (2.1).

Observation 1. *Following is an equivalent formulation of* (2.1)*:*

$$\begin{aligned}
\text{minimize}\quad & c(x) \\
\text{subject to}\quad & x(E(S)) \le |S| - 1, && \forall S \subsetneq V, \{s,t\} \not\subseteq S, S \neq \emptyset; \\
& x(E(S)) \le |S| - 2, && \forall S \subsetneq V, \{s,t\} \subseteq S; \\
& x(\delta(\{s\})) = x(\delta(\{t\})) = 1; \\
& x(\delta(\{v\})) = 2, && \forall v \in V \setminus \{s,t\}; \\
& x \ge 0.
\end{aligned}$$

Definition 2. *For $T \subset V$ and $J \subset E$, J is a T-*join *if the set of odd-degree vertices in $G' = (V, J)$ is T.*

Edmonds and Johnson [20] give a polyhedral characterization of T-joins: let $P_T(G)$ be the convex hull of the incidence vectors of the T-joins on $G = (V, E)$; $P_T(G) + \mathbb{R}_+^E$ is exactly characterized by

$$\begin{cases} y(\delta(S)) \ge 1, & \forall S \subset V, |S \cap T| \text{ odd}; \\ y \in \mathbb{R}_+^E. \end{cases} \tag{2.2}$$

We call a feasible solution to (2.2) a *fractional T-join dominator*.

Lastly, the polytope defined by the path-variant Held-Karp relaxation is contained in the spanning tree polytope of the same graph, as can be seen from Edmonds' characterization of spanning tree polytopes [19]; thus, given a feasible solution x^* to the path-variant Held-Karp relaxation, there exist spanning trees $\mathcal{T}_1, \ldots, \mathcal{T}_k$ and $\lambda_1, \ldots, \lambda_k \in \mathbb{R}_+$ such that $x^* = \sum_{i=1}^k \lambda_i \chi_{\mathcal{T}_i}$ and $\sum_{i=1}^k \lambda_i = 1$, where k is bounded by a polynomial. This follows from Grötschel, Lovász, and Schrijver [28].

2.2 Improving upon 5/3

We present the algorithm for the metric s-t path TSP and its analysis in this section.

Algorithm. Given a complete graph $G = (V, E)$ with cost function $c : E \to \mathbb{R}_+$ and the endpoints $s, t \in V$, the algorithm first computes an optimal solution x^* to the path-variant Held-Karp relaxation. Then it decomposes x^* into a convex combination $\sum \lambda_i \chi_{\mathscr{T}_i}$ of polynomially many spanning trees $\mathscr{T}_1, \ldots, \mathscr{T}_k$ with coefficients $\lambda_1, \ldots, \lambda_k \geq 0$; a spanning tree $\mathscr{T}$ is sampled among these spanning trees $\mathscr{T}_i$'s, choosing $\mathscr{T}_i$ with probability λ_i. This decomposition can be performed in polynomial time, as noted in Section 2.1. Let $T \subset V$ be the set of the vertices with the "wrong" parity of degree in $\mathscr{T}$: i.e., T is the set of odd-degree internal points and even-degree endpoints in $\mathscr{T}$. The algorithm finds a minimum T-join J and an s-t Eulerian path of the multigraph $\mathscr{T} \cup J$. This Eulerian path is shortcut to obtain a Hamiltonian path H between s and t; H is the output of the algorithm.

We note that this algorithm can be derandomized by trying each $\mathscr{T}_i$ instead of sampling $\mathscr{T}$. Observe that $\mathrm{E}[c(H)] \leq \rho c(x^*)$ implies that the derandomized algorithm is a deterministic ρ-approximation algorithm.

In the rest of this section, we prove the following theorem.

Theorem 1. *The present algorithm returns a Hamiltonian path between s and t whose expected cost is no more than $\frac{1+\sqrt{5}}{2} c(x^*)$. Therefore, there exists a deterministic $\left(\frac{1+\sqrt{5}}{2}\right)$-approximation algorithm for the s-t path TSP.*

Corollary 1. *The integrality gap of the path-variant Held-Karp relaxation is at most $\frac{1+\sqrt{5}}{2}$.*

Proof of 5/3-approximation. We first present a simple proof that the present algorithm is an (expected) 5/3-approximation algorithm; improved analyses are presented later, based on this simple proof.

We can understand the well-known 2-approximation algorithm for the circuit TSP and Christofides' 3/2-approximation algorithm as respectively using the minimum spanning tree and (half) the Held-Karp solution [54, 53] as a fractional T-join dominator. Let us consider whether $\chi_{\mathcal{T}}$ and x^* can be used to bound the cost of a minimum T-join in our case.

It can be seen from (2.1) that βx^* is a fractional T-join dominator for $\beta = 1$. If it were not for the s-t cuts, the same could be shown for $\beta = \frac{1}{2}$. However, an s-t cut may have capacity as low as 1, making it hard to establish the feasibility of βx^* for any $\beta < 1$.

$\alpha\chi_{\mathcal{T}}$ also is a fractional T-join dominator for $\alpha = 1$; in this case, however, s-t cuts do have some slack. Suppose that an s-t cut $(U, \bar{U})$ is odd with respect to T: i.e., $|U \cap T|$ is odd. Since U contains exactly one of s and t, U contains an even number of vertices that have odd degree in $\mathcal{T}$. $|\delta(U) \cap \mathcal{T}|$ is given as the sum of the degrees of the vertices in U minus twice the number of edges within U, and is therefore even. This shows $\chi_{\mathcal{T}}(\delta(U)) \geq 2$ and hence $\alpha\chi_{\mathcal{T}}$ for $\alpha = \frac{1}{2}$ does not violate (2.2) as far as s-t cuts are concerned. It is the nonseparating cuts that render it difficult to show the feasibility of $\alpha\chi_{\mathcal{T}}$ for $\alpha < 1$.

Given the difficulties in these two cases are complementary, it is natural to consider $\alpha\chi_{\mathcal{T}} + \beta x^*$ as a candidate for a fractional T-join dominator; Theorem 2 elaborates this observation.

Theorem 2. $\mathrm{E}[c(H)] \leq \frac{5}{3}c(x^*)$.

Proof. Let $y := \alpha\chi_{\mathscr{T}} + \beta x^*$ for some parameters $\alpha, \beta > 0$ to be chosen later. We examine a sufficient condition on α and β for y to be a fractional T-join dominator.

It is obvious that $y \geq 0$. Consider an odd cut $(U, \bar{U})$ with respect to T: i.e., $|U \cap T|$ is odd. We have $|\delta(U) \cap \mathscr{T}| > 0$ from the connectedness of $\mathscr{T}$. Suppose that $(U, \bar{U})$ is an s,t-cut; then $|\delta(U) \cap \mathscr{T}|$ is even as previously argued. Thus, $y(\delta(U)) = \alpha|\delta(U) \cap \mathscr{T}| + \beta x^*(\delta(U)) \geq 2\alpha + \beta$. Now suppose that $(U, \bar{U})$ is nonseparating; then we have $x^*(\delta(U)) \geq 2$ from the Held-Karp feasibility, and hence $y(\delta(U)) \geq \alpha|\delta(U) \cap \mathscr{T}| + \beta x^*(\delta(U)) \geq \alpha + 2\beta$. Therefore, if $2\alpha + \beta \geq 1$ and $\alpha + 2\beta \geq 1$ then y is feasible.

Now we bound the expected cost of H:

$$\begin{aligned} \mathrm{E}[c(H)] &\leq \mathrm{E}[c(\mathscr{T})] + \mathrm{E}[c(J)] \\ &\leq \mathrm{E}[c(\mathscr{T})] + \mathrm{E}[c(y)] \\ &= \mathrm{E}[c(\mathscr{T})] + \mathrm{E}[c(\alpha\chi_{\mathscr{T}})] + \mathrm{E}[c(\beta x^*)] \\ &= (1 + \alpha + \beta)c(x^*), \end{aligned}$$

where the second inequality holds since y is a fractional T-join dominator. Choose $\alpha = \frac{1}{3}$ and $\beta = \frac{1}{3}$. □

First Improvement upon 5/3**.** Now we demonstrate that the above analysis can be slightly improved.

Recall that the lower bound on the nonseparating cut capacities of y was given as $\alpha + 2\beta$ in the previous analysis; consider perturbing α and β by small amount while maintaining $\alpha + 2\beta = 1$. In particular, if we decrease α by 2ϵ and increase β by ϵ, we decrease the expected cost of y by $\epsilon c(x^*)$, without changing $\alpha+2\beta$; that is, if we can fix the possible deficiencies of y in s-t cuts with small cost, this perturbation will lead to an improvement in the performance guarantee.

Note that s-t cuts $(U, \bar{U})$ with large capacities are not a problem: $(\alpha\chi_{\mathscr{T}} + \beta x^*)(\delta(U)) \geq 2\alpha + \beta x^*(\delta(U))$ and thus, if $x^*(\delta(U))$ is large enough, the bound remains greater than one after a small perturbation.

On the other hand, cuts with $x^*(\delta(U)) = 1$ are also not a concern. $x^*(\delta(U)) = \mathrm{E}[|\delta(U) \cap \mathscr{T}|]$, and $|\delta(U) \cap \mathscr{T}| \geq 1$ from the connectedness of $\mathscr{T}$; hence $|\delta(U) \cap \mathscr{T}|$ is identically 1 and $|U \cap T|$ is always even. Formulation (2.2) constrains the capacities of only the cuts that are odd with respect to T, so the capacity of this particular cut $(U, \bar{U})$ will never be constrained. In fact, for an s-t cut $(U, \bar{U})$,

$$\Pr[|U \cap T| \text{ is odd}] \leq \Pr[|\delta(U) \cap \mathscr{T}| \geq 2] \leq \mathrm{E}[|\delta(U) \cap \mathscr{T}|] - 1 = x^*(\delta(U)) - 1. \quad (2.3)$$

We will begin with $y \leftarrow \alpha\chi_{\mathscr{T}} + \beta x^*$ for perturbed α and β, and ensure that y is a fractional T-join dominator by adding small fractions of the deficient odd s-t cuts. Yet, a cut being odd with small probability as shown by (2.3) does not directly connect to its edge being added with small probability, since an edge belongs to many s-t cuts. We address this issue by showing that the s-t cuts of small capacities are "almost" disjoint.

First, consider the s-t cuts $(U, \bar{U})$ whose capacities are not large enough for $2\alpha + \beta x^*(\delta(U))$ to be readily as large as 1; the following definition captures this idea. Let $\tau := \frac{1-2\alpha}{\beta} - 1$.

Definition 3. *For some* $0 < \tau \leq 1$, *an* s-t *cut* $(U, \bar{U})$ *is called* τ-narrow *if* $x^*(\delta(U)) < 1 + \tau$.

The following lemma shows that τ-narrow cuts do not cross.

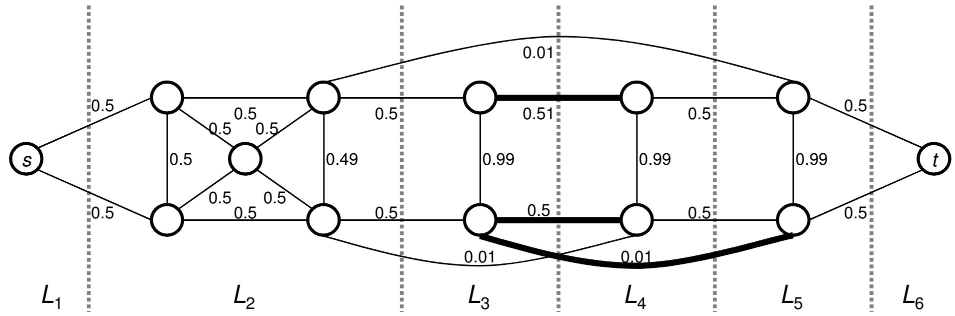

Figure 2.2: 0.05-narrow cuts of a feasible Held-Karp solution ($\ell = 6$).

Lemma 1. *Let $0 < \tau \leq 1$. For $U_1 \ni s$ and $U_2 \ni s$, if both $(U_1, \bar{U}_1)$ and $(U_2, \bar{U}_2)$ are τ-narrow, then $U_1 \subset U_2$ or $U_2 \subset U_1$.*

Proof. Suppose not. Then both $U_1 \setminus U_2$ and $U_2 \setminus U_1$ are nonempty and $x^*(\delta(U_1)) + x^*(\delta(U_2)) \geq x^*(\delta(U_1 \setminus U_2)) + x^*(\delta(U_2 \setminus U_1)) \geq 2 + 2 = 4$;on the other hand, $x^*(\delta(U_1)) + x^*(\delta(U_2)) < 2 + 2\tau \leq 4$, leading to a contradiction. □

Lemma 1 shows that the τ-narrow cuts constitute a layered structure, as illustrated in Figure 2.2:

Corollary 2. *There exists a partition $L_1, \ldots, L_\ell$ of V such that*

1. $L_1 = \{s\}$, $L_\ell = \{t\}$, *and*
2. $\{U | (U, \bar{U}) \text{ is } \tau\text{−narrow}, s \in U\} = \{U_i | 1 \leq i < \ell\}$, *where* $U_i := \cup_{k=1}^{i} L_k$.

Let $L_{\leq i}$ denote $\cup_{k=1}^{i} L_k$ and $L_{\geq i}$ denote $\cup_{k=i}^{\ell} L_k$. $U_i = L_{\leq i}$.

Now we show that τ-narrow cuts are almost disjoint: for each τ-narrow cut $(U_i, \bar{U}_i)$, we can choose $F_i \subset \delta(U_i)$ that occupies a large portion of $\delta(U_i)$ and mutually disjoint.

Definition 4. $F_i := E(L_i, L_{\geq i+1})$.

Lemma 2. *For each τ-narrow cut $(U_i, \bar{U}_i)$, $x^*(F_i) > \frac{1-\tau+x^*(\delta(U_i))}{2} \geq 1 - \frac{\tau}{2}$.*

Proof. The lemma holds trivially for $i = 1$. Suppose $2 \leq i \leq \ell - 1$. We have

$$x^*(E(L_i, L_{\geq i+1})) + x^*(E(L_{\leq i-1}, L_{\geq i+1})) = x^*(\delta(U_i))$$

and

$$x^*(E(L_{\leq i-1}, L_i)) + x^*(E(L_{\leq i-1}, L_{\geq i+1})) = x^*(\delta(U_{i-1})) < 1 + \tau;$$

subtracting the latter from the former yields

$$x^*(E(L_i, L_{\geq i+1})) - x^*(E(L_{\leq i-1}, L_i)) > x^*(\delta(U_i)) - 1 - \tau.$$

On the other hand,

$$x^*(E(L_i, L_{\geq i+1})) + x^*(E(L_{\leq i-1}, L_i)) = x^*(\delta(L_i)) \geq 2.$$

Thus, $x^*(F_i) = x^*(E(L_i, L_{\geq i+1})) > \dfrac{1 - \tau + x^*(\delta(U_i))}{2} \geq 1 - \dfrac{\tau}{2}$. □

It is obvious that F_i's are disjoint and $F_i \subset \delta(U_i)$. For each τ-narrow cut U_i, we define $f^*_{U_i}$ as

$$(f^*_{U_i})_e := \begin{cases} x^*_e & \text{if } e \in F_i, \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 3. $\mathrm{E}[c(H)] \leq 1.6577 c(x^*)$.

Proof. Let

$$y := \alpha \chi_{\mathcal{T}} + \beta x^* + \sum_{i:|U_i \cap T| \text{ is odd},\ 1 \leq i < \ell} \frac{1 - (2\alpha + \beta)}{1 - \frac{\tau}{2}} f^*_{U_i},$$

for $\alpha = 0.30$, $\beta = 0.35$ and $\tau = \frac{1-2\alpha}{\beta} - 1 = \frac{1}{7}$. We claim y is a fractional T-join dominator. It is obvious that $y \geq 0$, and we have argued that $y(\delta(U)) \geq 1$ for nonseparating $(U, \bar{U})$. Suppose $(U, \bar{U})$ is an s-t cut with $|U \cap T|$ odd. If $(U, \bar{U})$ is not τ-narrow, then

$$\begin{aligned} y(\delta(U)) &\geq \alpha|\delta(U) \cap \mathscr{T}| + \beta x^*(\delta(U)) \\ &\geq 2\alpha + \beta(1+\tau) \\ &= 1. \end{aligned}$$

If $(U, \bar{U})$ is τ-narrow, then

$$\begin{aligned} y(\delta(U)) &\geq \alpha|\delta(U) \cap \mathscr{T}| + \beta x^*(\delta(U)) + \frac{1-(2\alpha+\beta)}{1-\frac{\tau}{2}} f_U^*(\delta(U)) \\ &\geq 2\alpha + \beta + \frac{1-(2\alpha+\beta)}{1-\frac{\tau}{2}}\left(1-\frac{\tau}{2}\right) \\ &= 1. \end{aligned}$$

Thus y is a fractional T-join dominator. Now it remains to bound the expected cost of H. Let $A := \dfrac{1-(2\alpha+\beta)}{1-\frac{\tau}{2}}$.

$$\begin{aligned} \mathrm{E}[c(H)] &\leq \mathrm{E}[c(\mathscr{T})] + \mathrm{E}[c(J)] \\ &\leq \mathrm{E}[c(\mathscr{T})] + \mathrm{E}[c(y)] \\ &= \mathrm{E}[c(\mathscr{T})] + \mathrm{E}[c(\alpha\chi_{\mathscr{T}})] + \mathrm{E}[c(\beta x^*)] + \mathrm{E}\left[c\left(\sum_{i:|U_i\cap T| \text{ is odd},\ 1\leq i<\ell} A \cdot f_{U_i}^*\right)\right] \\ &= (1+\alpha+\beta)c(x^*) + c\left(\sum_{i=1}^{\ell-1} \Pr[|U_i \cap T| \text{ is odd}] \cdot A \cdot f_{U_i}^*\right). \end{aligned}$$

From (2.3), $\mathrm{E}[c(H)] \leq (1+\alpha+\beta)c(x^*) + \tau A c\left(\sum_{i=1}^{\ell-1} f_{U_i}^*\right) \leq (1+\alpha+\beta+\tau A)c(x^*)$, where the last inequality follows from the disjointness of F_i. Note that $1+\alpha+\beta+\tau A < 1.6577$. □

A Tighter Analysis. In the previous analysis, we separately bounded the probability that a τ-narrow cut is odd, the deficit of the cut, and $f_U^*(\delta(U))$; moreover, we used $1-\frac{\tau}{2}$ instead of $\frac{1-\tau+x^*(\delta(U_i))}{2}$ from Lemma 2. These observations lead to some improvement, as shown in the following theorem.

