On Minimum Cost Rainbow Structures

Patrick Bennett[*]   Quentin Dubroff[†]   Alan Frieze[‡]
Wesley Pegden[§]

December 15, 2025

Abstract

We discuss the expected minimum cost of rainbow spanning trees
and Hamilton cycles in randomly edge colored random graphs.

1   Introduction

The minimum length of structures like spanning trees and Hamilton cycles
among random points in Euclidean space has been a topic of interest for the
better part of a century. Beardwood, Halton and Hammersley [3] showed that
the minimum cost of a tour through n random points in the unit square is
asymptotic to $\beta_{TSP} n^{1/2}$ with high probability (we say an event $\mathcal{E}_n$ occurs <u>with</u>

high probability or w.h.p. if $\mathbb{P}(\mathcal{E}_n) \to 1$ as $n \to \infty$). Here β_{TSP} is a positive constant, still unknown after more than 60 years. Steele [22] generalised this result to Euclidean Functionals and showed that the minimum cost of a spanning tree on n random points is asymptotic to $\beta_{MST}n^{1/2}$, w.h.p. In both cases, the cost of edges is determined by the Euclidean distance between points. Steele's work shows that there are many instances of optimisation problems where the expected minimum cost grows like $\beta n^{1/2}$ for unknown constants β. Frieze and Pegden [12] showed the constants for lengths of lower bound structures for the tour (like the minimum spanning tree, or twice a matching) really are distinct from the constant for the TSP, even though the best known explicit bounds on the constants generally overlap.

When the costs are generated independently, the situation becomes much clearer. In the case of spanning trees where the edges of the complete graph K_n are given independent uniform $[0, 1]$ random costs, Frieze [8] showed that $\mathbb{E}(L_n) \sim \zeta(3) = \sum_{k=1}^{\infty} \frac{1}{k^3}$, where L_n denotes the (random) minimum cost of a spanning tree.

The expected cost of the minimum cost Hamilton cycle in the complete graph K_n is asymptotic to $\tau = \frac{1}{2} \int_{x=0}^{\infty} y(x)dx$ where y is the positive solution to $\left(1 + \frac{x}{2}\right) e^{-x} + \left(1 + \frac{y}{2}\right) e^{-y} = 1$. This was proved by Wästlünd [25]. Frieze [10] showed that the expected minimum cost of a Hamilton cycle is asymptotically equal to the expected minimum cost of a 2-factor.

If we orient the edges then it follows from the work of Karp [16] and Aldous [1] that the expected minimum cost of a directed Hamilton cycle is asymptotic to $\zeta(2) = \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$.

In this work we add a spot of color to the problems. We assume that the edges of the graphs in question have been given uniformly random colors and ask for minimum cost rainbow Hamilton cycles or spanning trees. In this context, a set S of edges is rainbow colored if all the colors in S are distinct. We do this within two contexts: either we have n random points in $[0, 1]^2$ and the cost of an edge is the Euclidean distance between its endpoints, or else we have independent uniform $[0, 1]$ random costs.

In the first context we let $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ where the $\mathbf{x}_i$ are independently chosen uniformly from the unit square $[0, 1]^2$. The cost of edge $\{\mathbf{x}_i, \mathbf{x}_j\}$

is the Euclidean length $|\mathbf{x}_i - \mathbf{x}_j|$. Our first results concern the length Z_{MST} of the minimum rainbow spanning tree in this model, when we randomly color labeled from a set of $n - 1$ colors. (Note that with fewer colors, no rainbow spanning tree would exist. Also, even if we have enough colors there is always some small probability that there is no rainbow spanning tree at all, in which case Z_{MST} is undefined.) We let Z_{MST}^* denote the length of the minimum spanning tree (without the rainbow constraint); recall that $Z_{\mathrm{MST}}^* \sim \beta_{\mathrm{MST}} \sqrt{n}$ w.h.p., for some constant β_{MST}.

Theorem 1. *In the Euclidean setting, with edges labeled randomly by $n - 1$ colors, we have $Z_{\mathrm{MST}} = \Theta(\sqrt{n})$ w.h.p. and $Z_{\mathrm{MST}} - Z_{\mathrm{MST}}^* = \Theta(\sqrt{n})$ w.h.p.*

In particular, Theorem 1 says that the minimum length doesn't change by more than a constant factor under a rainbow constraint, but that a constant-factor change does necessarily occur. To put Theorem 1 in context, there is an important sense in which it is much less obvious that the minimum rainbow spanning tree should have length $\Theta(\sqrt{n})$ than it is for the various optimization problems (matchings, 2-factors, linear programming relaxations of the TSP, etc) to which the general methods derived from Beardwood, Halton, and Hammersley and from Steele have been applied. In particular, the rainbow condition seems to preclude a straightforward application of subadditivity, which explains, for example, why we do not yet know that $Z_{\mathrm{MST}}/\sqrt{n}$ has a limit. In some sense, the success of subadditivity approaches to structure lengths in random point-sets reflects the sense in which the asymptotic structure lengths really depend only on the local geometry of the ambient space. (For example, in the Beardwood-Halton-Hammersley theorem, the unit square can be replaced by any bounded open set of $\mathbb{R}^2$ of measure 1, and the asymptotic length of a tour through n points remains the same, with the same unknown constant β_{TSP}. But the rainbow condition we impose is a global constraint on the spanning tree that cannot be easily decomposed into noninteracting parts.

For the minimum length Z_{TSP} of a rainbow Hamilton cycle in the Euclidean model, we do not achieve an optimal result up to constant factors unless we allow a slight excess in the number of available colors.

Theorem 2. *In the Euclidean setting, with edges labeled randomly by $q = (1 + \varepsilon)n$ colors, we have (a) $Z_{\mathrm{TSP}} = O(n^{1/2} \log n)$, and (b) $Z_{\mathrm{TSP}} - Z_{\mathrm{TSP}}^* = $*

3

$\Omega(\sqrt{n})$ *w.h.p. Indeed, (a) holds for any set of n points with randomly colored edges.*

Here like before the starred version Z^*_{TSP} represents the minimum length in the absence of the rainbow condition.

In the second scenario, $[0,1]^2$ is replaced by the complete graph K_n and the costs $C(i,j)$ are independent uniform $[0,1]$ random variables. We prove a general theorem (see Theorem 11) for cheap rainbow substructures in randomly colored structures with random costs. The proof is an adaptation of some of the recent work on thresholds in spread hypergraphs (see [5]) based on Frankston, Kahn, Narayanan, and Park's recent proof of the fractional Kahn–Kalai conjecture [7]. Our Theorem 11 implies the following:

Theorem 3. *The minimum cost solution of the following problems have value $O(1)$ w.h.p. The proofs are non-constructive.*

(a) *Traveling Salesperson Problem in K_n with at least n colors for the edges.*

(b) *Minimum cost perfect matching in complete k-uniform hypergraphs, $k \geq 2$ with at least n/k colors for the edges.*

(c) *Minimum cost classes of degree bounded spanning trees in K_n with at least $n-1$ colors for the edges e.g. spanning binary trees.*

The rest of the paper is organized as follows. In Section 2 we prove the first assertion of Theorem 1, i.e. that $Z_{\text{MST}} = \Theta(\sqrt{n})$ w.h.p. In Section 3 we prove the first assertion of Theorem 2, i.e. that $Z_{\text{TSP}} = O(\sqrt{n}\log n)$ for n colors. In Section 4 we prove the remaining assertions in Theorems 1 and 2, i.e. that $Z_{\text{MST}} - Z^*_{\text{MST}} = \Theta(\sqrt{n})$ for n colors and that $Z_{\text{TSP}} - Z^*_{\text{TSP}} = \Theta(\sqrt{n})$ for $(1+\varepsilon)n$ colors. In Section 5 we prove Theorem 3. Finally, we make some observations and state open questions in Section 6.

