Embedding rainbow trees with applications to graph labelling and decomposition

R. Montgomery[*], A. Pokrovskiy[†], and B. Sudakov[‡]

Abstract

A subgraph of an edge-coloured graph is called rainbow if all its edges have distinct colours. The study of rainbow subgraphs goes back more than two hundred years to the work of Euler on Latin squares. Since then rainbow structures have been the focus of extensive research and have found applications in the areas of graph labelling and decomposition. An edge-colouring is locally k-bounded if each vertex is contained in at most k edges of the same colour. In this paper we prove that any such edge-colouring of the complete graph K_n contains a rainbow copy of every tree with at most $(1 - o(1))n/k$ vertices. As a locally k-bounded edge-colouring of K_n may have only $(n-1)/k$ distinct colours, this is essentially tight.

As a corollary of this result we obtain asymptotic versions of two long-standing conjectures in graph theory. Firstly, we prove an asymptotic version of Ringel's conjecture from 1963, showing that any n-edge tree packs into the complete graph $K_{2n+o(n)}$ to cover all but $o(n^2)$ of its edges. Secondly, we show that all trees have an almost-harmonious labelling. The existence of such a labelling was conjectured by Graham and Sloane in 1980. We also discuss some additional applications.

1 Introduction

In this paper, we study the appearance of large rainbow trees in edge-coloured complete graphs and present applications of our result to several old open problems in graph theory.

A *rainbow* subgraph of an edge-coloured graph is one whose edges have different colours. The study of rainbow subgraphs of edge-coloured graphs goes back more than two hundred years to the work of Euler on Latin squares. A *Latin square of order n* is an $n \times n$ array filled with n symbols such that each symbol appears once in every row and column. A *partial transversal* is a collection of cells of the Latin square which do not share the same row, column or symbol. Starting with Euler (see [14, 28]), transversals in Latin squares have been extensively studied. The most famous open problem in this area is the Ryser-Brualdi-Stein Conjecture (see [11, 36, 37]), which says that every Latin square has a partial transversal of order $n-1$ and a full transversal (a partial transversal of order n) if n is odd. To every Latin square one can assign an edge-colouring of the complete bipartite graph $K_{n,n}$ by colouring the edge ij by the symbol in the cell (i, j). This is a *proper colouring*, i.e., one in which any edges which share a vertex have distinct colours. Identifying the cell (i, j) with the edge ij, a partial transversal corresponds to a rainbow matching. Thus, finding transversals is a special case of finding rainbow subgraphs. Another reason to study rainbow subgraphs arises in Ramsey theory, more precisely in the canonical version of Ramsey's theorem proved by Erdős and Rado [35]. Here the goal is to show that locally-bounded edge-colourings of the complete graph K_n contain rainbow copies of certain graphs. An edge-colouring is *locally k-bounded* if each vertex is in at most k edges of any one colour.

The most natural problem in the study of rainbow structures is to determine which graphs are guaranteed to have a rainbow copy in any properly coloured complete graph K_n. As, when n is even, K_n can be $(n-1)$-edge-coloured, we may ask in general only for rainbow subgraphs with at most $n-1$ edges. This leads to a natural question: which n-vertex trees have a rainbow copy in any proper colouring of K_n? Hahn [24] conjectured in 1980 the special case that there would always be a rainbow copy of the n-vertex path. Disproving this conjecture, Maamoun and Meyniel [31] constructed a proper colouring of K_n with no rainbow Hamilton path.

Nevertheless, it is widely believed that any properly coloured K_n contains a rainbow path covering all but exceptionally few vertices. In particular, Andersen [6] in 1989 conjectured that one can always find a rainbow path covering $n-1$ vertices. The progress on this conjecture was initially slow, despite the efforts of various researchers, see, for example, [2, 22, 23, 18, 12]. Only recently did Alon, Pokrovskiy and Sudakov [4] show that any properly coloured K_n contains a rainbow path with $n - O(n^{3/4})$ vertices (as part of a rainbow cycle).

Turning to more general trees, there are no previous general results to show that rainbow copies of large trees can be found in properly coloured complete graphs. The example of Maamoun and Meyniel [31] can be extended to show that there are proper colourings of K_n which do not contain rainbow copies of certain n-vertex trees (see [7]). However, it is still possible that nearly-spanning trees exist in all proper colourings of complete graphs. In this paper, we prove the first result of this type — we show that any properly coloured K_n contains a rainbow copy of any tree with $n - o(n)$ vertices. We will, in fact, prove the following more general result of which this is a special case. Recall that an edge-colouring of K_n is *locally k-bounded* if each vertex is in at most k edges of any one colour.

Theorem 1.1. *For $\varepsilon > 0, k \in \mathbb{N}$ and sufficiently large n, any locally k-bounded edge-colouring of K_n contains a rainbow copy of every tree with at most $(1-\varepsilon)n/k$ vertices.*

As a locally k-bounded edge-colouring of K_n may have only $(n-1)/k$ distinct colours, Theorem 1.1 is tight up to the constant ε for each k. Let us note that one of the distinguishing features of our result is that we place no conditions on the trees other than the number of vertices they may have. In comparison, all of the graph packing and labelling results which we mention below require a maximum degree bound which is conjectured to be unnecessary. The techniques we introduce for embedding trees with high degree vertices may offer new approaches to these problems.

An important tool in our methods is to demonstrate that, in properly coloured complete graphs, a large rainbow matching can typically be found in a random vertex set using a random set of colours chosen with the same density, in fact such a matching can cover almost all the vertices in the random vertex set. This allows most of a large tree to be embedded if it can be decomposed into certain large matchings. Contrastingly, we need deterministic methods to embed vertices in the tree with high degree. The interplay between the deterministic and random part of the embedding forms a key part of our methods. This is sketched in more detail in Section 2.

Applications to graph decompositions, labellings and orthogonal double-covers. Theorem 1.1 has applications in different areas of graph theory. We present three applications of it in the areas of graph decompositions, graph labellings, and orthogonal double-covers. With each application we prove an asymptotic form of a well-known conjecture. In each case, we consider some special colouring of K_n (coming from a graph-theoretic problem), and apply Theorem 1.1.

Graph decompositions. In graph decompositions one typically asks when the edge set of some graph G can be partitioned into disjoint copies of another graph H. This is a vast topic with many exciting results and conjectures (see, for example, [17, 39, 40]). One of the oldest and best known conjectures in this area, posed by Ringel in 1963 [33], deals with the decomposition of complete graphs into edge-disjoint copies of a tree.

Conjecture 1.2 (Ringel). *Any tree with $n+1$ vertices packs $2n+1$ times into the complete graph K_{2n+1}.*

This conjecture is known for many very special classes of trees such as caterpillars, trees with ≤ 4 leaves, firecrackers, diameter ≤ 5 trees, symmetrical trees, trees with ≤ 35 vertices, and olive trees (see Chapter 2 of [17] and the references therein). There are some partial general results in the direction of Conjecture 1.2. Typically, for these results, an extensive technical method is developed which is capable of almost-packing any appropriately-sized collection of certain sparse graphs, see, e.g., [9, 32, 15, 29]. In particular, Joos, Kim, Kühn and Osthus [27] proved the above conjecture for very large bounded-degree trees. Ferber and Samotij [16] obtained an almost-perfect packing of almost-spanning trees with maximum degree $O(n/\log n)$. A different proof of the approximate version of Ringel's conjecture for trees with maximum degree $O(n/\log n)$ was obtained by Adamaszek, Allen, Grosu, and Hladký [1], using graph labellings. Finally, Allen, Böttcher, Hladký and

Piguet [3] almost-perfectly packed arbitrary spanning graphs with maximum degree $O(n/\log n)$ and constant degeneracy [1] into large complete graphs.

Here we obtain the first asymptotic solution for this problem applicable with no degree restriction.

Theorem 1.3. *For $\varepsilon > 0$ and sufficiently large n, any $(n+1)$-vertex tree packs at least $2n+1$ times into the complete graph $K_{(2+\varepsilon)n}$.*

To see the connection between Theorem 1.1 and Conjecture 1.2 consider the following edge-colouring of the complete graph with vertex set $\{0, 1, \ldots, 2n\}$, which we call the *near distance (ND-) colouring*. Colour the edge ij by colour k, where $k \in [n]$, if either $i = j + k$ or $j = i + k$ with addition modulo $2n + 1$. It is easy to show (as below) that if an $(n + 1)$-vertex tree has a rainbow embedding into the ND-colouring of K_{2n+1} then Conjecture 1.2 holds for that tree.

Proof of Theorem 1.3. Let $\ell = (1 + \varepsilon/3)n$. Consider the ND-colouring of the complete graph $K_{2\ell+1}$, defined above. This is a locally 2-bounded colouring of $K_{2\ell+1}$, which thus, by Theorem 1.1, contains a rainbow copy of T, S_0 say. Now, for each $i \in [\ell]$, let S_i be the tree with vertex set $\{v + i : v \in V(S_0)\}$ and edge set $\{\{v + i, w + i\} : vw \in E(S_0)\}$.

Note that the colour of $vw \in E(K_{2\ell+1})$ is the same as the colour of the edge $\{v + i, w + i\}$ for each $i \in [2\ell]$, and under the translation $x \mapsto x + 1$ the edge vw moves around all $2\ell + 1$ edges with the same colour. Thus, each tree S_i is rainbow, and all the trees S_i are disjoint. $\qquad\square$

Graph labelling. Graph labelling originated in methods introduced by Rósa [34] in 1967 as a potential path towards proving Ringel's conjecture. In the intervening decades, a large body of work has steadily developed concerning different labellings and their applications (see [17]). One old, well-known, conjecture in this area concerns the *harmonious labelling*. This labelling was introduced by Graham and Sloane [19] and arose naturally out of the study of additive bases. Given an Abelian group Γ and a graph G, we say that a labelling $f : V(G) \to \Gamma$ is Γ-*harmonious* if the map $f' : E(G) \to \Gamma$ defined by $f'(xy) = f(x) + f(y)$ is injective. In the case when Γ is a group of integers modulo n we omit it from our notation and simply call such a labelling *harmonious*. In the particular case of an n-vertex tree, Graham and Sloane asked for a harmonious labelling using $\mathbb{Z}_{n-1}$ in which each label is used on some vertex, so that a single label is used on two vertices – where this exists we call the tree *harmonious*. More generally, we also call a graph with m edges and at most m vertices *harmonious* if it has an injective harmonious labelling with $\mathbb{Z}_m$. According to an unpublished result of Erdős [19], almost all graphs are not harmonious. On the other hand, Graham and Sloane [19] in 1980 made the following conjecture for trees.

Conjecture 1.4 (Graham and Sloane)**.** *All trees are harmonious.*

This conjecture is known for many very special classes of trees such as caterpillars, trees with ≤ 31 vertices, palm trees and fireworks (see Chapter 2 of [17] and the references therein). Żak conjectured [41] an asymptotic weakening of this conjecture — that every tree has an injective $\mathbb{Z}_{n+o(n)}$-harmonious labelling.

Note that, for any injective labelling of the vertices of the complete graph by elements of an Abelian group, the edge-colouring which is obtained by taking sums of labels of vertices is proper. Therefore we can use Theorem 1.1 to study such colourings. In particular, we can obtain the following general result which shows that every tree is almost harmonious.

Theorem 1.5. *Every n-vertex tree T has an injective Γ-harmonious labelling for any Abelian group Γ of order $n + o(n)$.*

When the group Γ is cyclic, this theorem proves Żak's conjecture from [41].

Proof of Theorem 1.5. Suppose $|\Gamma| = \ell = (1 + \varepsilon)n$ for some $\varepsilon > 0$ and let T be a tree on n vertices. Identify the vertices of K_ℓ with the elements of Γ and consider an edge-colouring that colours the edge ij by $i+j$. This is a proper colouring, so, when n is large, by Theorem 1.1 it contains a rainbow copy of T, which corresponds to a harmonious labelling. By taking $\Gamma = \mathbb{Z}_\ell$, we deduce that for any $\varepsilon > 0$ there exists n_0 such that any tree with $n \geq n_0$ vertices has an injective harmonious labelling with at most $(1 + \varepsilon)n$ labels. $\qquad\square$

---

[1] A graph is d-*degenerate* if all its induced subgraphs have a vertex of degree $\leq d$. Trees are exactly the 1-degenerate, connected graphs.

Orthogonal double covers. Theorem 1.1 can also be used to obtain an asymptotic solution for another old graph decomposition problem. An *orthogonal double cover* of the complete graph K_n by a graph G is a collection $G_1, \ldots, G_n$ of subgraphs of K_n such that each G_i is a copy of G, every edge of K_n belongs to exacly two of the copies and any two copies have exactly one edge in common. The study of orthogonal double covers was originally motivated by problems in statistical design theory (see [13], Chapter 2). Since the number of copies in the double cover is n, it follows that G must have $n-1$ edges. The central problem here is to determine for which graphs G there is a an orthogonal double cover of K_n by G. In this full generality, this extends the existence question for both biplanes and symmetric 2-designs (see [25]), and so must be considered difficult. About 20 years ago, Gronau, Mullin, and Rosa [20] made the following conjecture about trees.

Conjecture 1.6 (Gronau, Mullin, and Rosa). *Let T be an n-vertex tree which is not a path on 3 edges. Then, K_n has an orthogonal double cover by copies of T.*

This conjecture is known to hold for certain classes of trees including stars, trees with diameter ≤ 3, comets, and trees with ≤ 13 vertices (see [20, 30]). To see the connection between Conjecture 1.6 and Theorem 1.1, we will consider a colouring of a complete graph on 2^k vertices, where edges are coloured by the sum of their endpoints in the abelian group $\mathbb{Z}_2^k$. By considering such a colouring we can show that Conjecture 1.6 is asymptotically true whenever n is a power of 2.