Theorem 4. $\mathrm{E}[c(H)] \le \frac{9-\sqrt{33}}{2} c(x^*)$.

Proof. Let $b_i := \dfrac{1-\tau+x^*(\delta(U_i))}{2}$ denote the lower bound of $f^*_{U_i}(\delta(U_i))$ given by Lemma 2.

Let

$$y := \alpha\chi_{\mathcal{T}} + \beta x^* + \sum_{i:|U_i\cap T| \text{ is odd},\ 1\le i<\ell} \frac{1-\{2\alpha+\beta x^*(\delta(U_i))\}}{b_i} f^*_{U_i},$$

where α and β are to be chosen later; $\tau := \frac{1-2\alpha}{\beta} - 1$. As in the previous subsection, $\{U_i\}$ and $\{L_i\}$ denote the τ-narrow cuts and their layered structure. Assume $\frac{1}{3} \le \beta \le \frac{1}{2}$ and $1-2\beta \le \alpha \le \frac{1-\beta}{2}$.

A similar argument as in Theorem 3 proves that y is a fractional T-join dominator; it can also be shown that

$$\begin{aligned}
\mathrm{E}[c(H)] &\le (1+\alpha+\beta)c(x^*) + c\left(\sum_{i=1}^{\ell-1} \Pr[|U_i\cap T| \text{ is odd}] \frac{1-\{2\alpha+\beta x^*(\delta(U_i))\}}{b_i} f^*_{U_i}\right) \\
&\le (1+\alpha+\beta)c(x^*) + c\left(\sum_{i=1}^{\ell-1} \{x^*(\delta(U_i))-1\} \frac{1-\{2\alpha+\beta x^*(\delta(U_i))\}}{b_i} f^*_{U_i}\right) \\
&\le (1+\alpha+\beta)c(x^*) + \left[\max_{0\le\omega\le\tau}\left(\omega\frac{1-\{2\alpha+\beta(1+\omega)\}}{1-\frac{\tau}{2}+\frac{\omega}{2}}\right)\right] c\left(\sum_{i=1}^{\ell-1} f^*_{U_i}\right) \\
&\le \left\{1+\alpha+\beta+\max_{0\le\omega\le\tau}\left(\omega\frac{1-\{2\alpha+\beta(1+\omega)\}}{1-\frac{\tau}{2}+\frac{\omega}{2}}\right)\right\} c(x^*). \qquad (2.4)
\end{aligned}$$

Let $R(\omega) := \omega\dfrac{1-\{2\alpha+\beta(1+\omega)\}}{1-\frac{\tau}{2}+\frac{\omega}{2}} = \dfrac{\omega[1-\{2\alpha+\beta(1+\omega)\}]}{\frac{3}{2}-\frac{1}{2\beta}+\frac{\alpha}{\beta}+\frac{\omega}{2}}$. We have

$$R'(\omega) = \frac{-\frac{\beta}{2}\omega^2 + (1-2\alpha-3\beta)\omega + \left(2-4\alpha-\frac{3}{2}\beta-\frac{1}{2\beta}+\frac{2\alpha}{\beta}-\frac{2\alpha^2}{\beta}\right)}{\left(\frac{3}{2}-\frac{1}{2\beta}+\frac{\alpha}{\beta}+\frac{\omega}{2}\right)^2}$$

and the unique solution to

$$R'(\omega) = 0 \quad (0 \le \omega \le \frac{1-2\alpha}{\beta} - 1)$$

is

$$\omega = \omega_0 := \frac{1}{\beta}\left(1 - 2\alpha - 3\beta + \sqrt{(-2\beta)(1-2\alpha-3\beta)}\right).$$

Since $R(\omega) \ge 0$ for $0 \le \omega \le \frac{1-2\alpha}{\beta} - 1$ and $R(0) = R(\frac{1-2\alpha}{\beta} - 1) = 0$, $R(\omega)$ is maximized at $\omega = \omega_0$; hence, from (2.4),

$$E[c(H)] \le \left(5\alpha + 11\beta - 1 - 4\sqrt{(-2\beta)(1-2\alpha-3\beta)}\right) c(x^*).$$

Choose $\alpha = \frac{1}{\sqrt{33}}$, $\beta = \frac{1}{2} - \frac{1}{2\sqrt{33}}$ and we obtain

$$E[c(H)] \le \frac{9-\sqrt{33}}{2} c(x^*).$$

□

Proof of $\left(\frac{1+\sqrt{5}}{2}\right)$-approximation. Finally, we show that $\mathrm{E}[c(H)] \le \frac{1+\sqrt{5}}{2} c(x^*)$, proving Theorem 1 and Corollary 1.

In the previous analyses, F_i's serve as "representatives" of τ-narrow cuts. These representatives are useful since they have large weights while being disjoint. We improve the performance guarantee by introducing a new set of representatives that are "fractionally disjoint". Note that the three key properties of $\{f^*_{U_i}\}$ used in the proof of Theorem 4 are:

1. $f^*_{U_i} \ge 0$ for all i;
2. $\sum_{i=1}^{\ell-1} f^*_{U_i} \le x^*$; and
3. $f^*_{U_i}(\delta(U_i)) \ge \frac{1-\tau+x^*(\delta(U_i))}{2}$ for all i.

$\{f^*_{U_i}\}$ chosen in the previous analyses also satisfies that, for any given $e \in E$, $\left(f^*_{U_i}\right)_e \neq 0$ for at most one i. However, this was not a useful property in the analysis; Lemma 3 states that, by relaxing the definition of disjointness, we can choose $\{\hat{f}^*_{U_i}\}$ that have larger weights. The definitions of τ, $\{U_i\}$ and $\{L_i\}$ are unchanged.

Lemma 3. *There exists a set of vectors $\{\hat{f}^*_{U_i}\}_{i=1}^{\ell-1}$ satisfying:*

1. $\hat{f}^*_{U_i} \in \mathbb{R}^E_+$ *for all i;*
2. $\sum_{i=1}^{\ell-1} \hat{f}^*_{U_i} \leq x^*$*; and*
3. $\hat{f}^*_{U_i}(\delta(U_i)) \geq 1$ *for all i.*

This lemma is proven later; based on it, Lemma 4 proves the desired performance guarantee.

Lemma 4. $\mathrm{E}[c(H)] \leq \frac{1+\sqrt{5}}{2} c(x^*)$.

Proof. Let

$$y := \alpha \chi_{\mathscr{T}} + \beta x^* + \sum_{i:|U_i \cap T| \text{ is odd},\ 1 \leq i < \ell} \left[1 - \{2\alpha + \beta x^*(\delta(U_i))\}\right] \hat{f}^*_{U_i},$$

where α and β are parameters to be chosen later, satisfying

$$\frac{1}{3} \leq \beta \leq \frac{1}{2} \quad \text{and} \quad 1 - 2\beta \leq \alpha \leq \frac{1-\beta}{2}. \tag{2.5}$$

By following the same argument as in Theorem 4, we can easily show that y is a fractional T-join dominator; the only slight difference is when $(U, \bar{U})$ is τ-narrow and $|U \cap T|$ is odd, where we have

$$\begin{aligned} y(\delta(U)) &\geq \alpha|\delta(U) \cap \mathscr{T}| + \beta x^*(\delta(U)) + \left[1 - \{2\alpha + \beta x^*(\delta(U_i))\}\right] \hat{f}^*_U(\delta(U)) \\ &\geq 2\alpha + \beta x^*(\delta(U)) + \left[1 - \{2\alpha + \beta x^*(\delta(U_i))\}\right] \cdot 1 \\ &= 1, \end{aligned}$$

from the first and the third properties of Lemma 3. Hence, y is a fractional T-join dominator.

Now it remains to bound $\mathrm{E}[c(H)]$.

$$
\begin{aligned}
\mathrm{E}[c(H)] &\leq \mathrm{E}[c(\mathscr{T})] + \mathrm{E}[c(y)] \\
&= (1+\alpha+\beta)c(x^*) + c\left(\sum_{i=1}^{\ell-1} \Pr[|U_i \cap T| \text{ is odd}]\,[1 - \{2\alpha + \beta x^*(\delta(U_i))\}]\,\hat{f}^*_{U_i}\right) \\
&\leq (1+\alpha+\beta)c(x^*) + c\left(\sum_{i=1}^{\ell-1} \{x^*(\delta(U_i)) - 1\}\,[1 - \{2\alpha + \beta x^*(\delta(U_i))\}]\,\hat{f}^*_{U_i}\right) \\
&\leq (1+\alpha+\beta)c(x^*) + \left\{\max_{0\leq\omega\leq\tau} \omega\,[1 - \{2\alpha + \beta(1+\omega))\}]\right\} c\left(\sum_{i=1}^{\ell-1} \hat{f}^*_{U_i}\right). \qquad (2.6)
\end{aligned}
$$

From the second property of Lemma 3,

$$
\begin{aligned}
\mathrm{E}[c(H)] &\leq \left\{1 + \alpha + \beta + \max_{0\leq\omega\leq\tau} \omega\,[1 - \{2\alpha + \beta(1+\omega))\}]\right\} c(x^*) \\
&= \left\{1 + \alpha + \beta + \max_{0\leq\omega\leq\tau} \beta\omega(\tau - \omega)\right\} c(x^*) \\
&= \left\{1 + \alpha + \beta + \frac{(1 - 2\alpha - \beta)^2}{4\beta}\right\} c(x^*).
\end{aligned}
$$

We choose $\alpha = 1 - \frac{2}{\sqrt{5}}$ and $\beta = \frac{1}{\sqrt{5}}$. □

Proof of Lemma 3. Consider an auxiliary flow network illustrated in Figure 2.3, consisting of the source v^{source}, sink v^{sink}, a node v^{cut}_U for each τ-narrow cut U, and a node v^{edge}_e for each edge e in one or more τ-narrow cuts. The network has arcs of:

1. capacity 1 from v^{source} to v^{cut}_U for every τ-narrow cut U;
2. capacity ∞ from v^{cut}_U to v^{edge}_e for every $e \in \delta(U)$, for all U;
3. capacity x^*_e from v^{edge}_e to v^{sink} for every v^{edge}_e.

Let g be this capacity function.

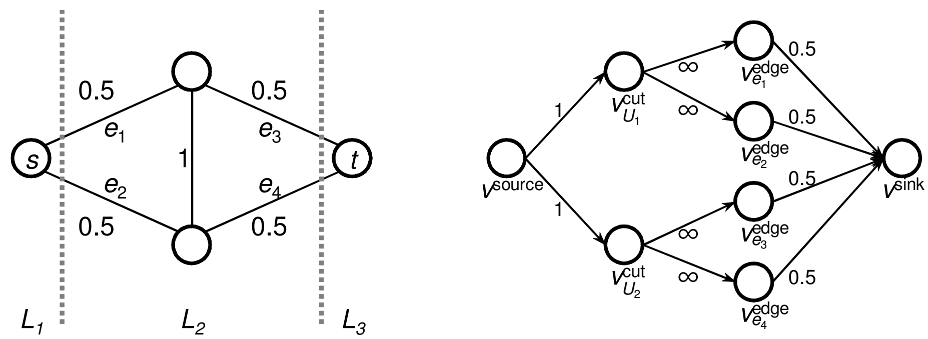

Figure 2.3: A feasible Held-Karp solution ($\ell = 3$) and its corresponding flow network.

Let $(S, \bar{S})$ be an arbitrary cut on this flow network, where $v^{\text{source}} \in S$. We claim the cut capacity of $(S, \bar{S})$ is at least $\ell - 1$.

Suppose there exists a τ-narrow cut U and $e \in \delta(U)$ such that $v^{\text{cut}}_U \in S$ and $v^{\text{edge}}_e \notin S$; the cut capacity is then ∞. So assume from now that (abusing the notation) every edge in any τ-narrow cut in S is also in S. Let $S \cap \{v^{\text{cut}}_{U_i} | 1 \leq i < \ell\} = \{v^{\text{cut}}_{U_{i_1}}, v^{\text{cut}}_{U_{i_2}}, \ldots, v^{\text{cut}}_{U_{i_k}}\}$ for some $1 \leq i_1 < i_2 < \ldots < i_k < \ell$. The cut capacity is then at least

$$\begin{aligned} & \sum_{v^{\text{cut}}_U \notin S} g(v^{\text{source}}, v^{\text{cut}}_U) + \sum_{e : \exists v^{\text{cut}}_U \in S \;\; e \in \delta(U)} g(v^{\text{edge}}_e, v^{\text{sink}}) \\ = \;\; & (\ell - 1 - k) + \sum_{e : \exists v^{\text{cut}}_U \in S \;\; e \in \delta(U)} x^*_e; \end{aligned}$$

if $k = 0$, the claim holds; the claim also holds for $k = 1$ since $x^*(\delta(U_{i_1})) \geq 1$. Suppose $k \geq 2$ (see Figure 2.4).

$$\begin{aligned} \sum_{e : \exists v^{\text{cut}}_U \in S \;\; e \in \delta(U)} x^*_e &= \frac{1}{2} \left[x^*(\delta(U_{i_1})) + \sum_{j=2}^{k} x^*(\delta(U_{i_j} \setminus U_{i_{j-1}})) + x^*(\delta(V \setminus U_{i_k})) \right] \\ &\geq \frac{1}{2} \left[1 + 2(k-1) + 1 \right] = k, \end{aligned}$$

proving the claim.

Thus the maximum flow on this flow network is of value at least $\ell - 1$. Con-

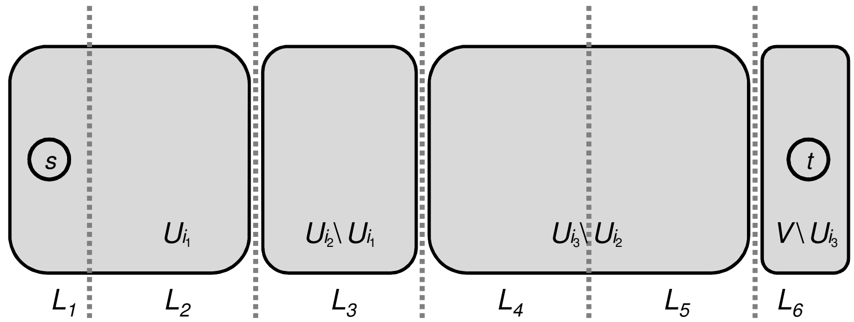

Figure 2.4: Schematic diagram: $\ell = 6$, $k = 3$, $i_1 = 2$, $i_2 = 3$, and $i_3 = 5$.

sider a maximum flow; this flow saturates all the edges from v^{source} to v_U^{cut}, since the cut capacity of $(\{v^{\mathsf{source}}\}, \overline{\{v^{\mathsf{source}}\}})$ is $\ell - 1$. Now, for each τ-narrow cut U, define $(\hat{f}_U^*)_e$ as the flow from v_U^{cut} to v_e^{edge} if $e \in \delta(U)$, and 0 otherwise. Then the first property is satisfied from the definition of flow; the second property is satisfied from the capacity constraints on v_e^{edge} to v^{sink}; lastly, the third property is satisfied since every edge from v^{source} to v_U^{cut} is saturated. □

2.3 Application to Other Problems

In this section, we exhibit how the present results can be applied to other problems to obtain approximation algorithms with better performance guarantees than the best known and improved LP integrality gap upper bounds.

2.3.1 Prize-collecting *s*-*t* Path Problem

The metric prize-collecting s-t path problem is, given a metric on vertices including s and t, and vertex prize defined on every vertex, to find a simple s-t path P that minimizes the sum of the path cost and the total prize "missed".

Archer et al. [6] use the path-variant Christofides' algorithm [33] as a black box to obtain a $\frac{241}{121}$-approximation algorithm for this problem; using the present algorithm as the black box readily produces an improvement, yielding a 1.9889-approximation algorithm. However, as the performance guarantee established by Theorem 1 is in terms of the Held-Karp optimum, the theorem, with the help of the parsimonious property [26, 23, 11], enables a further improvement via an analysis analogous to Goemans [25] based on an LP-rounding algorithm similar to Bienstock et al. [12]. This further improvement gives a 1.9535-approximation algorithm, and proves the same upper bound on the integrality gap of the linear program used.

Theorem 5. *There exists a* 1.9535*-approximation algorithm for the metric prize-collecting s-t path problem.*

Definition 5 (Metric prize-collecting s-t path problem)**.** *Given a complete graph* $G = (V, E)$ *with* $s, t \in V$*, metric edge cost function* $c : E \to \mathbb{R}_+$*, and vertex prize* $\pi : V \to \mathbb{R}_+$*, the metric prize-collecting s-t path problem is to find a simple s-t path P that minimizes the sum of the path cost and the total prize "missed", i.e.,* $c(P) + \pi(V \setminus V(P))$*.*

The s-t path TSP can be considered as a special case of the prize-collecting s-t path problem, where $\pi(v) = \infty$ for all $v \in V$.

Archer et al. [6] use the path-variant Christofides' algorithm [33] as a black box to obtain a $\frac{241}{121}$-approximation algorithm for the metric prize-collecting s-t path problem. $\frac{241}{121} < 1.9918$.

Theorem 6 (Archer et al. [6])**.** *Given a* ρ*-approximation algorithm* $\mathscr{A}$ *for the metric s-t path TSP, one can obtain a* $\left(2 - \left(\frac{2-\rho}{2+\rho}\right)^2\right)$*-approximation algorithm for the metric prize-collecting s-t path problem that uses* $\mathscr{A}$ *as a black box.*

This theorem, combined with Theorem 1, readily produces an improvement. Note that $\frac{1+4\sqrt{5}}{5} < 1.9889$.

Corollary 3. *There exists a* $\left(\frac{1+4\sqrt{5}}{5}\right)$*-approximation algorithm for the metric prize-collecting s-t path problem.*

However, as the performance guarantee established by Theorem 1 is in terms of the Held-Karp optimum, the theorem enables a further improvement via an analysis analogous to Goemans [25]. For the metric prize-collecting traveling salesman problem, Goemans [25] combines the LP rounding algorithm due to Bienstock et al. [12] and the primal-dual algorithm of Goemans & Williamson [27] (with the observation of [16] and [6]) to achieve the best performance guarantee known for the problem.

We start with the following LP relaxation of the problem:

$$\begin{array}{lll} \text{minimize} & c(x) + \pi(\mathbf{1} - y) & \\ \text{subject to} & x(\delta(S)) \geq 1, & \forall S \subsetneq V, |S \cap \{s,t\}| = 1; \\ & x(\delta(S)) \geq 2y_v, & \forall S \subsetneq V, S \cap \{s,t\} = \emptyset \quad \forall v \in S; \\ & x(\delta(\{s\})) = x(\delta(\{t\})) = 1; & \\ & x(\delta(\{v\})) = 2y_v, & \forall v \in V \setminus \{s,t\}; \\ & x_e \geq 0, & \forall e \in E; \\ & 0 \leq y_v \leq 1, & \forall v \in V \setminus \{s,t\}; \end{array} \tag{2.7}$$

where $\mathbf{1}$ denotes the all-1 vector in $V \in \mathbb{R}_+^{V \setminus \{s,t\}}$. It can be easily verified that this is a relaxation of the prize-collecting s-t path problem.

Given $V' \subset V \setminus \{s,t\}$, consider a related problem of finding a minimum s-t path on G that visits all the vertices in V', and only those vertices. The following

LP is a relaxation to this problem:

$$
\begin{aligned}
\text{minimize} \quad & c(x) & & \\
\text{subject to} \quad & x(\delta(S)) \geq 1, & & \forall S \subsetneq V, |S \cap \{s,t\}| = 1; \\
& x(\delta(S)) \geq 2, & & \forall S \subsetneq V, S \cap \{s,t\} = \emptyset, S \cap V' \neq \emptyset; \\
& x(\delta(\{s\})) = x(\delta(\{t\})) = 1; & & \\
& x(\delta(\{v\})) = 2, & & \forall v \in V'; \\
& x(\delta(\{v\})) = 0, & & \forall v \in V \setminus \{s,t\} \setminus V'; \\
& x_e \geq 0, & & \forall e \in E.
\end{aligned}
\tag{2.8}
$$

Observation 2. *Let $G' = (V' \cup \{s,t\}, E')$ be the subgraph of G induced by $V' \cup \{s,t\}$. Projecting a feasible solution to* (2.8) *to E' yields a feasible solution to the path-variant Held-Karp relaxation for G'.*

The following lemma, which also follows from Goemans [23] and Bertsimas & Teo [11], shows that we can use the parsimonious property. We give an alternative proof here.

Lemma 5. *The optimal solution value to* (2.8) *is equal to the optimal solution value to the following relaxation without the degree constraints:*

$$
\begin{aligned}
\textit{minimize} \quad & c(x) & & \\
\textit{subject to} \quad & x(\delta(S)) \geq 1, & & \forall S \subsetneq V, |S \cap \{s,t\}| = 1; \\
& x(\delta(S)) \geq 2, & & \forall S \subsetneq V, S \cap \{s,t\} = \emptyset, S \cap V' \neq \emptyset; \\
& x_e \geq 0, & & \forall e \in E.
\end{aligned}
\tag{2.9}
$$

Proof. Let $G = (V, E)$. It suffices to show that, given a feasible solution x^* to (2.9), how to construct a feasible solution to (2.8) whose cost is no greater than $c(x^*)$.

We will extend the graph (and x^*) so that the relaxation (almost) becomes a set of edge-connectivity requirements between pairs of vertices, and then use a similar approach as in Bienstock et al. [12], along with the following lemma:

Lemma 6 ([12])**.** *Let $G = (V, E)$ be an Eulerian multigraph. Suppose that, for some $U \subset V$ and $v \in V$, any two vertices in U other than v are k-edge-connected. Let x be an arbitrary neighbor of v; then, there exists a neighbor y of v such that*

1. *$x \neq y$; and*
2. *any two vertices in U other than v are still k-edge-connected after splitting (x, v) and (y, v): i.e., replacing (x, v) and (y, v) (one copy each) with (x, y).*

Without loss of generality, we can assume x^* is rational.

Now we add three new vertices to the graph: s', t' and u. We set $c(s', v) = c(s, v)$ and $c(t', v) = c(t, v)$ for all v; $c(s', s) = c(t', t) = 0$: s' and t' will be the "proxy" of s and t. We do not define the cost between u and other vertices: these costs do not affect the rest of the analysis. However, for notational convenience, we set these costs to be zero, potentially violating the triangle inequality. Let $\bar{G} = (\bar{V}, \bar{E})$ be this extended graph.

We extend x^* into $\bar{x}^*$ as well: $\bar{x}^*(s, s') = \bar{x}^*(s', u) = \bar{x}^*(u, t') = \bar{x}^*(t', t) = 1$, and all other newly added edges are set to zero. Note that the (fractional) degree of s', t' and u are 2.

Let $\bar{V}' := V' \cup \{s', t', u\}$; we claim that any two vertices in $\bar{V}'$ are 2-edge-connected.

Claim 1. *For any $S \subset \bar{V}$ such that $\bar{V}' \cap S \neq \emptyset$ and $\bar{V}' \setminus S \neq \emptyset$, $\bar{x}^*(\delta(S)) \geq 2$.*

Proof. Without loss of generality, assume $s \in S$. If $t \notin S$, then at least one edge of the path $P : s - s' - u - t' - t$ is in $\delta(S)$; thus,

$$\bar{x}^*(\delta(S)) \geq x^*(\delta_G(S \cap V)) + \bar{x}^*(\delta(S) \cap P) \geq 1 + 1.$$

Suppose $t \in S$. If $\{s', u, t'\} \setminus S \neq \emptyset$ then $|\delta(S) \cap P| \geq 2$; hence,

$$\bar{x}^*(\delta(S)) \geq \bar{x}^*(\delta(S) \cap P) \geq 2.$$

Otherwise, $V' \setminus S = \bar{V}' \setminus S \neq \emptyset$ and thus,

$$\bar{x}^*(\delta(S)) \geq x^*(\delta_G(S \cap V)) \geq 2,$$

since $(S \cap V) \cap V' \subsetneq V'$. □

Now scale $\bar{x}^*$ by some large constant C so that $\bar{z}^* := C\bar{x}^*$ is integral and, in the multigraph on $\bar{V}$ whose edge multiplicities are given by $\bar{z}^*$, the degree of every vertex is even. Note that any two vertices in $\bar{V}'$ are $2C$-edge-connected in this multigraph.

Let $\phi := \sum_{v \in \bar{V}'}[\bar{z}^*(\delta(v)) - 2C] + \sum_{v \in \bar{V} \setminus \bar{V}'} \bar{z}^*(\delta(v))$; ϕ is an even integer. We will modify $\bar{z}^*$ until ϕ reaches 0: in particular, we split two edges in the multigraph so that

(i) ϕ decreases by 2;

(ii) $c(\bar{z}^*)$ do not increase;

(iii) any two vertices in $\bar{V}'$ are $2C$-edge-connected;

(iv) the degrees of s', t' and u all remain $2C$;

(v) the only edges incident to u are (s', u) and (u, t'); and

(vi) every vertex has even degree and hence the connected component containing $\bar{V}'$ is Eulerian.

It is clear that the invariants (iii) through (vi) initially hold.

If there exists an edge that is not reachable from any vertex in $\bar{V}'$, we can remove all such edges without violating any of the conditions (ϕ may decrease by more than 2).

If there exists $v \in \bar{V} \setminus \bar{V}'$ such that $\bar{z}^*(\delta(v)) > 0$, then we apply Lemma 6 to pick two incident edges to split. Note that $v \notin \{s', t', u\}$ since $s', t', u \in \bar{V}'$. (iii) is maintained from the lemma. Splitting does not change the degree of any vertex other than v; hence (i), (iv) and (vi) are satisfied. Neither of the chosen edges is incident to u, as can be seen from (v); thus, (v) is maintained and (ii) follows from the triangle inequality.

Otherwise, we choose $v \in \bar{V}'$ such that $\bar{z}^*(\delta(v)) > 2C$. $\bar{z}^*(\delta(v)) \geq 2C + 2$ from (vi). Again $v \notin \{s', t', u\}$ from (iv); we can similarly verify all properties in this case as well.

Once ϕ reaches 0, we remove u and its incident edges. None of these edges got split during the process: this is the reason why the cost of these edges can be left undefined.

Note that the degree of s and t now are 0, whereas s' and t' are 1. Concatenate s and s', and t and t', respectively; we scale this multigraph back by $1/C$ to obtain a feasible solution to (2.8) whose cost is no greater than $c(x^*)$. □

We are now ready to apply the analyses of Goemans [25] and Bienstock et al. [12]. Let x^* and y^* be an optimal solution to (2.7).

Lemma 7. *Let $\mathscr{A}^\rho$ be an approximation algorithm for the s-t path TSP that produces a path of cost at most ρ times the Held-Karp optimum. Let $V_\gamma = \{v | y^*_v \geq \gamma\}$ for some $0 < \gamma \leq 1$. Running $\mathscr{A}^\rho$ on the subgraph G_γ induced by $V_\gamma \cup \{s, t\}$ yields a path P with $c(P) \leq \frac{\rho}{\gamma} c(x^*)$.*

Proof. The proof is basically the same as [12]. Observe that $\frac{x^*}{\gamma}$ is a feasible solution to (2.9), as can be seen from (2.7) and (2.9). From Lemma 5 and Observation 2, the Held-Karp optimum for G_γ is of cost no greater than $c(\frac{x^*}{\gamma})$. □

The primal-dual algorithm of Chaudhuri et al. [13] can be used to obtain the following performance guarantee for the metric prize-collecting *s*-*t* path problem.

Lemma 8 ([13, 6])**.** *There exists a polynomial-time algorithm* $\mathscr{A}_{\mathsf{PD}}$ *that produces an s-t path P satisfying*

$$c(P) + \pi(V \setminus V(P)) \leq 2c(x^*) + \pi(\mathbf{1} - y^*).$$

Now, the combined algorithm is as follows: let $a := e^{1-\frac{2}{\rho}}$ and $p := \frac{1+\rho \ln a}{2-a+\rho \ln a}$. The algorithm runs $\mathscr{A}_{\mathsf{PD}}$ with probability p; otherwise, it computes an optimal solution x^* and y^* to (2.7), samples γ uniformly at random from $(a, 1)$, and run $\mathscr{A}^\rho$ on the subgraph induced by $V_\gamma = \{v | y^*_v \geq \gamma\}$.

This algorithm can be derandomized since there are only $O(|V|)$ different V_γ's possible.

Theorem 7. *Let* $\mathscr{A}^\rho$ *be an approximation algorithm for the s-t path TSP that produces a path of cost at most* ρ *times the Held-Karp optimum, for some* $\frac{3}{2} \leq \rho < 2$*; then, there exists a* $\left(\dfrac{\rho}{\rho - e^{1-\frac{2}{\rho}}}\right)$*-approximation algorithm for the metric prize-collecting s-t path problem.*

Proof. The given algorithm is a polynomial-time algorithm. Let P denote the output path.

It can be easily verified that $0 < a < 1$ and $0 < p < 1$. From Lemma 7,

$$\begin{aligned}
\mathrm{E}[c(P)|\mathscr{A}^\rho \text{ is chosen}] &\leq \mathrm{E}[\frac{\rho}{\gamma}c(x^*)|\mathscr{A}^\rho \text{ is chosen}] \\
&= \rho c(x^*)\int_a^1 \frac{1}{1-a}\frac{1}{\gamma}d\gamma \\
&= \frac{-\ln a}{1-a}\rho c(x^*). \qquad (2.10)
\end{aligned}$$

We have

$$\begin{aligned}
\mathrm{E}[\pi(V \setminus V(P))|\mathscr{A}^\rho \text{ is chosen}] &= \sum_{v\in V\setminus\{s,t\}} \pi(v)\cdot \Pr[v \notin V_\gamma] \\
&= \sum_{v\in V\setminus\{s,t\}} \pi(v)\cdot \min\left(\frac{1-y_v^*}{1-a}, 1\right) \\
&\leq \frac{1}{1-a}\pi(\mathbf{1}-y^*). \qquad (2.11)
\end{aligned}$$

From (2.10), (2.11), and Lemma 8,

$$\begin{aligned}
&\mathrm{E}[c(P) + \pi(V \setminus V(P))] \\
=\ & p\left[2c(x^*) + \pi(\mathbf{1}-y^*)\right] + (1-p)\left[\frac{-\ln a}{1-a}\rho c(x^*) + \frac{1}{1-a}\pi(\mathbf{1}-y^*)\right] \\
=\ & \left[2p + (1-p)\frac{-\ln a}{1-a}\rho\right]c(x^*) + \left[p + (1-p)\frac{1}{1-a}\right]\pi(\mathbf{1}-y^*) \\
=\ & \frac{\rho}{\rho - e^{1-\frac{2}{\rho}}}\left[c(x^*) + \pi(\mathbf{1}-y^*)\right].
\end{aligned}$$

□

Theorem 7 along with Theorem 1 yields the following:

Corollary 4. *There exists a deterministic* 1.9535*-approximation algorithm for the metric prize-collecting s-t path problem.*

Corollary 5. *The integrality gap of* (2.7) *is smaller than* 1.9535.

2.3.2 Unit-weight Graphical Metrics

In this subsection, we study the s-t path TSP for the special case where the cost function is a shortest-path metric defined by an underlying undirected, unit-weight graph. Mucha [44] gives an improved analysis of the algorithm of Mömke and Svensson [43] for this problem, showing it to be a $(\frac{19}{12} + \epsilon)$-approximation algorithm for any $\epsilon > 0$; the critical case of this analysis is when the Held-Karp optimum is close to $|V| - 1$. Even though τ-narrow cuts function as a mere analytic tool in Section 2.2, we propose an algorithm that actually computes the τ-narrow cuts and uses them: once the τ-narrow cuts are computed, the algorithm constructs an s-t path that traverses from the first layer to the last, without "skipping" any layer in between. If the path is inexpensive, the number of τ-narrow cuts is also small, so the algorithm presented in Section 2.2 produces a good solution. If the path is expensive but the Held-Karp optimum is close to $|V|-1$, then we can show that the path already contains a large number of vertices and therefore can be augmented into a spanning Eulerian path with small additional cost. Lastly, if the Held-Karp optimum is bounded away from $|V| - 1$, then Mömke & Svensson's algorithm performs well, as can be seen from Mucha [44]. This gives a 1.5780-approximation algorithm for the s-t path TSP under the unit-weight graphical metric, and proves an upper bound of 1.6137 on the integrality gap of the path-variant Held-Karp relaxation under this special case.

Preliminaries. Let x^* be an optimal solution to the path-variant Held-Karp relaxation; let G_0 be the underlying unit-weight graph defining the cost function. G_0 is connected.

Mucha [44] gives an improved analysis of the 1.5858-approximation algorithm of Mömke and Svensson [43]; the following is from [44].

Lemma 9 (Mucha [44])**.** *There exists an algorithm $\mathscr{A}_0$ for the s-t path TSP under unit-weight graphical metrics, which returns a solution of cost at most*

$$\min\left(\frac{10}{9}c(x^*) + \frac{1}{3}c(s,t) + \frac{1}{3}|V| + \frac{4}{9}, 2|V| - 2 - c(s,t)\right).$$

This immediately gives a $(\frac{19}{12} + \epsilon)$-approximation algorithm for any $\epsilon > 0$. Note that $\frac{19}{12} < 1.5834$.

Theorem 8 (Mucha [44])**.** *There exists a $(\frac{19}{12} + \epsilon)$-approximation algorithm for the s-t path TSP under unit-weight graphical metrics, for any $\epsilon > 0$.*

Proof. Let P be the output of $\mathscr{A}_0$. From Lemma 9,

$$\begin{aligned} c(P) &\leq \frac{3}{4}\left(\frac{10}{9}c(x^*) + \frac{1}{3}c(s,t) + \frac{1}{3}|V| + \frac{4}{9}\right) + \frac{1}{4}\left(2|V| - 2 - c(s,t)\right) \\ &= \frac{5}{6}c(x^*) + \frac{3}{4}(|V| - 1) + \frac{7}{12} \\ &\leq \frac{5}{6}c(x^*) + \frac{3}{4}c(x^*) + \frac{7}{12}, \end{aligned}$$

where the last line holds since $c(e) \geq 1$ for all e.

Thus, there exists n_0 such that $c(P) \leq (\frac{19}{12} + \epsilon)c(x^*)$ for each input that has n_0 or more vertices. Smaller instances can be solved separately. □

It can be observed from Lemma 9 and Theorem 8 that the "critical case" determining the proven performance guarantee is when $c(x^*) \approx |V|$.

Algorithm. The algorithm gives three different constructions of Hamiltonian paths carrying performance analyses with complementary critical cases. Algorithm 1 shows the entire algorithm (except the separate handling of small instances); $\theta \in (0, 1)$ is a parameter to be chosen later. Let $\eta : E \to \mathbb{Z}_{\geq 0}$ be a function such that $\eta(e) := c(e) - 1$. For $U \subset V$, $G(U)$ denotes the subgraph of G induced by U. Suppose $|V| \geq 3$; this implies $\ell \geq 3$. Even though τ-narrow cuts function as a mere analytic tool in Section 2.2, this algorithm actually computes τ-narrow cuts and uses them.

The present algorithm obtains the first Hamiltonian path H_A using the algorithm of Mömke and Svensson [43]; H_B is obtained using the algorithm for the general metric problem, taken from Section 2.2. In order to obtain the last Hamiltonian path H_C, the algorithm first finds the layered structure induced by the $(1 - \theta)$-narrow cuts. Then it obtains an s-t path P_{LT} that traverses from the first layer to the last, without skipping over any layer. In particular, P_{LT} uses the cheapest possible edge (p_i, q_{i+1}) to move from one layer L_i to the next layer L_{i+1}, and these "inter-layer" edges are concatenated into a connected path P_{LT} by taking an "intra-layer" path P_i as the shortest path from q_i to p_i with respect to η.

Analysis. We first show that Algorithm 1 is a polynomial-time algorithm, and Step 14 of the algorithm is well-defined.

Lemma 10. *Algorithm 1 is a well-defined, polynomial-time algorithm.*

Proof. Steps 13-16 start with an s-t path, and augment it into a spanning multigraph that has an Eulerian path between s and t. This follows from the preservation of the parity of degree. The choice of (u, v) satisfying $c(u, v) = 1$ is always possible since G_0 is connected.

Algorithm 1: Algorithm for s-t path TSP under unit-weight graph. metric.

Input: Complete graph $G = (V, E)$ with cost $c : E \to \mathbb{Z}_{>0}$; endpoints $s, t \in V$.

Output: Hamiltonian path between s and t.

Run $\mathscr{A}_0$; let H_A be the output Hamiltonian path.
$x^* \leftarrow$an optimal solution to the path-variant Held-Karp relaxation
Run the algorithm from Section 2.2; let H_B be the output Hamiltonian path.
Compute the partition $L_1, \dots L_\ell$ induced by the $(1-\theta)$-narrow cuts U_i.
for $1 \leq i < \ell$ **do**
    Let (p_i, q_{i+1}) be the shortest edge in $E(L_i, L_{i+1})$, where $p_i \in L_i$ and $q_{i+1} \in L_{i+1}$.
end for
for $1 < i < \ell$ **do**
    Let P_i be shortest path from q_i to p_i within $G(L_i)$, under cost given by η.
end for
Let P_{LT} be an s-t path obtained by concatenating $(s, q_2), P_2, (p_2, q_3), P_3, \dots, P_{\ell-1}, (p_{\ell-1}, t)$.
$G_E \leftarrow (V, P_{\mathsf{LT}})$
while the multigraph G_E is not spanning **do**
    Choose (u, v) such that: $c(u, v) = 1$, u is isolated in G_E, and v is not.
    Add two copies of (u, v) to G_E.
end while
Shortcut an Eulerian path of G_E to obtain a Hamiltonian path H_C.
Let H_{out} be the best among H_A, H_B and H_C; output H_{out}.

P_{LT} is an s-t path since $L_1 = \{s\}$ and $L_\ell = \{t\}$. Note that some of the P_i's may be a length-0 path.

Step 4, unlike the algorithm from Section 2.2, actually computes the layered structure of $(1-\theta)$-narrow cuts, whereas this structure was only for the sake of analysis in Section 2.2. Yet, the layers can in fact be identified via a polynomial number of min-cut calculations; hence, the algorithm is a polynomial-time algorithm. $\square$

If P_{LT} is inexpensive, the number of $(1-\theta)$-narrow cuts is small, since P_{LT} does not skip over any layer; the algorithm from Section 2.2 provides a good solution in this case. If $c(x^*) \gg |V| - 1$, then Mömke & Svensson's algorithm performs well provided that the graph is sufficiently large. Lastly, if P_{LT} is expensive and $c(x^*) \approx |V|$, we prove that P_{LT} already contains a large number of vertices and therefore can be augmented into a spanning Eulerian path by adding a small number of edges. This follows from the fact that, in the Held-Karp solution, each layer is θ-edge-connected and that θ fractional edges lie between every two consecutive layers; the following lemmas establish these facts.

Lemma 11. *$x^*(E(L_1, L_2)) > \theta$.*

Proof. We have

$$x^*(E(L_1, L_2) + x^*(E(L_2, L_{\geq 3})) = x^*(\delta(L_2)) \geq 2 \tag{2.12}$$

and

$$x^*(E(L_1, L_{\geq 3})) + x^*(E(L_2, L_{\geq 3})) = x^*(\delta(U_2)) < 1 + (1-\theta); \tag{2.13}$$

from (2.12) and (2.13),

$$x^*(E(L_1, L_2)) - x^*(E(L_1, L_{\geq 3})) > \theta.$$

□

By symmetry, $x^*(E(L_{\ell-1}, L_\ell)) > \theta$.

Lemma 12. *For any* $i \geq 1$, $j \leq \ell$, $V_1 \neq \emptyset$ *and* $V_2 \neq \emptyset$ *such that*

1. $i + 2 \leq j$,
2. $V_1 \cup V_2 = \cup_{k=i+1}^{j-1} L_k$, *and*
3. $V_1 \cap V_2 = \emptyset$,

$x^*(E(V_1, V_2)) > \theta$ *holds.*

Proof. We have

$$x^*(E(L_{\leq i}, V_1)) + x^*(E(L_{\leq i}, V_2)) + x^*(E(L_{\leq i}, L_{\geq j})) = x^*(\delta(L_{\leq i})) < 1 + (1 - \theta); \tag{2.14}$$

by symmetry,

$$x^*(E(L_{\leq i}, L_{\geq j})) + x^*(E(V_1, L_{\geq j})) + x^*(E(V_2, L_{\geq j})) < 1 + (1 - \theta); \tag{2.15}$$

$$x^*(E(L_{\leq i}, V_1)) + x^*(E(V_1, V_2)) + x^*(E(V_1, L_{\geq j})) = x^*(\delta(V_1)) \geq 2; \tag{2.16}$$

again by symmetry,

$$x^*(E(L_{\leq i}, V_2)) + x^*(E(V_1, V_2)) + x^*(E(V_2, L_{\geq j})) \geq 2. \tag{2.17}$$

From (2.14) through (2.17), $2x^*(E(V_1, V_2)) - 2x^*(E(L_{\leq i}, L_{\geq j})) > 2\theta$. □

Corollary 6. *For all* $1 \leq i < \ell$, $x^*(E(L_i, L_{i+1})) > \theta$.

Proof. From Lemma 11 and Lemma 12 applied for $j - i = 3$. □

Corollary 7. *For all i, $G(L_i)$ weighted by (the projection of) x^* is θ-edge-connected.*

Proof. L_1 and L_ℓ are singletons; every cut in any other nonsingleton layer subgraphs are of capacity at least θ from Lemma 12, applied for $j - i = 2$. □

Let $\sigma, \kappa \geq 0$ be some parameters to be chosen later.

Lemma 13.

$$c(H_{\mathsf{out}}) \leq \max \begin{Bmatrix} \left(\frac{5}{6} + \frac{3}{4(1+\sigma)}\right) c(x^*) + \frac{7}{12} \\ \\ \left(2 - \kappa + \frac{2\sigma}{\theta}\right) c(x^*) \\ \\ \left[\frac{3+2\theta}{2+\theta} + \frac{(1-\theta)^2}{4(2+\theta)}\kappa\right] c(x^*) \end{Bmatrix}.$$

Proof. Suppose $c(x^*) \geq (1+\sigma)(|V|-1)$; from the proof of Theorem 8,

$$\begin{aligned} c(H_{\mathsf{out}}) &\leq c(H_A) \\ &\leq \frac{5}{6}c(x^*) + \frac{3}{4}(|V|-1) + \frac{7}{12} \\ &\leq \left(\frac{5}{6} + \frac{3}{4(1+\sigma)}\right) c(x^*) + \frac{7}{12}; \end{aligned}$$

thus, we can assume from now that $c(x^*) < (1+\sigma)(|V|-1)$.

Case 1.

$$c(P_{\mathsf{LT}}) \geq \kappa(|V|-1). \tag{2.18}$$

From Corollary 6 and the choice of (p_i, q_{i+1}),

$$\theta \cdot \eta(p_i, q_{i+1}) \leq (\eta * x^*)(E(L_i, L_{i+1})). \tag{2.19}$$

For each layer L_i with $1 < i < \ell$, consider a bidirected flow network on $G(L_i)$ whose capacities are given by x^*. From Corollary 7, we can route flow of θ from

q_i to p_i. This flow can be decomposed into cycles and paths from q_i to p_i; thus, by the choice of P_i,

$$\theta \cdot \eta(P_i) \leq (\eta * x^*)(E(L_i)). \tag{2.20}$$

From (2.19) and (2.20),

$$\begin{aligned}
\theta \cdot \eta(P_{\mathsf{LT}}) &= \sum_{1 \leq i < \ell} \theta \cdot \eta(p_i, q_{i+1}) + \sum_{1 < i < \ell} \theta \cdot \eta(P_i) \\
&\leq \sum_{1 \leq i < \ell} (\eta * x^*)(E(L_i, L_{i+1})) + \sum_{1 < i < \ell} (\eta * x^*)(E(L_i)) \\
&\leq (\eta * x^*)(E) \\
&= c(x^*) - x^*(E) \\
&< \sigma(|V| - 1). \qquad (2.21)
\end{aligned}$$

Let $|P_{\mathsf{LT}}|$ denote the number of edges on P_{LT}. We have

$$\begin{aligned}
c(H_{\mathsf{out}}) &\leq c(H_C) \\
&\leq c(P_{\mathsf{LT}}) + 2\left[(|V| - 1) - |P_{\mathsf{LT}}|\right] \\
&= c(P_{\mathsf{LT}}) + 2\left[(|V| - 1) - \{c(P_{\mathsf{LT}}) - \eta(P_{\mathsf{LT}})\}\right] \\
&\leq \left[2 - \kappa + \frac{2\sigma}{\theta}\right] \cdot (|V| - 1) \\
&\leq \left(2 - \kappa + \frac{2\sigma}{\theta}\right) c(x^*),
\end{aligned}$$

where the second-to-last line follows from (2.18) and (2.21); the last from $c(x^*) \geq |V| - 1$.

Case 2.

$$c(P_{\mathsf{LT}}) < \kappa(|V| - 1). \tag{2.22}$$

Note that, from the construction of P_{LT}, $\ell - 1 \leq |P_{\mathsf{LT}}|$; hence we have

$$\ell - 1 \leq |P_{\mathsf{LT}}| \leq c(P_{\mathsf{LT}}) < \kappa(|V| - 1).$$

From each $(1-\theta)$-narrow cut $(U_i, \bar{U}_i)$, we can pick an edge $d_i \in \delta(U_i)$ with $c(d_i) = 1$ due to the connectedness of G_0. Let $\hat{f}^*_{U_i} := \mathbf{e}_{d_i}$, $\alpha := \frac{\theta}{2+\theta}$, $\beta := \frac{1}{2+\theta}$, and $\tau = \frac{1-2\alpha}{\beta} - 1 = 1 - \theta$. Note that this choice of α and β satisfies (2.5). Since the second condition on $\{\hat{f}^*_{U_i}\}_{i=1}^{\ell-1}$ of Lemma 3 is not used to derive (2.6) (it is used in the later part of the proof), we have

$$\begin{aligned} c(H_{\text{out}}) &\leq c(H_B) \\ &\leq (1+\alpha+\beta)c(x^*) + \left\{\max_{0\leq\omega\leq\tau} \omega\left[1 - \{2\alpha + \beta(1+\omega)\}\right]\right\} c\left(\sum_{i=1}^{\ell-1} \hat{f}^*_{U_i}\right) \\ &= \frac{3+2\theta}{2+\theta}c(x^*) + \frac{(1-\theta)^2}{4(2+\theta)} c\left(\sum_{i=1}^{\ell-1} \hat{f}^*_{U_i}\right). \end{aligned}$$

As $c(d_i) = 1$ for all i,

$$\begin{aligned} c(H_{\text{out}}) &\leq \frac{3+2\theta}{2+\theta}c(x^*) + \frac{(1-\theta)^2}{4(2+\theta)}(\ell - 1) \\ &\leq \left[\frac{3+2\theta}{2+\theta} + \frac{(1-\theta)^2}{4(2+\theta)}\kappa\right] c(x^*). \end{aligned}$$

□

Corollary 8. *Let* $\rho := \max\left\{\frac{5}{6} + \frac{3}{4(1+\sigma)}, 2 - \kappa + \frac{2\sigma}{\theta}, \frac{3+2\theta}{2+\theta} + \frac{(1-\theta)^2}{4(2+\theta)}\kappa\right\}$. *There exists a* $(\rho+\epsilon)$-*approximation algorithm for the s-t path TSP under unit-weight graphical metrics, for any* $\epsilon > 0$.

Corollary 9. *There exists a* 1.5780-*approximation algorithm for the s-t path TSP under unit-weight graphical metrics.*

Proof. Directly follows from Corollary 8: if we choose, for example, $\theta = 1.2297 \times 10^{-1}$, $\sigma = 7.2774 \times 10^{-3}$, and $\kappa = 5.4045 \times 10^{-1}$, we have $\rho < 1.5780$. □

Corollary 10. *The integrality gap of the path-variant Held-Karp relaxation under the unit-weight graphical metric is smaller than* 1.6137.

Proof. Trivial for $|V| = 2$. Let OPT denote the optimal (integral) solution value.

Suppose $3 \leq |V| \leq 6$. From a similar argument as in the proof of Lemma 10, if there exists a simple s-t path with m edges in G_0, $\mathsf{OPT} \leq m + 2(|V| - 1 - m) = 2|V| - 2 - m$. Thus, if there exists a simple s-t path with at least two edges,

$$\frac{\mathsf{OPT}}{c(x^*)} \leq \frac{2|V| - 4}{|V| - 1} \leq \frac{8}{5} < 1.6137.$$

Suppose there does not exist a simple s-t path with more than one edge; then $(s,t) \in G_0$ and (s,t) is a bridge of G_0. Let $(U, \bar{U})$ be the s-t cut defined by the removal of (s,t) from G_0. $x^*(s,t) = 0$ since $2x^*(s,t) = x^*(\delta(\{s\})) + x^*(\delta(\{t\})) - x^*(\delta(\{s,t\})) \leq 1 + 1 - 2 = 0$; therefore,

$$\begin{aligned} c(x^*) &= (c * x^*)(\delta(U)) + (c * x^*)(E \setminus \delta(U)) \\ &= (c * x^*)(\delta(U) \setminus \{s,t\}) + (c * x^*)(E \setminus \delta(U)) \\ &\geq 2x^*(\delta(U) \setminus \{s,t\}) + x^*(E \setminus \delta(U)) \\ &= x^*(\delta(U)) + x^*(E) \\ &\geq |V| \end{aligned}$$

and

$$\frac{\mathsf{OPT}}{c(x^*)} \leq \frac{2|V| - 3}{|V|} \leq \frac{3}{2} < 1.6137.$$

Suppose $|V| \geq 7$. Choose $\theta = 3.7304 \times 10^{-1}$, $\sigma = 8.5757 \times 10^{-2}$, and $\kappa =$

8.4614×10^{-1}; from the proof of Lemma 13,

$$
\begin{aligned}
c(H_{\mathsf{out}}) &\le \max\left\{\begin{array}{l}
\left(\dfrac{5}{6}+\dfrac{3}{4(1+\sigma)}+\dfrac{7}{12(|V|-1)(1+\sigma)}\right)c(x^*) \\[2em]
\left(2-\kappa+\dfrac{2\sigma}{\theta}\right)c(x^*) \\[2em]
\left[\dfrac{3+2\theta}{2+\theta}+\dfrac{(1-\theta)^2}{4(2+\theta)}\kappa\right]c(x^*)
\end{array}\right\} \\
&\le Qc(x^*),
\end{aligned}
$$

for some $Q < 1.6137$. □

2.3.3 Unit-weight Graphical Metrics, an Alternative Approach

In this subsection, we present a new analysis of the path-variant Christofides' algorithm [15, 33] for the metric s-t path TSP, and show how the critical case characterized by this analysis can lead to an improvement in the special case of unit-weight graphical metrics. The analysis compares the output solution value to the LP optimum of the path-variant Held-Karp relaxation, thereby proving the upper bound of 5/3 on the integrality gap of the path-variant Held-Karp relaxation. We note that the LP optimum is never computed by the algorithm.

