A (Slightly) Improved Approximation Algorithm for Metric TSP

Anna R. Karlin
karlin@cs.washington.edu
University of Washington
USA

Nathan Klein
nwklein@cs.washington.edu
University of Washington
USA

Shayan Oveis Gharan
shayan@cs.washington.edu
University of Washington
USA

ABSTRACT

For some $\epsilon > 10^{-36}$ we give a randomized $3/2 - \epsilon$ approximation algorithm for metric TSP.

CCS CONCEPTS

• **Theory of computation** → **Routing and network design problems**.

KEYWORDS

Traveling Salesperson Problem, Approximation Algorithms, Randomized Rounding, Max Entropy, Strongly Rayleigh, Near Minimum Cuts

ACM Reference Format:
Anna R. Karlin, Nathan Klein, and Shayan Oveis Gharan. 2021. A (Slightly) Improved Approximation Algorithm for Metric TSP. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (STOC '21), June 21–25, 2021, Virtual, Italy.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3406325.3451009

1 INTRODUCTION

One of the most fundamental problems in combinatorial optimization is the traveling salesperson problem (TSP), formalized as early as 1832 (c.f. [2, Ch 1]). In an instance of TSP we are given a set of n cities V along with their pairwise symmetric distances, $c : V \times V \to \mathbb{R}_{\geq 0}$. The goal is to find a Hamiltonian cycle of minimum cost. In the metric TSP problem, which we study here, the distances satisfy the triangle inequality. Therefore, the problem is equivalent to finding a closed Eulerian connected walk of minimum cost.[1]

It is NP-hard to approximate TSP within a factor of $\frac{123}{122}$ [34]. An algorithm of Christofides-Serdyukov [14, 43] from four decades ago gives a $\frac{3}{2}$-approximation for TSP (see [47] for a historical note about TSP). This remains the best known approximation algorithm for the general case of the problem despite significant work, e.g., [10–13, 22, 24, 28, 29, 33, 40, 44, 48].

[1]Given such an Eulerian cycle, we can use the triangle inequality to shortcut vertices visited more than once to get a Hamiltonian cycle.

STOC '21, June 21–25, 2021, Virtual, Italy

ACM ISBN 978-1-4503-8053-9/21/06…$15.00
https://doi.org/10.1145/3406325.3451009

In contrast, there have been major improvements to this algorithm for a number of special cases of TSP. For example, polynomial-time approximation schemes (PTAS) have been found for Euclidean [3, 36], planar [4, 25, 35], and low-genus metric [17] instances. In addition, the case of graph metrics has received significant attention. In 2011, the third author, Saberi, and Singh [39] found a $\frac{3}{2} - \epsilon_0$ approximation for this case. Mömke and Svensson [37] then obtained a combinatorial algorithm for graphic TSP with an approximation ratio of 1.461. This ratio was later improved by Mucha [38] to $\frac{13}{9} \approx 1.444$, and then by Sebö and Vygen [42] to 1.4.

In this paper we prove[2] the following theorem:

Theorem 1.1. *For some absolute constant $\epsilon > 10^{-36}$, there is a randomized algorithm that outputs a tour with expected cost at most $\frac{3}{2} - \epsilon$ times the cost of the optimum solution.*

We note that while the algorithm makes use of the Held-Karp relaxation, we do not prove that the integrality gap of this polytope is bounded away from 3/2. We also remark that although our approximation factor is only slightly better than Christofides-Serdyukov, we are not aware of any example where the approximation ratio of the algorithm we analyze exceeds 4/3 in expectation.

Following a new exciting result of Traub, Vygen, Zenklusen [46] we also get the following theorem.

Theorem 1.2. *For some absolute constant $\epsilon > 0$ there is a randomized algorithm that outputs a TSP path with expected cost at most $\frac{3}{2} - \epsilon$ times the cost of the optimum solution.*

1.1 Algorithm

First, we recall the classical Christofides-Serdyukov algorithm: Given an instance of TSP, choose a minimum spanning tree and then add the minimum cost matching on the odd degree vertices of the tree. The algorithm we study is very similar, except we choose a random spanning tree based on the standard linear programming relaxation of TSP.

Let x^0 be an optimum solution of the following TSP linear program relaxation [15, 31]:

$$\begin{aligned}
\min \quad & \sum_{u,v} x_{(u,v)} c(u,v) & & \\
\text{s.t.,} \quad & \sum_{u} x_{(u,v)} = 2 & & \forall v \in V, \\
& \sum_{u \in S, v \notin S} x_{(u,v)} \geq 2, & & \forall S \subsetneq V, \\
& x_{(u,v)} \geq 0 & & \forall u, v \in V.
\end{aligned} \tag{1}$$

[2]Due to space limitations, in this extended abstract we are only able to outline the key technical tools and ideas of our proof. The full version can be found at https://arxiv.org/abs/2007.01409.

Given x^0, we pick an arbitrary node, u, split it into two nodes u_0, v_0 and set $x_{(u_0,v_0)} = 1, c(u_0, v_0) = 0$ and we assign half of every edge incident to u to u_0 and the other half to v_0. This allows us to assume without loss of generality that x^0 has an edge $e_0 = (u_0, v_0)$ such that $x_{e_0} = 1, c(e_0) = 0$.

Let $E_0 = E \cup \{e_0\}$ be the support of x^0 and let x be x^0 restricted to E and $G = (V, E)$. x^0 restricted to E is in the spanning tree polytope (3).

For a vector $\lambda : E \to \mathbb{R}_{\geq 0}$, a λ-uniform distribution μ_λ over spanning trees of $G = (V, E)$ is a distribution where for every spanning tree $T \subseteq E$, $\mathbb{P}_\mu[T] = \frac{\prod_{e\in T}\lambda_e}{\sum_{T'}\prod_{e\in T'}\lambda_e}$. Now, find a vector λ such that for every edge $e \in E$, $\mathbb{P}_{\mu_\lambda}[e \in T] = x_e(1 \pm \epsilon)$, for some $\epsilon < 2^{-n}$. Such a vector λ can be found using the multiplicative weight update algorithm [5] or by applying interior point methods [42] or the ellipsoid method [5]. (We note that the multiplicative weight update method can only guarantee $\epsilon < 1/\text{poly}(n)$ in polynomial time.)

Theorem 1.3 ([5])**.** *Let z be a point in the spanning tree polytope (see (3)) of a graph $G = (V, E)$. For any $\epsilon > 0$, a vector $\lambda : E \to \mathbb{R}_{\geq 0}$ can be found such that the corresponding λ-uniform spanning tree distribution, μ_λ, satisfies*

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

i.e., the marginals are approximately preserved. In the above $\mathcal{T}$ is the set of all spanning trees of (V, E). The running time is polynomial in $n = |V|$, $-\log \min_{e\in E} z_e$ and $\log(1/\epsilon)$.

Finally, we sample a tree $T \sim \mu_\lambda$ and then add the minimum cost matching on the odd degree vertices of T. The above algorithm is a

Algorithm 1 An Improved Approximation Algorithm for TSP

Find an optimum solution x^0 of Eq. (1), and let $e_0 = (u_0, v_0)$ be an edge with $x^0_{e_0} = 1, c(e_0) = 0$.
Let $E_0 = E \cup \{e_0\}$ be the support of x^0 and x be x^0 restricted to E and $G = (V, E)$.
Find a vector $\lambda : E \to \mathbb{R}_{\geq 0}$ such that for any $e \in E$, $\mathbb{P}_{\mu_\lambda}[e] = x_e(1 \pm 2^{-n})$.
Sample a tree $T \sim \mu_\lambda$.
Let M be the minimum cost matching on odd degree vertices of T.
Output $T \cup M$.

slight modification of the algorithm proposed in [39]. We refer the interested reader to exciting work of Genova and Williamson [23] on the empirical performance of the max-entropy rounding algorithm. We also remark that although the algorithm implemented in [23] is slightly different from the above algorithm, we expect the performance to be similar.

1.2 New Techniques

Here we discuss new machinery and technical tools that we developed for this result which could be of independent interest.

1.2.1 Polygon Structure for Near Minimum Cuts Crossed on one Side. Let $G = (V, E, x)$ be an undirected graph equipped with a weight function $x : E \to \mathbb{R}_{\geq 0}$ such that for any cut $(S, \overline{S})$ such that $u_0, v_0 \notin S$, $x(\delta(S)) \geq 2$ (recall u_0, v_0 is the edge of fraction 1 which appears in x^0 but not x).

For some (small) $\eta \geq 0$, consider the family of η-near min cuts of G. Let $\mathcal{C}$ be a connected component of crossing η-near min cuts. Given $\mathcal{C}$ we can partition vertices of G into sets $a_0, \dots, a_{m-1}$ (called atoms); this is the coarsest partition such that for each a_i, and each $(S, \overline{S}) \in \mathcal{C}$, we have $a_i \subseteq S$ or $a_i \subseteq \overline{S}$. Here a_0 is the atom that contains u_0, v_0.

There has been several works studying the structure of edges between these atoms and the structure of cuts in $\mathcal{C}$ w.r.t. the a_i's. The *cactus structure* (see [18]) shows that if $\eta = 0$, then we can arrange the a_i's around a cycle, say $a_1, \dots, a_m$ (after renaming), such that $x(E(a_i, a_{i+1})) = 1$ for all i.

Benczúr and Goemans [6, 8] studied the case when $\eta \leq 6/5$ and introduced the notion of *polygon representation*, in which case atoms can be placed on the sides of an equilateral polygon and some atoms placed inside the polygon, such that every cut in $\mathcal{C}$ can be represented by a diagonal of this polygon. Later, [39] studied the structure of edges of G in this polygon when $\eta < 1/100$.

In this paper, we show it suffices to study the structure of edges in a special family of polygon representations:

Theorem 1.4 (Informal version of the polygon structure theorem)**.** *Suppose we have a polygon representation for a connected component $\mathcal{C}$ of η-near min cuts of G such that*

- *No atom is mapped inside,*
- *If we identify each cut $(S, \overline{S}) \in \mathcal{C}$ with the interval along the polygon that does not contain a_0, then any interval is only crossed on one side (only on the left or only on the right).*

Then, we have:

- *For any atom a_i, $x(\delta(a_i)) \leq 2 + O(\delta)$,*
- *For any pair of atoms a_i, a_{i+1}, $x(E(a_i, a_{i+1})) \geq 1 - \Omega(\eta)$.*

We expect to see further applications of our theorem in studying variants of TSP.

1.2.2 Generalized Gurvits' Lemma. Given a real stable polynomial $p \in \mathbb{R}_{\geq 0}[z_1, \dots, z_n]$ (with non-negative coefficients), Gurvits proved the following inequality [26, 27]

$$\frac{n!}{n^n} \inf_{z>0} \frac{p(z_1,\dots,z_n)}{z_1 \dots z_n} \leq \partial_{z_1} \dots \partial_{z_n} p|_{z=0} \leq \inf_{z>0} \frac{p(z_1,\dots,z_n)}{z_1 \dots z_n}. \tag{2}$$

As an immediate consequence, one can prove the following theorem about strongly Rayleigh (SR) distributions.

Theorem 1.5. *Let $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$ be SR and $A_1, \dots, A_m$ be random variables corresponding to the number of elements sampled in m disjoint subsets of $[n]$ such that $\mathbb{E}[A_i] = n_i$ for all i. If $n_i = 1$ for all $1 \leq i \leq n$, then $\mathbb{P}[\forall i, A_i = 1] \geq \frac{m!}{m^m}$.*

One can ask what happens if the vector $\vec{n} = (n_1, \dots, n_m)$ in the above theorem is not equal but close to the all ones vector, $\mathbf{1}$.

A related theorem was proved in [39].

Theorem 1.6. *Let $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$ be SR and A, B be random variables corresponding to the number of elements sampled in two* *disjoint sets. If* $\mathbb{P}[A+B=2] \geq \epsilon$*,* $\mathbb{P}[A \leq 1], \mathbb{P}[B \leq 1] \geq \alpha$ *and* $\mathbb{P}[A \geq 1], \mathbb{P}[B \geq 1] \geq \beta$ *then* $\mathbb{P}[A=B=1] \geq \epsilon\alpha\beta/3$.

We prove a generalization of both of the above statements; roughly speaking, we show that as long as $\|\vec{n}-\mathbf{1}\|_1 < 1-\epsilon$ then $\mathbb{P}[\forall i, A_i = 1] \geq f(\epsilon, m)$ where $f(\epsilon, m)$ has no dependence on n, the number of underlying elements in the support of μ.