Theorem 1.7. *Let $n = 2^k$ and let T be a tree on $n - o(n)$ vertices. Then K_n contains n copies of T such that every edge of K_n belongs to at most two copies and any two copies have at most one edge in common.*

Proof. Identify $V(K_n)$ with the group $\mathbb{Z}_2^k$. Colour each edge ij with $i + j \in \mathbb{Z}_2^k$. By Theorem 1.1, K_n has a rainbow copy S of T. For all $x \in \mathbb{Z}_2^k$, define a permutation $\phi_x : V(K_n) \to V(K_n)$ by $\phi_x(v) = x + v$ (with addition in $\mathbb{Z}_2^k$). Use $\phi_x(S)$ to denote the subgraph of K_n with edges $\{\phi_x(a)\phi_x(b) : ab \in E(S)\}$. Notice that, since ϕ_x is a permutation of $V(K_n)$, $\phi_x(S)$ is a tree isomorphic to T. We claim that the family of n trees $\{\phi_x(S) : x \in \mathbb{Z}_2^k\}$ satisfies the theorem.

Notice that since $\phi_x(a) + \phi_x(b) = a + x + b + x = a + b$, the permutations ϕ_x preserve the colours of edges. This implies that the trees T_x are all rainbow. Notice that the only edges fixed by the permutations ϕ_x are those coloured by x. Finally, notice that $\phi_{x-y} \circ \phi_y = \phi_x$. Combining these, we have that if a colour c edge is in two trees $\phi_x(S)$ and $\phi_y(S)$, then $x - y = c$ in $\mathbb{Z}_2^k$. This implies that the trees $\phi_x(S)$ cover any edge at most twice, and any pair of them have at most one edge in common. $\qquad\square$

The rest of this paper is organized as follows. In the next section we sketch the proof of Theorem 1.1. In Section 3 we define our notation, and recall some probabilistic results. In Sections 4 – 10 we prove Theorem 1.1.

2 Sketch of the proof of Theorem 1.1

In this section we sketch a proof of Theorem 1.1, before giving a worked example in Section 2.1. For simplicity assume that K_n is properly coloured, i.e. that $k = 1$. The proof splits the tree into a sequence of subforests, and iteratively extends the embedding to each subforest. The splitting of the tree goes as follows. We show that every tree has a short sequence $T = T_\ell \supseteq T_{\ell-1} \supseteq \cdots \supseteq T_1 \supseteq T_0$ where $|T_0| = o(n)$, and each forest T_i is constructed from the previous one using one of the following three operations:

   (1) Add large stars whose centers are in T_{i-1}.

   (2) Add paths of length 3 whose endvertices are in T_{i-1}.

   (3) Add a large matching, one side of which is in T_{i-1}.

There are two additional properties we can ensure: that (1) only needs to be performed once (when going from T_0 to T_1), and in (2), the total number of vertices contained in all the paths is $o(n)$. See Lemma 4 for a precise statement of this splitting.

To find a rainbow embedding of T, we start with a rainbow embedding of T_0 and iteratively extend it to rainbow embeddings of $T_1, \ldots, T_\ell$ by performing one of the operations (1) – (3). We use different proof

techniques for performing (1) – (3): (1) is done deterministically, whereas (2) and (3) are done probabilistically. This interplay between deterministic and probabilistic techniques is one of the main new ideas introduced in this paper.

Stars: To find large stars in (1), we use the deterministic technique of "switchings". The particular technique that we use originated in the papers of Woolbright [38] and Brouwer, de Vries and Wieringa [10] about transversals in Latin squares. The idea is to consider a maximal rainbow family of stars, and show that it must be large enough for our purposes. If the family of stars is not large enough, then it is possible to perform local manipulations to turn it into a larger rainbow family of stars (contradicting the maximality). See Section 8 for further details.

Paths and matchings: To find paths and matchings in (2) and (3), we use the probabilistic method. That is, we choose a random set of vertices X and a random set of colours C, and then try to find the paths/matching using only vertices in X and colours in C. Here "random set" means that we choose every vertex/colour in K_n independently at random with some fixed probability p.

To show that it is possible to find rainbow paths/matchings using random sets X and C, we show that with high probability the subgraph of K_n on X and C has certain pseudorandom properties. There are two relevant pseudorandom properties. Firstly, we show that the subgraph G of K_n of edges with colours in C has roughly the same edge-distribution as an Erdős-Renyi random graph. Specifically, every vertex has degree about pn in G and every pair of disjoint vertex sets A and B have around $p|A||B|$ edges between them. Such a result was first proved by Alon, Pokrovskiy, and Sudakov in [4]. We use a generalization of their result to locally bounded colourings (see Lemma 5.1). Secondly, we show that for random sets $X \subseteq V(K_n)$ and $C \subseteq C(K_n)$, the number of colours of C between X and any set $A \subseteq V(K_n)$ is at least $(1 - o(1))|A|$. See Lemma 6.1 for the precise statement.

Once we have established these pseudorandom properties, we use them to embed the paths and matchings for (2) and (3). Embedding paths is easy — since we only look for $o(n)$ of them, there is enough room to find them greedily (see Lemma 9.2). Embedding matchings is harder, since there is less extra room. Matchings are embedded using a switching argument (like that used for stars), but one which exploits our pseudorandom properties (see Lemma 7.2).

Combining: There is one difficulty left — how do we combine the deterministic arguments for stars with the random ones for paths/matchings? The issue here is that when we deterministically embed the stars, we may not have control over which colours and vertices we use. The vertices and colours used for (2) and (3) need to be disjoint from those used for the stars in (1) and need to be random subsets of $V(K_n)$ and $C(K_n)$ respectively. These two requirements are incompatible with our aim to choose the vertices and colours in (1) deterministically.

We get around this issue by randomizing the stars we build in (1). Specifically, we deterministically embed stars which are bigger than we need, and then randomly delete each vertex with fixed probability. The result is that we find the stars we want, with sufficient randomness in the unused vertices and colours. There is a complication that arises with this argument — there will be a dependency between the colours used on the stars and their vertices. Because of this the setting of several of the lemmas in this paper is the following: in a properly coloured K_n, we randomly choose sets X and C of vertices and colours. The vertices are chosen independently of each other with probability p. The colours are chosen independently of each other with probability p. However, there may be arbitrary dependencies between the vertices and the colours. That is, the random choices of vertices and colours are dependent, but the marginal distributions of the vertices, and the colours, are those for independent choices. It turns out that the methods we use for (3) still work with these added dependencies, allowing us to combine (1) – (3) to embed the whole tree T. More detail is given on this in Section 10.2.

2.1 Worked example

In this section we show how to find a rainbow copy of a particular tree in a properly $(n - 1)$-edge-coloured K_n. The goal is to give simplified versions of our proofs for this tree, and hence highlight the new ideas of this paper. The tree that we focus on is called a *t-spider with size* $(1 - \varepsilon)n$, for arbitrarily small $\varepsilon > 0$. It is constructed from a star with $(1 - \varepsilon)n - t$ vertices by adding a matching of size t joined to the leaves of the star.

Notice that Theorem 1.1 predicts that for every t, every properly coloured K_n contains a t-spider of size $(1 - o(1))n$. In this section we will prove this. The proof we present is not the simplest method of finding a rainbow t-spider in a properly coloured K_n. Rather the purpose of this section is demonstrate the proof techniques of Theorem 1.1 using a very simple tree as an example.

Why is the t-spider a good example of a tree to demonstrate our techniques? Notice that the n-vertex t-spider consists of two parts: a "high degree" part consisting of the star of size $n - \varepsilon n - t$ and a "low degree" part consisting of the matching of size t. This paper gives the first general result about embedding trees with unbounded degrees — because of this a tree with both high and low degree vertices is a illuminating example for our new techniques.

We start with a lemma used for embedding the high-degree part of a t-spider, i.e. embedding a star. The key is for the embedding to be randomized in a suitable way.

Lemma 2.1. *For $\varepsilon, p > 0$, let n be sufficiently large. Let S be a star and $|S| \leq (1 - \varepsilon)pn$, let K_n be properly $(n - 1)$-edge-coloured, and let $v \in V(K_n)$. Then there is a random subgraph S' in K_n containing v such that*

(i) $V(S')$ is a random subset of $V(K_n) \setminus \{v\}$ with each vertex chosen independently with probability p.

(ii) $C(S')$ is a random subset of $C(K_n)$ with each colour chosen independently with probability p.

(iii) With probability $1 - o(1)$, the graph S' contains a rainbow copy of S centered at v.

The sets $V(S')$ and $C(S')$ are allowed to depend on each other.

Proof. Let U be a random set of vertices in $V(K_n) \setminus \{v\}$ with each vertex chosen independently with probability p. Let S' be the star consisting of edges vu for $u \in U$. Notice that since every colour occurs between v and $V(K_n) \setminus \{v\}$, the set $C(S')$ is a random subset of $C(K_n)$ with each colour chosen independently with probability p. By Chernoff's Bound, we have that with probability $1 - o(1)$, $|U| \geq |S|$, and hence S' contains a copy of S. This copy is necessarily rainbow since all stars in properly edge-coloured complete graphs are rainbow. $\square$

It is worth noticing that in the above lemma, one cannot ask for the sets $V(S')$ and $C(S')$ to be independent of each other. Indeed, if one chooses two sets $V \subseteq V(K_n)$ and $C \subseteq C(K_n)$, with all choices made independently at random with probability $p < 1$, then the expected degree of the subgraph formed on the chosen colours and vertices is $p^2(n - 1)$. With high probability, this subgraph will not contain a star of size pn. This demonstrates why we allow the sets $V(S')$ and $C(S')$ to depend on each other. A consequence of this is that when embedding the low-degree part of the tree, we embed it into random sets of vertices and colours allowing arbitrary dependencies between vertices and colours.

For embedding the low-degree part of the tree we will use a lemma about embedding rainbow matchings. It is a special case of Lemma 7.1, which is the main tool used for embedding low-degree parts of trees in this paper.

Lemma 2.2. *For $\varepsilon, q > 0$, let n be sufficiently large. Let K_n be properly edge-coloured. Let $X \subset V(K_n)$ and $C \subset C(K_n)$ be random subsets with the following properties:*

(a) For each colour the events "$c \in C$" have probability q and are independent of each other.

(b) For each vertex the events "$v \in X$" have probability q and are independent of each other.

No assumptions are made about dependencies between the choices of colours in C and vertices in X. Then with probability $1 - o(1)$, the following holds for any set $A \subset V(K_n) \setminus X$ with size at most qn, there is a C-rainbow matching in K_n of size at least $|A| - \varepsilon n$ from A into X.

The above lemma is a special case of Lemma 7.1. The proof techniques involve establishing pseudorandomness properties of the subgraph of K_n on A, X, C, and then showing that suitable pseudorandom coloured graphs contain rainbow matchings. See Section 7 for a proof of this lemma.

An easy combination of the above two lemmas allows us to find nearly-spanning t-spiders in properly $(n - 1)$-edge-coloured complete graphs.

Theorem 2.3. *For each $0 < \varepsilon < 1/10$, there is some n_0 such that the following holds for all $n \geq n_0$ and $t \leq n/3$. Every properly $(n-1)$-edge-coloured K_n contains a rainbow t-spider T of size $(1-2\varepsilon)n$.*

Proof. Choose p so that $|T| - t = (1-\varepsilon)pn$, noting that, as $n/3 \leq |T| \leq (1-2\varepsilon)n$, we have $1/3 \leq p < 1$. Let S be a star of size $|T| - t$. Fix any vertex $v \in V(K_n)$. By Lemma 2.1, there is a subgraph S' of K_n containing v which satisfies (i), (ii) and (iii). Let $X = V(K_n) \setminus V(S')$, and notice that every vertex of $K_n \setminus \{v\}$ ends up in X independently at random with probability $1 - p$ by (i). Let $C = C(K_n) \setminus C(S')$, and notice that every colour ends up in C independently at random with probability $1 - p$ by (ii). That is, (a) and (b) are satisfied in Lemma 2.2 for $K_n \setminus \{v\}$ with $q = 1 - p$. Therefore, by that lemma, with probability $1 - o(1)$, any set $A \subset V(K_n) \setminus \{v\}$ with size at most $(1-p)(n-1)$ has a C-rainbow matching from A into X with size at least $|A| - \varepsilon n$.

By (iii), S' contains a rainbow copy, S'' say, of S. Let L be a set of $t + \varepsilon n$ leaves of S, noting that this possible as S'' has $n - 2\varepsilon n - t - 1 \geq t + \varepsilon n$ leaves. Note that $t + \varepsilon n = |T| - (1-\varepsilon)pn + \varepsilon n = (1-\varepsilon)(1-p)n \leq q(n-1)$. By the property from Lemma 2.2, there is a C-rainbow matching with at least t leaves from L into X. Taking t edges in such a matching and adding them to S'' produces a $(n - 2\varepsilon n)$-vertex t-spider, as required. We have shown that, with high probability, the required rainbow t-spider can be found, and, therefore, if n is sufficiently large, such a t-spider must always exist. $\square$

3 Preliminaries

In this section, we first give some definitions and some notation, and then recall the concentration inequalities that we will use.