First we recall the following definition of the circuit-variant Held-Karp relaxation:

Definition 6 ([31])**.** *The* circuit-variant Held-Karp relaxation *is the following:*

$$
\begin{aligned}
\textit{minimize} \quad & c(x) \\
\textit{subject to} \quad & x(\delta(S)) \geq 2, \quad \forall S \subsetneq V, S \neq \emptyset; \\
& x(\delta(\{v\})) = 2, \quad \forall v \in V; \\
& x \geq 0.
\end{aligned}
\tag{2.23}
$$

Let $G = (V, E)$ be the input complete graph with cost function $c : E \to \mathbb{R}_+$ and the endpoints $s, t \in V$. The path-variant Christofides' algorithm first finds a minimum spanning tree $\mathscr{T}_{\min}$ of G; it then computes a minimum T-join J, where $T \subset V$ is the set of the vertices with the "wrong" parity of degree in $\mathscr{T}_{\min}$: i.e., T is the set of odd-degree internal points and even-degree endpoints in $\mathscr{T}_{\min}$. Lastly, the algorithm shortcuts an Eulerian path of the multigraph $\mathscr{T}_{\min} \cup J$ to obtain the output Hamiltonian path H.

We give two different bounds on the cost of J, which together will establish the performance guarantee. Let $x^* \in \mathbb{R}^E$ be the LP optimum of the path-variant Held-Karp relaxation.

Lemma 14. $c(\mathscr{T}_{\min}) \leq c(x^*)$.

Proof. As can be seen from Observation 1, the path-variant Held-Karp polytope is contained in the spanning tree polytope. The lemma follows from this observation, since $\mathscr{T}_{\min}$ is a minimum spanning tree. □

Lemmas 15 and 16 give the two bounds.

Lemma 15. $c(J) \leq \frac{1}{2}\{c(x^*) + c(s, t)\}$.

Proof. Let $x^*_{\text{circuit}} := x^* + \mathbf{e}_{(s,t)}$: i.e., x^*_{circuit} is obtained by "adding" the edge (s, t) to x^*. Then x^*_{circuit} is a feasible solution to the circuit-variant Held-Karp relaxation (see

(2.1) and (2.23)). Let $\mathrm{HK}_{\mathrm{circuit}}$ be the optimal value of the circuit-variant Held-Karp relaxation and we have

$$\begin{aligned} c(J) &\leq \frac{1}{2}\mathrm{HK}_{\mathrm{circuit}} \\ &\leq \frac{1}{2}c(x^*_{\mathrm{circuit}}) \\ &= \frac{1}{2}\left\{c(x^*) + c(s,t)\right\}, \end{aligned}$$

where the first inequality follows from [54, 53]. □

Lemma 16. *$c(J) \leq c(x^*) - c(s,t)$.*

Proof. Let $P_{st}^{\mathscr{T}_{\min}}$ be the path between s and t on $\mathscr{T}_{\min}$. Consider an edge set $J' := \mathscr{T}_{\min} \setminus P_{st}^{\mathscr{T}_{\min}}$. Note that J' is a T-join: $v \in V$ has even degree in $P_{st}^{\mathscr{T}_{\min}}$ if and only if v is internal; thus, v has even degree in the multigraph $\mathscr{T}_{\min} \cup J' = (\mathscr{T}_{\min} \cup \mathscr{T}_{\min}) \setminus P_{st}^{\mathscr{T}_{\min}}$ if and only if v is an internal point, and this shows that v has odd degree in J' if and only if $v \in T$.

We have

$$\begin{aligned} c(J) &\leq c(J') \\ &= c(\mathscr{T}_{\min}) - c(P_{st}^{\mathscr{T}_{\min}}) \\ &\leq c(x^*) - c(s,t). \end{aligned}$$

The last inequality follows from Lemma 14 and the triangle inequality. □

Theorem 9. *$c(H) \leq \frac{5}{3}c(x^*)$; therefore, the path-variant Christofides' algorithm is a 5/3-approximation algorithm, and the integrality gap of the path-variant Held-Karp relaxation is at most 5/3.*

Proof. We have

$$\begin{aligned}
c(H) &\leq c(\mathscr{T}_{\min}) + c(J) \\
&\leq c(x^*) + \min\left[\frac{1}{2}\left\{c(x^*) + c(s,t)\right\}, c(x^*) - c(s,t)\right] \\
&= \frac{5}{3}c(x^*) + \min\left[\frac{1}{2}\left\{-\frac{1}{3}c(x^*) + c(s,t)\right\}, \frac{1}{3}c(x^*) - c(s,t)\right] \\
&\leq \frac{5}{3}c(x^*), \qquad (2.24)
\end{aligned}$$

where the second inequality follows from Lemmas 14, 15 and 16. □

We observe that the equality of (2.24) is achieved when $c(s,t) = \frac{1}{3}c(x^*)$, and this is the critical case of this analysis that determines the performance guarantee proven. Hence, if we can improve the performance guarantee only near this critical case, such an improvement would lead to a better approximation ratio. We demonstrate this approach, by presenting how this analysis combines with the results of Oveis Gharan et al. [46] on the unit-weight graphical metric TSP to yield a comparable result in the *s*-*t* path TSP.

Now we consider the *s*-*t* path TSP under the unit-weight graphical metric; we show how to modify the algorithm of Oveis Gharan et al. for the path case and that, when $c(s,t)$ is close to $\frac{1}{3}c(x^*)$, this modified algorithm carries a performance guarantee that is slightly better than 5/3.

First we review the results in Oveis Gharan et al. [46]. In the following, the parameters $\epsilon_1, \epsilon_2, \gamma, \delta$ and ρ can be chosen as follows: $\epsilon_1 = 1.875 \cdot 10^{-12}$, $\epsilon_2 = 5 \cdot 10^{-2}$, $\gamma = 10^{-7}$, $\delta = 6.25 \cdot 10^{-16}$, $\rho = 1.5 \cdot 10^{-24}$, and n denotes $|V|$.

Definition 7 (Nearly integral edges)**.** *An edge e is* nearly integral *with respect to $x \in \mathbb{R}^E$ if $x_e \geq 1 - \gamma$.*

Definition 8. *For some constant* $\nu \leq \frac{1}{5}$ *and* $k \geq 2$, *a* maximum entropy distribution over spanning trees with approximate marginal $x \in \mathbb{R}^E$ *is a probability distribution* μ *defined by* $\lambda \in \mathbb{R}^E$ *such that* $\mu(\mathcal{T}) \propto \prod_{e \in T} \lambda_e$ *for every spanning tree* $\mathcal{T}$ *and the marginal probability of every edge* e *is no greater than* $(1 + \frac{\nu}{n^k})x_e$.

Definition 9 (Good edges). *A cut is* $(1+\delta)$*-near-minimum if its weight is at most* $(1+\delta)$ *times the minimum cut weight. An edge* e *is* even *with respect to* $F \subset E$ *if every* $(1+\delta)$*-near-minimum cut containing* e *has even number of edges intersecting with* F.

For a circuit-variant Held-Karp feasible solution x^*_{circuit}, *consider* x^*_{circuit} *as the edge weight and let* F *be a spanning tree sampled from a maximum entropy distribution with approximate marginal* $(1 - \frac{1}{n})x^*_{\text{circuit}}$. *We say an edge* e *is* good *with respect to* x^*_{circuit} *if the probability that* e *is even with respect to* F *is at least* ρ.

Theorem 10 (Structure Theorem). *Let* x^*_{circuit} *be a feasible solution to the circuit-variant Held-Karp relaxation, and let* μ *be a maximum entropy distribution over spanning trees with approximate marginal* $(1-\frac{1}{n})x^*_{\text{circuit}}$. *There exist small constants* $\epsilon_1, \epsilon_2 > 0$ *such that at least one of the following is true:*

1. *there exists a set* $E^* \subset E$ *such that* $x(E^*) \geq \epsilon_1 n$ *and every edge in* E^* *is good with respect to* x^*_{circuit};
2. *there exist at least* $(1 - \epsilon_2)n$ *edges that are nearly integral with respect to* x^*_{circuit}.

Lemma 17. *Suppose that Case 1 of Theorem 10 holds and* $\mathcal{T}$ *is sampled from* μ. *Let* T *be the set of odd-degree vertices in* $\mathcal{T}$, *then a minimum* T*-join* J *satisfies*

$$\mathrm{E}[c(J)] \leq c(x^*_{\text{circuit}})(\frac{1}{2} - \frac{\epsilon_1 \delta \rho}{4(1+\delta)}).$$

We are now ready to present the algorithm. Algorithm 2 describes the entire algorithm for the *s*-*t* path TSP under the unit-weight graphical metric. It first computes the LP optimum x^*. If $c(s,t)$ is close to $\frac{1}{3}c(x^*)$, we run a modified version of Oveis Gharan, Saberi, and Singh's algorithm (Cases A1 and A2); otherwise, we invoke Christofides' algorithm (Case B). Parameters σ_l, σ_u and ϵ_2' are to be chosen later.

First we show that we can have a Structure Theorem analogous to Theorem 10 by adjusting ϵ_2 and replacing n with $(n-1)$ in Case 2. The following corollary states that either there are good edges of significant weight with respect to x^*_{circuit} or there are many nearly integral edges with respect to x^*.

Corollary 11. *Let x^* be a feasible solution to the path-variant Held-Karp relaxation and $x^*_{\text{circuit}} := x^* + \mathbf{e}_{(s,t)}$. Let μ be a maximum entropy distribution over spanning trees with approximate marginal $(1-\frac{1}{n})x^*_{\text{circuit}}$. There exist small constants $\epsilon_1, \epsilon_2' > 0$ such that at least one of the following is true:*

1. *there exists a set $E^* \subset E$ such that $x(E^*) \geq \epsilon_1 n$ and every edge in E^* is good with respect to x^*_{circuit};*
2. *there exist at least $(1-\epsilon_2')(n-1)$ edges that are nearly integral with respect to x^*.*

Proof. By Theorem 10, at least one of the two cases of Theorem 10 holds. Case 1 of Theorem 10 and Case 1 of this corollary are identical, so consider when Case 2 of Theorem 10 holds.

Recall that ϵ_2 was chosen as $5 \cdot 10^{-2}$; we choose $\epsilon_2' = 6 \cdot 10^{-2}$.

Suppose $n \leq 19$. x^*_{circuit} has at least $(1-\epsilon_2)n$ nearly integral edges; thus, x^* has at least $\lceil (1-\epsilon_2)n \rceil - 1 = n-1 \geq (1-\epsilon_2')(n-1)$ nearly integral edges.

Algorithm 2: Algorithm for s-t path TSP under unit-weight graph. metric.

Input: Complete graph $G = (V, E)$ with cost $c : E \to \mathbb{Z}_{>0}$; endpoints $s, t \in V$.

Output: Hamiltonian path between s and t.

$x^* \leftarrow$optimal solution to the path-variant Held-Karp relaxation

if $c(s,t) = (\frac{1}{3} + \alpha)c(x^*)$ for $\alpha \in [-\sigma_l, \sigma_u]$ **then**

  if at least $(1 - \epsilon_2')(n-1)$ edges are nearly integral w.r.t. x^* **then** {Case A1}

    Find min spanning subgraph F' containing all nearly integral edges.

    Find min spanning tree $\mathscr{T}$ of F'.

    Let T be set of odd-degree internal pts and even-degree endpts in $\mathscr{T}$.

    Compute a minimum T-join J; $\mathscr{L} \leftarrow \mathscr{T} \cup J$.

  else {Case A2}

    $x^*_{\text{circuit}} := x^* + \mathbf{e}_{(s,t)}$

    Sample a spanning tree $\mathscr{T}$ from the maximum entropy distribution with approximate marginal $(1 - \frac{1}{n})x^*_{\text{circuit}}$.

    Let T be the set of odd-degree vertices in $\mathscr{T}$.

    Compute a minimum T-join J; $\mathscr{L}_0 \leftarrow \mathscr{T} \cup J$.

    if $(s,t) \in \mathscr{L}_0$ **then** $\mathscr{L} \leftarrow \mathscr{L}_0 \setminus \{(s,t)\}$ **else** $\mathscr{L} \leftarrow \mathscr{L}_0 \cup \{(s,t)\}$ **end if**

  end if

else {Case B}

  Find a minimum spanning tree $\mathscr{T}$ of G.

  Let T be set of odd-degree internal points and even-degree endpts in $\mathscr{T}$.

  Compute a minimum T-join J; $\mathscr{L} \leftarrow \mathscr{T} \cup J$.

end if

Shortcut Eulerian path of multigraph $\mathscr{L}$ to obtain Ham. path H; output it.

Suppose $n \geq 20$. x^* has at least

$$\begin{aligned}(1-\epsilon_2)n - 1 &= (1-\epsilon_2)(n-1) - \epsilon_2 \\ &\geq (1 - \frac{20}{19}\epsilon_2)(n-1) \\ &\geq (1-\epsilon_2')(n-1)\end{aligned}$$

nearly integral edges. □

Lemma 18. *In Case A1, $c(H) \leq (\frac{5}{3} - C_{A1})c(x^*)$ for some $c_{A1} > 0$.*

Proof. The following proof is adapted from [46] and modified for the path case.

Let S' be the set of nearly integral edges. Since the metric is defined by an unweighted connected graph, $c(F') = c(S') + |F' \setminus S'| \leq \frac{(c*x^*)(S')}{1-\gamma} + |F' \setminus S'|$. From $\gamma < \frac{1}{3}$, we know that S' is a union of disjoint cycles and paths and the lengths of cycles are at least $\frac{1}{\gamma}$. Thus, $|\mathscr{T} \cap S'| \geq (n-1)(1-\epsilon_2')(1-\gamma)$ and $|\mathscr{T} \setminus S'| \leq (n-1)(\epsilon_2' + \gamma) \leq c(x^*)(\epsilon_2' + \gamma)$. Let $S = S' \cap \mathscr{T}$.

We construct a fractional T-join dominator y as follows.

$$y_e = \begin{cases} 1 & \text{if } e \in \mathscr{T} \setminus S \\ x_e^* & \text{if } e \in E \setminus \mathscr{T} \\ \frac{x_e^*}{2(1-\gamma)} & \text{if } e \in S \end{cases}$$

We claim that y is a fractional T-join dominator. Let $(U, \bar{U})$ be any cut that has an odd number of vertices in T on one side. If there exists an edge $e \in (\mathscr{T} \setminus S) \cap \delta(U)$, then $y(\delta(U)) \geq y_e = 1$. So suppose from now on that $\delta(U) \cap \mathscr{T} \subset S$. Then $\delta(U) \cap S = \delta(U) \cap \mathscr{T}$.