Theorem 1.7 (Informal version of the generalized Gurvits lemma)**.** *Let* $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$ *be SR and let* $A_1, \dots, A_m$ *be random variables corresponding to the number of elements sampled in m disjoint subsets of* $[n]$*. Suppose that there are integers* $n_1, \dots, n_m$ *such that for any set* $S \subseteq [m]$*,* $\mathbb{P}[\sum_{i \in S} A_i = \sum_{i \in S} n_i] \geq \epsilon$*. Then,*

$$\mathbb{P}[\forall i, A_i = n_i] \geq f(\epsilon, m).$$

The above statement is even stronger than Theorem 1.5 as we only require $\mathbb{P}[\sum_{i \in S} A_i = \sum_{i \in S} n_i]$ to be bounded away from 0 for any set $S \subseteq [m]$ and we don't need a bound on the expectation. Our proof of the above theorem has double exponential dependence on ϵ. We leave it an open problem to find the optimum dependency on ϵ. Furthermore, our proof of the above theorem is probabilistic in nature; we expect that an algebraic proof based on the theory of real stable polynomials will provide a significantly improved lower bound. Unlike the above theorem, such a proof may possibly extend to the more general class of completely log-concave distributions [1].

1.2.3 Conditioning while Preserving Marginals. Consider a SR distribution $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$ and let $x : [n] \to \mathbb{R}_{\geq 0}$, where for all i, $x_i = \mathbb{P}_{T \sim \mu}[i \in T]$, be the marginals.

Let $A, B \subseteq [n]$ be two disjoint sets such that $\mathbb{E}[A_T], \mathbb{E}[B_T] \approx 1$. It follows from Theorem 1.7 that $\mathbb{P}[A_T = B_T = 1] \geq \Omega(1)$. Here, however, we are interested in a stronger event; let $\nu = \mu | A_T = B_T = 1$ and let $y_i = \mathbb{P}_{T \sim \mu}[i \in T]$. It turns out that the y vector can be very different from the x vector, in particular, for some i's we can have $|y_i - x_i|$ bounded away from 0. We show that there is an event of non-negligible probability that is a subset of $A_T = B_T = 1$ under which the marginals of elements in A, B are almost preserved.

Theorem 1.8 (Informal version of the max flow theorem)**.** *Let* $\mu : 2^{[n]} \to \mathbb{R}_{\geq 0}$ *be a SR distribution and let* $A, B \subseteq [n]$ *be two disjoint subsets such that* $\mathbb{E}[A_T], \mathbb{E}[B_T] \approx 1$*. For any* $\alpha \ll 1$ *there is an event* $\mathcal{E}_{A,B}$ *such that* $\mathbb{P}[\mathcal{E}_{A,B}] \geq \Omega(\alpha^2)$ *and*

- $\mathbb{P}[A_T = B_T = 1 | \mathcal{E}_{A,B}] = 1,$
- $\sum_{i \in A} |\mathbb{P}[i] - \mathbb{P}[i|\mathcal{E}]| \leq \alpha,$
- $\sum_{i \in B} |\mathbb{P}[i] - \mathbb{P}[i|\mathcal{E}]| \leq \alpha.$

We remark that the quadratic lower bound on α is necessary in the above theorem for a sufficiently small $\alpha > 0$. The above theorem can be seen as a generalization of Theorem 1.5 in the special case of two sets.

We leave it an open problem to extend the above theorem to arbitrary k disjoint sets. We suspect that in such a case the ideal event $\mathcal{E}_{A_1,\dots,A_k}$ occurs with probability $\Omega(\alpha)^k$ and preserves all marginals of elements in each of the sets $A_1, \dots, A_k$ up to a total variation distance of α.

2 PRELIMINARIES

2.1 Notation

We write $[n] := \{1, \dots, n\}$ to denote the set of integers from 1 to n. For a set of edges $A \subseteq E$ and (a tree) $T \subseteq E$, we write

$$A_T = |A \cap T|.$$

For a set $S \subseteq V$, we write

$$E(S) = \{(u,v) \in E : u, v \in S\}$$

to denote the set of edges in S and we write

$$\delta(S) = \{(u,v) \in E : |\{u,v\} \cap S| = 1\}$$

to denote the set of edges that leave S. For two *disjoint* sets of vertices $A, B \subseteq V$, we write

$$E(A,B) = \{(u,v) \in E : u \in A, v \in B\}.$$

For a set $A \subseteq E$ and a function $x : E \to \mathbb{R}$ we write

$$x(A) := \sum_{e \in A} x_e.$$

For two sets $A, B \subseteq V$, we say A *crosses* B if all of the following sets are non-empty:

$$A \cap B, A \smallsetminus B, B \smallsetminus A, \overline{A \cup B}.$$

We write $G = (V, E, x)$ to denote an (undirected) graph G together with special vertices u_0, v_0 and a weight function $x : E \to \mathbb{R}_{\geq 0}$ such that

$$x(\delta(S)) \geq 2, \qquad \forall S \subsetneq V : u_0, v_0 \notin S.$$

For such a graph, we say a cut $S \subseteq V$ is an η*-near min cut w.r.t.,* x (or simply η-near min cut when x is understood) if $x(\delta(S)) \leq 2 + \eta$. Unless otherwise specified, in any statement about a cut $(S, \overline{S})$ in G, we assume $u_0, v_0 \notin S$.

2.2 Polyhedral Background

For any graph $G = (V, E)$, Edmonds [19] gave the following description for the convex hull of spanning trees of a graph $G = (V, E)$, known as the *spanning tree polytope.*

$$\begin{aligned} z(E) &= |V| - 1 & & \\ z(E(S)) &\leq |S| - 1 & \forall S &\subseteq V \\ z_e &\geq 0 & \forall e &\in E. \end{aligned} \tag{3}$$

Edmonds [19] proved that the extreme point solutions of this polytope are the characteristic vectors of the spanning trees of G.

Fact 2.1. *Let* x^0 *be a feasible solution of* (1) *such that* $x^0_{e_0} = 1$ *with support* $E_0 = E \cup \{e_0\}$*. Let* x *be* x^0 *restricted to* E*; then* x *is in the spanning tree polytope of* $G = (V, E)$*.*

Proof. For any set $S \subseteq V$ such that $u_0, v_0 \notin S$, $x(E(S)) = \frac{2|S| - x^0(\delta(S))}{2} \leq |S| - 1$. If $u_0 \in S, v_0 \notin S$, then

$$x(E(S)) = \frac{2|S| - 1 - (x^0(\delta(S)) - 1)}{2} \leq |S| - 1.$$

Finally, if $u_0, v_0 \in S$, then

$$x(E(S)) = \frac{2|S| - 2 - x^0(\delta(S))}{2} \leq |S| - 2$$

The claim follows because $x(E) = x^0(E_0) - 1 = n - 1$. □

Since $c(e_0) = 0$, the following fact is immediate.

Fact 2.2. *Let $G = (V, E, x)$ where x is in the spanning tree polytope. Let μ be any distribution of spanning trees with marginals x, then $\mathbb{E}_{T\sim\mu}[c(T \cup e_0)] = c(x)$.*

To bound the cost of the min-cost matching on the set O of odd degree vertices of the tree T, we use the following characterization, which dominates the O-join polytope[3] due to Edmonds and Johnson [20].

Proposition 2.3. *For any graph $G = (V, E)$, cost function $c : E \to \mathbb{R}_+$, and a set $O \subseteq V$ with an even number of vertices, the minimum weight of an O-join equals the optimum value of the following integral linear program.*

$$\begin{aligned} \min \quad & c(y) \\ \text{s.t.} \quad & y(\delta(S)) \geq 1 \quad \forall S \subseteq V, |S \cap O| \text{ odd} \\ & y_e \geq 0 \qquad\qquad \forall e \in E \end{aligned} \tag{4}$$

Definition 2.4 (Satisfied cuts). *For a set $S \subseteq V$ such that $u_0, v_0 \notin S$ and a spanning tree $T \subseteq E$ we say a vector $y : E \to \mathbb{R}_{\geq 0}$ satisfies S if one of the following holds:*

- *$\delta(S)_T$ is even, or*
- *$y(\delta(S)) \geq 1$.*

To analyze our algorithm, we will see that the main challenge is to construct a (random) vector y that satisfies all cuts and $\mathbb{E}[c(y)] \leq (1/2 - \epsilon)OPT$.

2.3 Structure of Near Minimum Cuts

Lemma 2.5 ([39]). *For $G = (V, E, x)$, let $A, B \subsetneq V$ be two crossing ϵ_A, ϵ_B near min cuts respectively. Then, $A \cap B, A \cup B, A \smallsetminus B, B \smallsetminus A$ are $\epsilon_A + \epsilon_B$ near min cuts.*

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

$$x(\delta(A \cap B)) + x(\delta(A \cup B)) \leq x(\delta(A)) + x(\delta(B)) \leq 4 + \epsilon_A + \epsilon_B.$$

Since $x(\delta(A \cup B)) \geq 2$, we have $x(\delta(A \cap B)) \leq 2 + \epsilon_A + \epsilon_B$, as desired. □

The following lemma is proved in [7]:

Lemma 2.6 ([7, Lem 5.3.5]). *For $G = (V, E, x)$, let $A, B \subsetneq V$ be two crossing ϵ-near minimum cuts. Then, $x(E(A \cap B, A - B)), x(E(A \cap B, B - A)), x(E(\overline{A \cup B}, A - B))$, and $x(E(\overline{A \cup B}, B - A))$ are all at least $(1 - \epsilon/2)$.*

Lemma 2.7. *For $G = (V, E, x)$, let $A, B \subsetneq V$ be two ϵ near min cuts such that $A \subsetneq B$. Then*

$$x(\delta(A) \cap \delta(B)) = x(E(A, \overline{B})) \leq 1 + \epsilon, \text{ and}$$
$$x(E(\delta(A) \smallsetminus \delta(B))) \geq 1 - \epsilon/2.$$

Proof. Notice

$$2 + \epsilon \geq x(\delta(A)) = x(E(A, B \smallsetminus A)) + x(E(A, \overline{B}))$$
$$2 + \epsilon \geq x(\delta(B)) = x(E(B \smallsetminus A, \overline{B})) + x(E(A, \overline{B}))$$

Summing these up, we get

$$\begin{aligned} & 2x(E(A, \overline{B})) + x(E(A, B \smallsetminus A)) + x(E(B \smallsetminus A, \overline{B})) \\ & = 2x(E(A, \overline{B})) + x(\delta(B \smallsetminus A)) \leq 4 + 2\epsilon. \end{aligned}$$

Since $B \smallsetminus A$ is non-empty, $x(\delta(B \smallsetminus A)) \geq 2$, which implies the first inequality. To see the second one, let $C = B \smallsetminus A$ and note

$$\begin{aligned} 4 & \leq x(\delta(A)) + x(\delta(C)) = 2x(E(A, C)) + x(\delta(B)) \\ & \leq 2x(E(A, C)) + 2 + \epsilon \end{aligned}$$

which implies $x(E(A, C)) \geq 1 - \epsilon/2$. □

2.4 Strongly Rayleigh Distributions and λ-uniform Spanning Tree Distributions

Let $\mathcal{B}_E$ be the set of all probability measures on the Boolean algebra 2^E. Let $\mu \in \mathcal{B}_E$. The generating polynomial $g_\mu : \mathbb{R}[\{z_e\}_{e\in E}]$ of μ is defined as follows:

$$g_\mu(z) := \sum_S \mu(S) \prod_{e\in S} z_e.$$

We say μ is a strongly Rayleigh distribution if $g_\mu \neq 0$ over all $\{y_e\}_{e\in E} \in \mathbb{C}^E$ where $\text{Im}(z_e) > 0$ for all $e \in E$. We say μ is d-homogenous if for any $\lambda \in \mathbb{R}$, $g_\mu(\lambda \mathbf{z}) = \lambda^d g_\mu(\mathbf{z})$. Strongly Rayleigh (SR) distributions were defined in [9] where it was shown any λ-uniform spanning tree distribution is strongly Rayleigh. In this subsection we recall several properties of SR distributions proved in [9, 39] which will be useful to us.

Closure Operations of SR Distributions. SR distributions are closed under the following operations.

- **Projection.** For any $\mu \in \mathcal{B}_E$, and any $F \subseteq E$, the projection of μ onto F is the measure μ_F where for any $A \subseteq F$,
$$\mu_F(A) = \sum_{S:S\cap F=A} \mu(S).$$
- **Conditioning.** For any $e \in E$, $\{\mu | e \text{ out}\}$ and $\{\mu | e \text{ in}\}$.
- **Truncation.** For any integer $k \geq 0$ and $\mu \in \mathcal{B}_E$, truncation of μ to k, is the measure μ_k where for any $A \subseteq E$,
$$\mu_k(A) = \begin{cases} \frac{\mu(A)}{\sum_{S:|S|=k} \mu(S)} & \text{if } |A| = k \\ 0 & \text{otherwise.} \end{cases}$$
- **Product.** For any two disjoint sets E, F, and $\mu_E \in \mathcal{B}_E, \mu_F \in \mathcal{B}_F$ the product measure $\mu_{E\times F}$ is the measure where for any $A \subseteq E, B \subseteq F$, $\mu_{E\times F}(A \cup B) = \mu_E(A)\mu_F(B)$.

Throughout this paper we will repeatedly apply the above operations. We remark that SR distributions are *not* necessarily closed under truncation of a subset, i.e., if we require exactly k elements from $F \subsetneq E$.

Since λ-uniform spanning tree distributions are special classes of SR distributions, if we perform any of the above operations on a λ-uniform spanning tree distribution μ we get another SR distribution. Below, we see that by performing the following particular operations we still have a λ-uniform spanning tree distribution (perhaps with a different λ).

[3] The standard name for this is the T-join polytope. Because we reserve T to represent our tree, we call this the O-join polytope, where O represents the set of odd vertices in the tree.

Closure Operations of λ-uniform Spanning Tree Distributions.

- **Conditioning**. For any $e \in E$, $\{\mu \mid e \text{ out}\}, \{\mu \mid e \text{ in}\}$.
- **Tree Conditioning**. For $G = (V, E)$, a spanning tree distribution $\mu \in \mathcal{B}_E$, and $S \subseteq V$, $\{\mu \mid S \text{ tree}\}$.

Note that arbitrary spanning tree distributions are not necessarily closed under truncation and projection. We remark that SR measures are also closed under an analogue of tree conditioning, i.e., for a set $F \subseteq E$, let $k = \max_{S \in \text{supp } \mu} |S \cap F|$. Then, $\{\mu | |S \cap F| = k\}$ is SR. But if μ is a spanning tree distribution we get an extra *independence* property. The following independence is crucial to several of our proofs.

Fact 2.8. *For a graph $G = (V, E)$, and a vector $\lambda(G) : E \to \mathbb{R}_{\geq 0}$, let $\mu_{\lambda(G)}$ be the corresponding λ-uniform spanning tree distribution. Then for any $S \subsetneq V$,*

$$\{\mu_{\lambda(G)} \mid S \text{ tree}\} = \mu_{\lambda(G[S])} \times \mu_{\lambda(G/S)}.$$

Proof. Intuitively, this holds because in the max entropy distribution, conditioned on S being a tree, any tree chosen inside S can be composed with any tree chosen on G/S to obtain a spanning tree on G. So, to maximize the entropy these trees should be chosen independently. More formally for any $T_1 \in G[S]$ and $T_2 \in G/S$, we have the following equality for $\mathbb{P}[T = T_1 \cup T_2 \mid S \text{ is a tree}]$:

$$\begin{aligned} &= \frac{\lambda^{T_1}\lambda^{T_2}}{\sum_{T_1' \in G[S], T_2' \in G/S} \lambda^{T_1'}\lambda^{T_2'}} \\ &= \frac{\lambda^{T_1}}{\sum_{T_1' \in G[S]} \lambda^{T_1'}} \cdot \frac{\lambda^{T_2}}{\sum_{T_2' \in G/S} \lambda^{T_2'}} \\ &= \mathbb{P}_{T_1' \sim G[S]}\left[T_1' = T_1\right] \mathbb{P}_{T_2' \sim G/S}\left[T_2' = T_2\right], \end{aligned}$$

giving independence. □

Negative Dependence Properties. An *upward event*, $\mathcal{A}$, on 2^E is a collection of subsets of E that is closed under upward containment, i.e. if $A \in \mathcal{A}$ and $A \subseteq B \subseteq E$, then $B \in \mathcal{A}$. Similarly, a *downward event* is closed under downward containment. An *increasing function* $f : 2^E \to \mathbb{R}$, is a function where for any $A \subseteq B \subseteq E$, we have $f(A) \leq f(B)$. We also say $f : 2^E \to \mathbb{R}$ is a *decreasing function* if $-f$ is an increasing function. So, an indicator of an upward event is an increasing function. For example, if E is the set of edges of a graph G, then the existence of a Hamiltonian cycle is an increasing function, and the 3-colorability of G is a decreasing function.

Definition 2.9 (Negative Association). *A measure $\mu \in \mathcal{B}_E$ is* negatively associated *if for any increasing functions $f, g : 2^E \to \mathbb{R}$, that depend on* disjoint *sets of edges,*

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

It is shown in [9, 21] that strongly Rayleigh measures are negatively associated.

Stochastic Dominance. For two measures $\mu, \nu : 2^E \to \mathbb{R}_{\geq 0}$, we say $\mu \preceq \nu$ if there exists a *coupling* $\rho : 2^E \times 2^E \to \mathbb{R}_{\geq 0}$ such that

$$\begin{aligned} \sum_B \rho(A, B) &= \mu(A), \forall A \in 2^E, \\ \sum_A \rho(A, B) &= \nu(B), \forall B \in 2^E, \end{aligned}$$

and for all A, B such that $\rho(A, B) > 0$ we have $A \subseteq B$ (coordinate-wise).

Theorem 2.10 (BBL). *If μ is strongly Rayleigh and μ_k, μ_{k+1} are well-defined, then $\mu_k \preceq \mu_{k+1}$.*

Note that in the above particular case the coupling ρ satisfies the following: For any $A, B \subseteq E$ where $\rho(A, B) > 0$, $B \supseteq A$ and $|B \smallsetminus A| = 1$, i.e., B has exactly one more element.

Let μ be a strongly Rayleigh measure on edges of G. Recall that for a set $A \subseteq E$, we write $A_T = |A \cap T|$ to denote the random variable indicating the number of edges in A chosen in a random sample T of μ. The following facts immediately follow from the negative association and stochastic dominance properties. We will use these facts repeatedly in this paper.

Fact 2.11. *Let μ be any SR distribution on E, then for any $F \subset E$, and any integer k*

1. *(Negative Association) If $e \notin F$, then $\mathbb{P}_\mu[e \mid F_T \geq k] \leq \mathbb{P}_\mu[e]$ and $\mathbb{P}_\mu[e \mid F_T \leq k] \geq \mathbb{P}_\mu[e]$*
2. *(Stochastic Dominance) If $e \in F$, then $\mathbb{P}_\mu[e \mid F_T \geq k] \geq \mathbb{P}_\mu[e]$ and $\mathbb{P}_\mu[e \mid F_T \leq k] \leq \mathbb{P}_\mu[e]$.*

Fact 2.12. *Let μ be a homogenous SR distribution on E. Then,*

- *(Negative association with homogeneity) For any $A \subseteq E$, and any $B \subseteq \overline{A}$*

$$\mathbb{E}_\mu[B_T \mid A_T = 0] \leq \mathbb{E}_\mu[B_T] + \mathbb{E}_\mu[A_T] \tag{5}$$

- *Suppose that μ is a spanning tree distribution. For $S \subseteq V$, let $q := |S| - 1 - \mathbb{E}_\mu[E(S)_T]$. We will say S is a tree if $S_T = |S| - 1$, as this implies $E(S)$ is a spanning tree of the vertices in S. For any $A \subseteq E(S)$, $B \subseteq \overline{E(S)}$,*

$$\mathbb{E}_\mu[B_T] - q \leq \mathbb{E}_\mu[B_T \mid S \text{ is a tree}] \leq \mathbb{E}_\mu[B_T]$$
(Negative association and homogeneity)
$$\mathbb{E}_\mu[A_T] \leq \mathbb{E}_\mu[A_T \mid S \text{ is a tree}] \leq \mathbb{E}_\mu[A_T] + q$$
(Stochastic dominance and tree)

Rank Sequence. The *rank sequence* of μ is the sequence

$$\mathbb{P}[|S| = 0], \mathbb{P}[|S| = 1], \ldots, \mathbb{P}[|S| = m],$$

where $S \sim \mu$. Let $g_\mu(\mathbf{z})$ be the generating polynomial of μ. The *diagonal specialization* of μ is the univariate polynomial

$$\bar{g}_\mu(z) := g_\mu(z, z, \ldots, z).$$

Observe that $\bar{g}(.)$ is the generating polynomial of the rank sequence of μ. It follows that if μ is SR then $\bar{g_\mu}$ is real rooted.

It is not hard to see that the rank sequence of μ corresponds to sum of independent Bernoullis iff $\bar{g_\mu}$ is real rooted. It follows that the rank sequence of an SR distributions has the law of a sum of independent Bernoullis. As a consequence, it follows (see [9, 16, 30]) that the rank sequence of any strongly Rayleigh measure is log concave (see below for the definition), unimodal, and its mode differs from the mean by less than 1.

Definition 2.13 (Log-concavity [9, Definition 2.8]). *A real sequence $\{a_k\}_{k=0}^m$ is log-concave if $a_k^2 \geq a_{k-1} \cdot a_{k+1}$ for all $1 \leq k \leq m-1$, and it is said to have no internal zeros if the indices of its non-zero terms form an interval (of non-negative integers).*

2.5 Sum of Bernoullis

In this section, we collect a number of properties of sums of Bernoulli random variables, omitting many for brevity.

Definition 2.14 (Bernoulli Sum Random Variable). *We say $BS(q)$ is a* Bernoulli-Sum *random variable if it has the law of a sum of independent Bernoulli random variables, say $B_1 + B_2 + \ldots + B_n$ for some $n \geq 1$, with $\mathbb{E}[B_1 + \cdots + B_n] = q$.*

We start with the following theorem of Hoeffding.

Theorem 2.15 ([32, Corollary 2.1]). *Let $g : \{0, 1, \ldots, n\} \to \mathbb{R}$ and $0 \leq q \leq n$ for some integer $n \geq 0$. Let $B_1, \ldots, B_n$ be n independent Bernoulli random variables with success probabilities $p_1, \ldots, p_n$, where $\sum_{i=1}^n p_n = q$ that minimizes (or maximizes)*

$$\mathbb{E}[g(B_1 + \cdots + B_n)]$$

over all such distributions. Then, $p_1, \ldots, p_n \in \{0, x, 1\}$ for some $0 < x < 1$. In particular, if only m of p_i's are nonzero and ℓ of p_i's are 1, then the rest of the $m - \ell$ are $\frac{q-\ell}{m-\ell}$.

Fact 2.16. *Let $B_1, \ldots, B_n$ be independent Bernoulli random variables each with expectation $0 \leq p \leq 1$. Then*

$$\mathbb{P}\left[\sum_i B_i \text{ even}\right] = \frac{1}{2}(1 + (1-2p)^n)$$

Proof. Note that

$$(p + (1-p))^n = \sum_{k=0}^{n} p^k (1-p)^{n-k} \binom{n}{k}$$
$$((1-p) - p)^n = \sum_{k=0}^{n} (-p)^k (1-p)^{n-k} \binom{n}{k}$$

Summing them up we get,

$$1 + (1-2p)^n = \sum_{0 \leq k \leq n,\, k \text{ even}} 2p^k (1-p)^{n-k} \binom{n}{k}.$$

□

Corollary 2.17. *Given a $BS(q)$ random variable with $0 < q \leq 1.2$, then*

$$\mathbb{P}[BS(q) \text{ even}] \leq \frac{1}{2}(1 + e^{-2q})$$

Proof. First, if $q \leq 1$, then by Hoeffding's theorem we can write $BS(q)$ as sum of n Bernoullis with success probability $p = q/n$. If $n = 1$, then the statement obviously holds. Otherwise, by the previous fact, we have (for some n),

$$\mathbb{P}[BS(q) \text{ even}] \leq \frac{1}{2}(1 + (1-2p)^n)) \leq \frac{1}{2}(1 + e^{-2q})$$

where we used that $|1 - 2p| \leq e^{-2p}$ for $p \leq 1/2$.

So, now assume $q > 1$. Write $BS(q)$ as the sum of n Bernoullis, each with success probabilities 1 or p. First assume we have no ones. Then, either we only have two non-zero Bernoullis with success probability $q/2$ in which case $\mathbb{P}[BS(q) \text{ even}] \leq 0.6^2 + 0.4^2$ and we are done. Otherwise, $n \geq 3$ so $p \leq 1/2$ and similar to the previous case we get $\mathbb{P}[BS(q) \text{ even}] \leq \frac{1}{2}(1 + e^{-2q})$.