3.1 Definitions and notation

We use standard graph theory notation. Particularly the set of vertices and edges of a graph are denoted by $V(G)$ and $E(G)$. We use $C(G)$ to denote the set of colours of a coloured graph. The number of vertices of a graph is denoted by $|G| = |V(G)|$. For a set of colors C, we say a graph is C-*rainbow* if each of its edges has a different colour in C. A C-*edge* is one with colour in C. We say a vertex x is a *colour-C neighbour* of v in a graph G if $x \in N_G(v)$, the neighbourhood of x in G, and xv has colour in C. We denote the set of colour-C neighbours of v in G by $N_C(v)$, and define $N_C(A) = (\cup_{x \in A} N_C(x)) \setminus A$ for each $A \subset V(G)$. For a colour c, we make $N_c(x) = N_{\{c\}}(x)$ and other similar abbreviations. Let $C(G)$ be the set of the colours of the edges of G. For an edge $e \in E(G)$, let $c(e)$ be the colour of e.

We also use the following two important definitions when splitting trees.

Definition 3.1. A *bare path* P in a tree T is a path whose interior vertices all have degree 2 in T. Where P is a bare path in T, $T - P$ is the graph T with the edges of P and the interior vertices of P deleted.

Definition 3.2. In a tree T, we say L is a set of *non-neighbouring leaves* if L is a set of leaves which pairwise share no neighbours. In other words, removing L from T removes a matching.

We use common asymptotic notation for any strictly positive real functions f and g, as follows. If $f(n)/g(n) \to 0$ as $n \to \infty$, then we say $f = o(g)$ and $g = \omega(f)$. If there is some constant C for which $|f(n)| \leq C|g(n)|$ for all n, then we say $f = O(g)$ and $g = \Omega(f)$. In addition, we use the *hierachy* $x \gg y \gg z$ if there exists some non-zero decreasing functions f and g such that if $y \leq f(x)$ and $z \leq g(y)$ then all the subsequent inequalities we need concerning x, y and z hold. Where we say a property almost surely holds or holds with probability $1 - o(1)$ in conjunction with a hierachy $x \gg y \gg 1/n$, we mean that, for each $\varepsilon > 0$, and x and y with $x \ll y$, for all sufficiently large n the property holds with probability at least $1 - \varepsilon$. In all our lemmas we assume that n is an integer which is sufficiently large (e.g. bigger than 10^6 is sufficient).

3.2 Concentration inequalities

We make use of Chernoff's lemma in the following form (see, for example, [26]).

Lemma 3.3. *If X is a binomial variable with standard parameters n and p, denoted $X = \mathrm{Bin}(n, p)$, and ε satisfies $0 < \varepsilon \le 3/2$, then*

$$\mathbb{P}(|X - \mathbb{E}X| \ge \varepsilon \mathbb{E}X) \le 2 \exp\left(-\varepsilon^2 \mathbb{E}X/3\right).$$

Often, the random variables we consider will depend on both random vertices and colours, with some dependencies between the vertices and colours. Here, we will use a version of Azuma's inequality by McDiarmid (see Remark 2.28 from [26]), for which we need the following definitions. For a collection of probability spaces $\Omega_1, \ldots, \Omega_n$, the product space $\Omega = \prod_{i=1}^n \Omega_i$ is defined by setting $\mathbb{P}(\omega_1, \ldots, \omega_n) = \prod_{i=1}^n \mathbb{P}(\omega_i)$. For a product space $\Omega = \prod_{i=1}^n \Omega_i$, and a random variable $X : \Omega \to \mathbb{R}$ we say that X is k-Lipschitz if changing $\omega \in \Omega$ in any one coordinate changes $X(\omega)$ by at most k.

Lemma 3.4 (McDiarmid's inequality). *Suppose that $X : \prod_{i=1}^n \Omega_i \to \mathbb{R}$ is k-Lipschitz. Then*

$$\mathbb{P}\left(|X - \mathbb{E}X| > t\right) \le 2 \exp(-t^2/k^2 n).$$

4 Tree splitting

In this section, prove that every tree can be split into nested subforests $T_0 \subseteq T_1 \subseteq \cdots \subseteq T_\ell = T$, such that each subforest T_i is constructed from T_{i-1} by performing one of three simple operations:

(1) Add large stars whose centers are in T_{i-1}.

(2) Add paths of length 3 whose endvertices are in T_{i-1}.

(3) Add a large matching, whose every edge has at most one of its vertices in T_{i-1}.

The sequence of forests also has two additional properties: the tree T_0 is small, and the length of the sequence ℓ is short. This result is proved in Lemma 4.2. The proof idea is to work backwards. We start with the tree $T_\ell = T$ and see if it has a subtree $T_{\ell-1}$ such that T_ℓ is formed from $T_{\ell-1}$ by performing one of the operations (1) – (3). If there is such a subtree, we pass to it and repeat. Eventually we prove that we end up with a small tree after a small number of steps. To be able to prove that the final tree is small, we need the following standard lemma asserting that any tree either has many leaves or many bare paths.

Lemma 4.1. *Let $t, m \ge 2$. Suppose T is a tree with at most t leaves. Then, there is some s and some vertex-disjoint bare paths P_i, $i \in [s]$, in T with length m so that $|T - P_1 - \ldots - P_s| \le 6mt + 2|T|/(m+1)$.*

Proof. For the appropriate r, let $Q_1, \ldots, Q_r$ be the maximal bare paths in T. Note that each edge is in some path Q_i, and that these paths are edge-disjoint. Replacing each path Q_i by an edge gives a tree, S say, with r edges, no vertices of degree 2, and at most t leaves. Since S has $r + 1 - t$ vertices of degree ≥ 3, the sum of the degrees gives $t + 3(r + 1 - t) \le \sum_{v \in V(S)} d_S(v) = 2r$, which implies $r \le 2t$.

For each $i \in [r]$, find within Q_i as many vertex-disjoint length m subpaths as possible while avoiding the endvertices of Q_i (thus the subpaths from different Q_i are vertex-disjoint), say the paths $Q_{i,j}$, with $j \in [j_i]$, where due to the maximality $j_i = \lfloor (|Q_i| - 2)/(m+1) \rfloor \ge (|Q_i| - m - 2)/(m+1)$. Removing the internal vertices of a subpath with length m from Q_i removes $m - 1$ vertices. Thus, for each $i \in [r]$, we have

$$|Q_i - \cup_{j \in [j_i]} Q_{i,j}| = |Q_i| - (m-1)j_i \le \frac{(m+1)|Q_i| - (m-1)(|Q_i| - m - 2)}{m+1} \le m + \frac{2|Q_i|}{m+1}. \qquad (1)$$

Note that

$$\left| T - \bigcup_{i \in [r]} \bigcup_{j \in [j_i]} Q_{i,j} \right| \le \sum_{i \in [r]} \left| Q_i - \bigcup_{j \in [j_i]} Q_{i,j} \right| \le \sum_{i \in [r]} (m + 2|Q_i|/(m+1))$$

$$\le mr + 2(|T| + r)/(m+1) \le 6mt + 2|T|/(m+1).$$

The first inequality comes from $T = \cup_{i \in [r]} Q_i$, while the second comes from (1). The third inequality comes from $e(T) = \sum_{i \in [r]} (|Q_i| - 1) = |T| - 1$, and the last inequality comes from $r \le 2t$. Thus, the set of paths $\{Q_{i,j} : i \in [r], j \in [j_i]\}$ have the property required. $\qquad \square$

Using Lemma 4.1, we can now find the desired splitting of an arbitrary tree.

Lemma 4.2. *Given integers $n, D \geq 2$, $\mu > 0$ and a tree T with at most n vertices, there are integers $\ell \leq 10^4 D \mu^{-2}$ and $j \in \{2, \ldots, \ell\}$ and a sequence of subgraphs $T_0 \subset T_1 \subset \ldots \subset T_\ell = T$ such that*

P1 *for each $i \in [\ell] \setminus \{1, j\}$, T_i is formed from T_{i-1} by adding non-neighbouring leaves,*

P2 *T_j is formed from T_{j-1} by adding at most μn vertex-disjoint bare paths with length 3,*

P3 *T_1 is formed from T_0 by adding vertex-disjoint stars with at least D leaves each, and*

P4 *$|T_0| \leq \mu n$.*

Proof. Let $m = \lceil 100/\mu \rceil$. Construct three forests T_{match}, T_{paths}, and T_{stars} as follows:

- Iteratively remove sets of $\mu n/16mD$ non-neighbouring leaves from T as many times as possible in order to obtain T_{match}.

- For every vertex $v \in V(T_{\text{match}})$ which has $\geq D + 3$ leaves in T_{match}, remove all but 3 of the leaves of v in order to obtain T_{stars}.

- Remove a maximum size collection of vertex-disjoint length m bare paths from T_{stars} in order to obtain the graph T_{paths}.

We claim that the final tree is small.

Claim 4.3. $|T_{\text{paths}}| \leq \mu n$.

Proof. By the choice of T_{match}, at most $\mu n/16mD$ vertices in T_{match} have some incident leaf. When forming T_{stars}, any vertex with leaves removed has at least 3 remaining leaves. Thus, every leaf in T_{stars} is a leaf in T_{match}. Also by the definition of T_{stars}, every vertex in T_{stars} is adjacent to at most $D + 2$ leaves. Thus, T_{stars} has at most $(\mu n/16mD) \cdot (D + 2) \leq \mu n/8m$ leaves.

By Lemma 4.1, there are vertex-disjoint bare paths $P_1, \ldots, P_r$, for some r, with length m, so that $|T_{\text{stars}} - P_1 - \ldots - P_r| \leq 6\mu n/8 + 2n/m \leq \mu n$. Since the collection of paths removed to get T_{paths} has maximum size, we also have $|T_{\text{paths}}| \leq \mu n$. $\square$

Let $\mathcal{P}$ be the collection of bare paths deleted from T_{stars} to form T_{paths}. Suppose that $\mathcal{P}$ has p paths. Notice that $p \leq n/m \leq \mu n/2$. Suppose for now that $p \geq 1$. Notice that any degree 2 vertex of T_{stars} is also a degree 2 vertex of T_{match}. Therefore, $\mathcal{P}$ is also a collection of bare paths in T_{match}.

Let $T_0 = T_{\text{paths}}$, $T_1 = T_{\text{match}} - \mathcal{P}$ and $T_{m-4} = T_{\text{match}}$. Let T_{m-5} be the subgraph of T_{m-4} formed by deleting a length 3 path from each end of each path in $\mathcal{P}$. Note that the total number of paths of length 3 we delete is $2p \leq \mu n$ and T_{m-5} consist of T_1 together with p vertex-disjoint paths of length $m - 6$. For $i = 2, \ldots, m - 4$, let T_i be the forest consisting of T_1 together with p vertex-disjoint paths of length $i - 1$ so that, for each $i = 2, \ldots, m - 5$, T_{i-1} can be obtained from T_i by deleting an edge from one end of each of the p paths. Since the paths are disjoint, this implies that for $i = 2, \ldots, m - 6$, T_i is obtained from T_{i-1} by adding non-neighbouring leaves. Finally, for the appropriate ℓ, define trees $T_{m-3}, \ldots, T_\ell$ with $T_\ell = T$ so that, for each $i = m - 3, \ldots, \ell$, T_i is constructed from T_{i-1} by adding $\mu n/16mD$ non-neighbouring leaves (this is possible by construction as $T_{m-4} = T_{\text{match}}$).

We claim that $T_0, \ldots, T_\ell$ satisfy the conclusion of the lemma with $j = m - 4$. Conditions **P1** and **P2** are immediate from the construction. Condition **P3** holds because $T_0 = T_{\text{stars}} - \mathcal{P}$ and $T_1 = T_{\text{match}} - \mathcal{P}$ so T_1 is formed from T_0 by adding exactly the same stars that are added when forming T_{match} from T_{stars}. Condition **P4** comes from Claim 4.3. Finally $\ell \leq 10^4 D \mu^{-2}$ holds because $\ell \leq m + n(\mu n/16mD)^{-1} \leq 10^4 D \mu^{-2}$. This concludes the proof when $p \geq 1$. The proof when $p = 0$ is the same, except that since $\mathcal{P} = \emptyset$, the steps constructing paths with length m can be omitted.

$\square$

5 Choosing random colours

Consider a properly coloured complete graph K_n, and form a random set of colours by choosing every colour independently with probability p. Let G be the subgraph of K_n consisting of the edges of the chosen colours. Alon, Pokrovskiy and Sudakov [4] showed that, with high probability, the subgraph G has a similar edge distribution to the binomial random graph. Specifically they showed that, for any two sets of vertices A and B, the number of edges in the chosen colours between A and B is concentrated around the expectation $p|A||B|$. In this section we will prove a generalization of this result on proper colourings to colourings of K_n which are locally k-bounded. The method we use is similar to the proof from [4].

Lemma 5.1. *Let $\varepsilon > 0$ and $k \in \mathbb{N}$ be constant and let $p \geq n^{-1/100}$. Let K_n have a locally k-bounded colouring and suppose G is a subgraph of K_n chosen by including the edges of each colour independently at random with probability p. Then, with probability $1 - o(n^{-1})$, for any disjoint sets $A, B \subset V(G)$, with $|A|, |B| \geq n^{3/4}$,*

$$\big|e_G(A, B) - p|A||B|\big| \leq \varepsilon p|A||B|.$$

Proof. We will first show that, with high probability, the property we want holds for pairs of sets with many colours between them.

Claim 5.2. *With probability $1 - o(n^{-1})$ the following holds.*

P *For any disjoint sets $A, B \subset V(K_n)$, with $|A| \geq |B| \geq n^{1/10}$, which have at least $(1 - \varepsilon p/8)|A||B|$ different colours between A and B in K_n, we have $|e_G(A, B) - p|A||B|| \leq \varepsilon p|A||B|/2$.*

Proof. For any such sets A and B, we can select a rainbow subgraph R of $K_n[A, B]$ with $(1 - \varepsilon p/8)|A||B|$ edges. Notice that $e(R \cap G) \sim \text{Bin}((1 - \varepsilon p/8)|A||B|, p)$. By Lemma 3.3 applied with $\varepsilon/8$ for ε, with probability at least $1 - \exp(-\varepsilon^2 p|A||B|/10^3)$, we have $(1 - \varepsilon p/4)p|A||B| \leq e(R \cap G) \leq (1 + \varepsilon p/4)p|A||B|$, which in combination with $e(K_n[A, B] - R) \leq \varepsilon p|A||B|/8$ implies that **P** holds for A and B. Note that, for such sets, we have $p^3|B| \geq n^{7/100} = \omega(\log n)$, so when $|A| \geq |B| \geq n^{1/10}$, condition **P** holds for A and B with probability at least $1 - \exp(-|A| \cdot \omega(\log n))$.

Thus, **P** holds with probability at least

$$1 - \sum_{b=n^{1/10}}^{n} \sum_{a=b}^{n} \binom{n}{a}\binom{n}{b} \exp(-a \cdot \omega(\log n)) = 1 - o(n^{-1}). \qquad \square$$

Assuming that **P** holds, we now show that the property in the lemma holds. Let $A, B \subset V(K_n)$ be disjoint sets with $|A| \geq |B| \geq n^{3/4}$. Let $\ell = \lceil 10\sqrt{kn}/\varepsilon p \rceil \leq n^{0.6+o(1)}$.

Claim 5.3. *There are partitions $A = A_1 \cup \ldots \cup A_\ell$ and $B = B_1 \cup \ldots \cup B_\ell$ such there are at most $\varepsilon^2 p^2|A||B|/100$ edges ab between A and B for which there is another edge $a'b'$ with $c(ab) = c(a'b')$ and $ab, a'b' \in E(K_n[A_i, B_j])$ for some $i, j \in [\ell]$.*

Proof. Pick random partitions $A = A_1 \cup \ldots \cup A_\ell$ and $B = B_1 \cup \ldots \cup B_\ell$, by choosing the part of each element independently and uniformly at random. Fix any colour-c edge e, and let A_i, B_j be the classes it goes between. The probability that there is another colour-c edge between A_i and B_j sharing a vertex with e is $\leq 2(k-1)/\ell$ (since e touches at most $2(k-1)$ other colour-c edges). The probability there is a colour-c edge disjoint from e between A_i and B_j is $\leq kn/2\ell^2$ (since that are at most $kn/2$ colour-c edges in total). Combining these, the probability there is another colour-c edge between A_i and B_j is at most

$$\frac{2(k-1)}{\ell} + \frac{kn}{2\ell^2} \leq \frac{kn}{\ell^2} \leq \frac{\varepsilon^2 p^2}{100}.$$

Thus, the expected number of edges which have a non-unique colour across their classes is at most $\varepsilon^2 p^2|A||B|/100$. Fix such a partition then for which there are at most $\varepsilon^2 p^2|A||B|/100$ such edges. $\qquad \square$

Fix a partition $A = A_1 \cup \ldots \cup A_\ell$ and $B = B_1 \cup \ldots \cup B_\ell$ from the above claim.

Claim 5.4. *Let H be the subgraph of $K_n[A, B]$ with all the edges between any pair A_i, B_j which does not satisfy $\boldsymbol{P}$ removed. Then*

$$|e_H(A, B) - |A||B|| \leq \varepsilon p|A||B|/4. \tag{2}$$

Proof. First, delete any edge adjacent to a class A_i or B_j which contains at most $n^{1/10}$ vertices. For large n, the number of edges deleted in this first stage is at most

$$\ell \cdot n^{1/10}(|A| + |B|) \leq n^{0.7+o(1)}(|A| + |B|) \leq \varepsilon p|A||B|/8. \tag{3}$$

Secondly, delete edges between any class A_i and B_j if there are at most $(1 - \varepsilon p/8)|A_i||B_j|$ different colours between A_i and B_j in K_n. Note that, at each such deletion, we lose $|A_i||B_j|$ edges, at least $\varepsilon p|A_i||B_j|/8$ of which have non-unique colours between their classes. Thus, if I is the set of pairs (i, j) for which we deleted edges between A_i and B_j for this reason, we have

$$\sum_{(i,j)\in I} \varepsilon p|A_i||B_j|/8 \leq \varepsilon^2 p^2|A||B|/100.$$

Thus, we have deleted at most $\sum_{(i,j)\in I} |A_i||B_j| \leq \varepsilon p|A||B|/8$ edges in this second stage, and, by (3), at most $\varepsilon p|A||B|/4$ edges in total, proving (2). $\qquad\square$

Let I' be the set of pairs (i, j) for which there remain edges between A_i and B_j in H. By $\boldsymbol{P}$, for each $(i, j) \in I'$, we have

$$|e_{G\cap H}(A, B) - p \cdot e_H(A, B)| = \left| \sum_{(i,j)\in I'} e_G(A_i, B_j) - p|A_i||B_j| \right| \leq \sum_{(i,j)\in I'} |e_G(A_i, B_j) - p|A_i||B_j||$$

$$\leq \sum_{(i,j)\in I'} \varepsilon p|A_i||B_j|/2 \leq \varepsilon p|A||B|/2. \tag{4}$$

The first inequality is the triangle inequality. The second inequality is $\boldsymbol{P}$. The third inequality follows as $A_1 \cup \ldots \cup A_\ell$ and $B_1 \cup \ldots \cup B_\ell$ partition A and B respectively. The required property then holds for A and B, as follows.

$$|e_G(A, B) - p|A||B|| \leq |e_G(A, B) - e_{G\cap H}(A, B)| + |e_{G\cap H}(A, B) - p \cdot e_H(A, B)| + p|e_H(A, B) - |A||B||$$

$$\leq (1 + p)|e_H(A, B) - |A||B|| + \varepsilon p|A||B|/2 \leq \varepsilon p|A||B|.$$

The first inequality is the triangle inequality. The third inequality comes from (2). The second inequality comes from (4) and as $E_G(A, B) \setminus E_{G\cap H}(A, B) \subseteq E_{K_n}(A, B) \setminus E_H(A, B)$. $\qquad\square$

6   Colours and random vertex sets

Consider a locally k-bounded coloured complete graph K_n, and let C be a random set of colours chosen independently with probability p. Consider two sets of vertices A and B. In the previous section, we showed that the number of colour-C edges between A and B is about $p|A||B|$. Since every colour can occur $\leq k|A|$ times between A and B, this implies that the number of colours of C occuring between A and B is at least $(1 - o(1))p|B|/k$. Our goal here is to prove that if B is a *random set of vertices*, then this number increases to $(1 - o(1))|B|/k$. In fact, the following main result of this section shows that this holds for every large subset B of a random set of vertices X.

Lemma 6.1. *Let $k \in \mathbb{N}$ and $\varepsilon > 0$ be constant, and $p \geq n^{-1/10^4}$. Let K_n have a locally k-bounded colouring. Let $X \subset V(K_n)$ and $C \subset C(K_n)$ be random subsets where, for each $x \in V(K_n)$ and $c \in C(K_n)$, $\mathbb{P}(x \in X) = \mathbb{P}(c \in C) = p$, all the events $\{x \in X\}$ are independent and all the events $\{c \in C\}$ are independent (but the event $\{x \in X\}$ might depend on the events $\{c \in C\}$). Then, with probability $1 - o(n^{-1})$, for each $A \subset V(K_n) \setminus X$ and $B \subset X$ with $|A| \geq n^{3/4}$ and $|B| \geq \varepsilon pn$, there are at least $(1 - \varepsilon)|B|/k$ colours in C which appear between A and B in K_n.*

To prove this lemma we consider a set A and a random set X, and analyse how many times each colour can appear between them. Without any assumptions on X, we know that every colour appears $\leq k|A|$ times between A and X (from local k-boundedness). The following lemma shows that if X is chosen randomly with probability p, then *most* colours appear only at most $(1 + o(1))pk|A|$ times between A and X.

Lemma 6.2. *Let k be constant and $\varepsilon, p \geq n^{-1/100}$. Let K_n have a locally k-bounded colouring and let X be a random subset of $V(K_n)$ with each vertex included independently at random with probability p. Then, with probability $1 - o(n^{-1})$, for each $A \subset V(K_n)$ with $n^{1/20} \leq |A| \leq n^{1/4}$, for all but at most εn colours there are at most $(1 + \varepsilon)pk|A|$ edges of that colour between A and X.*

Proof. Let $\ell = \lceil kn^{1/2} \rceil$. Fix $A \subset V(K_n)$ with $n^{1/20} \leq |A| \leq n^{1/4}$. For each $c \in C(K_n)$, let $A_c = N_c(A)$, so that $|A_c| \leq k|A|$. Let $C'(K_n)$ be the colours with more than $pk|A|$ edges between A and $V(K_n)$ in K_n, so that $|C'(K_n)| \leq n/pk$.

Claim 6.3. *There is a partition $C'(K_n) = C_1 \cup \ldots \cup C_\ell$ so that, for each $i \in [\ell]$ and $a, b \in C_i$, A_a and A_b are disjoint.*

Proof. Create an auxillary graph with vertex set $C'(K_n)$ where ab is an edge exactly if $A_a \cap A_b \neq \emptyset$. For any colour $c \in C(K_n)$, there are at most $k|A|^2 \leq \ell$ edges between A and A_c, and hence at most $\ell - 1$ different colours from $C'(K_n) \setminus \{c\}$ on such edges. The auxillary graph then must have maximum degree at most $\ell - 1$, and is thus ℓ-colourable, so the claim holds. ☐

Now, consider X, the random subset of $V(K_n)$ with each vertex included independently at random with probability p. For each $i \in [\ell]$, let

$$B_i = \{c \in C_i : |A_c \cap X| > (1 + \varepsilon)pk|A|\}$$

and $B = \cup_{i \in [\ell]} B_i$. To show that $|B| \leq \varepsilon n$, we will use the following claim.

Claim 6.4. *For each $i \in [\ell]$, if $|C_i| \geq \varepsilon n/2\ell$, then $\mathbb{P}(|B_i| \geq \varepsilon p|C_i|/2) \leq \exp(-\varepsilon^2 pn^{1/2}/200k)$.*

Proof. Recall that $|A| \geq n^{1/20}$ and $\varepsilon, p \geq n^{-1/100}$, so that $\varepsilon^2 pk|A| = \omega(-\log(\varepsilon p))$. Note that $|A_c \cap X| \sim$ Bin$(|A_c|, p)$. By Lemma 3.3, and as $|A_c| \leq k|A|$, for each $c \in C_i$,

$$\mathbb{P}(c \in B_i) \leq 2\exp(-\varepsilon^2 pk|A|/3) \leq \varepsilon p/4,$$

for sufficiently large n. Furthermore, the events $\{c \in B_i\}$, $c \in C_i$, are independent (by the disjointness of the sets A_c for $c \in C_i$). This implies that $|B_i|$ is stochastically dominated by Bin$(|C_i|, \varepsilon p/4)$. By Lemma 3.3 applied with $1/2$ for ε, if $|C_i| \geq \varepsilon n/2\ell$, then

$$\mathbb{P}(|B_i| \geq \varepsilon p|C_i|/2) \leq 2\exp(-(\varepsilon p|C_i|/4)/12) \leq \exp(-\varepsilon^2 pn/100\ell) \leq \exp(-\varepsilon^2 pn^{1/2}/200k). \qquad ☐$$

Note that from our choice of parameters we have that $\varepsilon^2 pn^{1/2}/200k > n^{2/5}$. Therefore the probability that for some subset $A \subset V(K_n)$, with $n^{1/20} \leq |A| \leq n^{1/4}$, there is an index $i \in [\ell]$ such that $|C_i| \geq \varepsilon n/2\ell$ and $|B_i| \geq \varepsilon p|C_i|/2$ is at most

$$\sum_{a=n^{1/20}}^{n^{1/4}} \binom{n}{a} \cdot \ell \cdot \exp(-n^{2/5}) = o(n^{-1}).$$

Therefore, with high probability, we can assume that for every subset A and every C_i with $|C_i| \geq \varepsilon n/2\ell$ the corresponding B_i satisfies $|B_i| \leq \varepsilon p|C_i|/2$. Then,

$$
\begin{aligned}
|B| &= \sum_{i \in [\ell]} |B_i| \leq \sum_{i:|C_i| \geq \varepsilon n/2\ell} |B_i| + \sum_{i:|C_i| < \varepsilon n/2\ell} |C_i| \\
&\leq \frac{\varepsilon p}{2} \sum_{i \in [\ell]} |C_i| + \ell \cdot \frac{\varepsilon n}{2\ell} \leq \frac{\varepsilon n}{2k} + \frac{\varepsilon n}{2} < \varepsilon n,
\end{aligned}
$$

where we have used that $\sum_{i \in [\ell]} |C_i| = |C'(K_n)| \leq n/pk$. ☐

We can now show that the property in Lemma 6.2 is likely also to hold for any subset A which is not too small.

Corollary 6.5. *Let k be constant and $\varepsilon, p \geq n^{-1/10^3}$. Let K_n have a locally k-bounded colouring and let X be a random subset of $V(K_n)$ with each vertex included independently at random with probability p. Then, with probability $1 - o(n^{-1})$, for each $A \subset V(K_n) \setminus X$ with $|A| \geq n^{1/4}$, for all but at most εn colours there are at most $(1+\varepsilon)pk|A|$ edges of that colour between A and X.*

Proof. By Lemma 6.2 applied with $\varepsilon^2 p/4$ for ε, with probability $1 - o(n^{-1})$, for each subset $B \subset V(K_n)$ with $n^{1/20} \leq |B| \leq n^{1/4}$, for all but at most $\varepsilon^2 pn/4$ colours there are at most $(1 + \varepsilon^2)pk|B|$ edges of that colour between B and X.

Let $A \subset V(K_n)$ satisfy $|A| \geq n^{1/4}$ and choose $A = A_1 \cup \ldots \cup A_\ell$ with $\ell = \lfloor |A|/n^{1/20} \rfloor$ and $n^{1/20} \leq |A_i| \leq 2n^{1/20}$ for each $i \in [\ell]$. For each $i \in [\ell]$, let C_i be the set of colours for which there are more than $(1 + \varepsilon^2)pk|A_i|$ edges of that colour between A_i and X, so that $|C_i| \leq \varepsilon^2 pn/4$. Let C' be the set of colours for which there are more than $(1 + \varepsilon)pk|A|$ edges of that colour between A and X. We need then to show that $|C'| \leq \varepsilon n$.

Note that, if $c \in C'$, then

$$(1+\varepsilon)pk|A| \leq \sum_{a \in A} |N_c(a) \cap X| \leq \sum_{i:c \in C_i} k|A_i| + \sum_{i:c \notin C_i} (1+\varepsilon^2)pk|A_i|$$
$$\leq |\{i : c \in C_i\}| \cdot 2kn^{1/20} + (1+\varepsilon^2)pk|A|, \tag{5}$$

The first inequality comes from $c \in C'$. The second inequality comes from the fact that the number of colour-c edges between A_i and X is at most $(1 + \varepsilon^2)pk|A_i|$ for $i \notin C_i$, and at most $k|A_i|$ for all other i (by the local k-boundedness of K_n). The third inequality comes from $|A_i| \leq 2n^{1/20}$ and $\sum_{i \in [\ell]} |A_i| = |A|$. This implies that

$$\varepsilon pk|A|/2 \leq (\varepsilon - \varepsilon^2)pk|A| \overset{(5)}{\leq} |\{i : c \in C_i\}| \cdot 2kn^{1/20} \leq |\{i : c \in C_i\}| \cdot 2k|A|/\ell.$$

Thus, if $c \in C'$, then $|\{i : c \in C_i\}| \geq \varepsilon p\ell/4$. This gives

$$|C'| \cdot \varepsilon p\ell/4 \leq \sum_{c \in C'} |\{i : c \in C_i\}| = \sum_{i \in [\ell]} |C' \cap C_i| \leq \sum_{i \in [\ell]} |C_i|.$$

Therefore,

$$|C'| \leq \frac{\sum_{i \in [\ell]} |C_i|}{\varepsilon p\ell/4} \leq \frac{\ell \cdot \varepsilon^2 pn/4}{\varepsilon p\ell/4} = \varepsilon n,$$

as required. $\square$

Combining the property in Corollary 6.5 with Lemma 5.1, we can now complete the proof of the main result of this section.

Proof of Lemma 6.1 . The desired property in the lemma strengthens as ε decreases, so we may assume that $\varepsilon \leq 1/2$. Let G be the subgraph of K_n pn $V(K_n)$ consisting of edges of colours in C. By Corollary 6.5 applied with $\varepsilon^2 p^2/4k$ for ε, with probability $1 - o(n^{-1})$, for each $A \subset V(G)$ with $|A| \geq n^{3/4}$, for all but at most $\varepsilon^2 p^2 n/4k$ colours there are at most $(1 + \varepsilon^2)pk|A|$ edges of that colour between A and X in K_n. With probability $1 - o(n^{-1})$, by Lemma 5.1 applied with ε^2 for ε, for every two disjoint subsets $A, B \subset V(G)$ with $|A|, |B| \geq n^{3/4}$, we have $e_G(A, B) \geq (1 - \varepsilon^2)p|A||B|$.

Now, for each $A \subset V(K_n) \setminus X$ and $B \subset X$ with $|A| \geq n^{3/4}$ and $|B| \geq \varepsilon pn \geq n^{3/4}$, there are at least $(1 - \varepsilon^2)p|A||B|$ edges between A and B in G. Delete all edges between A and B in G whose colour appears more than $(1+\varepsilon^2)pk|A|$ times between A and B. As each colour appears between A and B in K_n at most $k|A|$ times, this removes at most $k|A| \cdot \varepsilon^2 p^2 n/4k$ edges.

Each remaining colour between A and B in G occurs between A and B at most $(1 + \varepsilon^2)pk|A|$ times in G. Therefore, between A and B in G the number of different remaining colours is at least

$$\frac{(1 - \varepsilon^2)p|A||B| - \varepsilon^2 p^2 n|A|/4}{(1 + \varepsilon^2)pk|A|} \geq (1 - \varepsilon/2)|B|/k - \varepsilon^2 pn/2k \geq (1 - \varepsilon)|B|/k,$$

as required. The first inequality uses $\varepsilon \leq 1/2$, while the last inequality uses $|B| \geq \varepsilon pn$. $\square$

7 Finding a large matching

In this section we show that one can find an almost spanning rainbow matching between any set A with appropriate size and a random set X. Note that in the following lemma the random set of colours and the random set of vertices are not required to be independent of each other. This is important for the application of this statement in our methods. More detail on this is given in Section 10.2.

Lemma 7.1. *Let $k \in \mathbb{N}$ and $\varepsilon > 0$ be constant, and suppose K_n has a locally k-bounded colouring and $p \geq n^{-1/10^4}$. Let $X \subset V(K_n)$ and $C \subset C(K_n)$ be random subsets where, for each $x \in V(K_n)$ and $c \in C(K_n)$, $\mathbb{P}(x \in X) = \mathbb{P}(c \in C) = p$, and all the events $\{x \in X\}$ are independent and all the events $\{c \in C\}$ are independent. Then, with probability $1 - o(n^{-1})$, for each set $A \subset V(K_n) \setminus X$ with $|A| \leq pn/k$, there is a C-rainbow matching in K_n of size at least $|A| - \varepsilon pn$ from A into X.*

To prove this lemma we first show, using fairly standard edge swapping arguments, how to find a rainbow matching in a bipartite graph which has edges of many different colours between any two large subsets of vertices. Note that we established such a property in our setting in the previous section.

Lemma 7.2. *Let $\varepsilon \gg \eta \gg 1/n > 0$. Let G be a bipartite graph with classes X and Y, with $|X| = n$ and $|Y| = kn$, which has a locally k-bounded colouring. Suppose that between any two subsets $A \subset X$ and $B \subset Y$ with size at least ηn there are at least $(1 - \eta)|B|/k$ colours in G which appear between A and B. Then, there exists a rainbow matching in G with at least $(1 - \varepsilon)n$ edges.*

Proof. Let M be a maximal rainbow matching in G. Let $A_0 = X \setminus V(M)$ and $B_0 = Y \setminus V(M)$. Suppose, for later contradiction, that $|A_0| \geq \varepsilon n$.

Letting $\ell = \lceil 4/\varepsilon \rceil$ and $s = 5^{\ell+1}$, define a sequence of colour sets $C_0 \subset C_1 \subset \ldots \subset C_\ell$, and sequences of vertex sets $A_1 \subset \ldots \subset A_\ell \subseteq X$ and $B_1 \subset \ldots \subset B_\ell \subseteq Y$ recursively as follows, for each $i \geq 1$.

$$C_{i-1} = \{\text{colours with at least } s \text{ disjoint edges between } A_{i-1} \text{ and } B_{i-1}\}$$
$$A_i = A_0 \cup \{u \in X : \exists v \in Y \text{ s.t. } uv \text{ is an edge of } M \text{ with colour from } C_{i-1}\}$$
$$B_i = B_0 \cup \{v \in Y : \exists u \in X \text{ s.t. } uv \text{ is an edge of } M \text{ with colour from } C_{i-1}\}$$

Observe that, for each $0 \leq i \leq \ell$, $M - (A_i \cup B_i)$ is a matching, and so $k|A_i| \leq |B_i|$. Let $C_{-1} = \emptyset$.

Claim 7.3. *For each $0 \leq i \leq \ell$ and any set of vertices $U \subseteq A_i \cup B_i$ with $|U| \leq 5^{\ell+1-i}$, there is a rainbow matching M_U in G, with $V(M_U) \cap U = \emptyset$ and $C(M_U) = C(M)$, which contains every edge in M with colour outside C_{i-1}.*

Proof. The proof is by induction on i. For $i = 0$, the claim holds, taking $M_U = M$ for any U. Suppose then that $i > 0$ and the claim holds for $i - 1$.

Let $U \subseteq A_i \cup B_i$ with $|U| \leq 5^{\ell+1-i}$. Let $U_1 = U \cap (A_{i-1} \cup B_{i-1})$ and $U_2 = U \setminus U_1$. By the definition of A_i and B_i, there is a set of at most $|U_2|$ edges $M_2 \subseteq M$ such that $U_2 \subseteq V(M_2)$ and $C(M_2) \subseteq C_{i-1} \setminus C_{i-2}$. Note that $V(M_2) \cap (A_{i-1} \cup B_{i-1}) = \emptyset$.

By the definition of C_{i-1}, for every $m \in M_2$, there are at least s disjoint edges between A_{i-1} and B_{i-1} with colour the same as m. Since $s = 5^{\ell+1} \geq 5 \cdot 5^{\ell+1-i} \geq |U_1| + 2|M_2|$, we can choose an edge, e_m say, with colour the same as m, for all $m \in M_2$, such that $M_2' = \{e_m : m \in M_2\}$ is a matching, $V(M_2') \subset A_{i-1} \cup B_{i-1}$ and $V(M_2') \cap U_1 = \emptyset$. Note that $|M_2'| = |M_2|$ and $C(M_2') = C(M_2) \subseteq C_{i-1} \setminus C_{i-2}$.

Let $U' = V(M_2') \cup U_1$, so that $U' \subseteq A_{i-1} \cup B_{i-1}$ and $|U'| \leq |U_1| + 2|U_2| \leq 5^{\ell+2-i}$. By induction, there is a matching $M_{U'}$ with $C(M_{U'}) = C(M)$ which is disjoint from U' and contains every edge in M with colour outside C_{i-2}. Note that, therefore, $M_2 \subset M_{U'}$. Let $M_U = (M_{U'} \setminus M_2) \cup M_2'$. This replaces each edge $m \in M_2 \subset M_{U'}$ with an edge of the same colour, e_m, which is disjoint from U, while keeping the edges in the matching independent. Thus, $C(M_U) = C(M_{U'}) = C(M)$. As the edges moved had colour in C_{i-1}, M_U contains every edge in $M_{U'}$ with colour outside C_{i-1}, and thus every edge in M with colour outside C_{i-1}. $\square$

Claim 7.4. *For each $0 \leq i \leq \ell$, all colours between A_i and B_i occur on M.*

Proof. Suppose that there is an edge xy from A_i to B_i whose colour does not occur on M. Claim 7.3 applied with $U = \{x, y\}$ gives a maximal rainbow matching M_U in G with $C(M_U) = C(M)$ and $V(M_U) \cap \{x, y\} = \emptyset$. Letting $M' = M_U + xy$ gives a rainbow matching in G with size $|M| + 1$, contradicting the maximality of M. $\square$

Let $\mu = 1/2\ell$, which is a non-zero decreasing function of ε, so therefore $\mu \gg \eta$.

Claim 7.5. *Let $0 \leq i \leq \ell$ and let F be a set of at most $(1 - \mu)|A_i|$ colours. Then, there is a colour $c \notin F$ which occurs on at least s disjoint edges between A_i and B_i.*

Proof. Let $t = \lceil 2ks/\mu\varepsilon \rceil$, and note this is a function of ε and k. Noting that $|A_i| \geq |A_0| \geq \varepsilon n$, take in A_i disjoint sets $S_1, \ldots, S_t$, each with size at least ηn. By the assumption of the lemma, there are at least $(1 - \eta)|B_i|/k \geq (1 - \eta)|A_i|$ colours between S_j and B_i, for each $j \in [t]$. Therefore, there are at least $(1 - \eta)|A_i| - |F| \geq (\mu - \eta)|A_i| \geq \mu\varepsilon n/2$ colours outside F between S_j and B_i for each $j \in [t]$. By Claim 7.4, there are at most n colours between A_i and B_i. Therefore, there is a colour, c say, outside F which occurs between B_i and at least $(\mu\varepsilon n/2) \cdot t/n \geq ks$ different sets S_j. As the colouring of G is locally k-bounded and the sets S_j are disjoint, there are thus at least s disjoint edges with colour c between A_i and B_i. $\square$

Claim 7.6. *For each $0 \leq i \leq \ell$, $|C_i| \geq (1 - \mu)|A_i|$.*

Proof. If $|C_i| < (1 - \mu)|A_i|$, then, by Claim 7.5 with $F = C_i$, there is some colour $c \notin C_i$ occuring at least s times between A_i and B_i. This colour should be in C_i, a contradiction. $\square$

Note that all colours of C_{i-1} occur between A_{i-1} and B_{i-1}, and so, by Claim 7.4, occur on M. Since M is rainbow, the definition of A_i implies that for each $i \in [\ell]$ we have $|A_i| = |A_0| + |C_{i-1}|$. Combined with Claim 7.6, this gives $|A_i| \geq (1 - \mu)|A_{i-1}| + |A_0|$ for each $i \in [\ell]$, and hence $|A_\ell| \geq \sum_{j=0}^{\ell-1}(1 - \mu)^j|A_0|$. Note that, as $\mu = 1/2\ell$, $(1 - \mu)^j \geq 1/2$ for each $j \in [\ell]$, and therefore $|A_\ell| \geq \ell|A_0|/2$. Thus, $|A_\ell| \geq \ell \cdot \varepsilon n/2 > n$, a contradiction, so that the original matching M must satisfy the property in the lemma. $\square$

We now put this together with Lemma 6.1 to complete the proof of the main result of this section.

Proof of Lemma 7.1. Let η be a fixed constant not dependent on n which satisfies $\varepsilon \gg \eta > 0$. With probability $1 - o(n^{-1})$, by Lemma 6.1 applied with $\eta/2k$ for ε, for each $A \subset V(K_n) \setminus X$ and $B \subset X$ with $|A|, |B| \geq \eta pn/2k$, there are at least $(1 - \eta)|B|/k$ colours in C between A and B. With probability $1 - o(n^{-1})$, by Lemma 3.3, $(1 - \eta/2)pn \leq |X| \leq (1 + \eta/2)pn$.

We claim that the property in the lemma holds. Let $A \subset V(K_n) \setminus X$ with $|A| \leq pn/k$. Add vertices to A from $V(K_n) \setminus X$, or delete vertices from A, to get a set A' with $|A'| = \lfloor (1 - \eta/2)pn/k \rfloor =: m$ and $|A \setminus A'| \leq \eta pn/k \leq \varepsilon pn/2$. Let X' be a subset of X of size km. Note that for any subsets $A'' \subset A'$ and $B \subset X'$ with $|A''|, |B| \geq \eta m$, there are at least $(1 - \eta)|B|/k$ colours in C between A'' and B (since $\eta m \geq \eta pn/2k$). Thus, by Lemma 7.2, there is a C-rainbow matching with at least $(1 - \varepsilon/4)m \geq |A'| - \varepsilon pn/2$ edges between A' and X'. As $|A \setminus A'| \leq \varepsilon pn/2$ at least $|A| - \varepsilon pn$ of the edges in this C-rainbow matching must lie between A and X. $\square$

We remark that, in the particular case where K_n has an equal number of edges of each colour, Lemma 7.1 can be proved from Lemma 5.1 and Corollary 6.5 using a standard implementation of the Rödl nibble (see, for example, [5]).

8 Star finding

In this section we use a switching argument to find vertex-disjoint collectively-rainbow stars. The argument is based on one independently discovered by Woolbright [38] and Brouwer, de Vries, and Wieringa [10]. We then apply the resulting lemma through two corollaries to prove a form applicable to our colourings. The resulting Corollary 8.4 will be used to embed the part of our tree consisting of large stars.

Lemma 8.1. *Let $0 < \varepsilon < 1/100$ and $\ell \leq \varepsilon^2 n/2$. Let G be an n-vertex graph with minimum degree at least $(1-\varepsilon)n$ which contains an independent set on the distinct vertices $v_1, \ldots, v_\ell$. Let $d_1, \ldots, d_\ell \geq 1$ be integers satisfying $\sum_{i \in [\ell]} d_i \leq (1-3\varepsilon)n$. Let G be edge-coloured so that no two edges touching the same vertex v_i are coloured the same.*

Then, G contains vertex-disjoint stars $S_1, \ldots, S_\ell$ so that, for each $i \in [\ell]$, S_i is a star rooted at v_i with d_i leaves, and $\cup_{i \in [\ell]} S_i$ is rainbow.

Proof. Let $m = \lceil \varepsilon^{-1} \rceil + 1$. Choose disjoint collectively-rainbow stars $S_1, \ldots, S_\ell$ in G so that, for each $i \in [\ell]$, S_i is rooted at v_i and has at most $d_i + m$ leaves, and $\sum_{i \in [\ell]} |S_i|$ is maximised. Either these stars satisfy the lemma, or, without loss of generality, S_1 has fewer than d_1 leaves. Suppose then, that S_1 has fewer than d_1 leaves.

Let $A_0 = \emptyset$. Define the sets $C_1 \subset C_2 \subset \ldots \subset C_m \subset C(G)$ and $A_1 \subset A_2 \subset \ldots \subset A_m$ recursively by letting C_j be the set of colours in $C(G)$ which do not appear on $(\cup_{i \in [\ell]} S_i) - A_{j-1}$ and letting $A_j = N_{C_j}(v_1)$.

Claim 8.2. *For each $s \in [m]$, $A_s \subset \cup_{i \in [\ell]} V(S_i)$.*

Proof. Suppose, for contradiction, that there is some $u_1 \in A_s$ with $u_1 \notin \cup_{i \in [\ell]} V(S_i)$. By the definition of A_s, we have $c(v_1 u_1) = c_1$ for some colour $c_1 \in C_s$. Let $u_1, \ldots, u_j \in V(G)$ and $c_1, \ldots, c_j \in C(G)$ be longest sequences of distinct vertices and colours satisfying the following properties.

 (1) $c(v_1 u_k) = c_k$ for $k = 1, \ldots, j$.

 (2) $c(u_k v_r) = c_{k-1}$ for $k = 2, \ldots, j$, for some v_r with $v_r u_k \in E(S_r)$.

 (3) For each $k = 1, \ldots, j$, $c_k \in C_{s+1-k}$.

We claim that $j = s$, so that $c_j \in C_1$ and the colour c_j does not appear on $\cup_{i \in [\ell]} S_i$. Suppose otherwise. Then, by (3) we have $c_j \in C_{s+1-j}$. By the definitions of C_{s+1-j} and C_1, there is some r and some edge $u v_r \in \bigcup_{i \in [\ell]} E(S_i)$ with $u \in A_{s-j}$ and $c(u v_r) = c_j$. By the definition of A_{s-j}, the edge $v_1 u$ is present and has $c(v_1 u) \in C_{s-j}$. Notice that since $c_j \neq c_1, \ldots, c_{j-1}$ (by (1)) and $\bigcup_{i=1}^{\ell} S_i$ is rainbow, this implies that $u \notin \{u_1, \ldots, u_j\}$ (by (2)). Since the colours of the edges at v_1 are all distinct, we have $c(v_1 u) \notin \{c_1, \ldots, c_j\}$ (using (1)). Let $c_{j+1} = c(v_1 u)$ and $u_{j+1} = u$. This is a longer sequence satisfying (1) – (3), contradicting the maximality of j.

Note that (3) implies that $j \leq m+1$. For each $i \geq 2$, let $S_i' = S_i - \{u_2, \ldots, u_j\}$ and let S_1' be the star formed from S_1 by adding the edges $v_1 u_1, \ldots, v_1 u_j$. In total, to get the stars S_i', $i \in [\ell]$, we added an edge $v_1 u_j$ of colour c_j, a colour not appearing on $\cup_{i \in [\ell]} S_i$, and, for each colour c_i, $i \in \{1, \ldots, j-1\}$ we removed some edge (adjacent to u_{i+1}) of that colour from some star $S_{i'}$ and then added an edge of that colour at v_1 $(v_1 u_i)$. Thus, the new stars are still collectively rainbow. As S_1 had fewer than d_1 leaves, S_1' has at most $d_1 + m$ leaves. Therefore, as $\sum_{i \in [\ell]} |S_i'| = 1 + \sum_{i \in [\ell]} |S_i|$, this violates the choice of the stars S_i, $i \in [\ell]$. $\qquad\square$

For each $j \in [m]$, we have by Claim 8.2, that

$$|C(G) \setminus C_j| = |\cup_{i \in [\ell]} E(S_i)| - |A_{j-1}| \leq \sum_{i \in [\ell]} (d_i + m) - |A_{j-1}| \leq n - 3\varepsilon n + \ell m - |A_{j-1}| \leq n - 2\varepsilon n - |A_{j-1}|.$$

As v_1 has at least $(1-\varepsilon)n$ neighbours, and thus is adjacent to edges of at least $(1-\varepsilon)n$ different colours, we then have, for each $j \in [m]$,

$$|A_j| \geq (1-\varepsilon)n - |C(G) \setminus C_j| \geq |A_{j-1}| + \varepsilon n.$$

Therefore, $|A_m| \geq m \cdot \varepsilon n > n$, a contradiction. $\qquad\square$

Given a locally k-bounded colouring, we can create a new colouring based on this in order to apply Lemma 8.1, as follows.

Corollary 8.3. *Let $0 < \varepsilon < 1/100$ and $\ell \leq \varepsilon^2 n/2$. Let G be an n-vertex graph with minimum degree at least $(1 - \varepsilon)n$ which contains an independent set on the distinct vertices $v_1, \ldots, v_\ell$. Let $d_1, \ldots, d_\ell \geq 1$ be integers satisfying $\sum_{i \in [\ell]} d_i \leq (1 - 3\varepsilon)n$, and suppose G has a locally k-bounded edge-colouring.*

Then, G contains disjoint stars $S_1, \ldots, S_\ell$ so that, for each $i \in [\ell]$, S_i is a star rooted at v_i with d_i leaves, and $\cup_{i \in [\ell]} S_i$ has at most k edges of each colour.

Proof. For each $i \in [\ell]$, recolour the edges next to v_i in G by, for each colour $c \in C(G)$, letting the edges next to v_i with colour c have distinct colours from $\{(c, 1), \ldots, (c, k)\}$. In this new colouring, no two edges of the same colour can meet at some vertex v_i. Applying Lemma 8.1, find disjoint stars $S_1, \ldots, S_\ell$ so that, for each $i \in [\ell]$, S_i is a star rooted at v_i with d_i leaves, and the stars are collectively rainbow in the new colouring. Observing that each colour must appear at most k times on $\cup_{i \in [\ell]} S_i$ in the original colouring completes the proof. $\square$

We wish to find vertex-disjoint collectively-rainbow stars in a locally k-bounded colouring. We show they exist within the stars found by Corollary 8.3, as follows.

Corollary 8.4. *Let $0 < \varepsilon < 1/100$ and $\ell \leq \varepsilon^2 n/2$. Let G be an n-vertex graph with minimum degree at least $(1 - \varepsilon)n$ which contains an independent set on the distinct vertices $v_1, \ldots, v_\ell$. Let $d_1, \ldots, d_\ell \geq 1$ be integers satisfying $\sum_{i \in [\ell]} d_i \leq (1 - 3\varepsilon)n/k$, and suppose G has a locally k-bounded edge-colouring.*

Then, G contains disjoint stars $S_1, \ldots, S_\ell$ so that, for each $i \in [\ell]$, S_i is a star rooted at v_i with d_i leaves, and $\cup_{i \in [\ell]} S_i$ is rainbow.

Proof. By Corollary 8.3, G contains disjoint stars $S_1, \ldots, S_\ell$ so that, for each $i \in [\ell]$, S_i is a star rooted at v_i with kd_i leaves, and $\cup_{i \in [\ell]} S_i$ has at most k edges of each colour. Create an auxillary bipartite graph H with vertex classes $[\ell]$ and $C(G)$, where there is an edge between $i \in [\ell]$ and $c \in C(G)$ exactly when there is an edge with colour c in S_i.

Given any set $A \subset [\ell]$, there are $\sum_{i \in A} kd_i$ edges in $\cup_{i \in A} S_i$, and therefore at least $\sum_{i \in A} d_i$ different colours. Thus, for any set $A \subset [\ell]$, we have $|N_H(A)| \geq \sum_{i \in A} d_i$. As Hall's generalised matching condition (see, for example, [8, Chapter 3, Corollary 11]) is satisfied, there is a collection of disjoint stars S_i', $i \in [\ell]$, in H, where, for each $i \in [\ell]$, the star S_i' is rooted at i and has d_i leaves.

For each $i \in [\ell]$, pick a neighbour in S_i for each colour in S_i', and call the resulting star S_i''. The stars S_i'', $i \in [\ell]$, then satisfy the corollary. $\square$

9 Rainbow connecting paths

In this section we prove that if we choose a random set X and a random set of colours C, then, with high probability, we can connect any small collection of pairs of vertices by collectively-rainbow vertex-disjoint paths of length 3, whose edges have colours in C and whose intermediate vertices are in X. Note that unlike some previous sections here we do assume that the random choices for X and C are completely independent. First we need the following simple proposition.

Proposition 9.1. *Let $1/k, p, q \gg 1/n > 0$ and suppose K_n has a locally k-bounded colouring. Let $X \subset V(K_n)$ and $C \subset C(K_n)$ be subsets with each element chosen independently at random with probability p and q respectively. Almost surely, each vertex has at least $pqn/2$ colour-C neighbours in X.*

Proof. For each $x \in V(K_n)$, let d_x be the number of colour-C neighbours in X. Note that d_x is k-Lipschitz, and $\mathbb{E}d_x = pq(n - 1)$. Thus, by Azuma's inequality (Lemma 3.4) with $t = pqn/3$, we have $\mathbb{P}(d_x < pqn/2) \leq 2\exp(-p^2q^2n/9k^2) = o(n^{-1})$. Thus, $d_x \geq pqn/2$ for each $x \in V(K_n)$ with probability $1 - o(1)$. $\square$

Lemma 9.2. *Let $1/k, p \gg \mu \gg 1/n > 0$ and suppose K_n has a locally k-bounded colouring. Let $X \subset V(K_n)$ and $C \subset C(K_n)$ be subsets with each element chosen independently at random with probability p. Almost surely, for each pair of distinct vertices $u, v \in V(K_n) \setminus X$ there are at least μn internally vertex-disjoint collectively C-rainbow u, v-paths with length 3 and internal vertices in X.*

Proof. Create a random partition $C = C_1 \cup C_2$ by assigning each element to a class uniformly at random. By Proposition 9.1, we almost surely have the following property.

Q1 Each vertex has at least $100k^2\mu^{1/3}n$ colour C_1-neighbours in X.

Note that $10k\mu^{1/3}n \geq n^{3/4}$. Almost surely, by Lemma 5.1, we have the following property.

Q2 Between every pair of disjoint subsets $A, B \subset V(K_n)$ with $|A|, |B| \geq 10k\mu^{1/3}n$ there are at least $p|A||B|/2 \geq 4k\mu n^2$ colour-C_2 edges.

Suppose then, for contradiction, there are some pair of distinct vertices $u, v \in V(K_n)$ and at most μn internally vertex-disjoint collectively-C-rainbow u, v-paths with length 3 and internal vertices in X. Fixing a maximal set of such paths, $\mathcal{P}$, let $U \subset X$ be their set of internal vertices and C' their set of edge colours. Note that $|U| \leq 2\mu n$ and $|C'| \leq 3\mu n$.

By **Q1**, we have $|N_{C_1 \setminus C'}(u, X \setminus U)| \geq 100k^2\mu^{1/3}n - 2\mu n - 3k\mu n \geq 10k\mu^{1/3}n$. Let $A \subset N_{C_1 \setminus C'}(u, X \setminus U)$ satisfy $|A| = 10k\mu^{1/3}n$, and let C'' be the set of colours between u and A. Using **Q1** again, we have

$$|N_{C_1 \setminus (C' \cup C'')}(v, X \setminus (U \cup A))| \geq 100k^2\mu^{1/3}n - 2\mu n - 3k\mu n - |A| - k|A| \geq 10k\mu^{1/3}n.$$

Let $B \subset N_{C_1 \setminus (C' \cup C'')}(v, X \setminus (U \cup A))$ satisfy $|B| = 10k\mu^{1/3}n$. By **Q2**, there are at least $4k\mu n^2$ C_2-edges between A and B, at most $kn \cdot |C'| \leq 3k\mu n^2$ of which have their colour in C'. Thus, there is some $x \in A$ and $y \in B$ so that $uxyv$ is a $(C \setminus C')$-rainbow path with internal vertices in $X \setminus U$. This contradicts the choice of $\mathcal{P}$. $\qquad\square$

10 Almost-spanning trees

We have now developed the tools that we will need to take an almost-spanning tree T and embed it into a locally k-bounded edge-coloured K_n. We will do this using (carefully chosen) random partitions $V(K_n) = X_0 \cup \ldots \cup X_\ell$ and $C(K_n) = C_0 \cup \ldots \cup C_\ell$, with both ℓ and the distributions depending on T. The vertices in X_0 and the colours in C_0 are used at various points in the embedding to find small parts of the tree T. This will be possible greedily, but to ease the checking for these parts of the proof we give appropriate embedding results for small trees (or forests) in Section 10.1. The two random partitions will not be independent, and depend both on T and each other, and on certain almost-sure properties holding. This complicates our use of the probabilistic method to find the partition we use, and so we explain these particulars carefully in Section 10.2. We then put together our proof of Theorem 1.1 in Section 10.3.

10.1 Embedding small parts of T

When there are many spare colours and vertices, we can construct small rainbow trees, finish rainbow matchings and find collectively-rainbow vertex-disjoint connecting paths, as we do in the following three propositions.

Proposition 10.1. *Suppose we have an m-vertex tree T and a graph G with a locally k-bounded colouring in which $\delta(G) \geq 3km$. Then, there is a rainbow copy of T in G.*

Proof. Let T' be a maximal subtree of T which has a rainbow copy, S' say, in G. Suppose, for contradiction, that $|S'| < m$ so that S' has a vertex, s say, to which a leaf can be appended to find a copy of a larger subtree than T' in T. The edges of S' have at most m colours collectively, so s has at most km colour-$C(S')$ neighbours. Thus, s must have at least $3km - |S'| - km > 0$ colour-$(C(G) \setminus C(S'))$ neighbours in $V(G) \setminus V(S')$. Such a neighbour allows S' to be extended to a larger rainbow copy of a subtree of T than S', a contradiction. $\qquad\square$

Proposition 10.2. *Suppose we have a graph G with a locally k-bounded colouring and disjoint sets $X, Y, Z \subset V(G)$ and disjoint sets of colours $C, C' \subset C(G)$, such that there is a rainbow C-matching with at least $|X| - m$ edges from X into Y, and each vertex in G has at least $2km$ colour-C' neighbours in Z.*

Then, there is a $(C \cup C')$-rainbow matching with $|X|$ edges from X into $Y \cup Z$ which uses at most m colours in C' and at most m vertices in Z.

Proof. Let M_0 be a C-rainbow matching with $|X| - m$ edges from X into Y. Greedily, pick matchings $M_0 \subset M_1 \subset \ldots \subset M_m$ so that, for each $i \in \{0, \ldots, m\}$, M_i is a $(C \cup C')$-rainbow matching with $|X| - m + i$ edges from X into $Y \cup Z$ which uses at most i colours in C' and at most i vertices in Z.

Note this is possible, as, for each $i \in [m]$, if we have a satisfactory matching M_{i-1}, then choosing $x \in X \setminus V(M_{i-1})$ we have that x has at most $k(i-1)$ colour-$(C' \cap C(M_{i-1}))$ neighbours in Z. Thus, x has at least $2km - k(i-1) \geq km$ colour-$(C' \setminus C(M_{i-1}))$ neighbours in Z, at most $i - 1 \leq m - 1$ of which can be in $V(M_{i-1})$. Therefore, we can pick a colour-$(C' \setminus C(M_{i-1}))$ neighbour y of x in $Z \setminus V(M_{i-1})$ and let $M_i = M_{i-1} \cup \{xy\}$.

Thus, we find a matching, M_m, as required. $\qquad\square$

Proposition 10.3. *Suppose we have a graph G with a locally k-bounded colouring containing the disjoint vertex sets $X = \{x_1, \ldots, x_m, x_1', \ldots, x_m'\}$ and Y such that, for each $i \in [m]$, there are at least $10m$ internally vertex-disjoint collectively-rainbow x_i, x_i'-paths of length three with interior vertices in Y. Then, there is a vertex disjoint set of collectively rainbow x_i, x_i'-paths, P_i, $i \in [m]$, of length three with interior vertices in Y.*

Proof. Let $I \subset [m]$ be a maximal subset for which there are vertex-disjoint collectively-rainbow x_i, x_i'-paths, P_i, $i \in I$, of length three with interior vertices in Y. Suppose, for contradiction, that $I \neq [m]$, and pick $j \in [m] \setminus I$.

Consider a collection $\mathcal{Q}$ of $10m$ internally vertex-disjoint collectively-rainbow x_j, x_j'-paths of length three with interior vertices in Y. Let $P = \cup_{i \in I} P_i$. Note that $|C(P)| = 3|I| \leq 3m$ and $|P| \leq 4m$. Thus, there can be at most $7m$ paths in $\mathcal{Q}$ with an edge with colour in $C(P)$ or a vertex in $V(P)$. Therefore, we can pick a path $P_j \in \mathcal{Q}$ so that P_i, $i \in I \cup \{j\}$, are vertex-disjoint collectively-rainbow x_i, x_i'-paths of length three with interior vertices in Y, a contradiction. $\qquad\square$

10.2   Dependence and the probabilistic method

To find a rainbow copy of an almost-spanning tree T, we will split the tree into pieces, find a random partition of the vertices and colours of K_n, and show that the properties we need to embed the tree (**R1-R5** as listed later) almost surely hold. Thus, there will be some partitions for which these properties hold, and using these we can then embed the tree T. Our implementation of the probabilistic method here is complicated by the dependence of some of the random sets in the partitions on each other, and furthermore **R4** may only hold if **R1-R3** hold, while **R5** holds only if the other properties all hold. Therefore, we discuss this here in detail to clarify this aspect of our proof, and explicitly give the simple formalities we later pass over.

In total, for some integer $\ell = O(\log^{10} n)$ and probabilities $p_0, \ldots, p_\ell$ depending on T we will pick random partitions

$$V(K_n) = X_0 \cup \ldots \cup X_\ell \quad \text{and} \quad C(K_n) = C_0 \cup \ldots \cup C_\ell$$

such that

(i) for each $0 \leq i \leq \ell$, $x \in V(K_n)$ and $c \in C(K_n)$, we have $\mathbb{P}(x \in X_i) = \mathbb{P}(c \in C_i) = p_i$,

(ii) the choice of the set containing each $x \in V(K_n)$ is made independently of the choice for each other vertex in $V(K_n)$,

(iii) the choice of the set containing each $c \in C(K_n)$ is made independently of the choice for each other vertex in $C(K_n)$, and

(iv) the choice of X_0 is made independently of the choice of C_0.

We do this by first selecting vertices for X_0 and colours for C_0 independently at random with probability p_0. Thus, (iv) holds and (i)-(iii) hold for $i = 0$. The properties **R1-R3** will depend only on X_0 and C_0, and will almost surely hold.

If **R1-R3** hold, we find a C_0-rainbow copy S_0 of part of the tree T using vertices in X_0, and then, depending on S_0, pair off some vertices in $V(K_n) \setminus X_0$ with colours in $C(K_n) \setminus C_0$. Formally, if **R1-R3** do not hold then we let $S_0 = \emptyset$ and take no such pairs. We then take a vertex set X_1 by selecting vertices in $V(K_n) \setminus X_0$ independently at random with probability $p_1/(1 - p_0)$. Each $x \in V(K_n)$ thus appears in X_1 with probability p_1 independently of each other vertex. Almost surely, if **R1-R3** hold, then **R4** will hold (a property depending only on X_0, C_0 and X_1). If any of **R1-R3** do not hold then we say **R4** does not hold.

Now we define the set C_1 which is disjoint from C_0. For each colour paired with a vertex, we take it in C_1 precisely if its paired vertex is in X_1. For each colour not paired with a vertex, we take it in C_1 independently at

random with probability $p_1/(1-p_0)$. Each colour is paired to at most one vertex, and each vertex is paired to at most one colour, so colours paired with a vertex appear in C_1 uniformly at random with probability $p_1/(1-p_0)$. Thus, whether it is paired to a vertex or not, each colour in $C(K_n)\setminus C_0$ appears in C_1 independently at random with probability $p_1/(1-p_0)$. Therefore, each colour in $C(K_n)$ appears in C_1 independently at random with probability p_1, completing the requirements of (i)–(iii) for $i=1$.

We then take random partitions

$$V(K_n)\setminus(X_0\cup X_1)=X_2\cup\ldots\cup X_\ell \quad\text{and}\quad C(K_n)\setminus(C_0\cup C_1)=C_2\cup\ldots\cup C_\ell,$$

by selecting the location of each $x\in V(K_n)\setminus(X_0\cup X_1)$ and $c\in C(K_n)\setminus(C_0\cup C_1)$ independently at random so that, for each $2\le i\le\ell$, $\mathbb{P}(x\in X_i)=\mathbb{P}(c\in C_i)=p_i/(1-p_0-p_1)$.

We can simply observe that, for each $x\in V(K_n)$ and $i\in[\ell]$, $\mathbb{P}(x\in X_i)=p_i$, and, furthermore, this is independent of the location of any other vertex, so that (i) holds for each vertex, and (ii) holds. Similarly, we can see that (i) holds for each colour, and (iii) holds.

If **R1**–**R4** hold, then our fifth property, **R5**, will hold for each $2\le i\le\ell$ with probability $1-o(\ell n^{-1})=1-o(1)$ by applying to X_i and C_i a result (Lemma 7.1) which needs precisely that (i)-(iii) hold for i but asks for no independence between C_i and X_i. If any of **R1**–**R4** do not hold, then we say that **R5** does not hold.

Finally, as $\mathbb{P}(\mathbf{R1}-\mathbf{R3}\text{ hold})=1-o(1)$, $\mathbb{P}(\mathbf{R4}\text{ holds}|\mathbf{R1}-\mathbf{R3}\text{ hold})=1-o(1)$ and $\mathbb{P}(\mathbf{R5}\text{ holds}|\mathbf{R1}-\mathbf{R4}\text{ hold})=1-o(1)$, we have that $\mathbb{P}(\mathbf{R1}-\mathbf{R5}\text{ hold})=1-o(1)$. Thus, there must be some partitions $V(K_n)=X_0\cup\ldots\cup X_\ell$ and $C(K_n)=C_0\cup\ldots\cup C_\ell$ such that **R1**–**R5** hold with the copy S_0 of part of T, and we will then complete the copy of T using such partitions, starting with the tree S_0.

10.3   Proof of Theorem 1.1

Proof of Theorem 1.1. Let μ satisfy $\varepsilon, 1/k\gg\mu\gg 1/n>0$ and let $D=\lceil\log^{10}n\rceil$. Let T be a tree with at most $(1-\varepsilon)n/k$ vertices and let K_n have a locally k-bounded colouring.

Split T. Using Lemma 4.2, find integers $\ell\le 10^4 D\mu^{-2}$ and $j\in[\ell]$ and a sequence of subgraphs $T_0\subset T_1\subset\ldots\subset T_\ell=T$ such that, for each $i\in[\ell]\setminus\{1,j\}$, T_i is formed from T_{i-1} by adding non-neighbouring leaves, T_j is formed from T_{j-1} by adding at most μn vertex disjoint bare paths with length 3, T_1 is formed from T_0 by adding vertex disjoint stars with at least D leaves each, and $|T_0|\le\mu n$.

Choose 'greedy' vertices and colours. Pick random subsets $X_0\subset V(K_n)$ and $C_0\subset C(K_n)$ by selecting each element uniformly at random with probability $p_0:=\varepsilon/200k$. By Proposition 9.1 and Lemma 9.2, we almost surely have the following properties (For more details on what happens if they, or any subsequent properties, do not hold, see Section 10.2).

R1 Each vertex in $V(K_n)$ has at least $10k\mu n$ colour-C_0 neighbours in X_0.

R2 For each pair of vertices $u,v\in V(K_n)$, there are at least $20\mu n$ internally vertex-disjoint collectively C_0-rainbow u,v-paths with length 3 and interior vertices in X_0.

Furthermore, by Lemma 3.3, almost surely we have $|X_0|,|C_0|\le\varepsilon n/100k$, and hence any vertex is contained in at most $\varepsilon n/100$ C_0-edges. Thus, the following almost surely holds.

R3 If G is the subgraph of K_n of the edges with colour in $C(K_n)\setminus C_0$, with any edges inside X_0 removed, then $\delta(G)\ge(1-\varepsilon/50)n$.

Embed T_0 and find rainbow stars. Using **R1** and Proposition 10.1, pick a C_0-rainbow copy, S_0 say, of T_0 in X_0. Then, for the appropriate integers $m\le n/D$ and $d_1,\ldots,d_m\ge D$, let $v_1,\ldots,v_m\in V(S_0)$ be such that S_0 can be made into a copy of T_1 by adding d_i new leaves at v_i, for each $i\in[m]$. Let $d=\sum_{i\in[m]}d_i=|T_1|-|T_0|\le(1-\varepsilon)n/k$. For each $i\in[m]$, let $n_i=\lceil(1-\varepsilon/8)nd_i/kd\rceil$. Note that $\sum_{i\in[m]}n_i\le(1-\varepsilon/8)n/k+m\le(1-\varepsilon/10)n/k$. Using **R3** and Corollary 8.4, find disjoint subsets $Y_i\subset V(K_n)\setminus X_0$, $i\in[m]$, so that $|Y_i|=n_i$, and $\{v_iy:i\in[m],y\in Y_i\}$ is $(C(K_n)\setminus C_0)$-rainbow.

For each vertex x in some set Y_i, pair x with the colour c of $v_i x$, noting that, as $\{v_i y : i \in [m], y \in Y_i\}$ is rainbow, each colour or vertex is in at most 1 pair. We will later choose a random vertex set X_1 and a random colour set C_1 so that, for such a pairing, $x \in X_1$ if and only if $c \in C_1$. For more details on this, see Section 10.2.

Choose probabilities p_i. For each $i \in [\ell]$, let $m_i = |T_i| - |T_{i-1}|$, and note that $m_1 = d$. For each $i \in [\ell-1]$, let

$$p_i = (1 + \varepsilon/4) k m_i / n + \varepsilon/4\ell \geq n^{-1/10^4}, \tag{6}$$

where the inequality follows as $\ell \leq 10^4 D \mu^{-2} = O(\log^{10} n)$. Let

$$p_\ell = 1 - p_0 - \sum_{i \in [\ell-1]} p_i = 1 - \frac{\varepsilon}{200k} - (1 + \varepsilon/4) k \cdot \frac{|T| - m_\ell - |T_0|}{n} - \frac{\varepsilon(\ell - 1)}{4\ell}$$

$$\geq 1 - (1 + \varepsilon/4) k \cdot \frac{(1 - \varepsilon) n/k - m_\ell}{n} - \frac{\varepsilon}{4} = (1 + \varepsilon/4) k m_\ell / n + \varepsilon \left(\frac{1}{2} + \frac{\varepsilon}{4} \right) \geq n^{-1/10^4}. \tag{7}$$

Choose X_1. Pick $X_1 \subset V(K_n) \setminus X_0$ by including each vertex independently at random with probability $p_1 / (1 - p_0)$. Recall that $m_1 = d$. By (6), we have, for each $i \in [m]$, that

$$p_1 n_i \geq (1 + \varepsilon/4) k m_1 / n \cdot (1 - \varepsilon/8) n d_i / kd$$
$$= (1 + \varepsilon/4) \cdot (1 - \varepsilon/8) d_i \geq (1 + \varepsilon/16) d_i \geq \log^{10} n.$$

Thus, by Lemma 3.3, for each $i \in [m]$, $\mathbb{P}(|X_1 \cap Y_i| \geq d_i) = \exp(-\Omega(\varepsilon^2 \log^{10} n)) = o(n^{-1})$. Thus, almost surely, the following property holds.

R4 For each $i \in [m]$, $|X_1 \cap Y_i| \geq d_i$.

Note, for later, that each vertex $x \in V(K_n)$ appears in X_1 independently at random with probability p_1.

Choose C_1. Let C^{paired} be the set of colours which appear between v_i and Y_i for some $i \in [m]$, and let $C^{\text{unpaired}} = C \setminus C^{\text{paired}}$ be the set of colours which never appear between any v_i and Y_i. We define a random set of colours C_1 as follows. For any colour $c \in C^{\text{paired}}$, c is included in C_1 whenever the vertex paired with c is in X_1, i.e. when c appears between v_i and $X_1 \cap Y_i$ for some $i \in [m]$. For any colour $c \in C^{\text{unpaired}} \setminus C_0$, c is included in C_1 independently at random with probability $p_1 / (1 - p_0)$. Thus, C_1 contains each colour paired with a vertex in X_1 and each unpaired colour outside C_0 is included uniformly at random. Thus, each colour appears in C_1 independently at random with probability p_1.

Choose a random vertex partition. Randomly partition $V(K_n) \setminus (X_0 \cup X_1)$ as $X_2 \cup \ldots \cup X_\ell$ so that, for each $x \in V(K_n) \setminus (X_0 \cup X_1)$, the class of x is chosen independently at random with $\mathbb{P}(x \in X_i) = p_i / (1 - p_0 - p_1)$ for each $2 \leq i \leq \ell$. Note that, for each $i \in \{0, 1, \ldots, \ell\}$, each $x \in V(K_n)$ appears in X_i independently at random with probability p_i, and the location of each vertex in $V(K_n)$ is independent of the location of all the other vertices.

Choose a random colour partition. Randomly partition $C(K_n) \setminus (C_0 \cup C_1)$ as $C_2 \cup \ldots \cup C_\ell$ so that, for each $c \in C \setminus (C_0 \cup C_1)$, the class of c is chosen independently at random with $\mathbb{P}(c \in C_i) = p_i / (1 - p_1 - p_0)$ for each $2 \leq i \leq \ell$. Note that, for each $0 \leq i \leq \ell$, each colour $c \in C(K_n)$ appears in C_i independently at random with probability p_i, and the location of each colour in $C(K_n)$ is independent of the location of all the other colours.

Rainbow matching properties. Note that, by (6) and (7), $m_i \leq p_i n / k$ for each $i \in [\ell]$. Therefore, from the properties of the random partitions of $C(K_n)$ and $V(K_n)$, and Lemma 7.1, we almost surely have the following property.

R5 For each $i \in [\ell]$ and subset $A \subset V(K_n) \setminus X_i$ with $|A| = m_i \leq p_i n / k$ there is a C_i-rainbow matching with at least $m_i - \mu p_i n$ edges from A into X_i.

As detailed in Section 10.2, we can thus fix deterministic partitions of $V(K_n)$ and $C(K_n)$, and the copy S_0 of T_0, for which **R1**–**R5** hold.

Extend to cover T_1. For each $i \in [m]$, use **R4** to add d_i leaves from $X_1 \cap Y_i$ to v_i in S_0 and call the resulting graph S_1. Note that these additions add leaves from X_1 using colours from C_1. Thus, $S_1 \subset K_n[X_0 \cup X_1]$ is a $(C_0 \cup C_1)$-rainbow copy of T_1 with at most μn colours in C_0 and at most μn vertices in X_0.

Iteratively, extend to cover $T_2, \dots, T_{j-1}$. Iteratively, for each $2 \le i \le j - 1$, extend S_{i-1} to $S_i \subset K_n[X_0 \cup \dots \cup X_i]$, a $(C_0 \cup \dots \cup C_i)$-rainbow copy of T_i with $|C(S_i) \cap C_0| \le \mu n + \sum_{i'=2}^{i} \mu p_{i'} n$ and $|V(S_i) \cap X_0| \le \mu n + \sum_{i'=2}^{i} \mu p_{i'} n$ (so that, certainly, $|C(S_i) \cap C_0| \le 2\mu n$ and $|V(S_i) \cap X_0| \le 2\mu n$). Note that T_i is obtained from T_{i-1} by adding a matching (i.e. a collection of non-neighbouring leaves). Let $A_i \subset S_{i-1}$ be the vertex set to which we need to attach the edges of the matching. Then we can first apply **R5** to sets A_i, X_i and the set of colours C_i to find a matching of size $|A_i| - \mu p_i n$ and then use **R1** and Proposition 10.2 with $G \subset K_n$ as the graph of colour-$((C_0 \setminus C(S_{i-1})) \cup C_i)$ edges, $C = C_i$, $C' = C_0 \setminus C(S_{i-1})$, $X = A_i, Y = X_i$ and $Z = X_0 \setminus V(S_{i-1})$ to find a rainbow matching covering the whole set A_i.

Extend to cover T_j. Using new vertices in X_0 and new colours in C_0, extend this to $S_j \subset K_n[X_0 \cup \dots \cup X_j]$, a $(C_0 \cup \dots \cup C_j)$-rainbow copy of T_j with $|C(S_j) \cap C_0| \le 4\mu n + \sum_{i=2}^{j} \mu p_i n$ and $|V(S_j) \cap X_0| \le 3\mu n + \sum_{i=2}^{j} \mu p_i n$. Note that, per path, we are using 3 additional colours from C_0 and 2 additional vertices from X_0, which explains the constants 4 and 3 in the last two inequalities. This is possible by **R2**, and Proposition 10.3 applied with $G \subset K_n$ as the graph of colour-$(C_0 \setminus C(S_{j-1}))$ edges and $Y = X_0 \setminus V(S_{j-1})$.

Iteratively, extend to cover $T_{j+1}, \dots, T_\ell$. Finally, for each $i \in \{j+1, \dots, \ell\}$, use **R1**, **R5** and Proposition 10.2 as before to extend S_{i-1} to $S_i \subset K_n[X_0 \cup \dots \cup X_i]$, a $(C_0 \cup \dots \cup C_i)$-rainbow copy of T_i with at most $4\mu n + \sum_{i'=2}^{i} \mu p_{i'} n$ colours in C_0 and at most $3\mu n + \sum_{i'=2}^{i} \mu p_{i'} n$ vertices in X_0. When this is finished, we have a rainbow copy of $T_\ell = T$, as required. $\square$

11   Concluding Remarks

- Our main theorem shows that properly coloured graphs have rainbow copies of every tree on $n - o(n)$ vertices. The most natural open problem is to ask how small the $o(n)$ term can be made. Note that we cannot take $o(n) = 0$ here, since there are proper colourings of K_n which do not have a rainbow copy of every n-vertex tree (see [31, 7]). This shows that the error term in Theorem 1.1 cannot be eliminated. However, we conjecture that it can be reduced to a constant.

  Conjecture 11.1. *There is a constant C so that every properly coloured K_n has a rainbow copy of every tree on $n - C$ vertices.*

  If true, this conjecture would lead to corresponding improvements to our applications for the conjectures of Graham-Sloane and Gronau-Mullin-Rosa.

- This paper gives a unified approach for attacking three conjectures about graph decomposition and labelling. In particular, we overcame one of the most significant barriers to progress towards solving these conjectures, i.e., we embedded all trees, rather than just trees with bounded degree. This "bounded degree" barrier exists in many other results about finding trees and more general subgraphs. Therefore, we expect that our methods might be useful to attack additional open problems in graph theory. One particular candidate is the Gyárfás Tree Packing Conjecture [21].

  Conjecture 11.2 (Gyárfás). *Let $T_1, \dots, T_{n-1}$ be trees with $|T_i| = i$ for each $i \in [n-1]$. The edges of K_n can be decomposed into $n - 1$ trees which are isomorphic to $T_1, \dots, T_{n-1}$ respectively.*

  Notice the strong parallels between this conjecture and Ringel's Conjecture — both conjectures concern the existence of decompositions of the complete graph into trees. Research on both these conjectures has progressed in parallel. That is, Conjecture 11.2 has also recently been proved for bounded degree trees by Joos, Kim, Kühn and Osthus [27] (see also [9, 32, 15, 29] for other results). There is no asymptotic version

of Conjecture 11.2 known for trees with arbitrary degrees, so it would be interesting to see whether our methods can be used here.

Acknowledgement

References

[1] A. Adamaszek, P. Allen, C. Grosu, and J. Hladký. Almost all trees are almost graceful. *arXiv preprint arXiv:1608.01577*, 2016.

[2] S. Akbari, O. Etesami, H. Mahini, and M. Mahmoody. On rainbow cycles in edge colored complete graphs. *Australasian Journal of Combinatorics*, 37:33, 2007.

[3] P. Allen, J. Böttcher, J. Hladký, and D. Piguet. Packing degenerate graphs. *arXiv preprint arXiv:1711.04869*, 2017.

[4] N. Alon, A. Pokrovskiy, and B. Sudakov. Random subgraphs of properly edge-coloured complete graphs and long rainbow cycles. *Israel Journal of Mathematics*, 222(1):317–331, 2017.

[5] N. Alon and J. Spencer. **The probabilistic method**. John Wiley & Sons, 2004.

[6] L. Andersen. Hamilton circuits with many colours in properly edge-coloured complete graphs. *Mathematica Scandinavica*, pages 5–14, 1989.

[7] F. Benzing, A. