Rainbow spanning trees in complete graphs colored by one-factorizations

Paul Horn*

Abstract

Brualdi and Hollingsworth conjectured that, for even n, in a proper edge coloring of K_n using precisely $n-1$ colors, the edge set can be partitioned into $\frac{n}{2}$ spanning trees which are rainbow (and hence, precisely one edge from each color class is in each spanning tree). They proved that there always are two edge disjoint rainbow spanning trees. Kaneko, Kano and Suzuki improved this to three edge disjoint rainbow spanning trees. Recently, Carraher, Hartke and the author proved a theorem improving this to $\epsilon\frac{n}{\log n}$ rainbow spanning trees, even when more general edge colorings of K_n are considered. In this paper, we show that if K_n is properly edge colored with $n-1$ colors, a positive fraction of the edges can be covered by edge disjoint rainbow spanning trees.
Keywords: Rainbow spanning trees, One-factorizations
AMS classification: Primary: 05C15, Secondary: 05C05, 05C70.

A one-factorization of the complete graph on n vertices, K_n, is a decomposition of the edge set of K_n into perfect matchings. Such a decomposition, which clearly exists only when n is even, yields a proper edge coloring of K_n with precisely $n-1$ colors where each color appears precisely $\frac{n}{2}$ times. Given any edge coloring of graph G (not necessarily proper), a spanning tree T of G is called *rainbow* if the color every edge of T is distinct. If we take G to be a properly edge colored K_n, then G clearly contains many rainbow spanning trees. Indeed, the $K_{1,n-1}$ based at any vertex is rainbow due to the proper coloring. Brualdi and Hollingsworth, in [3], conjectured that, in the special case where K_n is colored by a one-factorization, much more is true. In particular, they conjectured that the edge set of a K_n, where $n \geq 6$ and even, colored by a one-factorization can be *partitioned* into $\frac{n}{2}$ edge disjoint rainbow spanning trees.

In support of their conjecture they proved in [3], using Rado's matroid theorem, that every one-factorization coloring of the complete graph contains two edge disjoint rainbow spanning trees. This was slightly improved by Kaneko, Kano and Suzuki [6] to three edge disjoint rainbow spanning trees. The approach of Kaneko, Kano and Suzuki from was inductive,

*Department of Mathematics, University of Denver, Aspen Hall, Room 717, 2280 S. Vine St., Denver, CO 80208 paul.horn@du.edu

and led them to conjecture a strengthening of the Brualdi-Hollingsworth conjecture. In particular, they conjectured that every proper edge coloring (not necessarily arising from a one-factorization) of K_n can have its edge set partitioned into rainbow spanning trees. This remained, until recently, the best known result to either conjecture, but recently Carraher, Hartke, and the author [5] proved that every coloring of K_n (not necessarily proper) where every color is used as most $\frac{n}{2}$ times contains $\epsilon \frac{n}{\log n}$ edge disjoint rainbow spanning trees (for some absolute constant $\epsilon > 0$). This work also improved earlier results of Akbari and Alipour which showed that if K_n is edge colored so that no color appears more than $n/2$ times contains at least two rainbow spanning trees.

In this paper, we show that in the context of the original conjecture of Brualdi and Hollingsworth we can do significantly better. Indeed, we show that if K_n is colored with a one-factorization, then a positive fraction of the edge set can be covered by edge disjoint rainbow spanning trees.

Theorem 1. *There exist constants $\epsilon, n_0 > 0$, so that if $n > n_0$ with n even, and G is a K_n colored by a one-factorization, then G contains at least ϵn edge disjoint rainbow spanning trees.*

Note that we make no real effort to find an optimal ϵ that our proof will give – largely because it is clear that our proof will only work for ϵ fairly small and it is clear that new ideas are needed to prove the Brualdi-Hollingsworth conjecture. Also note that the restriction that $n \geq n_0$ is not serious – since it is already known that for $n \geq 6$ at least three rainbow spanning trees are already present, the condition $n \geq n_0$ may be removed (possibly by taking $3/n_0$ as a new, smaller, ϵ). None the less, as it is convenient to assume n is large in many computations we state the theorem for n sufficiently large.

Before we give the proof of Theorem 1, let us give a sketch of the strategy of the proof, to motivate our approach. A rather natural approach to find edge disjoint rainbow spanning trees is the following: *partition* the edge set of the colored graph G into t edge disjoint graphs $G_1, G_2, \ldots, G_t$ in such a way that each G_i can be guaranteed to have a rainbow spanning tree. One reasonable way to attempt this is to choose edges to be in G_i at random. The most naive way to attempt this would be to select each edge to be in G_i independently with probability $\frac{1}{t}$; thus each graph would have the law of the Erdős-Rényi $G(n, \frac{1}{t})$ random graph.

This is precisely the approach of Carraher, Hartke and the author in [5]. There is a natural bottleneck to this approach, however. If we are to find rainbow spanning trees in G_i, then certainly G_i must be connected. As is well known (see [2]), the random graph $G(n,p)$ is a.a.s.[1] disconnected if $p < (1-\epsilon)\frac{\log n}{n}$. This roadblock was the primary reason why, in [5], it was only possible to assert the existence of $\epsilon \frac{n}{\log n}$ rainbow spanning trees in graphs where each color class has size at most $\frac{n}{2}$.