Finally, if $q > 1$ and one of the Bernoullis is always 1, i.e. $BS(q) = BS(q-1) + 1$, then we get

$$\begin{aligned}\mathbb{P}[BS(q) \text{ even}] &= \mathbb{P}[BS(q-1) \text{ odd}] \\ &= \frac{1}{2}(1 - (1-2p)^{n-1}) \\ &\leq \frac{1}{2}(1 - e^{-4(q-1)}) \leq 0.3\end{aligned}$$

where we used that $1 - x \geq e^{-2x}$ for $0 \leq x \leq 0.2$. □

2.6 Random Spanning Trees

Lemma 2.18. *Let $G = (V, E, x)$, and let μ be any distribution over spanning trees with marginals x. For any ϵ-near min cut $S \subseteq V$ (such that none of the endpoints of $e_0 = (u_0, v_0)$ are in S), we have*

$$\mathbb{P}_{T \sim \mu}[T \cap E(S) \text{ is a tree}] \geq 1 - \epsilon/2.$$

Moreover, if μ is a max-entropy distribution with marginals x, then for any set of edges $A \subseteq E(S)$ and $B \subseteq E \smallsetminus E(S)$,

$$\mathbb{E}[A_T] \leq \mathbb{E}[A_T \mid S \text{ is a tree}] \leq \mathbb{E}[A_T] + \epsilon/2,$$
$$\mathbb{E}[B_T] - \epsilon/2 \leq \mathbb{E}[B_T \mid S \text{ is a tree}] \leq \mathbb{E}[B_T].$$

Proof. First, observe that

$$\mathbb{E}[E(S)_T] = x(E(S)) \geq \frac{2|S| - x(\delta(S))}{2} \geq |S| - 1 - \epsilon/2,$$

where we used that since $u_0, v_0 \notin S$, and that for any $v \in S$, $\mathbb{E}[\delta(v)_T)] = x(\delta(v)) = 2$.

Let $p_S = \mathbb{P}[S \text{ is tree}]$. Then, we must have

$$\begin{aligned}|S| - 1 - (1 - p_S) &= p_S(|S| - 1) + (1 - p_S)(|S| - 2) \\ &\geq \mathbb{E}[E(S)_T] \geq |S| - 1 - \epsilon/2.\end{aligned}$$

Therefore, $p_S \geq 1 - \epsilon/2$.

The second part of the claim follows from Fact 2.12. □

Corollary 2.19. *Let $A, B \subseteq V$ be disjoint sets such that $A, B, A \cup B$ are $\epsilon_A, \epsilon_B, \epsilon_{A \cup B}$-near minimum cuts w.r.t., x respectively, where none of them contain endpoints of e_0. Then for any distribution μ of spanning trees on E with marginals x,*

$$\mathbb{P}_{T \sim \mu}[E(A, B)_T = 1] \geq 1 - (\epsilon_A + \epsilon_B + \epsilon_{A \cup B})/2.$$

Proof. By the union bound, with probability at least $1 - (\epsilon_A + \epsilon_B + \epsilon_{A \cup B})/2$, A, B, and $A \cup B$ are trees. But this implies that we must have exactly one edge between A, B. □

The following simple fact also holds by the union bound.

Fact 2.20. *Let $G = (V, E, x)$ and let μ be a distribution over spanning trees with marginals x. For any set $A \subseteq E$, we have*

$$\mathbb{P}_{T \sim \mu}[T \cap A = \emptyset] \geq 1 - x(A).$$

For brevity omit the remaining facts about random spanning trees.

3 OVERVIEW OF PROOF

In the remainder of this extended abstract we will sketch the main ideas of the proof.

As alluded to earlier, the crux of the proof of Theorem 1.1 is to show that the expected cost of the minimum cost matching on the odd degree vertices of the sampled tree is at most $OPT(1/2-\epsilon)$. We do this by showing the existence of a cheap feasible O-join solution to (4).

First, recall that if we only wanted to get an O-join solution of value at most $OPT/2$, to satisfy all cuts, it is enough to set $y_e := x_e/2$ for each edge [48]. To do better, we want to take advantage of the fact that we only need to satisfy a constraint in the O-join for S when $\delta(S)_T$ is odd. Here, we are aided by the fact that the sampled tree is likely to have many even cuts because it is drawn from a Strong Rayleigh distribution.

If an edge e is exclusively on even cuts then y_e can be reduced below $x_e/2$. This, more or less, was the approach in [39] for graphic TSP, where it was shown that a constant fraction of LP edges will be exclusively on *even* near min cuts with constant probability. The difficulty in implementing this approach in the metric case comes from the fact that a high cost edge can be on many cuts and it may be exceedingly unlikely that *all* of these cuts will be even simultaneously. Overall, our approach to addressing this is to start with $y_e := x_e/2$ and then modify it with a random[4] *slack vector* $s : E \to \mathbb{R}$: When certain special (few) cuts that e is on are even we let $s_e = -x_e\eta/8$ (for a carefully chosen constant $\eta > 0$); for other cuts that contain e, whenever they are odd, we will increase the slack of other edges on that cut to satisfy them. The bulk of our effort is to show that we can do this while guaranteeing that $\mathbb{E}[s_e] < -\epsilon\eta x_e$ for some $\epsilon > 0$.

One thing we do not need to worry about if we perform the reduction just described is any cut S such that $x(\delta(S)) > 2(1+\eta)$. Since we always have $s_e \geq -x_e\eta/8$, any such cut is always satisfied, even if every edge in $\delta(S)$ is decreased and no edge is increased.

Let OPT be the optimum TSP tour, i.e., a Hamiltonian cycle, with set of edges E^*; throughout the paper, we write e^* to denote an edge in E^*. To bound the expected cost of the O-join for a random spanning tree $T \sim \mu_\lambda$, we also construct a random slack vector $s^* : E^* \to \mathbb{R}_{\geq 0}$ such that $(x+OPT)/4 + s + s^*$ is a feasible for Eq. (4) with probability 1. In Section 3.1 we explain how to use s^* to satisfy all but a linear number of near mincuts.

Theorem 3.1 (Main Technical Theorem). *Let x^0 be a solution of LP (1) with support $E_0 = E \cup \{e_0\}$, and x be x^0 restricted to E. Let $z := (x+OPT)/2$, $\eta \leq 10^{-12}$ and let μ be the max-entropy distribution with marginals x. Also, let E^* denote the support of OPT. There are two functions $s : E_0 \to \mathbb{R}$ and $s^* : E^* \to \mathbb{R}_{\geq 0}$ (as functions of $T \sim \mu$), , such that*

i) For each edge $e \in E$, $s_e \geq -x_e\eta/8$.
ii) For each η-near-min-cut S of z, if $\delta(S)_T$ is odd, then $s(\delta(S)) + s^*(\delta(S)) \geq 0$.
iii) For every OPT edge e^*, $\mathbb{E}\left[s^*_{e^*}\right] \leq 45\eta^2$ and for every LP edge $e \neq e_0$, $\mathbb{E}[s_e] \leq -x_e\epsilon_P\eta/2$ for some small ϵ_P.

[4]where the randomness comes from the random sampling of the tree

In the next subsection, we explain the main ideas needed to prove this technical theorem. But first, we show how our main theorem follows readily from Theorem 3.1.

Proof of Theorem 1.1. Let x^0 be an extreme point solution of LP (1), with support E_0 and let x be x^0 restricted to E. By Fact 2.1 x is in spanning tree polytope. Let $\mu = \mu_{\lambda^*}$ be the max entropy distribution with marginals x, and let s, s^* be as defined in Theorem 3.1. We will define $y : E_0 \to \mathbb{R}_{\geq 0}$ and $y^* : E^* \to \mathbb{R}_{\geq 0}$. Let

$$y_e = \begin{cases} x_e/4 + s_e & \text{if } e \in E \\ \infty & \text{if } e = e_0 \end{cases}$$

we also let $y^*_{e^*} = 1/4 + s^*_{e^*}$ for any edge $e^* \in E^*$. We will show that $y + y^*$ is a feasible solution[5] to (4). First, observe that for any S where $e_0 \in \delta(S)$, we have $y(\delta(S)) + y^*(\delta(S)) \geq 1$. Otherwise, we assume $u_0, v_0 \notin S$. If S is an η-near min cut w.r.t., z and $\delta(S)_T$ is odd, then by property (ii) of Theorem 3.1, we have

$$y(\delta(S)) + y^*(\delta(S)) = \frac{z(\delta(S))}{2} + s(\delta(S)) + s^*(\delta(S)) \geq 1.$$

On the other hand, if S is not an η-near min cut (w.r.t., z),

$$\begin{aligned} y(\delta(S)) + y^*(\delta(S)) &\geq \frac{z(\delta(S))}{2} - \frac{\eta}{8}x(\delta(S)) \\ &\geq \frac{z(\delta(S))}{2} - \frac{\eta}{8}2(z(\delta(S)) - 1) \\ &\geq z(\delta(S))(1/2 - \eta/4) + \eta/4 \\ &\geq (2+\eta)(1/2 - \eta/4) + \eta/4 \geq 1. \end{aligned}$$

where in the first inequality we used property (i) of Theorem 3.1 which says that $s_e \geq x_e\eta/8$ with probability 1 for all LP edges and that $s^*_{e^*} \geq 0$ with probability 1. In the second inequality we used that $z = (x + OPT)/2$, so, since $OPT \geq 2$ across any cut, $x(\delta(S)) \leq 2(z(\delta(S)) - 1)$. Therefore, $y + y^*$ is a feasible O-join solution.

Finally, using $c(e_0) = 0$ and part (iii) of Theorem 3.1,

$$\begin{aligned} \mathbb{E}\left[c(y) + c(y^*)\right] &= OPT/4 + c(x)/4 + \mathbb{E}\left[c(s) + c(s^*)\right] \\ &\leq OPT/4 + c(x)/4 + 45\eta^2 OPT - \epsilon_P\eta c(x)/2 \\ &\leq (1/2 - \epsilon_P\eta/4)OPT \end{aligned}$$

choosing η such that $45\eta = \epsilon_P/4.1$ and that $c(x) \leq OPT$.

Now, we are ready to bound approximation factor of our algorithm. First, since x^0 is an extreme point solution of (1), $\min_{e \in E_0} x^0_e \geq \frac{1}{n!}$. So, by Theorem 1.3, in polynomial time we can find $\lambda : E \to \mathbb{R}_{\geq 0}$ such that for any $e \in E$, $\mathbb{P}_{\mu_\lambda}[e] \leq x_e(1+\delta)$ for some δ that we fix later. It follows that

$$\sum_{e \in E} | \mathbb{P}_\mu[e] - \mathbb{P}_{\mu_\lambda}[e] | \leq n\delta.$$

By stability of maximum entropy distributions (see [45, Thm 4] and references therein), we have that $\|\mu - \mu_\lambda\|_1 \leq O(n^4\delta) =: q$. Therefore, for some $\delta \ll n^{-4}$ we get $\|\mu - \mu_\lambda\|_1 = q \leq \frac{\epsilon_P\eta}{100}$. That

[5]Recall that we merely need to prove the *existence* of a cheap O-join solution. The actual optimal O-join solution can be found in polynomial time.

means that

$$\mathbb{E}_{T\sim\mu_\lambda}[\text{min cost matching}] \le \mathbb{E}_{T\sim\mu}\left[c(y)+c(y^*)\right] + q(OPT/2) \le \left(\frac{1}{2} - \frac{\epsilon_P\eta}{4} + \frac{\epsilon_P\eta}{100}\right)OPT,$$

where we used that for any spanning tree the cost of the minimum cost matching on odd degree vertices is at most $OPT/2$. Finally, since $\mathbb{E}_{T\sim\mu_\lambda}[c(T)] \le OPT(1+\delta)$ and $\epsilon_P = 3.9\cdot 10^{-17}$ we get a $3/2 - 2\cdot 10^{-36}$ approximation algorithm for TSP. □

3.1 Ideas Underlying the Proof of Theorem 3.1

The first step of the proof is to show that it suffices to construct a slack vector s for a "cactus-like" structure of near min-cuts that we call a *hierarchy*. Informally, a hierarchy $\mathcal{H}$ is a laminar family of mincuts[6], consisting of two types of cuts: *triangle cuts* and *degree cuts*. A triangle S is the union of two min-cuts X and Y in $\mathcal{H}$ such that $x(E(X,Y)) = 1$. See Fig. 1 for an example of a hierarchy with three triangles.