If U is nonseparating, U contains odd number of odd-degree vertices, and thus $|\delta(U) \cap \mathscr{T}|$ is odd. We have $x^*(\delta(U)) \geq 2$ from the Held-Karp formulation and thus

$$\begin{cases} y(\delta(U)) \geq x^*(\delta(U) \setminus \mathscr{T}) \geq 1 & \text{if } |\delta(U) \cap \mathscr{T}| = 1 \\ y(\delta(U)) \geq y(\delta(U) \cap S) \geq 3\frac{1-\gamma}{2(1-\gamma)} > 1 & \text{if } |\delta(U) \cap S| \geq 3. \end{cases}$$

If $(U, \bar{U})$ is an s-t cut, then U contains even number of odd-degree vertices, and thus $|\delta(U) \cap \mathscr{T}|$ is even. We have $(\delta(U) \cap \mathscr{T}) \neq \emptyset$ since $\mathscr{T}$ is connected and

$$y(\delta(U)) \geq y(\delta(U) \cap S) \geq 2\frac{1-\gamma}{2(1-\gamma)} = 1.$$

Thus y is a fractional T-join dominator. Now,

$$\begin{aligned} c(H) &\leq c(\mathscr{T}) + c(y) \\ &\leq \frac{(c * x^*)(S)}{1-\gamma} + c(\mathscr{T} \setminus S) + c(\mathscr{T} \setminus S) + (c * x^*)(E \setminus \mathscr{T}) + \frac{(c * x^*)(S)}{2(1-\gamma)} \\ &\leq \frac{3(c * x^*)(S)}{2(1-\gamma)} + 2c(x^*)(\epsilon_2' + \gamma) + (c * x^*)(E \setminus S) \\ &\leq c(x^*)(\frac{3}{2(1-\gamma)} + 2\epsilon_2' + 2\gamma) \\ &\leq c(x^*)(\frac{5}{3} - C_{A1}) \end{aligned}$$

for some $C_{A1} > 0$. For example, we can choose $c_{A1} = 4 \cdot 10^{-2}$. □

Lemma 19. *In Case A2,* $\mathrm{E}[c(H)] \leq (\frac{5}{3} - C_{A2})c(x^*)$ *for some* $C_{A2} > 0$*.*

Proof. First we have

$$\begin{aligned} \mathrm{E}[c(\mathscr{T})] &\leq c\left((1 + \frac{\nu}{n^k})(1 - \frac{1}{n})x^*_{\text{circuit}}\right) \\ &\leq (1 + \frac{1}{5n^2})(1 - \frac{1}{n})(\frac{4}{3} + \alpha)c(x^*) \\ &\leq (1 - \frac{4}{5n})(\frac{4}{3} + \alpha)c(x^*). \end{aligned}$$

From Lemma 17,

$$\mathrm{E}[c(J)] \leq (\frac{4}{3} + \alpha)c(x^*)(\frac{1}{2} - \frac{\epsilon_1 \delta \rho}{4(1+\delta)}).$$

We have

$$\begin{aligned}
\Pr[(s,t) \in L_0] &\geq \Pr[(s,t) \in \mathscr{T}] \\
&= n - 1 - \mathrm{E}[|\mathscr{T} \setminus (s,t)|] \\
&\geq n - 1 - (n - 2 + \frac{1}{n})(1 + \frac{\nu}{n^k}) \\
&\geq n - 1 - (n - 2 + \frac{1}{n})(1 + \frac{1}{5n^2}) \\
&\geq 1 - \frac{7}{5n}
\end{aligned}$$

and hence

$$\begin{aligned}
\mathrm{E}[c(H)] &\leq \mathrm{E}[c(\mathscr{T})] + \mathrm{E}[c(J)] - (1 - \frac{7}{5n})c(s,t) + \frac{7}{5n}c(s,t) \\
&\leq c(x^*)\left\{(1 - \frac{4}{5n})(\frac{4}{3} + \alpha) + (\frac{4}{3} + \alpha)(\frac{1}{2} - \frac{\epsilon_1\delta\rho}{4(1+\delta)}) \right. \\
&\qquad \left. -(1 - \frac{7}{5n})(\frac{1}{3} + \alpha) + \frac{7}{5n}(\frac{1}{3} + \alpha)\right\} \\
&= c(x^*)\left\{(\frac{5}{3} - \frac{\epsilon_1\delta\rho}{3(1+\delta)}) + \alpha(\frac{1}{2} - \frac{\epsilon_1\delta\rho}{4(1+\delta)}) - \frac{1}{n}(\frac{2}{15} - 2\alpha)\right\} \\
&\leq c(x^*)(\frac{5}{3} - C_{A2})
\end{aligned}$$

for some $C_{A2} > 0$ by choosing sufficiently small $\sigma_l, \sigma_u > 0$. For example, we can choose $\sigma_l = 7.8 \cdot 10^{-52}$, $\sigma_u = 3.9 \cdot 10^{-52}$ and $C_{A2} = 3.9 \cdot 10^{-52}$. □

Lemma 20. *In Case B, $c(H) \leq (\frac{5}{3} - C_B)c(x^*)$ for some $C_B > 0$.*

Proof. Suppose that $c(s,t) < (\frac{1}{3} - \sigma_l)c(x^*)$. From Lemmas 14 and 15, it follows that

$$\begin{aligned}
c(H) &\leq c(\mathscr{T}) + c(J) \\
&< c(x^*) + \frac{1}{2}\left\{c(x^*) + (\frac{1}{3} - \sigma_l)c(x^*)\right\} \\
&= \left(\frac{5}{3} - \frac{\sigma_l}{2}\right)c(x^*).
\end{aligned}$$

Suppose $c(s,t) > (\frac{1}{3} + \sigma_u)c(x^*)$. From Lemmas 14 and 16,

$$\begin{aligned} c(H) &\leq c(\mathscr{T}) + c(J) \\ &< c(x^*) + \left\{ c(x^*) - (\frac{1}{3} + \sigma_u)c(x^*) \right\} \\ &= \left(\frac{5}{3} - \sigma_u \right) c(x^*). \end{aligned}$$

Now choose $C_B := \min(\frac{\sigma_l}{2}, \sigma_u)$. □

Lemmas 18, 19 and 20 yield the following theorem.

Theorem 11. *For some $\epsilon > 0$, Algorithm 2 is a $(\frac{5}{3} - \epsilon)$-approximation algorithm for the s-t path TSP under the unit-weight graphical metric.*

Proof. In Cases A1 and B, the multigraph $\mathscr{L}$ is the union of a spanning tree and a T-join where T is the set of the vertices with the wrong parity of degree. Thus, $\mathscr{L}$ has an Eulerian path between the two endpoints.

In Case A2, $\mathscr{L}_0$ is Eulerian and hence 2-edge-connected; $\mathscr{L} \supset \mathscr{L}_0 \setminus \{(s,t)\}$ is therefore connected and $\mathscr{L}$ has an Eulerian path between the two endpoints.

By choosing $\epsilon = \min\{C_{A1}, C_{A2}, C_B\}$, $\epsilon = 3.9 \cdot 10^{-52}$ for example, we have $\mathrm{E}[c(H)] \leq (\frac{5}{3} - \epsilon)c(x^*)$ from Lemmas 18, 19 and 20. Thus, Algorithm 2 is a $(\frac{5}{3} - \epsilon)$-approximation algorithm. □

2.4 Open Questions

An immediate open question is in improving the performance guarantee. The fractional T-join dominators constructed in the analyses are not directly derived from the algorithm; a different construction may lead to an improved performance guarantee. In fact, the worst input we know of is from computational experiments, and the algorithm produces only a 1.1255-approximate solution for this input. One related question is whether α and β can be chosen differently; yet, the following two theorems show that certain types of approaches to an improved analysis are not promising. As Lemma 3 can be interpreted as distributing $c(x^*)$ over the cuts of different capacities, we can consider an algorithm that first finds $\hat{f}_U^*$ for the 1-narrow cuts and adaptively choose α and β depending on this distribution specified. Another possible algorithm obliviously but randomly chooses α and β; this will be analyzed with respect to the same distribution.

Theorem 12. *Two approaches are equally powerful.*

Proof. Let $\rho(\alpha, \beta, \omega) := 1 + \alpha + \beta + \max\left[\omega\{1 - (2\alpha + \beta(1 + \omega))\}, 0\right]$. This corresponds to the performance guarantee when the cut capacity of every 1-narrow cut is $1 + \omega$.

The first approach can be reinterpreted as follows: the adversary chooses a distribution of ω (requiring the total probability to be 1 does not harm the adversary); then we choose α and β, and the performance guarantee is given as $\mathrm{E}_\omega[\rho(\alpha, \beta, \omega)]$. Cuts of different capacities linearly contribute to the performance guarantee, so this can be easily verified.

In the second approach, we choose a distribution over α and β and then the adversary chooses ω; the performance guarantee is $\mathrm{E}_{\alpha,\beta}[\rho(\alpha, \beta, \omega)]$.

Now the claim follows from Yao's lemma. □

Theorem 13. *In the reinterpreted version of the first approach,*

$$\max_{\Omega \in \Delta[0,1)} \min_{\alpha,\beta} \mathrm{E}_{\omega \sim \Omega}[\rho(\alpha,\beta,\omega)] \geq \frac{1+\sqrt{5}}{2}.$$

Proof. Suppose the adversary chooses $\omega = \frac{3-\sqrt{5}}{2}$ with weight (probability) 1. Then

$$\begin{aligned}
\rho(\alpha,\beta,\omega) &= 1+\alpha+\beta+\frac{3-\sqrt{5}}{2}\left\{1-(2\alpha+\beta(1+\frac{3-\sqrt{5}}{2})\right\} \\
&= \frac{5-\sqrt{5}}{2}+(-2+\sqrt{5})\alpha+(-4+2\sqrt{5})\beta \\
&\geq \frac{5-\sqrt{5}}{2}+(-2+\sqrt{5})(1-2\beta)+(-4+2\sqrt{5})\beta \\
&= \frac{1+\sqrt{5}}{2},
\end{aligned}$$

where the third line follows from $-2+\sqrt{5}>0$ and $\alpha \geq 1-2\beta$. □

A bigger open question is whether the techniques presented in this chapter can be extended to the circuit variant as well. Given the successful adaptation of the techniques devised in one variant to the other in the unit-weight graphical metric case, whether the present techniques can be extended to beat the long-standing 3/2 barrier of the general-metric circuit problem becomes an interesting question. It appears that the layered structure of τ-narrow cuts or the parity argument on them is less likely to directly extend to the circuit case, as the arguments rely on the characteristics of the path case; what could be more promising is the approach of repairing deficient cuts using a set of vectors obtained from an auxiliary flow network, since this approach might extend to work with some different type of "fragile cut structure".

2.5 Computational Evaluation

In this section, we computationally evaluate the performance of our algorithm. The optimal solution and the Held-Karp solution were computed using modified Concorde-03.12.19 [4] with IBM ILOG CPLEX Optimization Studio V12.4; the Held-Karp solution was decomposed into a convex combination of spanning trees by solving the following linear program using CPLEX with the column-generation method, where x^* is the Held-Karp solution and $\mathscr{S}$ is the set of spanning trees:

$$\begin{aligned} \text{minimize} \quad & \sum_{e \in E} s_e \\ \text{subject to} \quad & s_e + \sum_{T : e \in T, T \in \mathscr{S}} y_T = x^*_e, \quad \forall e \in E; \\ & s, y \geq 0. \end{aligned}$$

Computations have been performed for instances from TSPLIB [50] with up to 1000 vertices. For each dataset, three sets of endpoints have been considered: the furthest pair of vertices, the closest, and a random pair. For dataset lin318 which originally came from a path problem, the original pair of vertices has been considered instead of a random pair.

Table 2.1 summarizes the cost of the solution found by the original path-variant Christofides' algorithm and our algorithm. In 175 out of 189 cases, our algorithm produced a strictly better result compared to the original Christofides' algorithm; in fact, in 104 cases, the difference of our algorithm's cost from the optimum was less than half of the difference of Christofides' cost from the optimum.

Table 2.1: Computational results.

Dataset	Endpoints	Optimum	Original Christofides'		Our algorithm	
			Cost	Ratio	Cost	Ratio
burma14	6, 12	3304	3529	1.0681	3304	1.0000
	5, 10	2615	2799	1.0704	2615	1.0000
	8, 13	3050	3584	1.1751	3050	1.0000
ulysses16	13, 14	6807	7022	1.0316	6807	1.0000
	2, 11	4919	5062	1.0291	4919	1.0000
	4, 5	6050	6365	1.0521	6050	1.0000
ulysses22	20, 21	6999	7273	1.0391	6999	1.0000
	2, 11	5149	5232	1.0161	5149	1.0000
	8, 15	6507	6719	1.0326	6507	1.0000
bayg29	10, 20	1585	1670	1.0536	1585	1.0000
	3, 7	1473	1603	1.0883	1481	1.0054
	23, 25	1566	1617	1.0326	1566	1.0000
bays29	10, 20	1992	2104	1.0562	2011	1.0095
	3, 7	1804	1885	1.0449	1811	1.0039
	23, 25	1937	2165	1.1177	1951	1.0072
att48	19, 37	10586	11707	1.1059	10611	1.0024
	4, 17	9788	10862	1.1097	9863	1.0077
	4, 21	10065	10584	1.0516	10065	1.0000
eil51	46, 51	424	470	1.1085	454	1.0708
	36, 40	403	437	1.0844	408	1.0124
	48, 49	417	475	1.1391	440	1.0552
berlin52	35, 36	7527	8141	1.0816	7527	1.0000
	2, 52	7209	7670	1.0639	7741	1.0738
	25, 40	7539	7875	1.0446	7593	1.0072
st70	21, 34	674	739	1.0964	710	1.0534
	55, 64	631	696	1.1030	637	1.0095
	13, 22	668	749	1.1213	716	1.0719
eil76	34, 46	536	565	1.0541	541	1.0093
	55, 60	531	563	1.0603	563	1.0603
	8, 65	530	564	1.0642	530	1.0000
pr76	72, 73	107859	115277	1.0688	110656	1.0259
	70, 74	97584	101712	1.0423	102341	1.0487
	29, 48	108398	116239	1.0723	111037	1.0243
gr96	56, 57	55204	58771	1.0646	58931	1.0675
	3, 95	53628	58428	1.0895	54568	1.0175
	28, 57	54595	58648	1.0742	57916	1.0608

Dataset	Endpoints	Optimum	Original Christofides' Cost	Original Christofides' Ratio	Our algorithm Cost	Our algorithm Ratio
rat99	39, 48	1207	1319	1.0928	1247	1.0331
	1, 98	1203	1313	1.0914	1265	1.0515
	18, 19	1220	1337	1.0959	1295	1.0615
kroA100	15, 17	21269	23490	1.1044	21851	1.0274
	26, 41	20701	22932	1.1078	22500	1.0869
	53, 100	21102	23180	1.0985	22825	1.0817
kroB100	49, 86	22115	23897	1.0806	23158	1.0472
	51, 67	21059	23033	1.0937	21450	1.0186
	53, 100	21988	23095	1.0503	22488	1.0227
kroC100	3, 73	20731	22599	1.0901	22144	1.0682
	58, 87	20657	22348	1.0819	22439	1.0863
	53, 100	20638	22130	1.0723	21964	1.0643
kroD100	26, 87	21282	23102	1.0855	21780	1.0234
	29, 48	20957	23400	1.1166	21426	1.0224
	53, 100	21224	22720	1.0705	21638	1.0195
kroE100	15, 20	22046	23395	1.0612	24723	1.1214
	3, 11	21275	22889	1.0759	21275	1.0000
	53, 100	21724	22714	1.0456	24450	1.1255
rd100	46, 50	7905	8730	1.1044	8180	1.0348
	25, 70	7776	8807	1.1326	8272	1.0638
	53, 100	7738	8825	1.1405	7946	1.0269
eil101	37, 98	628	685	1.0908	628	1.0000
	38, 65	615	656	1.0667	646	1.0504
	23, 52	624	689	1.1042	646	1.0353
lin105	1, 2	14348	15752	1.0979	14668	1.0223
	1, 100	14074	15186	1.0790	14464	1.0277
	48, 100	14201	15238	1.0730	14404	1.0143
pr107	1, 3	44147	47855	1.0840	44147	1.0000
	52, 65	39544	43446	1.0987	39626	1.0021
	46, 73	40196	43512	1.0825	40529	1.0083
pr124	41, 42	58880	61377	1.0424	60127	1.0212
	13, 124	57690	60842	1.0546	61539	1.0667
	69, 92	58522	62629	1.0702	58810	1.0049
bier127	4, 22	118166	126924	1.0741	123114	1.0419
	98, 99	110393	117316	1.0627	111444	1.0095
	74, 100	117411	127178	1.0832	123242	1.0497

Dataset	Endpoints	Optimum	Original Christofides'		Our algorithm	
			Cost	Ratio	Cost	Ratio
ch130	12, 87	6109	6494	1.0630	6349	1.0393
	66, 111	5890	6486	1.1012	6080	1.0323
	28, 106	5974	6399	1.0711	6027	1.0089
pr136	13, 22	96602	104321	1.0799	99138	1.0263
	4, 119	94829	101779	1.0733	99233	1.0464
	38, 72	95646	103639	1.0836	97911	1.0237
gr137	123, 124	69784	74993	1.0746	72000	1.0318
	1, 108	66348	72027	1.0856	71022	1.0704
	51, 71	69679	75341	1.0813	71418	1.0250
pr144	3, 4	58437	64213	1.0988	59583	1.0196
	13, 144	55003	65553	1.1918	55003	1.0000
	32, 46	56160	65786	1.1714	57485	1.0236
ch150	49, 147	6526	7045	1.0795	6705	1.0274
	17, 82	6435	6960	1.0816	6767	1.0516
	86, 148	6504	6951	1.0687	6732	1.0351
kroA150	15, 17	26511	28535	1.0763	27223	1.0269
	41, 129	25950	27882	1.0745	27483	1.0591
	86, 148	26459	28436	1.0747	27658	1.0453
kroB150	51, 127	26122	29127	1.1150	27677	1.0595
	58, 87	25447	28295	1.1119	26274	1.0325
	86, 148	25975	28017	1.0786	27845	1.0720
pr152	63, 64	73607	75681	1.0282	75446	1.0250
	9, 152	63582	73813	1.1609	63763	1.0028
	14, 32	73377	76414	1.0414	75927	1.0348
u159	7, 152	41980	45191	1.0765	42834	1.0203
	26, 105	42614	46723	1.0964	45923	1.0777
	19, 122	41752	46084	1.1038	44181	1.0582
rat195	36, 37	2317	2569	1.1088	2384	1.0289
	1, 195	2260	2421	1.0712	2260	1.0000
	28, 41	2308	2489	1.0784	2382	1.0321
d198	67, 70	15757	16644	1.0563	16041	1.0180
	1, 193	12753	13557	1.0630	13174	1.0330
	122, 154	15648	16429	1.0499	16249	1.0384
kroA200	10, 175	29358	32275	1.0994	30622	1.0431
	87, 141	28799	32336	1.1228	30101	1.0452
	136, 198	29283	31928	1.0903	30472	1.0406