In order to find ϵn edge disjoint spanning trees, this would suggest decomposing the host G into $t = \epsilon n$ rainbow spanning trees. As mentioned above, the graphs $G(n, \frac{1}{\epsilon n})$ are not

[1]Recall that if $\mathcal{A}_n$ is a sequence of events (for instance $\mathcal{A}_n$ may be the event that $G \in G(n,p)$ is disconnected), we say that $\mathcal{A}_n$ occurs a.a.s. (or asymptotically almost surely) if $\mathbb{P}(\mathcal{A}_n) = 1 - o(1)$.

even connected – indeed, they have roughly an $e^{-1/\epsilon}$ fraction of isolated vertices. Beyond connectivity, a second requirement to finding rainbow spanning trees is having all $n-1$ colors present in each of the t subgraphs. If G_i is distributed as $G(n, \frac{1}{\epsilon n})$, however, and since each color class in G is of size $n/2$, there are roughly an $e^{-1/(2\epsilon)}$ proportion of colors missing from each G_i. Both of these serve as obstacles to G_i having a rainbow spanning tree which must be overcome.

Another reasonable way to choose the subgraphs G_i is to chose each to be a random *regular* subgraph of G. Ignoring, for a second, that partitioning K_n into edge disjoint subgraphs each having the law of a random regular graph seems a nightmarish technical challenge, such a decomposition would relieve the problem of the G_i having isolated vertices. Robinson and Wormald [9, 10] show a random d-regular graph, where $d \geq 3$ is hamiltonian a.a.s., and hence is certainly connected. A random d-regular subgraph of G will still have many missing colors however (again, with high probability and for fixed d), and thus will not contain a rainbow spanning tree. Alternately, one might choose G_i so that each color appears a fixed number of times. While this fixes the problem of the G_i not containing all of the colors, each subgraph will be disconnected a.a.s.

Ultimately, we combine the approaches of partitioning G into random regular graphs (and hence ensuring connectivety) and random color-regular graphs (and hence ensuring that each color is present.) We will proceed as follows. We initially randomly partition G into graphs $S_1 \cupdot S_2 \cupdot \cdots \cupdot S_t \cupdot T$, where the S_i are color-regular (that is, they contain the same number of edges of each color) and where the remainder T has certain nice properties. From T we further extract random expander graphs. (Note that due to the fact that random regular graphs are somewhat hard to generate, we do not actually partition T into random regular graphs, but it suffices for our arguments that we partition T into graphs which are sufficiently good expanders.) We then show that the unions of the color-uniform graphs and expander graphs have rainbow spanning trees, using a criterion for the existence of a rainbow spanning tree in a graph due to Schrijver.

The key to finding rainbow spanning trees is the following criteria, first established by Schrijver in [11] using matroid methods, but later given graph theoretical proofs by Suzuki [12] and also by Carraher and Hartke [4].

Proposition 1. *G has a rainbow spanning tree if and only if for every $2 \leq t \leq n$ and every partition of G with t parts, at least $t-1$ different colors are represented in edges between partition classes.*

Essentially, the color uniform graphs will allow us to guarantee sufficiently many colors between partitions with many parts and the expander graphs will allow us to guarantee sufficiently many colors between partitions with few parts, together covering the whole range.

The remainder of the paper is organized as follows. In Section 1, we will introduce some preliminaries making the above sketch more precise, and perform the initial partitioning into the S_i and T. In Section 2.2 we will show that we can extract the expanders which we need. Finally, we will complete the proof in Section 3.

1 Preliminaries

Throughout the paper, G is an edge colored K_n with color classes $\mathcal{C}_1, \ldots, \mathcal{C}_{n-1}$ which together yield a one-factorization of K_n. In particular, this implies that n is even, and each color class $\mathcal{C}_i$ satisfies $|\mathcal{C}_i| = \frac{n}{2}$. Also throughout C will denote a (large) constant, and t will denote a integer with $Ct < \frac{n}{4} - 10\sqrt{n}$. For concreteness and simplicity, we may take $t = \lceil \frac{n}{5C} \rceil$. Finally, we will assume throughout that n, and C are taken to be large enough so that certain inequalities hold.

In this paper, we will frequently use random permutations as a tool to generate graphs. A fundamental tool we will use is a version of Talagrand's inequality established by McDiarmid in [7] for random variables that are a function of independent random permutations along with a set of independent random variables. In the form we use it, it states the following:

Proposition 2. *Suppose $V_1, \ldots, V_k$ are sets and let $\pi = (\pi_1, \ldots, \pi_k)$ be a family of independent random permutations, so that π_i is chosen uniformly at random from the set $\mathrm{Sym}(V_i)$ of all permutations of V_i, and let $\Omega = \prod \mathrm{Sym}(V_i)$. Furthermore, let $\mathbf{X} = (X_1, X_2, \ldots, X_t)$ be a finite family of independent random variables so that X_j takes values in Ω_j, so that $\mathbf{X}$ takes values in $\Omega' = \prod_j \Omega_j$.*

Le c and r be positive constants, and suppose the nonnegative real-valued function $h : \Omega' \times \Omega \to \mathbb{R}$ satisfies the following condition for each $(\mathbf{x}, \sigma) \in \Omega' \times \Omega$:

- *Changing the value of a coordinate x_j can change the value of $h(\mathbf{x}, \sigma)$ by at most $2c$*
- *Swapping any two elements in any σ_i for $i = 1, 2, \ldots, k$ can change the value of $h(\mathbf{x}, \sigma)$ by at most c*
- *If $h(\sigma) = s$, then in order to certify that $h(\sigma) \geq s$, we need to specify at most rs coordinates of σ. That is, if $h(\mathbf{x}, \sigma) \geq s$, there exists a set of rs coordinates of $(\mathbf{x}, \sigma)$ so that any $(\mathbf{x}', \sigma')$ agreeing with $(\mathbf{x}, \sigma)$ on these rs coordinates also has $h(\mathbf{x}', \sigma') \geq s$.*