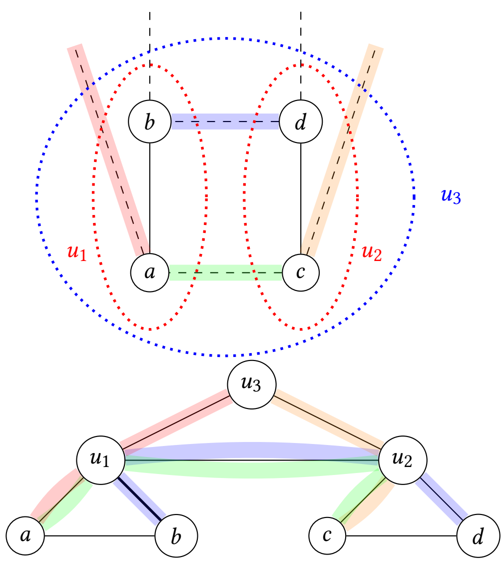


Figure 1: An example of part of a hierarchy with three triangles. The graph on the left shows part of a feasible LP solution where dashed (and sometimes colored) edges have fraction $1/2$ and solid edges have fraction 1. The dotted ellipses on the left show the min-cuts u_1, u_2, u_3 in the graph. (Each vertex is also a min-cut). On the right is a representation of the corresponding hierarchy. Triangle u_1 corresponds to the cut $\{a, b\}$, u_2 corresponds to $\{c, d\}$ and u_3 corresponds to $\{a, b, c, d\}$. Note that, for example, the edge (a, c), represented in green, is in $\delta(u_1)$, $\delta(u_3)$, and inside u_3. For triangle u_1, we have $A = \delta(a) \smallsetminus (a, b)$ and $B = \delta(b) \smallsetminus (b, d)$.

We will refer to the set of edges $E(X,\overline{S})$ (resp. $E(Y,\overline{S})$) as A (respectively B) for a triangle cut S. In addition, we say a triangle cut S is *happy* if A_T and B_T are both odd. All other cuts are called degree cuts. A degree cut S is *happy* if $\delta(S)_T$ is even.

[6] This is really a family of near-min-cuts, but for the purpose of this overview, assume $\eta = 0$

Theorem 3.2 (Main Payment Theorem (informal)). *Let $G = (V, E, x)$ for LP solution x and let μ be the max-entropy distribution with marginals x. Given a hierarchy $\mathcal{H}$, there is a slack vector $s : E \to \mathbb{R}$ such that*

i) For each edge $e \in E$, $s_e \ge -x_e\eta/8$.
ii) For each cut $S \in \mathcal{H}$ if $\delta(S)_T$ is not happy, then $s(\delta(S)) \ge 0$.
iii) For every LP edge $e \ne e_0$, $\mathbb{E}[s_e] \le -\eta\epsilon_P x_e$ for $\epsilon_P > 0$.

In the following subsection, we discuss how to prove this theorem. Here we explain at a high level how to define the hierarchy and reduce Theorem 3.1 to this theorem.

First, observe that, given Theorem 3.2, cuts in $\mathcal{H}$ will automatically satisfy (ii) of Theorem 3.1. The approach we take to satisfying all other cuts is to introduce additional slack, the vector s^*, on OPT edges.

Consider the set of all near-min-cuts of z, where $z := (x+OPT)/2$. Starting with z rather than x allows us to restrict attention to a significantly *more structured* collection of near-min-cuts. The key observation here is that in OPT, all min-cuts have value 2, and any non-min-cut has value *at least* 4. Therefore averaging x with OPT guarantees that every η-near min-cut of z must consist of a *contiguous sequence of vertices (an interval) along the OPT cycle.* Moreover, each of these cuts is a 2η-near min-cut of x. Arranging the vertices in the OPT cycle around a circle, we identify every such cut with the interval of vertices that does not contain (u_0, v_0). Also, we say that a cut is crossed on both sides if it is crossed on the left and on the right.

To ensure that any cut S that is *crossed on both sides* is satisfied, we first observe that S is odd with probability $O(\eta)$. To see this, let S_L and S_R be the cuts crossing S on the left and right with minimum intersection with S and consider the two (bad) events $\{E(S\cap S_L, S_L\smallsetminus S))_T \ne 1\}$ and $\{E(S\cap S_R, S_R\smallsetminus S))_T \ne 1\}$. Recall that if A, B and $A\cup B$ are all near-min-cuts, then $\mathbb{P}[E(A,B)_T \ne 1] = O(\eta)$ (see Corollary 2.19). Applying this fact to the two aforementioned bad events implies that each of them has probability $O(\eta)$. Therefore, we will let the two OPT edges in $\delta(S)$ be responsible for these two events, i.e., we will increase the slack s^* on these two OPT edges by $O(\eta)$ when the respective bad events happens. This gives $\mathbb{E}[s^*(e^*)] = O(\eta^2)$ for each OPT edge e^*. As we will see, this simple step will reduce the number of near-min-cuts of z that we need to worry about satisfying to $O(n)$.

Next, we consider the set of near-min-cuts of z that are crossed on at most one side. Partition these into maximal connected components of crossing cuts. Each such component corresponds to an interval along the OPT cycle and, by definition, these intervals form a laminar family.

A single connected component C of at least two crossing cuts is called a *polygon*. We prove the following structural theorem about the polygons induced by z:

Theorem 3.3 (Polygons look like cycles (Informal version of polygon structure theorem)). *Given a connected component C of near-min-cuts of z that are crossed on one side, consider the coarsest partition of vertices of the OPT cycle into a sequence $a_1, \ldots, a_{m-1}$ of sets called atoms (together with a_0 which is the set of vertices not contained in any cut of C). Then*

- *Every cut in C is the union of some number of consecutive atoms in $a_1, \ldots, a_{m-1}$.*

- *For each i such that $0 \le i < m-1$, $x(E(a_i, a_{i+1})) \approx 1$ and similarly $x(E(a_{m-1}, a_0)) \approx 1$.*
- *For each $i > 0$, $x(\delta(a_i)) \approx 2$.*

The main observation used to prove Theorem 3.3 is that the cuts in C crossed on one side can be partitioned into two laminar families $\mathcal{L}$ and $\mathcal{R}$, where $\mathcal{L}$ (resp. $\mathcal{R}$) is the set of cuts crossed on the left (resp. right). This immediately implies that $|C|$ is linear in m. Since cuts in $\mathcal{L}$ cannot cross each other (and similarly for $\mathcal{R}$), the proof boils down to understanding the interaction between $\mathcal{L}$ and $\mathcal{R}$.

The approximations in Theorem 3.3 are correct up to $O(\eta)$. Using additional slack in *OPT*, at the cost of an additional $O(\eta^2)$ for edge, we can treat these approximate equations as if they are exact. Observe that if $x(E(a_i, a_{i+1})) = 1$, and $x(\delta(a_i)) = x(\delta(a_{i+1})) = 2$ for $1 \le i \le m-2$, then with probability 1, $E(a_i, a_{i+1})_T = 1$. Therefore, any cut in C which doesn't include a_1 or a_{m-1} is even with probability 1. The cuts in C that contain a_1 are even precisely[7] when $E(a_0, a_1)_T = 1$ and similarly the cuts in C that contain a_{m-1} are even when $E(a_0, a_{m-1})_T = 1$. These observations are what allow us to imagine that each polygon is a triangle, i.e., assume $m = 3$.

The hierarchy $\mathcal{H}$ is the set of all η-near mincuts of z that are not crossed at all (these will be the degree cuts), together with a triangle for every polygon. In particular, for a connected component C of size more than 1, the corresponding triangle cut is $a_1 \cup \ldots \cup a_{m-1}$, with $A = E(a_0, a_1)$ and $B = E(a_0, a_{m-1})$. Observe that from the discussion above, when a triangle cut is happy, then all of the cuts in the corresponding polygon C are even.

Summarizing, we show that if we can construct a good slack vector s for a hierarchy of degree cuts and triangles, then there is a nonnegative slack vector s^*, that satisfies all near-minimum cuts of z not represented in the hierarchy, while maintaining slack for each OPT edge e^* such that $\mathbb{E}[s^*(e^*)] = O(\eta^2)$.

Remarks: The reduction that we sketched above only uses the fact that μ is an arbitrary distribution of spanning trees with marginals x and not necessarily a maximum-entropy distribution.

We also observe that to prove Theorem 1.1, we crucially used that $45\eta \ll \epsilon$. This forces us to take η very small, which is why we get only a "very slightly" improved approximation algorithm for TSP. Furthermore, since we use OPT edges in our construction, we don't get a new upper bound on the integrality gap. We leave it as an open problem to find a reduction to the "cactus" case that does not involve using a slack vector for OPT (or a completely different approach).

3.2 Proof Ideas for Theorem 3.2

We now address the problem of constructing a good slack vector s for a hierarchy of degree cuts and triangle cuts. For each LP edge f, consider the lowest cut in the hierarchy, that contains both endpoints of f. We call this cut $\mathrm{p}(f)$. If $\mathrm{p}(f)$ is a degree cut, then we call f a *top edge* and otherwise, it is a *bottom edge*[8]. We will see that bottom edges are easier to deal with, so we start by discussing the slack vector s for top edges.

Let S be a degree cut and let $\mathbf{e} = (u, v)$ (where u and v are children of S in $\mathcal{H}$) be the set of all top edges $f = (u', v')$ such that $u' \in u$ and $v' \in v$. We call $\mathbf{e}$ a *top edge bundle* and say that u and v are the *top cuts* of each $f \in \mathbf{e}$. We will also sometimes say that $\mathbf{e} \in S$.

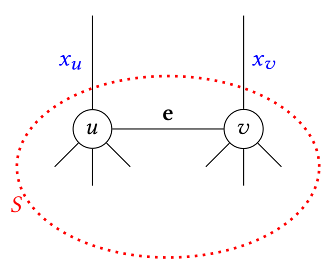
Ideally, our plan is to reduce the slack of every edge $f \in \mathbf{e}$ when it is *happy*, that is, both of its top cuts are even in T. Specifically, we will set $s_f := -\eta x_f$ when $\delta(u)_T$ and $\delta(v)_T$ are even. When this happens, we say that f is *reduced*, and refer to the event $\{\delta(u)_T, \delta(v)_T \text{ even}\}$ as the *reduction event* for f. Since this latter event doesn't depend on the actual endpoints of f, we view this as a simultaneous reduction of $s_{\mathbf{e}}$.

Now consider the situation from the perspective of the degree cut u (where $\mathrm{p}(u) = S$) and consider any incident edge bundle in S, e.g., $\mathbf{e} = (u, v)$. Either its top cuts are both even and $s_{\mathbf{e}} := -\eta x_{\mathbf{e}}$, or they aren't even, because, for example, $\delta(u)_T$ is odd. In this latter situation, edges in $\delta^{\uparrow}(u) := \delta(u) \cap \delta(S)$ might have been reduced (because *their* top two cuts are even), which a priori could leave $\delta(u)$ unsatisfied. In such a case, we *increase* $s_{\mathbf{e}}$ for edge bundles in $\delta^{\rightarrow}(u) := \delta(u) \smallsetminus \delta(S)$ to compensate for this reduction. Our main goal is then to prove is that for any edge bundle its expected reduction is greater than its expected increase. The next example shows this analysis in an ideal setting.

Example 3.4 (Simple case). Fix a top edge bundle $\mathbf{e} = (u, v)$ with $\mathrm{p}(\mathbf{e}) = S$. Let $x_u := x(\delta^{\uparrow}(u))$ and let $x_v := x(\delta^{\uparrow}(v))$. Suppose we have constructed a (fractional) *matching* between edges whose top two cuts are children of S in $\mathcal{H}$ and the edges in $\delta(S)$, and this matching satisfies the following three conditions: (a) $\mathbf{e} = (u, v) \in S$ is matched (only) to edges going higher from its top two cuts (i.e., to edges in $\delta^{\uparrow}(u)$ and $\delta^{\uparrow}(v)$), (b) $\mathbf{e}$ is matched to an $m_{\mathbf{e},u}$ fraction of every edge in $\delta^{\uparrow}(u)$ and to an $m_{\mathbf{e},v}$ fraction of each edge in $\delta^{\uparrow}(v)$, where

$$m_{\mathbf{e},u} + m_{\mathbf{e},v} = x_{\mathbf{e}},$$

and (c) the fractional value of edges in $\delta^{\rightarrow}(u) := \delta(u) \smallsetminus \delta^{\uparrow}(u)$ matched to edges in $\delta^{\uparrow}(u)$ is equal to x_u. That is, for each $u \in S$, $\sum_{\mathbf{f} \in \delta^{\rightarrow}(u)} m_{\mathbf{f},u} = x_u$.



The plan is for $\mathbf{e} \in S$ to be tasked with part of the responsibility for *fixing the cuts* $\delta(u)$ and $\delta(v)$ when they are odd and edges going higher are reduced. Specifically, $s_{\mathbf{e}}$ is increased to compensate for an $m_{\mathbf{e},u}$ fraction of the reductions in edges in $\delta^{\uparrow}(u)$ when $\delta(u)_T$ is odd. (And similarly for reductions in v.) Thus, we may compute

[7] Roughly, this corresponds to the definition of the polygon being left-happy.
[8] For example, in Fig. 1, $\mathrm{p}(a, c) = u_3$, and (a, c) is a bottom edge.

$\mathbb{E}[s_{\mathbf{e}}]$ as:

$$-\mathbb{P}[\mathbf{e} \text{ reduced}]\,\eta x_{\mathbf{e}} \quad (6)$$
$$+\, m_{\mathbf{e},u} \sum_{g\in\delta^{\uparrow}(u)} \mathbb{P}[\delta(u)_T \text{ odd} \mid g \text{ reduced}]\,\mathbb{P}[g \text{ reduced}]\,\eta \frac{x_g}{x(\delta^{\uparrow}(u))}$$
$$+\, m_{\mathbf{e},v} \sum_{g\in\delta^{\uparrow}(v)} \mathbb{P}[\delta(v)_T \text{ odd} \mid g \text{ reduced}]\,\mathbb{P}[g \text{ reduced}]\,\eta \frac{x_g}{x(\delta^{\uparrow}(v))} \quad (7)$$

We will lower bound $\mathbb{P}[\delta(u)_T \text{ even} \mid g \text{ reduced}]$. We can write this as

$$\mathbb{P}\left[\delta^{\rightarrow}(u)_T \text{ and } \delta^{\uparrow}(u)_T \text{ have same parity } \mid g \text{ reduced}\right].$$

Unfortunately, we do not currently have a good handle on the parity of $\delta^{\uparrow}(u)_T$ conditioned on g reduced. However, we can use the following simple but crucial property: Since $x(\delta(S)) = 2$, by Lemma 2.18, T consists of two independent trees, one on S and one on $V \smallsetminus S$, each with the corresponding marginals of x. Therefore, we can write

$$\mathbb{P}[\delta(u)_T \text{ even} \mid g \text{ reduced}]$$
$$\geq \min(\mathbb{P}\left[(\delta^{\rightarrow}(u))_T \text{ even}\right], \mathbb{P}\left[(\delta^{\rightarrow}(u))_T \text{ odd}\right]).$$

This gives us a reasonable bound when $\epsilon \leq x_u, x_v \leq 1-\epsilon$ since, because $x(\delta(u)) = x(\delta(v)) = 2$, by the SR property, $(\delta^{\rightarrow}(u))_T$ (and similarly $(\delta^{\rightarrow}(v))_T$) is the sum of Bernoulis with expectation in $[1+\epsilon, 2-\epsilon]$. From this it follows that

$$\min(\mathbb{P}\left[(\delta^{\rightarrow}(u))_T \text{ even}\right], \mathbb{P}\left[(\delta^{\rightarrow}(u))_T \text{ odd}\right]) = \Omega(\epsilon).$$

We can therefore conclude that $\mathbb{P}[\delta(u)_T \text{ odd} \mid g \text{ reduced}] \leq 1 - O(\epsilon)$.

The rest of the analysis of this special case follows from (a) the fact that our construction will guarantee that for *all* edges g, the probability that g is reduced is *exactly* p, i.e., it is the same for all edges, and (b) the fact that $m_{\mathbf{e},u}x_u + m_{\mathbf{e},v}x_v = x_{\mathbf{e}}$. Plugging these facts back into (7), gives

$$\mathbb{E}[s_{\mathbf{e}}] \leq -p\eta x_{\mathbf{e}} + m_{\mathbf{e},u}(1-\epsilon)p\eta + m_{\mathbf{e},v}(1-\epsilon)p\eta$$
$$\leq -p\eta x_{\mathbf{e}} + (1-\epsilon)p\eta x_{\mathbf{e}} = -\epsilon p\eta x_{\mathbf{e}}. \quad (8)$$

If we could prove (8) for *every* edge f in the support of x, that would complete the proof that the expected cost of the min O-join for a random spanning tree $T \sim \mu$ is at most $(1/2-\epsilon)OPT$.

Remark: Throughout this paper, we repeatedly use a mild generalization of the above "independent trees fact": that if S is a cut with $x(\delta(S)) \leq 2+\epsilon$, then S_T is very likely to be a tree. Conditioned on this fact, marginals inside S and outside S are nearly preserved and the trees inside S and outside S are sampled independently (see Lemma 2.18).

Ideal reduction: In the example, we were able to show that the quantity $\mathbb{P}[\delta(u)_T \text{ odd} \mid g \text{ reduced}]$ was bounded away from 1 for every edge $g \in \delta^{\uparrow}(u)$, and this is how we proved that the expected reduction for each edge was greater than the expected increase on each edge, yielding negative expected slack.

This motivates the following definition: A reduction for an edge g is k-*ideal* if, conditioned on g reduced, every cut S that is in the top k levels of cuts containing g is odd with probability that is bounded away from 1.

Moving away from an idealized setting: In Example 3.4, we over-simplified in four ways:

(a) We assumed that it would be possible to show that each top edge is *good*. That is, that its top two cuts are even *simultaneously* with constant probability.
(b) We considered only top edge bundles (i.e., edges whose top cuts were inside a degree cut).
(c) We assumed that $x_u, x_v \in [\epsilon, 1-\epsilon]$.
(d) We assumed the existence of a nice matching between edges whose top two cuts were children of S and the edges in $\delta(S)$.

Our proof needs to address all four anomalies that result from deviating from these assumptions.

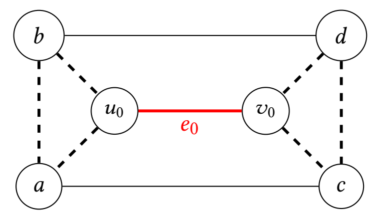


Figure 2: An Example with Bad Edges. A feasible solution of (1) is shown; dashed edges have fraction 1/2 and solid edges have fraction 1. Writing $E = E_0 \smallsetminus \{e_0\}$ as a maximum entropy distribution μ we get the following: Edges $(a,b), (c,d)$ must be completely negatively correlated (and independent of all other edges). So, $(b,u_0), (a,u_0)$ are also completely negatively correlated. This implies (a,b) is a bad edge.

Bad edges. Consider first (a). Unfortunately, it is not the case that all top edges are good. Indeed, some are *bad*. However, it turns out that bad edges are rare in the following senses: First, for an edge to be bad, it must be a half edge, where we say that an edge $\mathbf{e}$ is a half edge if $x_{\mathbf{e}} \in 1/2 \pm \epsilon_{1/2}$ for a suitably chosen constant $\epsilon_{1/2}$. Second, of any two half edge bundles sharing a common endpoint in the hierarchy, at least one is good. For example, in Fig. 2, (a,u_0) and (b,u_0) are good half-edge bundles. We advise the reader to ignore half edges in the first reading of the paper. Correspondingly, we note that our proofs would be much simpler if half-edge bundles never showed up in the hierarchy. It may not be a coincidence that half edges are hard to deal with, as it is conjectured that TSP instances with half-integral LP solutions are the hardest to round [40, 41].

Our solution is to *never* reduce bad edges. But this in turn poses two problems. First, it means that we need to address the possibility that the bad edges constitute most of the cost of the LP solution. Second, our objective is to get negative expected slack on each good edge and non-positive expected slack on bad edges. Therefore, if we never reduce bad edges, we can't increase them either, which means that the responsibility for fixing an odd cut with reduced edges going higher will have to be split amongst fewer edges (the incident good ones).

We deal with the first problem by showing that in every cut u in the hierarchy at least 3/4 of the fractional mass in $\delta(u)$ is good and these edges suffice to compensate for reductions on the edges going higher. Moreover, because there are sufficiently many good edges incident to each cut, we can show that either using the slack

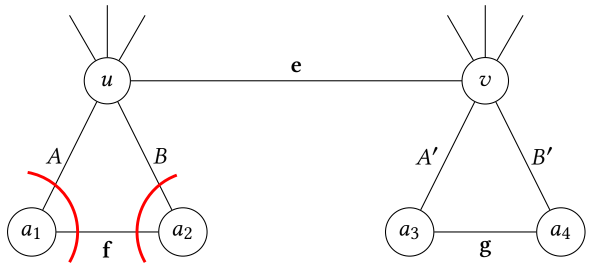


Figure 3: In the triangle u corresponding to the cut $\delta(a_1 \cup a_2)$, when A_T and B_T are odd, all 3 cuts ($\delta(a_1)_T, \delta(a_2)_T$ and $\delta(a_1 \cup a_2)_T = \delta(u)_T$ are odd (since $\mathbf{f}_T$ is always 1). (Recall also that the edges in the bundle e must have one endpoint in $\{a_1 \cup a_2\}$ and one endpoint in $\{a_3 \cup a_4\}$, as was the case, e.g., for the edge (a, c) in Fig. 1.)

vector $\{s_e\}$ gives us a low-cost O-join, or we can average it out with another O-join solution concentrated on bad edges to obtain a reduced cost matching of odd degree vertices.

We deal with the second problem by proving there is a matching between *good* edge bundles $\mathbf{e} = (u, v)$ and fractions $m_{\mathbf{e},u}, m_{\mathbf{e},v}$ of edges in $\delta^{\uparrow}(u), \delta^{\uparrow}(v)$ such that, roughly, $m_{\mathbf{e},u} + m_{\mathbf{e},v} = (1 + O(\epsilon_{1/2}))x_{\mathbf{e}}$.

Dealing with triangles. Turning to (b), consider a triangle cut S, for example $\delta(a_1 \cup a_2)$ in Fig. 3. Recall that in a triangle, we can assume that there is an edge of fractional value 1 connecting a_1 and a_2 in the tree, and this is why we defined the cut to be happy when A_T and B_T are odd: this guarantees that all 3 cuts defined by the triangle $(\delta(a_1), \delta(a_2), \delta(a_1 \cup a_2)$ are even.

Now suppose that $\mathbf{e} = (u, v)$ is a top edge bundle, where u and v are both triangles, as shown in Fig. 3. Then we'd like to reduce $s_{\mathbf{e}}$ when both cuts u and v are happy. But this would require more than simply both cuts being even. This would require *all* of A_T, B_T, A'_T, B'_T to be odd. Note that if, for whatever reason, $\mathbf{e}$ is reduced only when $\delta(u_1)_T$ and $\delta(u_2)_T$ are both even, then it could be, for example, that this only happens when A_T and B_T are both even. In this case, both $\delta(a_1)_T$ and $\delta(a_2)_T$ will be odd with probability 1 (recalling that $\mathbf{f}_T = 1$), which would then necessitate an increase in $s_{\mathbf{f}}$ whenever $\mathbf{e}$ is reduced. In other words, the reduction will not even be 1-ideal.

It turns out to be easier for us to get a 1-ideal reduction rule for $\mathbf{e}$ as follows: Say that $\mathbf{e}$ is *2-1-1 happy with respect to* u if $\delta(u)_T$ is even and both A'_T, B'_T are odd. We reduce $\mathbf{e}$ with probability $p/2$ when it is 2-1-1 happy with respect to u and with probability $p/2$ when it is 2-1-1 happy with respect to v. This means that when $\mathbf{e}$ is reduced, half of the time no increase in $s_{\mathbf{f}}$ is needed since u is happy. Similarly for v.

The 2-1-1 criterion for reduction introduces a new kind of bad edge: a half edge that is good, but not 2-1-1 good. We are able to show that non-half-edge bundles are 2-1-1 good, and that if there are two half edges which are both in A or are both in B, then at least one of them is 2-1-1 good. Finally, we show that if there are two half edges, where one is in A and the other is in B, and neither is 2-1-1 good, then we can apply a different reduction criterion that we call *2-2-2 good*. When the latter applies, we are guaranteed to decrease both of the half edge bundles simultaneously. All together, the various considerations discussed in this paragraph force us to come up with a relatively more complicated set of rules under which we reduce $s_{\mathbf{e}}$ for a top edge bundle $\mathbf{e}$ whose children are triangle cuts.