Dataset	Endpoints	Optimum	Original Christofides'		Our algorithm	
			Cost	Ratio	Cost	Ratio
kroB200	91, 129	29432	31582	1.0730	31743	1.0785
	51, 198	28927	31523	1.0897	29736	1.0280
	136, 198	29323	31328	1.0684	30977	1.0564
gr202	109, 110	40152	42113	1.0488	40389	1.0059
	76, 202	36978	39144	1.0586	37315	1.0091
	126, 136	40063	42365	1.0575	40220	1.0039
ts225	1, 2	126143	131426	1.0419	129444	1.0262
	1, 125	126799	132622	1.0459	130697	1.0307
	73, 86	127224	132804	1.0439	131099	1.0305
tsp225	28, 204	3909	4189	1.0716	4202	1.0750
	21, 178	3830	4091	1.0681	4007	1.0462
	73, 86	3893	4301	1.1048	4030	1.0352
pr226	2, 3	80269	87922	1.0953	82502	1.0278
	1, 155	79277	90029	1.1356	80897	1.0204
	95, 224	77066	88183	1.1443	78241	1.0152
gr229	194, 195	134558	141997	1.0553	141695	1.0530
	77, 228	131988	139587	1.0576	138686	1.0507
	147, 183	134323	140556	1.0464	143562	1.0688
gil262	5, 133	2377	2571	1.0816	2463	1.0362
	1, 159	2366	2522	1.0659	2497	1.0554
	49, 224	2370	2575	1.0865	2515	1.0612
pr264	1, 40	49035	51222	1.0446	49339	1.0062
	52, 261	45235	48672	1.0760	45973	1.0163
	62, 127	48729	50810	1.0427	49077	1.0071
a280	171, 172	2579	2738	1.0617	2706	1.0492
	1, 96	2583	2732	1.0577	2744	1.0623
	47, 118	2565	2789	1.0873	2686	1.0472
pr299	37, 38	48190	51395	1.0665	50664	1.0513
	54, 298	47278	52484	1.1101	50295	1.0638
	128, 248	47995	52116	1.0859	50429	1.0507
lin318	1, 2	41998	46148	1.0988	42928	1.0221
	1, 310	41432	44437	1.0725	42379	1.0229
	1, 214	41345	45644	1.1040	41962	1.0149
rd400	172, 267	15280	16605	1.0867	15860	1.0380
	346, 374	15190	16555	1.0899	15716	1.0346
	38, 207	15189	16483	1.0852	16010	1.0541

Dataset	Endpoints	Optimum	Original Christofides'		Our algorithm	
			Cost	Ratio	Cost	Ratio
fl417	2, 4	11853	12772	1.0775	12111	1.0218
	176, 400	11506	12445	1.0816	11788	1.0245
	211, 308	11554	12518	1.0834	11966	1.0357
gr431	109, 110	171406	179729	1.0486	179867	1.0494
	14, 417	169644	179981	1.0609	173983	1.0256
	138, 147	171207	179556	1.0488	179505	1.0485
pr439	371, 372	107127	115453	1.0777	109547	1.0226
	133, 434	104813	114079	1.0884	108637	1.0365
	147, 169	106888	115726	1.0827	110534	1.0341
pcb442	32, 376	50728	52774	1.0403	51652	1.0182
	375, 442	50287	52286	1.0398	52646	1.0469
	154, 183	50922	52775	1.0364	52697	1.0349
d493	9, 10	34984	37041	1.0588	36244	1.0360
	1, 94	32822	34860	1.0621	34346	1.0464
	69, 200	34825	37361	1.0728	36069	1.0357
att532	227, 228	27685	30200	1.0908	29142	1.0526
	1, 489	27421	29764	1.0854	28996	1.0574
	184, 456	27591	30174	1.0936	29473	1.0682
ali535	32, 459	202338	220111	1.0878	210821	1.0419
	15, 19	202027	222963	1.1036	212378	1.0512
	342, 502	202278	223422	1.1045	210700	1.0416
u574	490, 492	36902	40195	1.0892	38612	1.0463
	130, 439	36286	39686	1.0937	38560	1.0627
	44, 344	36551	40041	1.0955	38412	1.0509
rat575	372, 395	6771	7352	1.0858	7232	1.0681
	23, 553	6749	7341	1.0877	7186	1.0648
	147, 277	6765	7406	1.0948	7141	1.0556
p654	107, 108	34628	38555	1.1134	35007	1.0109
	16, 106	33090	35666	1.0778	33685	1.0180
	98, 654	33574	36228	1.0790	34792	1.0363
d657	33, 34	48895	53231	1.0887	51394	1.0511
	1, 624	47233	50555	1.0703	49475	1.0475
	413, 459	48884	53199	1.0883	51206	1.0475
gr666	194, 195	294353	313972	1.0667	309813	1.0525
	1, 666	290030	310551	1.0708	303428	1.0462
	72, 404	294029	316413	1.0761	309654	1.0531

Dataset	Endpoints	Optimum	Original Christofides'		Our algorithm	
			Cost	Ratio	Cost	Ratio
u724	368, 372	41907	46030	1.0984	43848	1.0463
	1, 425	41786	45803	1.0961	44175	1.0572
	428, 592	41883	45344	1.0826	44584	1.0645
rat783	24, 28	8805	9509	1.0800	9176	1.0421
	6, 778	8811	9534	1.0821	9353	1.0615
	449, 666	8805	9553	1.0850	9215	1.0466
dsj1000	637, 983	18659508	20250950	1.0853	19551556	1.0478
	439, 895	18328805	19902043	1.0858	19249871	1.0503
	572, 652	18631011	20182917	1.0833	19631049	1.0537

CHAPTER 3

APPROXIMATION ALGORITHMS FOR THE BOTTLENECK ASYMMETRIC TSP

In this chapter, we present an $O(\log n/\log\log n)$-approximation algorithm for the bottleneck asymmetric TSP.

For any bottleneck problem, one can immediately reduce the optimization problem with cost data to a more combinatorially defined question, since there is the trivial relationship that the optimal bottleneck solution is of objective function value at most T if and only if there exists a feasible solution that uses only those edges of cost at most T. Furthermore, there are only a polynomial number of potential thresholds T, and so a polynomial-time algorithm that answers this purely combinatorial decision question leads to a polynomial-time optimization algorithm. Similarly, for a ρ-approximation algorithm, it is sufficient for the algorithm to solve a "relaxed" decision question: either provide some certificate that no feasible solution exists, or produce a solution in which each edge used is of cost at most ρT. If G denotes the graph of all edges of cost at most T, then the triangle inequality implies that it is sufficient to find feasible solutions within G^ρ, the ρth power of G, in which we include an edge (u, v) whenever G contains a path from u to v with at most ρ edges. In the context of the TSP, this means that we either want to prove that G is not Hamiltonian, or else to produce a Hamiltonian cycle within, for example, the square of G (to yield a 2-approximation algorithm as in [40, 49]).

Unfortunately, the techniques invented in the context of the min-sum problem do not seem to be amenable to bottleneck objective function. For example, the analysis of the $O(\log n)$-approximation algorithm for the min-sum asymmetric TSP due to Kleinberg and Williamson [38] depends crucially on the monotonicity of the optimal value over the vertex-induced subgraphs, and the fact that shortcutting a circuit does not increase the objective. That fact clearly is not true in the bottleneck setting: shortcutting arbitrary subpaths of a circuit may result in a tour that is valid only in a higher-order power graph. The aforementioned monotonicity is also lost as it relies on this fact as well.

In order to resolve this difficulty, we devise a condition on Eulerian circuits under which we can limit the lengths of the paths that are shortcut to obtain a Hamiltonian cycle. We will present a polynomial-time constructive proof of this condition using Hall's Transversal Theorem [30]; this proof is directly used in the algorithm. One of the special cases of the condition particularly worth mentioning is a degree-bounded spanning circuit (equivalently, an Eulerian spanning subgraph of bounded degree). If there exists a bound k on the number of occurrences of any vertex in a spanning circuit, our theorem provides a bound of $2k - 1$ on the length of the shortcut paths.

We will then show how thin trees defined in Asadpour et al. [8] can be used to compute these degree-bounded spanning circuits. An α-thin tree with respect to a weighted graph G is a unit-weighted spanning tree of G whose cut weights are no more than α times the corresponding cut weights of G. The min-sum algorithm due to Asadpour et al. [8] augments an $O(\log n / \log \log n)$-thin tree with respect to a (scaled) directed Held-Karp solution (Held and Karp [31]) into a spanning Eulerian graph by solving a circulation problem. The directed Held-Karp relaxation consists of the equality constraints on the in- and out-degree of each vertex and the inequality constraints on the directed cut weights: the equality constraints set the degrees to one, and the inequality constraints ensure that the total weight of edges leaving S is at least 1 for each subset S. We introduce vertex capacities to the circulation problem to impose the desired degree bound without breaking the feasibility of the circulation problem. This leads to an algorithm that computes degree-bounded spanning circuits with an $O(\log n/\log\log n)$ bound.

Oveis Gharan and Saberi [45] gave an $O(1)$-approximation algorithm for the min-sum asymmetric TSP when the support of the Held-Karp solution can be embedded on an orientable surface with a bounded genus. They achieved this by showing how to extract an $O(1)$-thin tree in this special case. Our result can be combined with this to yield an $O(1)$-approximation algorithm for the bottleneck asymmetric TSP when the support of the Held-Karp solution has a bounded orientable genus. Chekuri, Vondrák, and Zenklusen [14] showed that an alternative sampling procedure can be used to find the thin tree in Asadpour et al. [8].

3.1 Preliminaries

We introduce some notation and review previous results in this section. Some notation was adopted from Asadpour et al. [8]

Let $G = (V, A)$ be a digraph and E be the underlying undirected edge set: $\{u, v\} \in E$ if and only if $\langle u, v\rangle \in A$ or $\langle v, u\rangle \in A$. For $S \subset V$, let

$$\begin{aligned}
\delta^{+}(S) &:= \{\langle u, v\rangle \in A \mid u \in S, v \notin S\}, \\
\delta^{-}(S) &:= \delta^{+}(V \setminus S), \\
\delta(S) &:= \{\{u, v\} \in E \mid |E \cap S| = 1\};
\end{aligned}$$

for $v \in V$,

$$\begin{aligned}\delta^+(v) &:= \delta^+(\{v\}),\\ \delta^-(v) &:= \delta^-(\{v\}),\\ \delta(v) &:= \delta(\{v\});\end{aligned}$$

for $B \subset A$ and $x \in \mathbb{R}^A$,

$$x(B) := \sum_{b \in B} x_b;$$

and similarly, for $F \subset E$ and $z \in \mathbb{R}^E$,

$$z(F) := \sum_{f \in F} z_f.$$

We need a notion of the non-Hamiltonicity certificate to solve the "relaxed" decision problem. We establish this certificate by solving the Held-Karp relaxation ([31]) in our algorithm. The Held-Karp relaxation to the asymmetric traveling salesman problem is the following linear program (we do not define an objective here):

$$\begin{cases} x(\delta^+(v)) = x(\delta^-(v)) = 1 & \forall v \in V \\ x(\delta^+(S)) \geq 1 & \forall S \subsetneq V, S \neq \emptyset \\ x \geq 0. \end{cases} \tag{3.1}$$

A graph is non-Hamiltonian if (3.1) is infeasible. This linear program can be solved in polynomial time [28].

A thin tree is defined as follows in Asadpour et al. [8].

Definition 10. *A spanning tree T is α-thin with respect to $z^* \in \mathbb{R}^E$ if $|T \cap \delta(U)| \leq \alpha z^*(\delta(U))$ for all $U \subset V$.*

Asadpour et al. [8] then prove Theorem 14: they show the thinness for $z^*_{uv} := \frac{n-1}{n}(x^*_{uv} + x^*_{vu})$ where $n = |V|$, and Theorem 14 is only weaker.

Theorem 14. *There exists a probabilistic algorithm that, given an extreme point solution $x^* \in \mathbb{R}^A$ to the Held-Karp relaxation, produces an α-thin tree T with respect to $z^*_{uv} := x^*_{uv} + x^*_{vu}$ with high probability, for $\alpha = \frac{4 \ln n}{\ln \ln n}$.*

Let $T_{\rightarrow}$ be a directed version of T, obtained by choosing the arcs in the support of x^*. If arcs exist in both directions, an arbitrary choice can be made. Consider a circulation problem on G: recall that the circulation problem requires, given a lower and upper bound on each arc, a set of flow values on arcs such that the sum of the incoming flows at every vertex matches the sum of outgoing, while honoring both bounds imposed on each arc. When all of the bounds are integers, an integral solution can be found in polynomial time unless the problem is infeasible [51]. Here we consider an instance where the lower bounds l and upper bounds u on the arcs are given as follows:

$$\begin{aligned} l(e) &= \begin{cases} 1 & \text{if } e \in T_{\rightarrow} \\ 0 & \text{otherwise} \end{cases} \\ u(e) &= \begin{cases} 1 + 2\alpha x^*_e & \text{if } e \in T_{\rightarrow} \\ 2\alpha x^*_e & \text{otherwise.} \end{cases} \end{aligned} \tag{3.2}$$

Asadpour et al. [8] show that this problem is feasible; the existence of an integral circulation under the rounded-up bounds follows from that.

Lemma 21. *The circulation problem defined by* (3.2) *is feasible.*

3.2 Algorithm

This section gives the $O(\frac{\log n}{\log \log n})$-approximation algorithm to the bottleneck asymmetric traveling salesman problem and its analysis. We present the lemmas to bound the lengths of the paths that are shortcut in the process of transforming a spanning circuit into a Hamiltonian cycle; we also show how a degree-bounded spanning circuit can be constructed.

Lemma 22. *Let $v_1, \ldots, v_m, v_1$ be a (non-simple) circuit that visits every vertex at least once. Partition $v_1, \ldots, v_m$ into contiguous subsequences of length k, except for the final subsequence whose length may be less than k. Denote the pieces of this partition by $P_1, \ldots, P_\ell$. If, for all t, the union of any t sets in $\{P_1, \ldots, P_\ell\}$ contains at least t distinct vertices, G^{2k-1} is Hamiltonian.*

Proof. From Hall's Transversal Theorem [30], if the given condition holds, $\{P_1, \ldots, P_\ell\}$ has a transversal: i.e., we can choose one vertex from each piece P_i such that no vertex is chosen more than once. If we take any subsequence of $v_1, \ldots, v_m$ that contains every vertex exactly once and includes all of the vertices in the transversal, this subsequence is a Hamiltonian cycle in G^{2k-1}. This is because any two contiguous vertices chosen in the transversal are at most $2k - 1$ arcs away. Since a transversal can be found in polynomial time (see Kleinberg and Tardos [37]), a Hamiltonian cycle can be constructed in polynomial time as well. □

Lemma 23 shows that a degree-bounded spanning circuit forms a special case of Lemma 22.

Lemma 23. *Given a circuit on G that visits every vertex at least once and at most k times, a Hamiltonian cycle on G^{2k-1} can be found in polynomial time.*

Proof. Consider $\{P_1, \ldots, P_\ell\}$ as defined in Lemma 22. For any t sets in $\{P_1, \ldots, P_\ell\}$, the sum of their cardinalities is strictly greater than $(t - 1)k$. If their union contained only $t - 1$ distinct vertices, then by the pigeonhole principle there would be some vertex that occurs at least $k+1$ times, violating the upper bound on the number of occurrences of any vertex in the circuit.

Thus, by Lemma 22, there exists a Hamiltonian cycle in G^{2k-1}, and this can be found in polynomial time. □

Now we show how to construct a degree-bounded spanning circuit.

Lemma 24. *Let x^* be a feasible solution to the Held-Karp relaxation. Given an α-thin tree T with respect to $z^*_{uv} := x^*_{uv} + x^*_{vu}$, a circuit on G with every vertex visited at least once and at most $\lceil 4\alpha \rceil$ times can be found in polynomial time.*

Proof. We modify the circulation problem defined in (3.2) by introducing vertex capacities to the vertices: every vertex v is split into two vertices v_i and v_o, where all the incoming edges are connected to v_i and the outgoing edges are from v_o. We set the vertex capacity $u(\langle v_i, v_o \rangle)$ as $\sum_{e:\mathrm{tail}(e)=v} u(e)$. (See Fig. 3.1.) It is easy to see that this modification does not change the feasibility; thus, from Lemma 21, this new circulation problem instance is also feasible.

Rounding up all u values of this instance preserves the feasibility and guarantees the existence of an integral solution. By contracting split vertices back in the integral solution, we obtain a spanning Eulerian subgraph of $G = (V, A)$ (with arcs duplicated) whose maximum indegree is at most $\max_{v \in V} \lceil \sum_{e:\mathrm{tail}(e)=v} u(e) \rceil$. Observe that, for any $v \in V$,

$$
\begin{aligned}
\sum_{e:\mathrm{tail}(e)=v} u(e) &= |\{e \in T_{\rightarrow} \mid \mathrm{tail}(e) = v\}| + \sum_{e:\mathrm{tail}(e)=v} 2\alpha x^*_e \\
&\leq \alpha z^*(\delta(v)) + 2\alpha x^*(\delta^+(v)) \\
&= 4\alpha.
\end{aligned}
$$

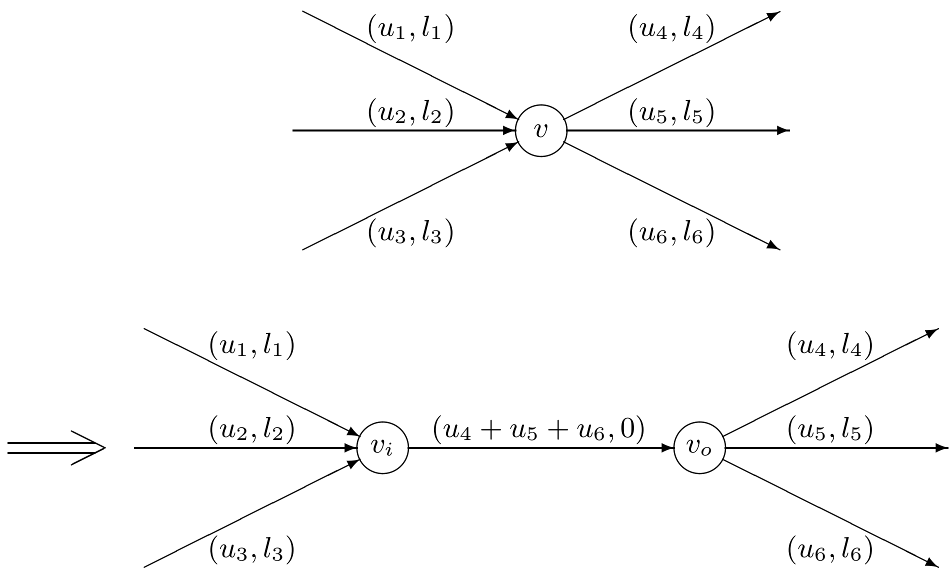

Figure 3.1: Introducing vertex capacities.