Then, if $Z = h(\mathbf{X}, \pi)$ and $\mathrm{Med}(Z)$ denotes the median of Z we have for each $0 \leq t \leq m$,

$$\mathbb{P}(|Z - \mathrm{Med}(Z)| \geq -\lambda) \leq 4 \exp\left(-\frac{\lambda^2}{32rc^2(\mathrm{Med}(Z))}\right).$$

Essentially, Talagrand's inequality implies that if we have a (not too large) deterministic bound on the change that occurs when we switch two entries of a permutation we get tight concentration. A key difference which separates it from martingale methods (such as the Azuma-Hoeffding inequality) is the dependence in the exponential on the length of the permutation. When applying Azuma-Hoeffding bounds, if X is the final stage of a martingale with n steps, the concentration is of the form

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \lambda) \leq \exp\left(-\frac{\lambda^2}{c^2 n}\right),$$

where c denotes a Lipschitz constant (analogous to the c in Talagrand's inequality). For a random permutation on n letters, the most natural filtration takes n steps. Meanwhile, McDiarmid's version of Talagrand's inequality yields concentration of the form

$$\mathbb{P}(|X - \mathrm{Med}(X)| \geq \lambda) \leq \exp\left(-\frac{\lambda^2}{rc^2 \mathrm{Med}(X)}\right).$$

In the case where $r\mathrm{Med}(x)$ is much smaller than n (as in our application here) Talagrand's inequality yields much tighter results.

A slight annoyance of Talagrand's inequality that this gives concentration around the median as opposed to the mean. However, in the situations where we will use it, the following will suffice for our purposes

Proposition 3. *Under the conditions of Talagrand's inequality,*

$$|\mathbb{E}[X] - \mathrm{Med}(X)| = O(c\sqrt{r\mathbb{E}[X]}).$$

For a simple proof, one can consult Molloy and Reed [8].

An advantage to using Talagrand's inequality over martingale methods is the removal of the dependence on the length of the permutations. A second advantage is merely that it is already well setup in terms of the machinery of random permutations used in the proof.

2 Partitioning

We begin the proof of Theorem 1 by finding an appropriate initial partitioning of the edges. Recall G is a complete graph colored by one-factorizations, and $\mathcal{C}_i$, where $1 \leq i \leq n-1$ denote the colors classes.

2.1 Initial Partitioning

Lemma 1. *There exists sets $S_1, S_2, \ldots, S_t$ and $T \subseteq E(G)$, and an orientation σ of the edges in T which satisfy the following properties:*

(i) $S_i \cap S_j = \emptyset$ *for* $1 \leq i < j \leq t$, *and* $S_i \cap T = \emptyset$ *for all* $1 \leq i \leq t$.

(ii) $|S_i \cap \mathcal{C}_j| = C$ *for all* $1 \leq i \leq t$ *and* $1 \leq j \leq n-1$.

(iii) For a set of edges $E' \subseteq E(G)$, let $V(E')$ denote the support of E'; that is the set of vertices in $V(G)$ which have positive degree in the subgraph induced by edges in E'. For all $1 \leq i \leq t$ and $\mathcal{I} \subseteq [n-1]$ with $|\mathcal{I}| \leq \frac{n}{100}$,

$$\left| V\left(\bigcup_{j \in \mathcal{I}} (S_i \cap \mathcal{C}_j) \right) \right| \geq 2|\mathcal{I}|.$$

(iv) For a set of edges $E' \subseteq E(G)$, let $V'(E')$ denote the set of vertices incident to at least $\log(C)$ edges in E'. Then for all $1 \leq i \leq t$ and every $\mathcal{I} \subseteq [n-1]$ with $|\mathcal{I}| \geq \frac{n}{100}$,

$$\left| V'\left(\bigcup_{j\in\mathcal{I}} (S_i \cap \mathcal{C}_j)\right)\right| \geq \left(1 - \frac{2}{\log C}\right) n.$$

(v) For all $S \subseteq V$ with $\frac{n}{C} \leq |S| \leq \frac{n}{2}$, and all $1 \leq i \leq t$,

$$e_{S_i}(S, \bar{S}) \geq \frac{C|S|}{4}.$$

(vi) $\deg^{out}_{T,\sigma}(v) \geq \frac{n}{4}$ for all $v \in V(G)$.

We construct the sets S_i and T whose existence is promised in Lemma 1 as follows:

For $1 \leq i \leq n-1$, π_i denotes a uniformly random permutation of the edges of $\mathcal{C}_i$, and σ denotes a uniformly random orientation of the edges of G. Sets $S_{i,j}$ $S_{i,j}$ are defined as follows: $S_{1,j}$ consists of the first C elements of π_j. $S_{2,j}$ consists of the second C elements, and so on. Then

$$S_i = \bigcup_{j=1}^{n-1} S_{i,j}$$
$$T = E(G) \setminus (\bigcup S_i).$$

Finally, we choose σ to be a uniformly chosen orientation of the edges G.

We claim that with positive probability the S_i and T, along with the restriction of σ to the edges in T, satisfy properties $(i)-(vi)$ of Lemma 1, and hence such a partitioning exists.

We begin by showing that assertion (vi) of Lemma 1 is satisfied by T with high probability.

Lemma 2. *Consider T and σ generated as described above, and let σ_T denote the restriction of σ to the edges in T. Then*

$$\deg^{out}_{T,\sigma|_T}(v) \geq \frac{n}{4} \text{ for all } v \in V(G).$$

with probability $1 - o(1)$.