4

2 Finding a rainbow spanning tree with $n-1$ available colors

In this section, we describe a simple polynomial time algorithm for finding a rainbow spanning tree of cost $O(n^{1/2})$. (We realise that finding a cheapest rainbow tree is solvable in polynomial time, see for example Lawler [17]. But the weighted matroid intersection algorithm is not so easy to analyse from a probabilistic point of view.) For $\mathbf{x} = (x, y) \in \mathcal{X}$ we let $\xi_1(\mathbf{x}) = x, \xi_2(\mathbf{x}) = y$. Assuming that $\mathbf{x}_1 = (x_1, y_1), \mathbf{x}_2 = (x_2, y_2), \ldots, \mathbf{x}_n = (x_n, y_n)$ are ordered in increasing value of ξ_2, let $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ and $\mathcal{X}_1 = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n-1}\}$.

We define a weighted bipartite (multi-)graph Γ with bipartition $\mathcal{X}_1 \cup C$ where $C = [n-1]$ is our set of colors. The edge set will be the union $E_1 \cup E_2$ of sets which we define next; for both sets E_1 and E_2, an edge $(\mathbf{x}, c)$ will correspond to some edge $(\mathbf{x}, \mathbf{x}')$ with $\xi_2(\mathbf{x}) < \xi_2(\mathbf{x}')$ in the embedded graph that has been assigned color c. This last property guarantees that a perfect matching in Γ corresponds to a rainbow spanning tree in the embedded graph. So to prove the first assertion in Theorem 1, it suffices to show that w.h.p.

$$\Gamma \text{ contains a perfect matching,} \tag{1}$$

and

$$\text{the total weight of all edges in } \Gamma \text{ is } O(n^{1/2}). \tag{2}$$

First we define the set of directed edges

$$E^+ = \{(\mathbf{x}, \mathbf{y}) \in \mathcal{X}_1 \times \mathcal{X} : \xi_2(\mathbf{y}) > \xi_2(\mathbf{x})\}.$$

For $\mathbf{x} \in \mathcal{X}$, we let $N(\mathbf{x}) = \{\mathbf{y} : (\mathbf{x}, \mathbf{y}) \in E^+\}$. We let K be a constant large enough to make the argument work and define E_1 as follows. For each $\mathbf{x} \in \mathcal{X}_1$, $c \in C$ there is an edge $e_{\mathbf{x},c} = \{\mathbf{x}, c\}$ in E_1 if c is one of the colors of the K shortest edges in $N(\mathbf{x})$ (noting that if $i > n - K$, then we include all edges $e_{\mathbf{x},c}$ where c is some color of an edge in $N(\mathbf{x}_i)$). We will prove that the total weight of the edges in E_1 is $O(n^{1/2})$ w.h.p. at the end of this section.

In addition, there is another set of edges E_2. We first define a set $A \subseteq \mathcal{X}_1$. For a large positive value B and $1 \leq j \leq L = \log^2 n$, we let

$$A_j = \left\{ \mathbf{x} \in \mathcal{X}_1 : \exists \mathbf{y} \text{ s.t.} (\mathbf{x}, \mathbf{y}) \in E^+ \text{ and } |\mathbf{x} - \mathbf{y}| \in \left[\frac{Bj^2}{\sqrt{n}}, \frac{B(j+1)^2}{\sqrt{n}} \right) \right\}.$$

We then let $A = \bigcap_{j=1}^{L} A_j$. We will show that A is most of $\mathcal{X}$, but first, we finish defining the edge set E_2.

Define a set E_A by choosing, for each $\mathbf{x} \in A$ and $j \in [L]$, one edge $(\mathbf{x}, \mathbf{y}) \in E^+$ incident to $\mathbf{x}$ of length in the interval

$$[Bj^2/\sqrt{n}, B(j+1)^2/\sqrt{n}). \tag{3}$$

The level of $e \in E_A$ is the unique j such that the length of e lies in $[Bj^2/\sqrt{n}, B(j+1)^2/\sqrt{n})$. We obtain E_2 by adding, for each color c, the edges $(\mathbf{x}, c)$, where $\mathbf{x}$ is a lower endpoint of one of the K lowest level edges of color c in E_A; here, we break ties by choosing a random ordering of the edges of each fixed level (and this random ordering is chosen independently for each distinct color). Note that if $|A|$, is large, as we will now show, then our choice of $L = \log^2 n$ guarantees that each color c has K incident edges in E_2 w.h.p.

We have that for any $j \in [L]$, if $\xi_2(\mathbf{x}) < 1 - n^{-2/5}$, then the probability that $\mathbf{x} \notin A_j$ is at most

$$\left(1 - \frac{2\pi B^2 j^3}{n}\right)^{n-1} \le e^{-6B^2 j^3}. \tag{4}$$

So,

$$\mathbb{P}(\mathbf{x} \notin A) \le \mathbb{P}(\mathbf{x} \notin A \,|\, \xi_2(\mathbf{x}) < 1 - n^{-2/5}) + n^{-2/5} \le \sum_j e^{-6B^2 j^3} + n^{-2/5}.$$

For large B (and n), we find

$$\mathbb{P}(\mathbf{x} \notin A) < e^{-5B^2}.$$

So

$$\mathbb{E}(|A|) \ge n - e^{-5B^2} n = n(1 - e^{-5B^2}). \tag{5}$$

We now use an inequality of Warnke to show that $|A|$ is large with very high probability.