Bottom edge reduction. Next, consider a bottom edge bundle $\mathbf{f} = (a_1, a_2)$ where $\mathrm{p}(a_1) = \mathrm{p}(a_2)$ is a triangle. Our plan is to reduce $s_{\mathbf{f}}$ (i.e., set it to $-\eta x_{\mathbf{f}}$) when the triangle is happy, that is, $A_T = B_T = 1$. The good news here is that every triangle is happy with constant probability. However, when a triangle is *not* happy, $s_{\mathbf{f}}$ may need to increase to make sure that the O-join constraint for $\delta(a_1)$ and $\delta(a_2)$ are satisfied, if edges in A and B going higher are reduced. Since $x_{\mathbf{f}} = x(A) = x(B) = 1$, this means that $\mathbf{f}$ may need to compensate at *twice* the rate at which it is getting reduced. This would result in $\mathbb{E}[s_{\mathbf{f}}] > 0$, which is the opposite of what we seek.

We use two key ideas to address this problem. First, we reduce top edges and bottom edges by different amounts: Specifically, when the relevant reduction event occurs, we reduce a bottom edge $\mathbf{f}$ by $\beta x_{\mathbf{f}}$ and top edges $\mathbf{e}$ by $\tau x_{\mathbf{e}}$, where $\beta > \tau$ (and τ is a multiple of η).

Thus, the expected reduction in $s_{\mathbf{f}}$ is $p\beta x_{\mathbf{f}} = p\beta$, whereas the expected increase (due to compensation of, say, top edges going higher) is $p\tau(x(A) + x(B))q = p\tau 2q$, where

$$q = \mathbb{P}[\text{ triangle happy } | \text{ reductions in } A \text{ and } B] .$$

Thus, so long as $2\tau q < \beta - \epsilon$, we get the expected reduction in $s_{\mathbf{f}}$ that we seek.

The discussion so far suggests that we need to take τ smaller than $\beta/2q$, which is $\beta/2$ if q is 1, for example. On the other hand, if $\tau = \beta/2$, then when a top edge needs to fix a cut due to reductions on bottom edges, we have the opposite problem – their expected increase will be greater than their expected reduction, and we are back to square one.

Coming to our aid is the second key idea, already discussed in Section 1.2.3. We reduce bottom edges only when $A_T = B_T = 1$ *and* the marginals of edges in A, B are approximately preserved (conditioned on $A_T = B_T = 1$). This allows us to get much stronger upper bounds on the probability that a lower cut a bottom edge is on is odd, given that the bottom edge is reduced, and enables us to show that bottom edge reduction is ∞-ideal.

It turns out that the combined effects of (a) choosing $\tau = 0.571\beta$, and (b) getting better bounds on the probability that a lower cut is even given that a bottom edge is reduced, suffice to deal with the interaction between the reductions and the increases in slack for top and bottom edges.

Example 3.5. [Bottom-bottom case] To see how preserving marginals helps us handle the interaction between bottom edges at consecutive levels, consider a triangle cut $a'_1 = \{a_1, a_2\}$ whose parent cut $\hat{S} = \{a'_1, a'_2\}$ is also a triangle cut (as shown in Fig. 4). Let's analyze $\mathbb{E}[s_{\mathbf{f}}]$ where $\mathbf{f} = (a_1, a_2)$. Observe first that $A^{\rightarrow} \cup B^{\rightarrow}$ is a bottom edge bundle in the triangle $\hat{S}$ and all edges in this bundle are reduced simultaneously when $\hat{A}_T = \hat{B}_T = 1$ and marginals of all edges in $\hat{A} \cup \hat{B}$ are approximately preserved. (For the purposes of this overview, we'll assume they are preserved exactly). Let $x(A^{\uparrow}) = \alpha$. Then since $A = A^{\uparrow} \cup A^{\rightarrow}$ and $x(A) = 1$, we have $x(A^{\rightarrow}) = 1 - \alpha$. Moreover, since $\hat{A} = A^{\uparrow} \cup B^{\uparrow}$ and $x(\hat{A}) = 1$, we also have $x(B^{\uparrow}) = 1 - \alpha$ and $x(B^{\rightarrow}) = \alpha$.

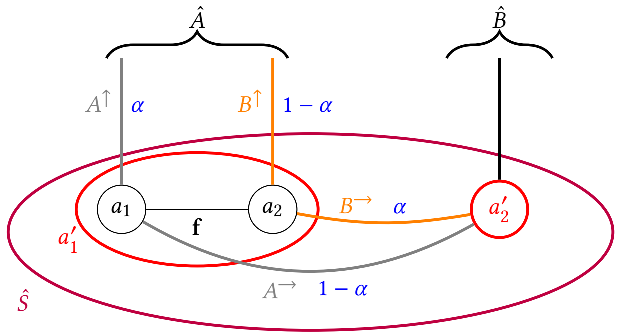


Figure 4: Setting of Example 3.5. Note that the set $A = \delta(a_1) \cap \delta(a_1')$ decomposes into two sets of edges, $A^{\uparrow}$, those that are also in $\delta(S)$, and the rest, which we call $A^{\rightarrow}$. Similarly for B.

Therefore, using the fact that when $A^{\rightarrow} \cup B^{\rightarrow}$ is reduced, exactly one edge in $A^{\uparrow} \cup B^{\uparrow}$ is selected (and also exactly one edge in $A^{\rightarrow} \cup B^{\rightarrow}$ is selected since it is a bottom edge bundle), and marginals are preserved given the reduction, we conclude that

$$\mathbb{P}\left[a_1' \text{ happy} \mid A^{\rightarrow} \cup B^{\rightarrow} \text{ reduced}\right]$$
$$= \mathbb{P}\left[A_T = B_T = 1 \mid A^{\rightarrow} \cup B^{\rightarrow} \text{ reduced}\right] = \alpha^2 + (1-\alpha)^2.$$

Now, we calculate $\mathbb{E}[s_{\mathbf{f}}]$. First, note that $\mathbf{f}$ may have to increase to compensate either for reduced edges in $A^{\uparrow} \cup B\uparrow$ or in $A^{\rightarrow} \cup B^{\rightarrow}$. For the sake of this discussion, suppose that $A^{\uparrow} \cup B^{\uparrow}$ is a set of top edges. Then, in the worst case we need to increase $\mathbf{f}$ by $p\tau$ in expectation to fix the cuts a_1, a_2 due to the reduction in $A^{\uparrow} \cup B^{\uparrow}$. Now, we calculate the expected increase due to the reduction in $A^{\rightarrow} \cup B^{\rightarrow}$. The crucial observation is that edges in $A^{\rightarrow} \cup B^{\rightarrow}$ are reduced simultaneously, so both cuts $\delta(a_1)$ and $\delta(a_2)$ can be fixed simultaneously by an increase in $s_{\mathbf{f}}$. Therefore, when they are both odd, it suffices for $\mathbf{f}$ to increase by

$$\max\{x(A^{\rightarrow}), x(B^{\rightarrow})\}\beta = \max\{\alpha, 1-\alpha\}\beta,$$

to fix cuts a_1, a_2. Putting this together, we get

$$\begin{aligned}\mathbb{E}[s_{\mathbf{f}}] &= -p\beta + \mathbb{E}\left[\text{increase due to } A^{\rightarrow} \cup B^{\rightarrow}\right] \\ &\quad + \mathbb{E}\left[\text{increase due to } A^{\uparrow} \cup B^{\uparrow}\right] \\ &\le -p\beta + p\beta \max_{\alpha \in [1/2,1]} \alpha[1-\alpha^2-(1-\alpha)^2] + p\tau\end{aligned}$$

which, since $\max_{\alpha\in[1/2,1]} \alpha[1-\alpha^2-(1-\alpha)^2] = 8/27$ and $\tau = 0.571\beta$ is

$$= p\beta(-1 + \frac{8}{27} + 0.571) = -0.13p\beta.$$

Dealing with x_u close to 1. [9] Now, suppose that $\mathbf{e} = (u, v)$ is a top edge bundle with $x_u := x(\delta^{\uparrow}(u))$ is close to 1. Then, the analysis in Example 3.4, bounding $r := \mathbb{P}[\delta(u)_T \text{ odd} \mid g \text{ reduced}]$ away from 1 for an edge $g \in \delta^{\uparrow}(u)$ doesn't hold. To address this, we consider two cases: The first case, is that the edges in $\delta^{\uparrow}(u)$ break up into many groups that end at different levels in the hierarchy. In this case, we can analyze r separately for the edges that end at any given level, taking advantage of the independence between the trees chosen at different levels of the hierarchy.

The second case is when nearly all of the edges in $\delta^{\uparrow}(u)$ end at the same level, for example, they are all in $\delta^{\rightarrow}(u')$ where p(u') is a degree cut. In this case, we introduce a more complex (2-1-1) reduction rule for these edges. The observation is that from the perspective of these edges u' is a "pseudo-triangle". That is, it looks like a triangle cut, with atoms u and $u' \smallsetminus u$ where $\delta(u) \cap \delta(u')$ corresponds to the "A"-side of the triangle.

Now, we define this more complex 2-1-1 reduction rule: Consider a top edge $\mathbf{f} = (u', v') \in \delta^{\rightarrow}(u')$. So far, we only considered the following reduction rule for $\mathbf{f}$: If both u', v' are degree cuts, $\mathbf{f}$ reduces when they are both even in the tree; otherwise if say u' is a triangle cut, $\mathbf{f}$ reduces when it is 2-1-1 good w.r.t., u' (and similarly for v'). But clearly these rules ignore the pseudo triangle. The simplest adjustment is, if u' is a pseudo triangle with partition $(u, u' \smallsetminus u)$, to require $\mathbf{f}$ to reduce when $A_T = B_T = 1$ and v' is happy. However, as stated, it is not clear that the sets A and B are well-defined. For example, u' could be an actual triangle or there could be multiple ways to see u' as a pseudo triangle only one of which is $(u, u' \smallsetminus u)$. Our solution is to find the *smallest* disjoint pair of cuts $a, b \subset u'$ in the hierarchy such that $x(\delta(a) \cap \delta(u')), x(\delta(b) \cap \delta(u')) \ge 1 - \epsilon_{1/1}$, where $\epsilon_{1/1}$ is a fixed universal constant, and then let $A = \delta(a) \cap \delta(u'), B = \delta(b) \cap \delta(u')$ and $C = \delta(u') \smallsetminus A \smallsetminus B$ (see Fig. 5 for an example). Then, we say $\mathbf{f}$ is 2-1-1 happy w.r.t., u' if $A_T = B_T = 1$ and $C_T = 0$.

A few observations are in order:

- Since u is a candidate for, say a, it must be that a is a descendent of u in the hierarchy (or equal to u). In addition, b cannot simultaneously be in u, since $a \cap b = \emptyset$ and $x(\delta(u) \cap \delta(u')) \le 1$ by Lemma 2.7. So, when $\mathbf{f}$ is 2-1-1 happy w.r.t. u' we get $(\delta(u) \cap \delta(u'))_T = 1$.
- If $u' = (X, Y)$ is a actual triangle cut, then we must have $a \subseteq X, b \subseteq Y$. So, when $\mathbf{f}$ is 2-1-1 happy w.r.t. u', we know that u' is a happy triangle, i.e., $(\delta(X) \cap \delta(u'))_T = 1$ and $(\delta(Y) \cap \delta(u'))_T = 1$.

Now, suppose for simplicity that all top edges in $\delta(u')$ are 2-1-1 good w.r.t. u'. Then, when an edge $g \in \delta(u) \cap \delta(u')$ is reduced, $(\delta(u) \cap \delta(u'))_T = 1$, so $\mathbb{P}[\delta(u)_T \text{ odd} \mid g \text{ reduced}]$ is at most

$$\le \mathbb{P}\left[E(u, u' \smallsetminus u)_T \text{ even} \mid g \text{ reduced}\right] \le 0.57,$$

since edges in $E(u, u' \smallsetminus u)$ are in the tree independent of the reduction and $\mathbb{E}[E(u, u' \smallsetminus u)_T] \approx 1$.

Dealing with x_u close to 0 and the matching. We already discussed how the matching is modified to handle the existence of bad edges. We now observe that we can handle the case $x_u \approx 0$ by further modifying the matching. The key observation is that in this case, $x(\delta^{\rightarrow}(u)) \gg x(\delta^{\uparrow}(u))$. Roughly speaking, this enables us to find a matching in which each edge in $\delta^{\rightarrow}(u)$ has to increase about half as much as would normally be expected to fix the cut of u. This eliminates the need to prove a nontrivial bound on $\mathbb{P}[\delta(u)_T \text{ odd} \mid g \text{ reduced}]$.

This completes the proof sketch.

[9] Some portions of this discussion might be easier to understand after reading the rest of the paper.

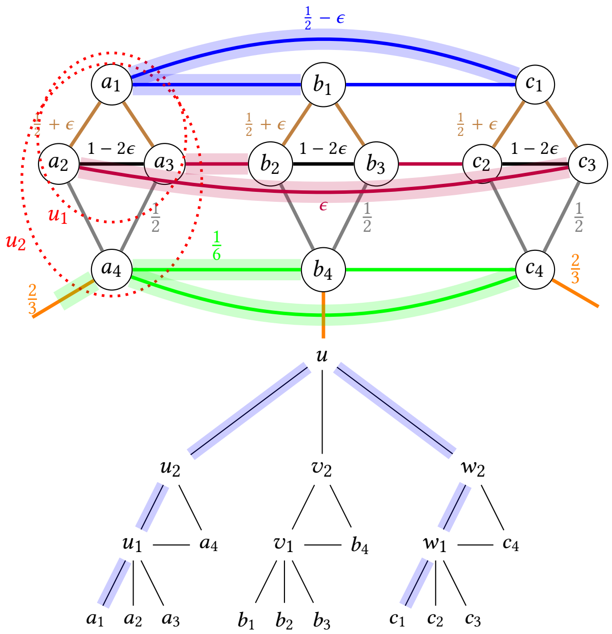


Figure 5: Part of the hierarchy of the graph is shown on top. Edges of the same color have the same fraction and $\epsilon \gg \eta$ is a small constant. u_1 corresponds to the degree cut $\{a_1, a_2, a_3\}$, u_2 corresponds to the triangle cut $\{u_1, a_4\}$ and u corresponds to the degree cut containing all of the vertices shown. Observe that edges in $\delta^{\uparrow}(a_1)$ are top edges in the degree cut u. If $\epsilon < \frac{1}{2}\epsilon_{1/1}$ then the (A, B, C)-degree partitioning of edges in $\delta(u_2)$ is as follows: $A = \delta(a_1) \cap \delta(u_2)$ are the blue highlighted edges each of fractional value $1/2 - \epsilon$, $B = \delta(a_4) \cap \delta(u_2)$ are the green highlighted edges of total fractional value 1, and C are the red highlighted edges each of fractional value ϵ. The cuts that contain edge (a_1, c_1) are highlighted in the hierarchy at the bottom.

ACKNOWLEDGEMENTS

The research of the first author was supported in part by NSF grant CCF-1813135 and the Air Force Office of Scientific Research grant FA9550-20-1-0212. The research of the second author was supported in part by NSF grants CCF-1813135, CCF-1552097 and CCF-1907845. The research of the third author was supported in part by Air Force Office of Scientific Research grant FA9550-20-1-0212, NSF grants CCF-1552097, CCF-1907845, ONR YIP grant N00014-17-1-2429, and a Sloan fellowship.

REFERENCES

[1] Nima Anari, Shayan Oveis Gharan, and Cynthia Vinzant. 2018. Log-Concave Polynomials, Entropy, and a Deterministic Approximation Algorithm for Counting Bases of Matroids. In *FOCS*, Mikkel Thorup (Ed.). IEEE Computer Society, 35–46.

[2] David L. Applegate, Robert E. Bixby, Vasek Chvatal, and William J. Cook. 2007. *The Traveling Salesman Problem: A Computational Study (Princeton Series in Applied Mathematics)*. Princeton University Press, Princeton, NJ, USA.

[3] Sanjeev Arora. 1996. Polynomial Time Approximation Schemes for Euclidean TSP and Other Geometric Problems. In *FOCS*. 2–11.

[4] Sanjeev Arora, Michelangelo Grigni, David Karger, Philip Klein, and Andrzej Woloszyn. 1998. A polynomial-time approximation scheme for weighted planar graph TSP. In *SODA*. 33–41.

[5] Arash Asadpour, Michel X. Goemans, Aleksander Madry, Shayan Oveis Gharan, and Amin Saberi. 2010. An O(log n/ log log n)-approximation Algorithm for the Asymmetric Traveling Salesman Problem. In *SODA*. 379–389.

[6] András A. Benczúr. 1995. A Representation of Cuts within 6/5 Times the Edge Connectivity with Applications. In *FOCS*. 92–102.

[7] Andras A. Benczúr. 1997. *Cut structures and randomized algorithms in edge-connectivity problems.* Ph.D. Dissertation. MIT.

[8] András A. Benczúr and Michel X. Goemans. 2008. Deformable Polygon Representation and Near-Mincuts. *Building Bridges: Between Mathematics and Computer Science, M. Groetschel and G.O.H. Katona, Eds., Bolyai Society Mathematical Studies* 19 (2008), 103–135.

[9] Julius Borcea, Petter Branden, and Thomas M. Liggett. 2009. Negative dependence and the geometry of polynomials. *Journal of American Mathematical Society* 22 (2009), 521–567.

[10] Sylvia Boyd and Robert Carr. 2011. Finding low cost TSP and 2-matching solutions using certain half-integer subtour vertices. *Discrete Optimization* 8, 4 (2011), 525 – 539. http://www.sciencedirect.com/science/article/pii/S1572528611000302

[11] S. Boyd and P. Elliott-Magwood. 2010. Structure of the extreme points of the subtour elimination polytope of the STSP. In *Combinatorial Optimization and Discrete Algorithms*, Vol. B23. 33–47.

[12] S. C. Boyd and William R. Pulleyblank. 1991. Optimizing over the subtour polytope of the travelling salesman problem. *Math. Program.* 49 (1991), 163–187.

[13] Robert D. Carr and Santosh Vempala. 2000. Towards a 4/3 approximation for the asymmetric traveling salesman problem. In *SODA*. 116–125.

[14] Nicos Christofides. 1976. *Worst Case Analysis of a New Heuristic for the Traveling Salesman Problem.* Report 388. Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, PA.

[15] G.B. Dantzig, D.R. Fulkerson, and S. Johnson. 1959. On a Linear Programming Combinatorial Approach to the Traveling Salesman Problem. *OR* 7 (1959), 58–66.

[16] J. N. Darroch. 1964. On the distribution of the number of successes in independent trials. *Ann. Math. Stat.* 36 (1964), 1317–1321.

[17] Erik D. Demaine, MohammadTaghi Hajiaghayi, and Bojan Mohar. 2007. Approximation algorithms via contraction decomposition. In *SODA*. 278–287.

[18] E.A. Dinits, A.V. Karzanov, and M.V. Lomonosov. 1976. On the structure of a family of minimal weighted cuts in graphs. *Studies in Discrete Mathematics (in Russian), ed. A.A. Fridman, 290-306, Nauka (Moskva)* (1976).

[19] Jack Edmonds. 1970. Submodular functions, matroids and certain polyhedra. In *Combinatorial Structures and Their Applications*. Gordon and Breach, New York, NY, USA, 69–87.

[20] Jack Edmonds and Ellis L. Johnson. 1973. Matching, Euler tours and the Chinese postman. *Mathematical Programming* 5, 1 (1973), 88–124.

[21] Tomás Feder and Milena Mihail. 1992. Balanced matroids. In *Proceedings of the twenty-fourth annual ACM symposium on Theory of Computing* (Victoria, British Columbia, Canada). ACM, New York, NY, USA, 26–38.

[22] David Gamarnik, Moshe Lewenstein, and Maxim Sviridenko. 2005. An improved upper bound for the TSP in cubic 3-edge-connected graphs. *Oper. Res. Lett.* 33, 5 (Sept. 2005), 467–474.

[23] Kyle Genova and David P. Williamson. 2017. An Experimental Evaluation of the Best-of-Many Christofides' Algorithm for the Traveling Salesman Problem. *Algorithmica* 78, 4 (2017), 1109–1130.

[24] Michel X. Goemans. 1995. Worst-Case Comparison of Valid Inequalities for the TSP. *MATH. PROG* 69 (1995), 335–349.

[25] M. Grigni, E. Koutsoupias, and C. Papadimitriou. 1995. An approximation scheme for planar graph TSP. In *FOCS '95: Proceedings of the 36th Annual Symposium on Foundations of Computer Science*. IEEE Computer Society, Washington, DC, USA, 640.

[26] Leonid Gurvits. 2006. Hyperbolic polynomials approach to Van der Waerden/Schrijver-Valiant like conjectures: sharper bounds, simpler proofs and algorithmic applications. In *STOC*, Jon M. Kleinberg (Ed.). ACM, 417–426.

[27] Leonid Gurvits. 2008. Van der Waerden/Schrijver-Valiant like Conjectures and Stable (aka Hyperbolic) Homogeneous Polynomials: One Theorem for all. *Electr. J. Comb.* 15, 1 (2008). http://www.combinatorics.org/Volume_15/Abstracts/v15i1r66.html

[28] Arash Haddadan and Alantha Newman. 2019. Towards Improving Christofides Algorithm for Half-Integer TSP. In *ESA (LIPIcs, Vol. 144)*, Michael A. Bender, Ola Svensson, and Grzegorz Herman (Eds.). Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 56:1–56:12.

[29] Arash Haddadan, Alantha Newman, and R. Ravi. 2017. Cover and Conquer: Augmenting Decompositions for Connectivity Problems. (2017). http://arxiv.org/abs/1707.05387 abs/1707.05387.

[30] G. H. Hardy, J. E. Littlewood, and G. Polya. 1952. *Inequalities.* Cambridge Univ. Press.

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

[32] W. Hoeffding. 1956. On the distribution of the number of successes in independent trials. *Ann. Math. Statist.* 27 (1956), 713–721.

[33] Anna R. Karlin, Nathan Klein, and Shayan Oveis Gharan. 2020. An improved approximation algorithm for TSP in the half integral case. In *STOC*, Konstantin Makarychev, Yury Makarychev, Madhur Tulsiani, Gautam Kamath, and Julia Chuzhoy (Eds.). ACM, 28–39.

[34] Marek Karpinski, Michael Lampis, and Richard Schmied. 2015. New inapproximability bounds for TSP. *J. Comput. System Sci.* 81, 8 (2015), 1665 – 1677.

[35] Philip N. Klein. 2005. A linear-time approximation scheme for planar weighted TSP. In *FOCS*. 647–657.

[36] Joseph SB Mitchell. 1999. Guillotine subdivisions approximate polygonal subdivisions: A simple polynomial-time approximation scheme for geometric TSP, k-MST, and related problems. *SIAM J. Comput.* 28, 4 (1999), 1298–1309.

[37] Tobias Moemke and Ola Svensson. 2011. Approximating Graphic TSP by Matchings. In *FOCS*. 560–569.

[38] M Mucha. 2012. $\frac{13}{9}$-approximation for graphic TSP.. In *STACS*. 30–41.

[39] Shayan Oveis Gharan, Amin Saberi, and Mohit Singh. 2011. A Randomized Rounding Approach to the Traveling Salesman Problem. In *FOCS*. IEEE Computer Society, 550–559.

[40] Frans Schalekamp, David P. Williamson, and Anke van Zuylen. 2012. A proof of the Boyd-Carr conjecture. In *SODA*. 1477–1486.

[41] Frans Schalekamp, David P. Williamson, and Anke van Zuylen. 2013. 2-Matchings, the Traveling Salesman Problem, and the Subtour LP: A Proof of the Boyd-Carr Conjecture. *Mathematics of Operations Research* 39, 2 (2013), 403–417.

[42] András Sebö and Jens Vygen. 2012. Shorter Tours by Nicer Ears:. (2012). CoRR abs/1201.1870.

[43] A. I. Serdyukov. 1978. O nekotorykh ekstremal'nykh obkhodakh v grafakh. *Upravlyaemye sistemy* 17 (1978), 76–79. http://nas1.math.nsc.ru/aim/journals/us/us17/us17_007.pdf

[44] D. B. Shmoys and D. P. Williamson. 1990. Analyzing the Held-Karp TSP bound: a monotonicity property with application. *Inf. Process. Lett.* 35, 6 (Sept. 1990), 281–285.

[45] Damian Straszak and Nisheeth K. Vishnoi. 2019. Maximum Entropy Distributions: Bit Complexity and Stability. In *COLT (Proceedings of Machine Learning Research, Vol. 99)*, Alina Beygelzimer and Daniel Hsu (Eds.). PMLR, 2861–2891.

[46] Vera Traub, Jens Vygen, and Rico Zenklusen. 2020. Reducing path TSP to TSP. In *STOC*, Konstantin Makarychev, Yury Makarychev, Madhur Tulsiani, Gautam Kamath, and Julia Chuzhoy (Eds.). ACM, 14–27.

[47] René van Bevern and Viktoriia A. Slugina. 2020. A historical note on the 3/2-approximation algorithm for the metric traveling salesman problem. (2020). https://arxiv.org/abs/2004.02437

[48] Laurence A. Wolsey. 1980. Heuristic analysis, linear programming and branch and bound. In *Combinatorial Optimization II*. Mathematical Programming Studies, Vol. 13. Springer Berlin Heidelberg, 121–134.