Proof. Recall, all but Ct edges are chosen to be in T, and $t = \lceil \frac{n}{5C} \rceil$. Since each edge is equally likely to be in T, for a fixed edge $e \in G$, the probability that $e \in T$ is $\frac{n/2 - Ct}{n/2} \geq \frac{3}{5} - \frac{2C}{n}$. Therefore for any fixed $v \in G$, noting that each edge incident to v is oriented out with probability $\frac{1}{2}$

$$\mathbb{E}[\deg^{out}_{T,\sigma}(v)] \geq \frac{3}{10}(n-1) - C.$$

We will apply Talagrand's inequality to show concentration. First, note that a transposition in any π_i changes $\deg^{out}_{T,\sigma}(v)$ by at most one. Likewise, the change of an orientation changes $\deg^{out}_{T,\sigma}$ by at most one so we may take $c = 1$ in our application. Also note that in order to certify that $\deg^{out}_{T,\sigma}(v) \geq s$, it suffices to exhibit s elements of $(\pi_1, \pi_2, \ldots, \pi_{n-1})$ corresponding to those s edges being in T, along with s elements of σ verifying their orientation. Thus we may take $r = 2$ in our application. Applying Talagrand's inequality with $c = 1$ and $r = 2$ and $\lambda = 9\sqrt{n \log n}$, we have that

$$\mathbb{P}(|\deg^{out}_{T,\sigma}(v) - \mathrm{Med}(X)| \geq 9\sqrt{n \log n}) \leq 4 \exp\left(-\frac{81 n \log n}{64 \mathrm{Med}(X)}\right) < 4n^{-1.2},$$

where in the last inequality follows simply by bounding $\mathrm{Med}(X) \leq n$.

By Proposition 3 we actually have $|\mathrm{Med}(X) - \mathbb{E}[X]| = O(\sqrt{n})$. Thus,

$$\begin{aligned} |\deg^{out}_{T,\sigma}(v) - \mathbb{E}[X]| &\leq |\deg_{H_1}(v) - \mathrm{Med}(X)| + |\mathrm{Med}(X) - \mathbb{E}[X]| \\ &\leq |\deg_{H_1}(v) - \mathrm{Med}(X)| + O(\sqrt{n}), \end{aligned}$$

for n sufficiently large. Again assuming n is sufficiently large, we have

$$\mathbb{E}[X] - \frac{n}{4} \geq \left(\frac{3(n-1)}{10} - C\right) - \frac{n}{4} > 10\sqrt{n \log n}.$$

This yields

$$\mathbb{P}(\deg^{out}_{T,\sigma}(v) < \frac{n}{4}) < \frac{4}{n^{1.2}},$$

and a union bound over vertices completes the proof.

□

The next two lemmas, which verify that the sets constructed above satisfy properties (iii) and (iv) of Lemma 1 establish a sort of *color expansion* in the graphs induced by the S_i. That is, they show that the graphs spanned by a small collection of colors (say $\leq \frac{n}{100}$ colors) have many more non-isolated vertices than the number of represented colors. For very small numbers of colors this is, of course, obvious but for larger number of colors this requires that each of the color classes is sufficiently spread out. Indeed, this is one of the locations where we use the fact that the color classes in the original complete graphs are matchings in a non-trivial way.

For larger number of colors we require more: if we consider the graph induced by a large collection of colors, then we want to say that not only are many vertices incident to colors from the color classes but that many vertices are incident to *many* edges from the color class. This is one point that actually requires the parameter C to be rather large, and one reason why this approach won't find close to $\frac{n}{2}$ spanning trees: this simply won't be true if the S_i contain few colors from each color class.

We now show that, with high probability, (iii) of Lemma 1 is satisfied by our partitioning.

Lemma 3. *For all* $1 \le i \le t$ *and* $\mathcal{I} \subseteq [n-1]$ *with* $|\mathcal{I}| \le \frac{n}{100}$,

$$\left| V\left(\bigcup_{j\in\mathcal{I}} (S_i \cap \mathcal{C}_j)\right)\right| \ge 2|\mathcal{I}|.$$

with probability $1 - o(1)$.

Proof. Fix $1 \le i \le t$ and $\mathcal{I} \subseteq [n-1]$. As an abuse of notation, let $V(\mathcal{I}) = V\left(\bigcup_{j\in\mathcal{I}}(S_i \cap \mathcal{C}_j)\right)$ denote the set of vertices which are incident in S_i to at least one of the color classes indexed by $\mathcal{I}$.

Note that if $|V(\mathcal{I})| \le 2|\mathcal{I}|$, then there exists a set $U \subseteq V(G)$ of size $2|\mathcal{I}|$ so that all edges in $\mathcal{C}_j \cap S_i$ have both ends in U. Further note that $S_{i,j}$ is equally likely to be any set of C edges of $\mathcal{C}_j$, and for any U of size $2|\mathcal{I}|$,

$$|U \cap \mathcal{S}_j| \le |U|/2 = |\mathcal{I}|.$$

As the edges of each color are chosen independently, the probability that there exists such a set is at most

$$\begin{aligned}
\binom{n}{2|\mathcal{I}|}\left(\frac{\binom{|\mathcal{I}|}{C}}{\binom{n/2}{C}}\right)^{|\mathcal{I}|} &\le \left(\frac{ne}{2|\mathcal{I}|}\right)^{2|\mathcal{I}|}\left(\left(\frac{2e|\mathcal{I}|}{n}\right)^{C}\right)^{|\mathcal{I}|} \\
&= \exp\left(2|\mathcal{I}|\log\left(\frac{n}{|\mathcal{I}|}\right) + 2|\mathcal{I}|\log\left(\frac{e}{2}\right) - C|\mathcal{I}|\log\left(\frac{n}{|\mathcal{I}|}\right) + C|\mathcal{I}|\log(2e)\right) \\
&\le \exp\left((2-C)|\mathcal{I}|\log\left(\frac{n}{|\mathcal{I}|}\right) + (2+C)|\mathcal{I}|\log(2e)\right).
\end{aligned}$$