Lemma 4 (Warnke [24]). *Let $W = (W_1, W_2, \ldots, W_n)$ be a family of independent random variables with W_i taking values in a set Λ_i. Let $\Omega = \prod_{i \in [n]} \Lambda_i$*

and suppose that $\mathcal{G} \subseteq \Omega$ and $f : \Omega \to \mathbb{R}$ are given. Suppose also that whenever $\mathbf{x}, \mathbf{x}' \in \Omega$ differ only in the i-th coordinate

$$|f(\mathbf{x}) - f(\mathbf{x}')| \leq \begin{cases} c_i & if\ \mathbf{x} \in \mathcal{G}. \\ d_i & otherwise. \end{cases}$$

If $Y = f(X)$, then for all reals $\gamma_i > 0$,

$$\mathbb{P}(Y \geq \mathbb{E}(Y) + t) \leq \exp\left\{-\frac{t^2}{2\sum_{i\in[n]}(c_i + \gamma_i(d_i - c_i))^2}\right\} + \mathbb{P}(W \notin \mathcal{G})\sum_{i\in[n]}\gamma_i^{-1}.$$
(6)

To use this lemma we need to define a "good" event $\mathcal{G}$. So, we let $\mathcal{G}$ be the event that for all $\mathbf{x}$,

$$\nu_{\mathbf{x}} = |\{\mathbf{y} : |\mathbf{x} - \mathbf{y}| \leq 2B\log^4 n/n^{1/2}\}| < 15B^2 \log^8 n.$$

The Chernoff bounds imply that

$$\mathbb{P}(\neg\mathcal{G}) = e^{-\Omega(\log^8 n)}.$$
(7)

To apply Lemma 4 we let $W_i = \mathbf{x}_i$ and $f(\mathcal{X}) = n - |A|$ and take $c_i = 30B^2 \log^8 n$ and $d_i = n$ for $i \in [n]$. Putting $\gamma_i = n^{-3}$ and our bound on $\mathbb{P}(\neg\mathcal{G})$ in (6), we see that $|A| > \mathbb{E}(|A|) - n^{2/3}$ w.h.p. We conclude that w.h.p.

$$|A| > (1 - e^{-5B^2})n.$$
(8)

The purpose of this construction of the edge set E_2 is to guarantee the following property.

Observation 5. *The (multi)-set of neighbors $\mathbf{x}$ for a color c in E_2 is symmetric with respect to permutation of labels of A.*

Proof. This follows from construction of the set E_2, since each vertex $\mathbf{x}$ is the lower endpoint of exactly one edge in $E_A(x)$ for each interval length. $\square$

To summarize, letting $N_1(\cdot)$ and $N_2(\cdot)$ refer to neighborhoods using edges from E_1 and E_2 respectively, the neighborhood sets $N_1(\mathbf{x}_i)$ ($i \in [n-1]$) are a collection of independent uniform random (multi)sets of C (of size $\min\{K, n-i\}$). And the edge (multi)sets $N_2(c)$ ($c \in C$) are uniform in the sense of Observation 5, and we will show in Claim 1 that they exhibit some approximate correlation in our favor.

We now show using Hall's theorem that (1) holds. We first show that $S \subseteq \mathcal{X}_1$ implies that $|N(S)| \geq |N_1(S)| \geq |S|$ for $|S| \leq n_0 := e^{-4/K}(n-1)$. We do this by finding a perfect matching between $\mathcal{X}' = \{x_{n-K}, \ldots, x_{n-1}\}$ and a set $Y \subset C$ and subsequently showing that for every S with $S \cap \mathcal{X}' = \emptyset$ and $|S| \leq n_0$, we have $|N_1(S) \setminus Y| \geq |S|$. Finding a matching covering $\mathcal{X}'$ just requires checking Hall's condition on $\mathcal{X}'$: each $S \subseteq \mathcal{X}'$ of size s has at least $\binom{s+1}{2}$ outgoing E_1 edges, so the probability there is a set $S \subseteq \mathcal{X}'$ violating Hall's condition is at most

$$\sum_{s=1}^{K} \binom{K}{s} \binom{|C|}{s-1} \left(\frac{s-1}{|C|}\right)^{\binom{s+1}{2}} \leq 2^K \sum_{s=2}^{K} \left(\frac{e|C|}{s-1} \left[\frac{s-1}{|C|}\right]^{(s+1)/2}\right)^s,$$

which is $O(1/n)$. Thus Hall's theorem implies the existence of a set $Y \subseteq C$ with a perfect matching to $\mathcal{X}'$. For $S \subset \mathcal{X}_1 \setminus \mathcal{X}'$, let $n_1 = n - K - 1$, note

$$\mathbb{P}(\exists S \subseteq \mathcal{X}_1 \setminus \mathcal{X}', |S| \leq n_0, |N_1(S) \setminus Y| \leq |S|) \leq \sum_{s=K}^{n_0} \binom{n_1}{s} \binom{n-1}{s} \left(\frac{s+K}{n-1}\right)^{Ks}$$

$$\leq \sum_{s=K}^{n_0} \left(\frac{(n-1)e}{s}\right)^{2s} \left(\frac{s+K}{n-1}\right)^{Ks}$$

$$\leq e^{K^2} \sum_{s=K}^{n_0} \left[e^2 \left(\frac{s}{n-1}\right)^{(K-2)}\right]^s$$

$$= O(n^{-(K-3)}),$$

where in the second-to-last line we used $(1 + K/s)^{Ks} \leq e^{K^2}$.

To check Halls condition for S with $|S| > n_0$, we look at the sizes of the E_2 neighborhoods of subsets of C. If $S \subseteq \mathcal{X}$ satisfies $|N(S)| \leq |S| - 1$ then for $T = C \setminus N(S)$ we have

$$|N_2(T)| \leq |N(T)| \leq n - 1 - |S| \leq n - 1 - |N(S)| - 1 = |T| - 1. \qquad (9)$$

Note also that
$$|T| \le n - 1 - n_0 \le n_2 := 4(n-1)/K. \tag{10}$$

Let the positions of the points $\mathcal{X}$ be denoted by $\mathcal{P}$. Assume now that $\mathcal{P}$ is given and that $|A| \ge (1 - e^{-5B^2})n$.

Claim 1. *Let $\mathcal{E}_c = \mathcal{E}_c(S)$ be the event that color c's E_2-neighbors are in some set $S \subseteq A$. Then for any fixed S, T, P,*

$$\mathbb{P}\left(\bigcap_{c \in T} \mathcal{E}_c(S) \ \middle| \ \mathcal{P} = P \right) \le e^{9|T|} \prod_{c \in T} \mathbb{P}(\mathcal{E}_c(S) \mid \mathcal{P} = P).$$

Proof. If $T = \{c_1, c_2, \ldots, c_\tau\}$ then

$$\mathbb{P}\left(\bigcap_{c \in T} \mathcal{E}_c \middle| \mathcal{P} = P \right) = \prod_{j=1}^{\tau} \mathbb{P}\left(\mathcal{E}_{c_j} \ \middle| \ \bigcap_{k<j} \mathcal{E}_{c_k}, \mathcal{P} = P \right).$$

Now write

$$\mathbb{P}\left(\mathcal{E}_{c_j} \ \middle| \ \bigcap_{k<j} \mathcal{E}_{c_k}, \mathcal{P} = P \right) = \sum_{\mathcal{H} \in H} \mathbb{P}(\mathcal{E}_{c_j} \mid \mathcal{H}, \mathcal{P} = P)\mathbb{P}\left(\mathcal{H} \mid \cap_{k<j}\mathcal{E}_{c_k}, \mathcal{P} = P \right),$$

where each <u>history</u> $\mathcal{H} \in H$ is an event (defined following (11)) consisting of the choices for the $\le K$ lowest level edges in each of the colors $c_1, \ldots, c_{j-1}$ (under each color's collection of random orderings of edges of each fixed level).

To prove our claim it suffices to show that for all choices of the event $\mathcal{H} \in H$,

$$\mathbb{P}(\mathcal{E}_{c_j} \mid \mathcal{H}, \mathcal{P} = P) \le e^9 \mathbb{P}(\mathcal{E}_{c_j} \mid \mathcal{P} = P) \tag{11}$$

For this, we write $\mathcal{H} = \cap_{i<j}\mathcal{H}_i$, where $\mathcal{H}_i$ is color i's random orderings of the edges of each fixed level as well as its first K edges in E_A under this ordering (which are all edges with lower endpoint in S). Note that each $\mathcal{H}_i$ introduces a set G_i of K edges in E_A which are colored by c_i and an additional set F_i of (usually linearly many) edges that may not receive the color c_i. We now let $L = (\mathbf{x}_i, \mathbf{y}_i), i = 1, 2, \ldots, |E_A|$ be a listing of the edges in E_A in increasing order of level, where edges of the same level are ordered uniformly at random, and we consider some history $\mathcal{H}_j$ that determines the event $\mathcal{E}_{c_j}$.

9

The probability $\mathbb{P}(\mathcal{H}_j \mid \mathcal{H}, \mathcal{P} = P)$ is determined by the outcome of an experiment $\Lambda_{\mathcal{H}} = \Lambda_{\mathcal{H}}(P)$ (the <u>conditioned experiment</u>) in which we consider the edges in L one at a time and choose its color according to the constraints imposed by the event $\mathcal{H}$: when we reach the tth edge, if $t \in G_i$ for some $i < j$, the edge is colored by c_i, and otherwise the tth edge is uniform over $C \setminus I_t$, where $I_t := \{c_i : t \in F_i\}$. We compare this conditioned experiment to the <u>auxiliary experiment</u> $\Lambda = \Lambda(P)$ in which we simply randomly assign random colors from C to edges in E_A; note that

$$\sum_{\mathcal{H}_j} \mathbb{P}(\mathcal{H}_j \mid \mathcal{P} = P) = \mathbb{P}(\mathcal{E}_{c_j} \mid \mathcal{P} = P),$$

so for (11), it suffices to show

$$\mathbb{P}(\mathcal{H}_j \mid \mathcal{H}, \mathcal{P} = P) \le e^9 \mathbb{P}(\mathcal{H}_j \mid \mathcal{P} = P) \qquad (12)$$

for every $\mathcal{H}_j$.

Defining $p = (n-1)^{-1}$, we have

$$\mathbb{P}(\mathcal{H}_j \mid \mathcal{P} = P) = p^K (1-p)^{|F_j|}.$$

We similarly calculate $\mathbb{P}(\mathcal{H}_j \mid \mathcal{H}, \mathcal{P} = P)$, the probability in the conditioned experiment, but now we have parameters $a_t = |I_t|$ and p_t, the probability that the tth edge receives color c_j in the conditioned experiment. The important fact is that $p_t = 0$ if $t \in \cup_{i<j} G_i$ and otherwise

$$p_t = (n-1-a_t)^{-1} \le (n-1-n_2)^{-1} < e^{4.5/K} p$$

where the last inequality uses our definition of n_2 in (10) and the fact that K is large. We calculate

$$\mathbb{P}(\mathcal{H}_j \mid \mathcal{H}, \mathcal{P} = P) = \prod_{t \in G_j} p_t \prod_{t \in F_j} (1-p_t) \le e^{4.5} p^K \prod_{t \in F_j} (1-p_t).$$

We finish the proof by noting that $1 - p_t \le 1 - p$ if $t \notin \cup_{i<j} G_j$, and if $t \in \cup_{i<j} G_j$ then $1 - p_t = 1$ on this set of size $(j-1)K < n_2 K$, so

$$\prod_{t \in F_j} (1-p_t) \le (1-p)^{|F_j|}(1-p)^{-n_2 K} < (1-p)^{|F_j|} e^{4.5}$$

for large enough K and n.

This implies (12) and completes the proof of Claim 1. $\qquad\square$

Claim 2. *If the event $\mathcal{G}$ holds (where $\mathcal{G}$ is defined above (7)), then for any fixed $S \subseteq A$ of size s, (under the probability of choosing random colors for each edge)*

$$\mathbb{P}(\mathcal{E}_c(S) \mid A) \le \left(\frac{s}{|A|}\right)^{\sqrt{K}} + O(n^{-\sqrt{K}+2} \log^{8\sqrt{K}} n).$$

Proof. Fix $\mathbf{x} \in A$ and recall the definition of $\nu_{\mathbf{x}}$ above (7). Let $\mathcal{A}$ be the event that for each $\mathbf{x} \in A$, each color appears fewer than $\sqrt{K}$ times on edges of length at most $2Bn^{-1/2} \log^4 n$. The probability of this event failing for fixed $\mathbf{x} \in A$ is at most

$$|C|\binom{\nu_{\mathbf{x}}}{\sqrt{K}}|C|^{-\sqrt{K}},$$

and taking a union bound over A and using (7) gives the second bound in the claim.

So we may assume that $\mathcal{A}$ holds. Since all edges of E_2 have length at most $2Bn^{-1/2} \log^4 n$, this implies that each $c \in C$ has at least $\sqrt{K}$ neighbors in E_2. As the distribution of the (multi)-set of neighbors along E_2 edges of c is symmetric with respect to permutation of the vertex set (Observation 5), we have that

$$\mathbb{P}(\mathcal{E}_c(S) \mid A, \mathcal{A}) \le \frac{\binom{s}{\sqrt{K}}}{\binom{|A|}{\sqrt{K}}} \le \left(\frac{s}{|A|}\right)^{\sqrt{K}}.$$

$\square$

Note that by (8), if K is large, then the bound in Claim 2 implies

$$\mathbb{P}(\mathcal{E}_c(S) \mid A) \le \left(\frac{2s}{n}\right)^{\sqrt{K/2}}.$$

We can now finish the proof that Γ has a perfect matching. Recalling (9), we just need to show that $|N_2(T)| \ge |T|$ for every $T \subseteq C$ with $|T| \le n_2$. But the probability of this event is at most

$$\mathbb{P}(\exists S \subseteq A, T \subseteq C, |S| = |T| = t : N_2(T) \subseteq S \mid \mathcal{P} = P) \le \binom{n}{t}^2 \left(\frac{2e^9 t}{n}\right)^{t\sqrt{K/2}}.$$

11

Summing over t (and recalling $n_2 = 4(n-1)/K$ for a large constant K), we get

$$\mathbb{P}(\exists S, T, |S| = |T| \le n_2 : N_2(T) \subseteq S) \le o(1) + \sum_{t=1}^{n_2} \binom{n}{t}^2 \left(\frac{2e^9 t}{n}\right)^{t\sqrt{K/2}} = o(1).$$

This completes the proof that Γ has a perfect matching M w.h.p. This matching M gives us an acyclic rainbow set of $n-1$ edges and so defines a rainbow spanning tree.

We now crudely bound the cost of M by the cost of all the edges in Γ. We begin with the cost of E_2. We show that under our assumption that $|A| > n/2$ (see (8)) the expected cost of the edges in E_2 is $O(n^{1/2})$. Moreover, given A, the cost of edges from E_2 is an $O(n^{-1/2} \log^4 n)$-Lipschitz function of the colors of E_A (the Lipschitz constant coming from our definition of E_A), so we can convert the bound on the expectation to a probability bound using Warnke's inequality, Lemma 4. To bound the expected cost given A, we fix a color c and define Y as the index of the Kth lowest index edge of color c in E_A. Since edges of index j have weight $O(j^2 n^{-1/2})$, it suffices to show that

$$\mathbb{P}(Y > j) < \exp[-\Omega(j)]. \tag{13}$$

The event $Y > j$ means that color c appeared fewer than K times among the edges of E_A with index at most j. And under our assumption that $|A| > n/2$, this is at most the probability that $X < K$ where X is distributed as Binomial$(jn/2, 1/n)$. If $j > 4K$, the Chernoff bounds immediately give (13) (and for smaller j, we may take the implicit constant in (13) small enough to make the inequality trivial), completing the proof.

We now turn to the cost of the edges in E_1. First, if $i > n - 2n^{-1/2}$, then we use the trivial bound that the K shortest edges incident to $\mathbf{x}_i$ each have length at most 1; so the total length of the E_1 edges incident to $\mathbf{x}_i$ with $i > n - 2n^{-1/2}$ is at most $2Kn^{1/2}$. For smaller i, we define $Z = Z_i$ as the expected value of the Kth shortest edge incident to $\mathbf{x}_i$ and show

$$\mathbb{E}Z = O(n^{-1/2}).$$

(As in the previous case of E_2 edges, we can sum this bound over $i \le n - 2n^{1/2}$ to get a bound on the expected cost of E_1 edges incident to $\mathbf{x}_i$'s with $i \le$

$n - 2n^{1/2}$, and this expectation bound can be converted to a probability bound using Lemma 4.) For this bound, we show that for large enough D_0, if $D_0 < D < n^{1/2}$, then

$$\mathbb{P}(Z > Dn^{-1/2}) < \exp[-D/8]. \tag{14}$$

We prove this bound for $i = n - 2n^{1/2}$ noting that the argument is the same or easier for smaller i. First, Chernoff bounds imply that $\xi_2(\mathbf{x}_i) \in [1 - 4n^{-1/2}, 1 - n^{-1/2}]$ with probability at most $\exp[-n^{1/2}/4]$. Now, suppose $\xi_2(\mathbf{x}_i) = y \in [1 - 4n^{-1/2}, 1 - n^{-1/2}]$ and (without loss of generality) that $x := \xi_1(\mathbf{x}_i) \leq 1/2$. Then the event $Z > Dn^{-1/2}$ is contained (for moderately large D) in the event that the rectangle $[x, x + Dn^{-1/2}/2] \times [y, 1]$ has at most $K - 1$ points from the set $\mathbf{x}_{i+1}, \ldots, \mathbf{x}_n$. Since the locations of these points are uniform in the rectangle $[0, 1] \times [y, 1]$ (under our condition that $\xi_2(\mathbf{x}_i) = y$), the probability of this event is exactly $\mathbb{P}(X < K)$ where X is distributed as $\text{Binomial}(2n^{1/2}, Dn^{-1/2}/2)$. Noting that $\mathbb{E}X = D$, the Chernoff bounds immediately imply (14) for large enough D, completing the proof. $\square$

3 Proof of Theorem 2(a)

We start by revealing the positions of the vertices and setting aside a set R of r arbitrary vertices where $r = Cn^{1/2}$ for some constant C large enough to make the following argument work. Set $N' = [n] \setminus R$ and $n' = n - r$. We then run the following greedy algorithm on N' for $k_0 := n' - r$ steps: we begin with $A = \emptyset$ and let $G_A = (N', A)$ be the graph induced by the selected edges A (we will maintain that G_A is a disjoint set of paths). We consider the natural greedy algorithm which iteratively adds the cheapest edge to A that (i) does not create a cycle in G_A, (ii) does not create a vertex of degree 3 in G_A, and (iii) does not repeat a color. The algorithm is implemented as follows: we start with all edges available and the edges naturally ordered by length; at stage $i + 1$, we choose the cheapest available edge and reveal its color c; if this color appears already in G_A, then we move to the next cheapest available edge and repeat the process until the revealed color does not appear in G_A; we add this edge to G_A; we then remove from the available edges any edge whose addition to A would violate condition (i) or (ii) above.

13

After k iterations, say there are m nontrivial paths of lengths $\ell_1, \ell_2, \ldots, \ell_m \geq 1$ where $\ell_1 + \cdots + \ell_m = k$. Let $\mathbf{x}_1, \ldots, \mathbf{x}_m$ be made up from one endpoint of each path and let $\mathbf{x}_{m+1}, \ldots, \mathbf{x}_{n'-k}$ be the isolated vertices (not in any edge of A). Let $X_k = \{\mathbf{x}_1, \ldots, \mathbf{x}_{n'-k}\}$.

We claim that with failure probability $o(1/n)$, there is an edge of X_k of length at most $\lambda_k = Cn^{1/2}/(n' - k)$ whose color does not appear in A. To see this, partition $[0, 1]^2$ into $m_k = 2/\lambda_k^2$ sub-squares of side $\lambda_k/2^{1/2}$, and let S be an arbitrary set of points of size $n' - k$. Let s_i denote the number of points of S in square i. Then, by convexity,

$$\sum_{i=1}^{m_k} \binom{s_i}{2} \geq m_k \binom{|X_k|/m_k}{2} \geq C^2 n/5.$$

Then, the probability that there is a set S of $n' - k$ points (from N') and a set Q of $q - k$ colors (from our palette of size $q := (1 + \varepsilon)n$) such that no sub-square contains a pair of points joined by an edge with color in Q can be bounded by

$$
\begin{aligned}
\binom{n'}{n'-k}\binom{q}{q-k}\left(\frac{k}{q}\right)^{C^2 n/5} &= \binom{n'}{k}\binom{q}{k}\left(\frac{k}{q}\right)^{C^2 n/5} \\
&\leq \left(\frac{ne}{k} \cdot \frac{qe}{k} \cdot \left(\frac{k}{q}\right)^{C^2 n/(5k)}\right)^k \\
&\leq \left(e^2(1+\varepsilon)\left(\frac{1}{1+\varepsilon}\right)^{C^2 n/(5k)-2}\right)^k = o(n^{-1}),
\end{aligned}
$$

for sufficiently large $C = C(\varepsilon)$.

At the end of the greedy algorithm, there are exactly r components of G_A since at each step the number of components drops by one. We claim that we can complete G_A to a rainbow Hamilton cycle using edges in a graph $\nabla(N', R)$, the edges with one end in N' and one end in R, noting that the colors of all edges in $\nabla(N', R)$ remain unrevealed. Use C_1 for the set of colors appearing in G_A.

The only edges in $\nabla(N', R)$ that are relevant for completing G_A to a Hamilton cycle are those incident to an endpoint of a (possibly trivial) path in G_A.

14

Supposing that there are p nontrivial paths in G_A, we think of the relevant edges of $\nabla(N', R)$ as a partially directed bipartite graph $B = (X, Y; E)$ where $|X| = |Y| = |R|$. X is the union of P, one vertex for each non-trivial path, and I, one vertex for each isolated vertex. Y is a copy of R and E is the union of the complete bipartite graph on $I \cup Y$ together with the complete directed graph on $P \cup Y$. Observe that a (partially directed) rainbow Hamilton cycle on B using colors from $C \setminus C_1$ together with the edges from G_A forms a rainbow Hamilton cycle on $[n]$. Moreover, since $|R| \approx \sqrt{n}$, the cost of any Hamilton cycle on B is $O(\sqrt{n})$. So to finish the proof, it suffices to find a rainbow Hamilton cycle on B using $C \setminus C_1$.

Reveal the colors on B, noting that an edge receives a color from $C \setminus C_1$ with probability $p_1 = \varepsilon/(1+\varepsilon)$. So let the edges of $\Gamma_1 = B_{p_1}$ be randomly colored using a palette of $q' \geq \varepsilon n$ colors. Note next that the expected number of times a particular color is used is at most $\lambda := 2r^2/q' = O(1)$. It is not difficult then to show using McDiarmid's inequality [19] that the set of colors C_2 used exactly once satisfies $|C_1| \gtrsim q'e^{-\lambda}$ with failure probability $o(1/n)$. Now consider the subgraph Γ_2 consisting of edges in Γ_1 whose colors are in C_2. Conditioning on $|\Gamma_2| = m$, Γ_2 will be distributed as (a supergraph of) B_m, a random set of m edges from B. All we need to prove now is the following:

Proposition 6. *With probability $1 - o(1/n)$, Γ_2 contains a Hamilton cycle.*

Proof. This follows from a result of Frieze [9] if B contains no directed edges. Moreover, Γ_2 can be coupled with B_{p_2}, the binomial random subgraph of B with $p_2 \sim e^{-\lambda}q'/n$ with failure probability $o(1/n)$. To finish, note that Mc-Diarmid's coupling [18] shows that B_{p_2} is more likely to contain a Hamilton cycle than K_{r,r,p_2} (which contains a Hamilton cycle with probability $1 - o(1/n)$ by [9]). $\square$

This completes the proof of Theorem 2.

4 Separating Z_{MST} from Z_{MST}^* and Z_{TSP} from Z_{TSP}^* − Theorem 2(b)

We can show the separations

$$Z_{\mathrm{MST}} - Z_{\mathrm{MST}}^* = \Omega(\sqrt{n}) \quad \text{w.h.p} \tag{15}$$

$$Z_{\mathrm{TSP}} - Z_{\mathrm{TSP}}^* = \Omega(\sqrt{n}) \quad \text{w.h.p.} \tag{16}$$

using the methods of [12] (which there were used to show, among other things, that $Z_{TSP}^* - Z_{MST}^* = \Omega(\sqrt{n})$).

We begin with a simple lemma.

Lemma 7. *Let S be a set of size $\alpha = \Theta(n)$ that is randomly colored using $\beta = \Theta(n)$ colors. Then w.h.p. there are $\Theta(n)$ colors that are used more than once.*

Proof. Let Z denote the number of colors that appear more than once. Then

$$\mathbb{E}(Z) = \beta \left(1 - \left(1 - \frac{1}{\beta} \right)^\alpha - \alpha \cdot \frac{1}{\beta} \left(1 - \frac{1}{\beta} \right)^{\alpha-1} \right) = \Theta(n).$$

Now changing the color of one edge changes Z by at most one. So, applying McDiarmid's inequality [19] we have

$$\mathbb{P}\left(Z \leq \mathbb{E}(Z)/2 \right) \leq \exp\left\{ -\frac{\mathbb{E}(Z)^2}{2\alpha} \right\} = e^{-\Omega(n)}.$$

$\square$

Now back to the main argument. For notational convenience, in this section we scale $\mathcal{X}$ so that we are instead working with a set $\mathcal{Y}_n$ of n random points in a square of side length $\sqrt{n}$; note that there is thus one point on average, per unit of area.

We first recall a definition and lemma used in [12]. Given $\varepsilon, D > 0$ and a finite set of points $S \subseteq \mathbb{R}^d$ and a universe Y, we say that $T \subseteq Y$ is an (ε, D)-copy of S if there is a bijection f between T and S such that $\|x - f(x)\| < \varepsilon$ for all $x \in T$, and such that T is at distance $> D$ from $Y \setminus T$. First we have:

Proposition 8 (see Observation 2.1 in [12]). *Given any finite point set S, any $\varepsilon > 0$, and any D, $\mathcal{Y}_n$ w.h.p contains at least $C^S_{\varepsilon,D} n$ (ε, D)-copies of S, for some constant $C^S_{\varepsilon,D} > 0$.*

This already suffices to show (15). Indeed, apply Proposition 8 to a set S consisting of two points at unit distance, with $\varepsilon = 1/4$ and $D = 4$. We learn that w.h.p. the number κ of $(\frac{1}{4}, 4)$-copies of S in $\mathcal{Y}_n$ satisfies $\kappa \geq Cn$ for a constant C. Since the colors of edges are independent of the geometry of the points, we can apply Lemma 7 with $\alpha = \kappa, \beta = q$ to show that w.h.p. there are $\Omega(\kappa)$ colors that are repeated among the κ edges which join the two points in a given copy. Thus for any rainbow MST, there are at least $\Omega(n)$ copies in which it does not use the edge between the two points. We can add this edge and delete one of the edges leaving the copy to produce a spanning tree shorter by at least unit distance. Doing this for each of these $\Omega(n)$ copies produces a spanning tree shorter than the rainbow MST by $\Omega(n)$. Upon rescaling to the unit square, this corresponds to the existence of a spanning tree of length $\Omega(\sqrt{n})$ less than the length of the rainbow MST.

To prove (16), we will additionally use the following results from [12]:

Proposition 9 (Observation 2.9 in [12]). *Suppose that $S_{\varepsilon,D}$ is an (ε, D) copy of a fixed set S for fixed ε and sufficiently large D, and that at least 2 pairs of edges of a shortest TSP tour $\mathcal{L}$ join $S_{\varepsilon,D}$ to $V \setminus S_{\varepsilon,D}$. Then the pairs are nearly straight (i.e., the angle for each pair is arbitrarily close to π as $\varepsilon \to 0$, and $k, D \to \infty$). Moreover, any tour without this property can be shortened by a length bounded below by a function just of ε, S, and D by changing just how it passes through $S_{\varepsilon,D}$.*

Proposition 10 (see Observation 2.10 in [12]). *Suppose that $S_{\varepsilon,D}$ is an (ε, D)-copy of any fixed set S for fixed ε and sufficiently large D. Then there are at most 2 pairs of edges in a shortest TSP tour which join $S_{\varepsilon,D}$ to $V \setminus S_{\varepsilon,D}$. Moreover, any tour with at least 3 such pairs can be shortened by a length bounded below by a function just of ε, S, and D by changing just how it passes through $S_{\varepsilon,D}$.*

We now let v_1, v_2, v_3 denote the vertices of a fixed unit equilateral triangle, and choose a large D, small δ and even smaller ε, so that $1 \gg \delta \gg \varepsilon > 0$, and apply 8 to a set S consisting of three pairs of distinct points $p_i, q_i,$

Figure 1: An instance of the set S used, together with an example of how an (uncolored) optimum tour may traverse this set in a way consistent with Propositions 10 and 9.

$i = 1, 2, 3$, where the points p_i, q_i are both within distance δ of v_i (Figure 1). In particular, we learn that there are $\kappa \geq Cn$ (ε, D)-copies of this set S in $\mathcal{Y}_n$.

Propositions 9 (applied both to S and to the individual pairs $\{p_i, q_i\}$ separately) and 10 (applied to S) now imply that for each such copy, any rainbow tour through $\mathcal{Y}_n$ tour either:

(a) Traverses the copy that in a way that uses all of the edges $\{p_i, q_i\}$, $i = 1, 2, 3$, or else

(b) Traverses the copy in a way that can be improved by at least a fixed distance $\gamma > 0$ by only changing the path(s) the tour takes through the copy.

Our goal is to show that there are a linear number of copies for which the optimum rainbow tour falls into case b; this then implies that in the (rescaled) unit square, the optimum rainbow tour is longer by an additive factor of $\Omega(\sqrt{n})$. Note that overall, there are $3\kappa \geq 3Cn$ edges $\{p_i, q_i\}$ over all $i = 1, 2, 3$ and all (ε, D)-copies of S. The colors of edges are independent of the geometry of the points. Applying Lemma 7 with $\alpha = 3\kappa, \beta = q$ we see that w.h.p. there are $\Omega(n)$ colors that are repeated in the (ε, D)-copies of S. Thus w.h.p. there are $\Omega(n)$ instances of case (b) with respect to the rainbow tour, as desired.

5 Proof of Theorem 3

Here we assume that the costs $C(i, j)$ are independent uniform $[0, 1]$. The general idea is to replace an edge (i, j) of color c with an edge $\{i, c, j\}$ in a

18

random 3-uniform hypergraph. There is a technical problem in that we are not allowed to have $\{i, c, j\}$, $\{i, c', j\}$ for $c \neq c'$. We deal with this by merging the arguments of Frankston, Kahn, Narayanan and Park [7] and Bell, Frieze and Marbach [5].

Let X be a finite set and let $\mathcal{H}$ be an r-bounded hypergraph on X. For $k \geq r$, let $X^* = X \times [k]$. Let $\mathcal{H}^*$ denote the family consisting of all possible rainbow copies of $\mathcal{H}$ using the color set $[k]$. Let ξ_x $(x \in X)$ be independent random variables, each uniform from $[0, 1]$, and set

$$\xi_{\mathcal{H}} = \min_{S \in \mathcal{H}^*} \sum_{x \in S} \xi_x.$$

We prove

Theorem 11. *Let $\mathcal{H}$ be a κ-spread, r-uniform hypergraph on vertex set X. Suppose that (i) $\kappa = \Omega(r)$ and (ii) $|X| \leq \kappa^2 r / \log^5 r$. If we randomly color the elements of $\mathcal{H}$ with $k \geq r$ colors, then $\xi_{\mathcal{H}} \leq 3C_0 r / \kappa$ w.h.p. (assuming $r \to \infty$) for some absolute constant $C_0 > 0$.*

(The paper [13] removes conditions (i) and (ii), but the proof doesn't adapt well to deal with the costs, as far as we can see.)

Proof. We let ℓ_0 be the least i such that $r/2^i < \log r$, and let ℓ be the least $i > \ell_0$ such that $r/2^i < 1$. We sort the elements of $X = \{x_1, x_2, \ldots, x_N\}$ so that $\xi_{x_1} < \xi_{x_2} < \cdots < \xi_{x_N}$. We then let $p_i = C_i/\kappa$ where

$$C_i = \begin{cases} C_0 & i \leq \ell_0. \\ \frac{\log r}{\log \log r} & \ell_0 < i \leq \ell. \end{cases}$$

Here C_0 is a large constant. Let $L_i = Np_i$ and define n_0 to be the smallest integer such that $2^{-n_0} \leq \log r$ and let n be the least $i > n_0$ such that $r/2^i < 1$. Then for $i \in [n]$ we let

$$a_i = \sum_{j=1}^{i} L_i \text{ and } \varepsilon_i = \frac{2a_i}{N}.$$

19

Then let $W_1^* = \{x_1, x_2, \ldots, x_{a_1}\}$, $W_2^* = \{x_{a_1+1}, x_{a_1+2}, \ldots, x_{a_2}\}$ etc. Note that $W_1^*, W_2^*, \ldots, W_n^*$ are randomly colored random sets of size L_i. The Chernoff bounds imply that

$$\mathbb{P}(\exists i : \max\{\xi_x : x \in W_i^*\} \geq \varepsilon_i) = \sum_{i=1}^{n} \mathbb{P}(Bin(N, \varepsilon_i) \leq \varepsilon_i N)$$

$$\leq \sum_{i=1}^{n} e^{-\varepsilon_i N/8} \leq 2e^{-a_1/4} = 2e^{-C_0 N/\kappa}. \quad (17)$$

In [5] (as we will explain shortly), it is proved that with probability $1 - \frac{\log \log r}{\log^{1/4} r}$, there is a rainbow colored edge e such that $|e \setminus \bigcup_{j \leq i} W_j^*| \leq r/2^{i-1}$. We will use the main theorem (and its proof which we discuss below) from [5]:

Theorem 12 (Theorem 2 in [5]). *Let $\mathcal{H}$ be an r-bounded, κ-spread hypergraph and let $X = V(\mathcal{H})$ be randomly colored from $Q = [q]$ where $q \geq r$. Suppose also that (i) $\kappa = \Omega(r)$, that is, there exists a constant $L > 0$ such that $\kappa \geq Lr$ for all valid r, and that (ii) $N \leq \kappa^2 r / \log^5 r$. Then there is a constant $C > 0$ such that if*

$$m \geq \frac{(C \log_2 r)|X|}{\kappa} \quad (18)$$

then X_m contains a rainbow colored edge of $\mathcal{H}$ w.h.p.

The proof of Theorem 12 in [5] proceeds by iteratively choosing random sets of colored elements W_i^* (just as we have in our current proof). Theorem 12 is proved by iteratively applying the lemma we state next. Here $\mathcal{H}^*$ is the set of possible rainbow colored edges of $\mathcal{H}$. For $H^* \in \mathcal{H}^*$, $T^*(W^*, H^*)$ is $G^* \setminus W^*$ for some $G^* \in \mathcal{H}^*$.

Lemma 13 (Lemma 5 in [5]). *Let $H^* \in \mathcal{H}^*$ be good with respect to W^* if $H^* \sim W^*$ and $|T^*(W^*, H^*)| < r/2$. Let*

$$\mathcal{S} \text{ be the event } |\{H^* \in \mathcal{H}^* : H^* \text{ is good}\}| < (1 - \varepsilon)|\mathcal{H}^*|(1 - p)^r.$$

Then

$$\mathbb{P}(\mathcal{S}) \leq \exp\left\{-\frac{\varepsilon^2 \kappa^2 q(1 - p)}{16e^3 N}\right\} + \frac{2K}{C_0^{r/3} \varepsilon (1 - p)^r}$$

In other words, so long as the unlikely event $\mathcal{S}$ does not occur, we have that for each rainbow colored edge H^*, there is some rainbow colored edge

20

G^* that is mostly contained in W^*. The proof of Theorem 12 proceeds by iteratively applying Lemma 13, each time replacing $\mathcal{H}^*$ with the hypergraph whose edges are the $T^*(W^*, H^*)$ (i.e. portions of edges of $\mathcal{H}^*$ which are not contained in W^*). In [5] they show that during all applications of the Lemma 13, the unlikely event $\mathcal{S}$ never happens w.h.p. This means that the rainbow edge H^* we obtain in the end was good for every application of Lemma 13. In particular, we have

$$|H^* \setminus \bigcup_{j \leq i} W_j^*| \leq \frac{r}{2^{i-1}}, \qquad \text{i=1, \ldots, } \ell. \tag{19}$$

In our setting, since W_i^* was the set of vertices of cost at most ε_i, the rainbow edge H^* we obtain from Theorem 12 costs at most

$$\sum_{i=1}^{n} \frac{r\varepsilon_i}{2^{i-1}} = \frac{r}{N} \sum_{i=1}^{n} \frac{1}{2^{i-1}} \sum_{j=1}^{i} N p_i$$

$$= r \left(\sum_{i=1}^{\ell_0} \frac{C_i}{2^{i-1}\kappa} + \sum_{i=\ell_0+1}^{\ell} \frac{C_{\ell_0} + (i - \ell_0)\log r}{2^{i-1}\kappa} \right) \leq \frac{3C_0 r}{\kappa} = O(1).$$

$\square$

All of the examples given in Theorem 3 have the requisite spread to justify the claim. For (1) – (3) we have $r = O(n)$ and $\kappa = \Omega(n^{k-1})$. (For (3), we have $k = 2$.)

This concludes the proof of Theorem 3.

6 Observations and open questions

We have estimated the effect of adding a rainbow constraint in the probabilistic analysis of some random combinatorial optimization problems. For the Euclidean case we are tight up to a constant if we are dealing with the minimum spanning tree problem. For the TSP and $(1 + \varepsilon)n$ colors, there is

a factor of $\log n$ between our lower and upper bounds. It is an open question as to whether this is actually necessary.

Karp [15] gave a polynomial time algorithm that is asymptotically optimal for the TSP in the unit square. It would be interesting to see if there is an analogous algorithm for the Rainbow TSP in the unit square. One also wonders if the algorithms of Arora [2] or Mitchell [20] can be adapted to the same problem.

Can we find the correct constants for any of the problems mentioned in Theorem 3?

References

[1] D. Aldous, Asymptotics in the random assignment problem, Probability Theory and Related Fields 93 (1992) 507-534.

[2] S. Arora, Polynomial time approximation schemes for Euclidean Traveling Salesman and other geometric problems, Journal of the Association for Computing Machinery 45 (1998) 753-782.

[3] J. Beardwood, J. H. Halton and J. M. Hammersley, The shortest path through many points, Mathematical Proceedings of the Cambridge Philosophical Society 55 (1959) 299-327.

[4] D. Bal and A.M. Frieze, Rainbow Matchings and Hamilton Cycles in Random Graphs, Random Structures and Algorithms 48 (2016) 503-523.

[5] T. Bell, A.M. Frieze and T. Marbach, Rainbow Thresholds, SIAM Journal on Discrete Mathematics 38 (2024) 2361-2369.

[6] L. Espig, A.M. Frieze and M. Krivelevich, Elegantly colored paths and cycles in edge colored random graphs, SIAM Journal on Discrete Mathematics 32 (2018) 1585-1618.

[7] K. Frankston, J. Kahn, B. Narayanan and J. Park, Thresholds versus fractional expectation thresholds, Annals of Mathematics 194 (2021) 475-495.

[8] A.M. Frieze, On the value of a random minimum spanning tree problem, Discrete Applied Mathematics 10 (1985) 47-56.

[9] A.M. Frieze, Limit distribution for the existence of hamiltonian cycles in random bipartite graphs, European Journal of Combinatorics 6 (1985) 327-334.

[10] A.M. Frieze, On random symmetric travelling salesman problems, Mathematics of Operations Research 29 (2004) 878-890.

[11] A.M. Frieze and M. Karoński, Introduction to Random Graphs, Cambridge University Press, 2015.

[12] A.M. Frieze and W. Pegden, Separating subadditive Euclidean functionals, Random Structures and Algorithms 51 (2017) 375-403.

[13] J. Han and X. Yuan, On rainbow thresholds.

[14] T. Harris A lower bound for the critical probability in a certain percolation process, Mathematical Proceedings of the Cambridge Philosophical Society 56 (1960) 1-20.

[15] R.M. Karp, Probabilistic Analysis of Partitioning Algorithms for the Traveling-Salesman Problem in the Plane, Mathematics of Operations Research 2 (1977) 209-244.

[16] R.M. Karp, A patching algorithm for the non-symmetric traveling salesman problem, SIAM Journal on Computing 8 (1979) 561-573.

[17] E. Lawler, Combinatorial Optimization: Networks and Matroids, Dover Publications, 1976.

[18] C. McDiarmid, Clutter percolation and random graphs, Mathematical Programming Studies 13 (1980) 17-25.

[19] C. McDiarmid, On the method of bounded differences, in Surveys in Combinatorics, ed. J. Siemons, London Mathematical Society Lecture Notes Series 141, Cambridge University Press, 1989.

[20] J. Mitchell, Guillotine Subdivisions Approximate Polygonal Subdivisions: A simple polynomial-time approximation scheme for geometric TSP, k-MST, and related problems, SIAM Journal on Computing 28 (1999) 1298-1309.

[21] J. Park and H. T. Pham, A proof of the Kahn-Kalai conjecture, J. Amer. Math. Soc. 37 (2024), no. 1, 235-243.

[22] J. Michael Steele, Probability Theory and Combinatorial Optimization, SIAM CBMS series, 1996.

[23] D. Walkup, Matchings in random regular bipartite graphs, Discrete Mathematics 31 (1980) 59-64.

[24] L. Warnke, On the method of typical bounded differences, Combinatorics, Probability and Computing **25** (2016), 269-299.

[25] J. Wastlünd, The mean field traveling salesman and related problems, Acta Mathematica 204 (2010) 91-150.