Thus, we can find a spanning Eulerian subgraph of $G = (V, A)$ whose maximum degree is at most $\lceil 4\alpha \rceil$, given the α-thin tree T. Any Eulerian circuit of this graph will satisfy the desired property. □

Theorem 14 and Lemmas 23 and 24 yield the algorithm.

Theorem 15. *There exists a probabilistic $O(\frac{\log n}{\log \log n})$-approximation algorithm for the bottleneck asymmetric traveling salesman problem under a metric cost.*

Proof. Let $A_{\leq \tau} := \{\langle u, v \rangle \mid c(u, v) \leq \tau\}$ and $G_{\leq \tau} := (V, A_{\leq \tau})$. The algorithm first determines the minimum τ such that the Held-Karp relaxation for $G_{\leq \tau}$ is feasible. Let τ^* be this minimum. If $\tau_1 \leq \tau_2$ and the Held-Karp relaxation for $G_{\leq \tau_2}$ is infeasible, the relaxation for $G_{\leq \tau_1}$ is also infeasible; therefore, τ^* can be discovered by binary search. Note that τ^* can serve as a lower bound on the optimal solution value.

Once τ^* is determined, we compute an extreme point solution x^* to the Held-

Karp relaxation for $G_{\leq\tau^*}$. Then we sample an α-thin tree T with respect to $z^*_{uv} := x^*_{uv} + x^*_{vu}$ for $\alpha = \frac{4\ln n}{\ln\ln n}$. By Theorem 14, this can be performed in polynomial time with high probability.

Then the algorithm constructs the circulation problem instance described in the proof of Lemma 24 and finds an integral solution. Lemma 24 shows that any Eulerian circuit of this integral solution is a spanning circuit where no vertex appears more than $\lceil 4\alpha \rceil$ times.

Let $\{P_1, \ldots, P_\ell\}$ be the partition of this spanning circuit as defined in Lemma 22 for $k = \lceil 4\alpha \rceil$. The algorithm computes a transversal of $\{P_1, \ldots, P_\ell\}$ and augments it into a Hamiltonian cycle C in $G^{2\lceil 4\alpha \rceil - 1}$. By the triangle inequality, the cost of C is at most $(2\lceil 4\alpha \rceil - 1) \cdot \tau^*$; thus, C is a $(2\lceil 4\alpha \rceil - 1)$-approximate solution to the given input. Note that $2\lceil 4\alpha \rceil - 1 = 2\lceil \frac{16\ln n}{\ln\ln n} \rceil - 1 = O(\frac{\log n}{\log\log n})$.

The foregoing is a probabilistic $O(\frac{\log n}{\log\log n})$-approximation algorithm for the bottleneck asymmetric traveling salesman problem under a metric cost. □

3.3 Special Case

In this section, we illustrate how our framework can be used together with other results to yield a stronger approximation guarantee in certain special cases. Lemmas 23 and 24 imply the following theorem.

Theorem 16. *If an $f(n)$-thin tree can be found in polynomial time for a certain class of metric, an $O(f(n))$-approximation algorithm exists for the bottleneck asymmetric traveling salesman problem under the same class of metric.*

In particular, Oveis Gharan and Saberi [45] investigate the case when the Held-Karp solution can be embedded on an orientable surface with a bounded genus; Oveis Gharan and Saberi [45], in addition to an $O(1)$-approximation algorithm for the min-sum problem, show the following:

Theorem 17. *Given a feasible solution $x^* \in \mathbb{R}^A$ to the Held-Karp relaxation, let $z^*_{uv} := x^*_{uv} + x^*_{vu}$. If the support of z^* can be embedded on an orientable surface with a bounded genus, an α-thin tree with respect to z^* can be found in polynomial time, where α is a constant that depends on the bound on the genus.*

Theorems 16 and 17 together imply the following.

Corollary 12. *There exists an $O(1)$-approximation algorithm for the bottleneck asymmetric traveling salesman problem when the support of the Held-Karp solution can be embedded on an orientable surface with a bounded genus.*

3.4 Open Questions

Given that the bottleneck symmetric TSP is 2-approximable [21, 40, 49], a naturally following question is if the asymmetric version also admits a 2-approximation algorithm. The algorithms for the symmetric case are based on the fact that the square of a 2-connected graph is Hamiltonian. One could regard the analogue of 2-connectedness of an undirected graph in a digraph as the following property: for any two vertices, there exists a simple directed cycle that includes both vertices. However, unfortunately, there exists such a graph that is non-Hamiltonian. In fact, for any constant k and p, the following can be shown:

Theorem 18. *For any constant $k, p \in \mathbb{N}$, there exists a digraph $G = (V, A)$ such that:*

(i) for all $u, v \in V$, there exist k paths $P_1, \ldots, P_k$ from u to v and k paths $Q_1, \ldots, Q_k$ from v to u such that $P_1, \ldots, P_k, Q_1, \ldots, Q_k$ are internally vertex-disjoint;

(ii) G^p is non-Hamiltonian.

As this approach appears unpromising, one could instead ask if some constant-order power of a graph whose Held-Karp relaxation is feasible is Hamiltonian.

Question 1. *Does there exist a constant p such that the pth power of any digraph with a feasible Held-Karp relaxation is Hamiltonian?*

One plausible way to affirmatively answer Question 1 is by proving that a graph whose Held-Karp relaxation is feasible contains a spanning circuit that satisfies the property of Lemma 22; Lemma 23 might be helpful in this. In particular, if there exists an efficient procedure that computes an $O(1)$-thin tree with respect to the Held-Karp solution, that would affirm Question 1.

Considering the undirected case, we can show that the set of graphs whose Held-Karp relaxation is feasible is a proper subset of the set of 2-connected graphs (see Theorem 19 for one direction); therefore, it is conceivable that one could attain a direct and simpler proof that the square of a graph whose Held-Karp relaxation is feasible is Hamiltonian. Such a proof may provide some inspiration for the asymmetric case.

Theorem 19. *For an undirected graph $G = (V, E)$, if the linear system*

$$\begin{cases} z(\delta(v)) = 2 & \forall v \in V \\ z(\delta(S)) \geq 2 & \forall S \subsetneq V, S \neq \emptyset \\ z \geq 0. \end{cases} \tag{3.3}$$

has a feasible solution $z^* \in \mathbb{R}^E$, G is 2-connected.

Proof. This proof borrows some idea from the proof of Lemma 24.

Let $G' = (V, A)$ be the digraph obtained from G by replacing each edge with two arcs in both directions. Consider a flow network on G', where the arc capacity is given as the z^* value of the underlying edge.

For any $u, v \in V$, a flow of 2 can be routed from u to v on this network. Let $f \in \mathbb{R}^A$ be this flow. Without loss of generality, we can assume that

$$\forall \{x, y\} \in E \quad f(x, y) = 0 \text{ or } f(y, x) = 0. \tag{3.4}$$

We drop the arcs on which the flow is zero from the network.

Let x be an arbitrary vertex other than u or v. Note that, from (3.4), the sum of the capacities of the arcs incident to/from x is at most 2. From the flow conservation, the incoming flow into x is at most 1; thus, introducing the vertex capacity of 1 to every vertex other than u and v does not break the feasibility of f.

Now we round up all of the capacities, and there exists an integral flow of value 2 from u to v on this flow network. This proves the existence of two vertex-disjoint paths from u to v. □

BIBLIOGRAPHY

[1] Hyung-Chan An, Robert Kleinberg, and David B. Shmoys. Improving Christofides' algorithm for the *s*-*t* path TSP. In *Proceedings of the 44th annual ACM symposium on Theory of computing*, STOC '12, pages 875–886, 2012.

[2] Hyung-Chan An, Robert D. Kleinberg, and David B. Shmoys. Approximation algorithms for the bottleneck asymmetric traveling salesman problem. APPROX/RANDOM '10, pages 1–11, 2010.

[3] Shoshana Anily, Julien Bramel, and Alain Hertz. A $\frac{5}{3}$-approximation algorithm for the clustered traveling salesman tour and path problems. *Operations Research Letters*, 24(1-2):29–35, 1999.

[4] D. Applegate, R. E. Bixby, V. Chvátal, and W. Cook. Concorde. Available at http://www.tsp.gatech.edu.

[5] Aaron Archer. Inapproximability of the asymmetric facility location and *k*-median problems. http://www2.research.att.com/~aarcher/Research/asym-hard.ps, 2000.

[6] Aaron Archer, MohammadHossein Bateni, MohammadTaghi Hajiaghayi, and Howard J. Karloff. Improved approximation algorithms for prize-collecting Steiner tree and TSP. *SIAM Journal on Computing*, 40(2):309–332, 2011.

[7] Vijay Arya, Naveen Garg, Rohit Khandekar, Adam Meyerson, Kamesh Munagala, and Vinayaka Pandit. Local search heuristics for *k*-median and facility location problems. *SIAM Journal on Computing*, 33(3):544–562, 2004.

[8] Arash Asadpour, Michel X. Goemans, Aleksander Mądry, Shayan Oveis Gharan, and Amin Saberi. An $O(\log n/\log\log n)$-approximation algorithm for the asymmetric traveling salesman problem. In *Proceedings of the 21st Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '10, pages 379–389, 2010.

[9] Jungyun Bae and Sivakumar Rathinam. Approximation algorithms for multiple terminal, Hamiltonian path problems. *Optimization Letters*, 6:69–85, 2012.

[10] Egon Balas. The prize collecting traveling salesman problem. *Networks*, 19(6):621–636, 1989.

[11] Dimitris Bertsimas and Chungpiaw Teo. The parsimonious property of cut covering problems and its applications. *Operations Research Letters*, 21(3):123–132, 1997.

[12] Daniel Bienstock, Michel X. Goemans, David Simchi-Levi, and David P. Williamson. A note on the prize collecting traveling salesman problem. *Mathematical Programming*, 59:413–420, 1993.

[13] Kamalika Chaudhuri, Brighten Godfrey, Satish Rao, and Kunal Talwar. Paths, trees, and minimum latency tours. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '03, pages 36–45, 2003.

[14] Chandra Chekuri, Jan Vondrák, and Rico Zenklusen. Dependent randomized rounding for matroid polytopes and applications. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science*, FOCS '10, pages 575–584, 2010.

[15] N. Christofides. Worst-case analysis of a new heuristic for the travelling salesman problem. Technical Report 388, Graduate School of Industrial Administration, CMU, 1976.

[16] Fabián A. Chudak, Tim Roughgarden, and David P. Williamson. Approximate k-MSTs and k-Steiner trees via the primal-dual method and Lagrangean relaxation. *Mathematical Programming*, 100(2):411–421, 2004.

[17] Julia Chuzhoy, Sudipto Guha, Eran Halperin, Sanjeev Khanna, Guy Kortsarz, Robert Krauthgamer, and Joseph Naor. Asymmetric k-center is $\log^* n$-hard to approximate. *Journal of the ACM*, 52(4):538–551, 2005.

[18] G. Dantzig, R. Fulkerson, and S. Johnson. Solution of a large-scale traveling-salesman problem. *Operations Research*, 2(4):393–410, 1954.

[19] Jack Edmonds. Matroids and the greedy algorithm. *Mathematical Programming*, 1:127–136, 1971.

[20] Jack Edmonds and Ellis L. Johnson. Matching, Euler tours and the Chinese postman. *Mathematical Programming*, 5:88–124, 1973.

[21] Herbert Fleischner. The square of every two-connected graph is Hamiltonian. *Journal of Combinatorial Theory, Series B*, 16(1):29–34, 1974.

[22] 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.

[23] Michel X. Goemans. Unpublished manuscript, 1990.

[24] Michel X. Goemans. Worst-case comparison of valid inequalities for the TSP. *Mathematical Programming*, 69:335–349, 1995.

[25] Michel X. Goemans. Combining approximation algorithms for the prize-collecting TSP. *CoRR*, abs/0910.0553, 2009.

[26] Michel X. Goemans and Dimitris Bertsimas. Survivable networks, linear programming relaxations and the parsimonious property. *Mathematical Programming*, 60:145–166, 1993.

[27] Michel X. Goemans and David P. Williamson. A general approximation technique for constrained forest problems. *SIAM Journal on Computing*, 24(2):296–317, 1995.

[28] Martin Grötschel, László Lovász, and Alexander Schrijver. The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica*, 1(2):169–197, 1981.

[29] N. Guttmann-Beck, R. Hassin, S. Khuller, and B. Raghavachari. Approximation algorithms with bounded performance guarantees for the clustered traveling salesman problem. *Algorithmica*, 28:422–437, 2000.

[30] P. Hall. On Representatives of Subsets. *Journal of the London Mathematical Society*, 10:26–30, 1935.

[31] Michael Held and Richard M. Karp. The traveling-salesman problem and minimum spanning trees. *Operations Research*, 18(6):1138–1162, 1970.

[32] Dorit S. Hochbaum and David B. Shmoys. A unified approach to approximation algorithms for bottleneck problems. *Journal of the ACM*, 33(3):533–550, 1986.

[33] J. A. Hoogeveen. Analysis of Christofides' heuristic: some paths are more difficult than cycles. *Operations Research Letters*, 10(5):291–295, 1991.

[34] Wen-Lian Hsu and George L. Nemhauser. Easy and hard bottleneck location problems. *Discrete Applied Mathematics*, 1(3):209 – 215, 1979.

[35] Kamal Jain, Mohammad Mahdian, Evangelos Markakis, Amin Saberi, and Vijay V. Vazirani. Greedy facility location algorithms analyzed using dual fitting with factor-revealing LP. *Journal of the ACM*, 50(6):795–824, 2003.

[36] Richard M. Karp. Reducibility among combinatorial problems. In R. E. Miller and J. W. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. Plenum Press, 1972.

[37] Jon Kleinberg and Éva Tardos. *Algorithm Design*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2005.

[38] Jon Kleinberg and David P. Williamson. Unpublished manuscript, 1998. Also see http://legacy.orie.cornell.edu/~dpw/cornell.ps, pages 124–126.

[39] Michael Lampis. Improved Inapproximability for TSP. APPROX /RANDOM '12, 2012. *To appear*.

[40] H. T. Lau. Finding EPS-Graphs. *Monatshefte für Mathematik*, 92(1):37–40, 1981.

[41] E. L. Lawler, J. K. Lenstra, A. H. G. Rinnooy Kan, and D. B. Shmoys, editors. *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization*. Wiley, 1985.

[42] Jyh-Han Lin and Jeffrey Scott Vitter. ϵ-approximations with minimum packing constraint violation (extended abstract). In *STOC '92: Proceedings of the 24th Annual ACM Symposium on Theory of Computing*, pages 771–782. ACM, 1992.

[43] Tobias Mömke and Ola Svensson. Approximating graphic TSP by matchings. In *Proceedings of the 52nd Annual IEEE Symposium on Foundations of Computer Science*, FOCS '11, pages 560–569, 2011.

[44] Marcin Mucha. 13/9-approximation for graphic TSP. In *Proceedings of the 29th International Symposium on Theoretical Aspects of Computer Science*, STACS '12, pages 30–41, 2012.

[45] Shayan Oveis Gharan and Amin Saberi. The asymmetric traveling salesman problem on graphs with bounded genus. In *Proceedings of the 22nd An-*

nual ACM-SIAM Symposium on Discrete Algorithms, SODA ’11, pages 967–975, 2011.

[46] Shayan Oveis Gharan, Amin Saberi, and Mohit Singh. A randomized rounding approach to the traveling salesman problem. In *Proceedings of the 52nd Annual IEEE Symposium on Foundations of Computer Science*, FOCS ’11, pages 550–559, 2011.

[47] Rina Panigrahy and Sundar Vishwanathan. An $O(\log^* n)$ approximation algorithm for the asymmetric p-center problem. *Journal of Algorithms*, 27(2):259–268, 1998.

[48] Christos H. Papadimitriou and Santosh Vempala. On the approximability of the traveling salesman problem. *Combinatorica*, 26(1):101–120, 2006.

[49] R. Gary Parker and Ronald L. Rardin. Guaranteed performance heuristics for the bottleneck traveling salesman problem. *Operations Research Letters*, 2(6):269 – 272, 1984.

[50] Gerhard Reinelt. TSPLIB—a traveling salesman problem library. *ORSA Journal on Computing*, 3:376–384, 1991.

[51] Alexander Schrijver. *Combinatorial Optimization: Polyhedra and Efficiency*. Springer, 2003.

[52] András Sebő and Jens Vygen. Shorter tours by nicer ears: 7/5-approximation for graphic TSP, 3/2 for the path version, and 4/3 for two-edge-connected subgraphs. *CoRR*, abs/1201.1870, 2012.

[53] David B. Shmoys and David P. Williamson. Analyzing the Held-Karp TSP bound: a monotonicity property with application. *Information Processing Letters*, 35(6):281–285, 1990.

[54] 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.