A union bound over all sets of size at most s implies that the the probability that there exists a such a 'bad' $\mathcal{I}$ of size at most s is at most

$$\begin{aligned}
&\sum_{w=1}^{s}\binom{n}{w}\exp\left((2-C)w\log\left(\frac{n}{w}\right) + (2+C)w\log(2e)\right) \\
&\qquad \le \sum_{w=1}^{s}\left(\frac{ne}{w}\right)^{w}\exp\left((2-C)w\log\left(\frac{n}{w}\right) + (2+C)w\log(2e)\right) \\
&\qquad \le \sum_{w=1}^{s}\exp\left((3-C)w\log\left(\frac{n}{w}\right) + (2+C)w\log(2e)\right) \\
&\qquad \le s \times \max_{1\le w\le s}\left\{\exp\left((3-C)w\log\left(\frac{n}{w}\right) + (2+C)w\log(2e)\right)\right\}.
\end{aligned}$$

Let

$$f(x) = (3-C)x\log(n/x) + (2+C)x\log(2e)$$

and $g(x) = \exp(f(x))$ denote the function maximized in the preceding display.

Then

$$g''(x) = f''(x)\exp(f(x)) + (f'(x))^2\exp(f(x)).$$

For $C > 3$, observe that $f''(x) > 0$, and hence $g''(x) > 0$ as well. This implies that

$$\max_{1\le x\le s} g(x) = \max\{g(1), g(s)\} \le \max\{\exp(f(1))), \exp(f(s))\}.$$

Evaluating both ends, we observe that:

$$g(1) = O(n^{3-C})$$

and

$$g\left(\frac{n}{100}\right) = \exp\left(\left[(2-C)\log(100) + (2+C)\log(2e)\right]\frac{n}{100}\right) = \exp(-\Omega(n)),$$

so long as C is sufficiently large. In particular, for $s = \frac{n}{100}$ and C large enough

$$s \times \max_{1\le w\le s}\left\{\exp\left((3-C)w\log\left(\frac{n}{w}\right) + (2+C)w\log(2e)\right)\right\} = O(n^{-2}).$$

Since this serves as a bound for the probability that some set $\mathcal{I}$ violating the conclusion of the lemma for a single S_i, combining this with a union bound over $1 \le i \le t$ completes the proof. □

Lemma 4. *For a set of edges $E' \subseteq E(G)$, let $V'(E')$ denote the set of vertices incident to at least* $\log(C)$ *edges in E'. Then for all $1 \le i \le t$ and $\mathcal{I} \subseteq [n-1]$ with $|\mathcal{I}| \ge \frac{n}{100}$,*

$$\left|V'\left(\bigcup_{j\in T}(S_i \cap \mathcal{C}_j)\right)\right| \ge \left(1 - \frac{2}{\log C}\right)n.$$

Proof. While we require that any set of at least $\frac{n}{100}$ colors has this property, we actually show that all collections of $\frac{n}{\sqrt{C}}$ colors already have the desired property. Naturally this is sufficient so long as we choose $C \ge 10,000$.

Fix a set $\mathcal{I}$ of size $\frac{n}{\sqrt{C}}$, and fix an $1 \le i \le t$. For $v \in V(G)$ let X_v denote the zero/one valued indicator random variable of whether v has at least $\log(C)$ incident colors in S_i, and let Y_v denote the number of colors incident to v.

Since all colors incident to v are distinct, note that Y_v is the sum of independent indicator variables. For a given edge, the probability that it is in S_i is $\frac{2C}{n}$, since C edges are chosen from each color class. Thus the expected number of colors incident to C_i in W is $2\sqrt{C}$. Since, for a given vertex, the number of incident colors is the sum of independent indicators, the Chernoff bounds yield that

$$\mathbb{P}(X_v = 0) = \mathbb{P}(Y_v \le \log C) \le \mathbb{P}(Y_v \le \mathbb{E}[Y_v]/2) \le e^{-\sqrt{C}/8}.$$

Therefore, if $X = \sum_v X_v$,

$$\mathbb{E}[X] = \sum_v \mathbb{E}[X_v] \geq (1 - e^{-\sqrt{C}/8})n.$$

Now we wish to apply Talagrand's inequality. Note that a single transposition of a permutation can affect at most two vertices and hence $c = 2$ in our application of Talagrand's inequality. Likewise, to verify that a vertex has $X_v = 1$, we must merely show that $\log(C)$ edges incident to V that are present in T. Thus we may take $r = \log C$. Note that by Proposition 3 that

$$|\mathrm{Med}(X) - \mathbb{E}[X]| = O(\sqrt{\log(C)\mathbb{E}[X]}) = O(\sqrt{n\log(C)}).$$

Therefore for C large enough,

$$\mathbb{P}\left(X \leq (1 - \frac{2}{\log(C)})n\right) \leq \mathbb{P}\left(X \leq \mathrm{Med}(X) - \frac{n}{\log(C)}\right),$$

and we can take $\lambda = \frac{n}{\log C}$ in our application. Applying Talagrand's inequality with these parameters, we see that

$$\mathbb{P}\left(X \leq (1 - \frac{2}{\log(C)})n\right) \leq 4\exp\left(-\frac{(n/\log(C))^2}{128\log(C)n}\right) \leq 4\exp\left(-\frac{n}{128\log^3(C)}\right).$$

Taking a union bound over all sets of size $\frac{n}{\sqrt{C}}$, and over $1 \leq i \leq t$ yields that the conclusion of the lemma fails to hold with probability at most

$$\begin{aligned} n \cdot \binom{n}{n/\sqrt{C}} \cdot \left(4\exp\left(-\frac{n}{128\log^3 C}\right)\right) &\leq 4n\left(\sqrt{eC}\right)^{n/\sqrt{C}} \exp\left(-\frac{n}{128\log^3 C}\right) \\ &\leq 4n\exp\left(\frac{n\log(e\sqrt{C})}{\sqrt{C}} - \frac{n}{128\log^3 C}\right) = \exp(-\Omega(n)), \end{aligned}$$

for C sufficiently large.

□

Finally, we verify that the sets selected satisfy property (v) of the lemma.

Lemma 5. *Fix $1 \leq i \leq t$. For all $S \subseteq V$ with $\frac{n}{C} \leq |S| \leq \frac{n}{2}$, $e_{S_i}(S, \bar{S}) \geq \frac{C|S|}{4}$.*

Proof. Fix $1 \leq i \leq t$, and $S \subseteq V$ with $|S| = s$. Note that each edge is in S_i with probability $\frac{2C}{n}$, so

$$\mathbb{E}[e_{S_i}(S, \bar{S})] = \frac{2Cs(n-s)}{n} \geq \frac{Cs}{n}.$$

For concentration note that we may take $c = t = 1$ in the hypothesis of Talagrand's inequality. Applying Talagrand's inequality with $\lambda = \text{Med}(X)/2$ gives

$$\mathbb{P}(X \le \text{Med}(X) - \lambda) \le 4\exp\left(-\frac{\text{Med}(X)}{128}\right).$$

Applying Proposition 3, we have that

$$\mathbb{P}(X \le \frac{Cs}{4}) \le \mathbb{P}(X \le \mathbb{E}[X]/4) \le \mathbb{P}(X \le \text{Med}(X) - \lambda) \le 4\exp\left(-\frac{\mathbb{E}[X]}{256}\right)$$
$$= 4\exp\left(-\frac{C}{128}s(n-s)\right) \le 4\exp\left(-\frac{C}{256}s\right).$$

A union bound over sets of size $s \le \frac{n}{2}$ and over $1 \le i \le t$ yields that the probability that the conclusion of the lemma fails to hold is at most

$$4n\sum_{s=n/C}^{n/2}\binom{n}{s}\exp\left(-\frac{C}{256}s\right) \le 4n^2 \max_{n/C \le s \le n/2} \exp\left(s\log(n/s) - \frac{C}{256}s\right) \tag{1}$$

Note that for $s \ge n/C$,

$$s\log(n/s) - \frac{C}{256}s \le s\left(\log(C) - \frac{C}{256}\right).$$

Thus, so long as C is chosen large enough so that $\frac{C}{256} > \log(C)$, the quantity of (1) is of the form $\exp(-\Omega(n))$ and this completes the proof of the Lemma.

□

Lemma 1 follows immediately from the preceeding Lemmas.

Proof of Lemma 1. Note that condition (i) and (ii) of Lemma 1 is automatically satisfied by the S_i and T as generated above. Since Lemmas 2-5 show that conditions (iii)-(vi) of Lemma 1 are satisfied by our random S_i, T, and σ a.a.s., for n sufficiently large there exists sets S_i and T, and σ satisfying all conditions of Lemma 1. □

For the remainder of the proof we fix $S_1, S_2, \ldots, S_t$ along with T and σ enjoying the properties guaranteed by Lemma 1.

2.2 Expander graphs from T

While the S_i have many properties that are helpful for finding rainbow spanning trees, they are lacking one crucial thing: they are not connected. In this section, we extract from the remainder set T graphs with expansion properties that, when combined with the S_i, will contain rainbow spanning trees. Recall from Lemma 1 (*vi*), that each vertex has $\deg^{out}_{T,\sigma}(v) > \frac{n}{4} > tC$. For each vertex v, choose (arbitrarily) a set of $\frac{n}{4}$ out-edges of v, denoting these sets T_v.

The random expanders T_i are extracted as follows: For each vertex v, independently and uniformly permute T_v, choose the first C elements to be in T_1, the next C to be in T_2, and so on.

The key expansion property satisfied by the T_i is the following:

Lemma 6. *For a set S, let $\mathrm{cdeg}_i(S)$ denote the number of colors represented in $e_{T_i}(S, \bar{S})$. Then for all sets S with $1 \le |S| \le n/75$ and for all T_i, $1 \le i \le t$ we have that $\mathrm{cdeg}_i(S) \ge 2|S|$.*

Proof. Fix a T_i and a set S with $|S| = s$. Fix a vertex $v \in S$, and a set of r colors. Let us denote by $\mathcal{C}$ the edges in those r chosen colors. The probability that the only edges between S and $\bar{S}$ are from the r chosen colors is at most

$$\frac{\binom{|\mathcal{C}\cap T_v|+|e_{T_v}(v,S)|}{C}}{\binom{n/4}{C}} \le \frac{\binom{r+s}{C}}{\binom{n/4}{C}} \le \left(\frac{4e(r+s)}{n}\right)^C,$$

since all sets of C neighbors of v from the $n/4$ in T_v are equally likely to be those in T_i.

Thus for a fixed set S, by independence of the out-neighbors of each vertex $v \in S$

$$\mathbb{P}(\mathrm{cdeg}_i(S) \le r) \le \binom{n}{r}\left(\frac{e(r+s)4}{n}\right)^{Cs}.$$

Taking a union bound over set S up to size $\frac{n}{75}$ and taking $r = 2s$ gives (for a fixed i)

$$\begin{aligned}
\mathbb{P}(\exists S : \mathrm{cdeg}_i(S) \le 2s) &\le \sum_{|S|=1}^{n/75} \binom{n}{s}\binom{n}{2s}\left(\frac{12es}{n}\right)^{Cs} \\
&\le \sum_{|S|=1}^{n/75} \left(\frac{en}{s}\right)^{s}\left(\frac{en}{2s}\right)^{2s}\left(\frac{12es}{n}\right)^{Cs} \\
&\le \sum_{|S|=1}^{n/75} \exp\big(s\log(n/s) + s + 2s\log(n/s) \\
&\qquad\qquad + 2s\log(e/2) - Cs\log(n/s) + Cs\log(12e)\big) \\
&\le \sum_{|S|=1}^{n/75} \exp\big((3-C)s\log(n/s) + (3+C)s\log(12e)\big).
\end{aligned}$$

Similarly to the proof of Lemma 4, this is $O(n^{-2})$ for C sufficiently large. Here we also observe that we needed to stop this at set of a size s so that $n/s > 12e$. A union bound over $1 \le i \le t$ the completes the proof. □

For the remainder of the proof, we fix $T_1, T_2, \ldots, T_t$ which satisfy the property guaranteed by Lemma 6, along with the $S_1, S_t 2 \ldots, S_t$ we have already fixed.

3 Proof of Theorem 1

In this section we combine the properties of the S_i and T_i proved in Lemmas 1 and 6 to show that the (edge disjoint) $G_i = S_i \cup T_i$ each contain rainbow spanning trees, thus completing the proof of Theorem 1.

Proof of Theorem 1. Let $G_i = S_i \cup T_i$ for $1 \leq i \leq t$ be so that S_i and T_i satisfy the properties guaranteed in Lemmas 1 and 6 respectively. We check that G_i satisfies the property of Proposition 1. To do this, fix an arbitrary $1 \leq i \leq t$ and an arbitrary partition $\mathcal{P} = \{P_1, P_2, \ldots, P_s\}$. We show that there are at least $s-1$ colors between parts

Case 1: $s \leq \frac{C}{4}$.

Suppose $\mathcal{P}$ is a partition into $s \leq \frac{C}{2}$ parts. If $\mathcal{P}$ contains a part $P_j \in \mathcal{P}$ of size $\frac{n}{C} \leq |P_j| \leq \frac{n}{2}$, then by Lemma 1 (*v*), $e_{S_i}(P_j, \bar{P}_j) > \frac{C|P_j|}{4}$ for every i. Thus there exists a $v \in P_j$ with $e_{S_i}(v, \bar{P}_j) \geq C/4 > s-1$. Since the colors incident to v_i are distinct, we are done in this case.

Otherwise, all parts in s have size at most $\frac{n}{C}$, and there is at most one larger part of size at least $n/2$. Let the parts of $\mathcal{P}$ be denoted $1 \leq |P_1| \leq |P_2| \leq \cdots \leq |P_s|$, with $|P_{s-1}| \leq \frac{n}{C}$. Then there exists a $k \leq s-1$ so that

$$Q = \bigcup_{j \leq k} P_j$$

satisfies $\frac{s-1}{2} \leq |Q| \leq 2\frac{n}{C}$. The conclusion of Lemma 6 applied to Q implies that there are at least $2|Q| \geq s-1$ colors represented between $\bar{Q}$ and Q and hence between parts.

Case 2: $\frac{C}{4} \leq s \leq n/100$.

This case is quite similar to the above. Suppose $\mathcal{P}$ is a partition into $\frac{C}{4} \leq s \leq n/100$ parts. If $\mathcal{P}$ contains a part P of size $\frac{s-1}{2} \leq |P| \leq n/75$, Lemma 6 guarantees that $\text{cdeg}(P) \geq 2|P| \geq s-1$, and we are done.

Otherwise, assume the parts of $\mathcal{P}$ are $\{P_1, P_2, \ldots, P_s\}$ with $|P_1| \leq |P_2| \leq \cdots \leq |P_s|$. Note that at most 75 of the parts can be of size at least $\frac{n}{75}$, the remainder are of size at most $\frac{s}{2}$. Let j denote the largest index so that $|P_j| \leq \frac{s}{2}$. Then $j \geq s-75$ so that $\left|\bigcup_{k \leq j} P_k\right| \geq s-75$. Note that for C sufficiently large, $s-75 > \frac{s}{2}$. This, along with the fact that $|P_k| \leq \frac{s}{2}$ for $1 \leq k \leq j$ implies that there exists some index $j' \leq j$ so that if $Q = \bigcup_{k \leq j'} P_k$

$$\frac{s}{2} \leq |Q| \leq s,$$

and as in the previous case satisfying the conclusion of Lemma 6 guarantees that the edges between Q and $\bar{Q}$ witness the necessary colors.

Note that the only substantive difference between this and the previous case is that in the previous case we handled the case in which parts were large (say of size $\approx n/s$) slightly differently: here we argued that their could not be so many of them so that the smaller parts still have a large enough union to find the set we want, while previously we argued that these large parts already witnessed enough colors between the parts.

Case 3: $\frac{n}{100} \leq s \leq \frac{99n}{100}$.

Suppose $\mathcal{P}$ is a partition into $\frac{n}{100} \leq s \leq \frac{99}{100}n$ parts. Suppose fewer than $s-1$ colors appear between the partition classes of $\mathcal{P}$ in our fixed G_i. Then there exists a set of $n-s+1$ color classes so that all edges of these colors are contained entirely within the P_i, say X. For s in this range, $n-s+1 \geq \frac{n}{100}$ so the conclusion of Lemma 1 (*iv*) holds for the colors in set X. Consider $\sum_{i=1}^{s}(|P_i|-1)$. On one hand,

$$\sum_{i=1}^{s}(|P_i|-1) = n-s.$$

Bounding the sum in a slightly different way, we see that

$$(n-s) = \sum_{i=1}^{s}(|P_i|-1) = \left(\sum_{i:|P_i|\geq 2}|P_i|\right) - |\{i : |P_i| \geq 2\}|, \tag{2}$$

But by Lemma 1 (*iv*), noting that certainly all vertices with $\log(C)$ colors represented in X are in parts of size at least 2,

$$\sum_{i:|P_i|\geq 2}|P_i| \geq (1-\frac{2}{\log(C)})n. \tag{3}$$

On the other hand, for any vertex with at least $\log(C)$ colors represented as neighbors it must be in a part of size at least $\log(C)$. Thus,

$$|\{i : |P_i| \geq 2\}| \leq \frac{n}{\log(C)} + \frac{2}{\log(C)} \cdot \frac{n}{2}, \tag{4}$$

with the first term coming from the parts including vertices with $\log(C)$ neighbors, and the second term covering the remainder. But then, combining equations (2), (3) and (4), we have

$$n-s \geq \left(1-\frac{4}{\log(C)}\right)n,$$

or

$$s \leq \frac{4}{\log(C)}n.$$

But since $s \geq \frac{n}{100}$, this is not true for C chosen sufficiently large.

Case 4: $s \geq \frac{99n}{100}$

Here the proof is similar to Case 3, but slightly easier.

Suppose $\mathcal{P}$ is a partition into $s \geq \frac{99}{100}n$ parts, with fewer than $s-1$ colors appearing between the partition classes of $\mathcal{P}$. Then there is a set X of $n-s+1 < \frac{n}{100}$ color classes so that all edges of these colors is entirely contained within the P_i. Again considering $\sum_{i=1}^{s}(|P_i|-1)$, we have

$$(n-s) = \sum_{i=1}^{s}(|P_i|-1) = \left(\sum_{i:|P_i|\geq 2}|P_i|\right) - |\{i : |P_i| \geq 2\}|. \tag{5}$$

Since the conclusion of Lemma 1 (iii) holds for this set X,

$$\sum_{i:|P_i|\geq 2}|P_i| \geq 2(n-s+1).$$

But then

$$\left(\sum_{i:|P_i|\geq 2}|P_i|\right) - |\{i : |P_i| \geq 2\}| \geq \frac{1}{2}\sum_{i:|P_i|\geq 2}|P_i| \geq (n-s+1),$$

where the inequality follows as $|P_i|-1 \geq \frac{1}{2}|P_i|$ if $|P_i| \geq 2$. Thus equation (5) yields

$$n-s \geq (n-s+1).$$

This obvious contradiction completes the proof of the theorem.

□

Acknowledgment: This problem was brought to the author's attendance through the Research Experience for Graduate Students (REGS) program which he had the pleasure of attending as a visiting faculty member. He gratefully acknowledges the support of NSF DMS 08-38434, "EMSW21-MCTP: Research Experience for Graduate Students", which enabled his visit.

References

[1] S. Akbari and A. Alipour. Multicolored trees in complete graphs. *J. Graph Theory*, 54(3):221–232, 2007.

[2] Béla Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.

[3] Richard A. Brualdi and Susan Hollingsworth. Multicolored trees in complete graphs. *J. Combin. Theory Ser. B*, 68(2):310–313, 1996.

[4] James Carraher and Stephen Hartke. Eulerian circuits with no monochromatic transitions. *preprint*.

[5] James Carraher, Stephen Hartke, and Paul Horn. Edge-disjoint rainbow spanning trees in complete graphs. *preprint*.

[6] Atsushi Kaneko, M. Kano, and Kazuhiro Suzuki. Three edge disjoint multicolored spanning trees in complete graphs. *preprint*, 2003.

[7] Colin McDiarmid. Concentration for independent permutations. *Combin. Probab. Comput.*, 11(2):163–178, 2002.

[8] Michael Molloy and Bruce Reed. *Graph colouring and the probabilistic method*, volume 23 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 2002.

[9] R. W. Robinson and N. C. Wormald. Almost all cubic graphs are Hamiltonian. *Random Structures Algorithms*, 3(2):117–125, 1992.

[10] R. W. Robinson and N. C. Wormald. Almost all regular graphs are Hamiltonian. *Random Structures Algorithms*, 5(2):363–374, 1994.

[11] Alexander Schrijver. *Combinatorial optimization. Polyhedra and efficiency. Vol. B*, volume 24 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 2003. Matroids, trees, stable sets, Chapters 39–69.

[12] Kazuhiro Suzuki. A necessary and sufficient condition for the existence of a heterochromatic spanning tree in a graph. *Graphs Combin.*, 22(2):261–269, 2006.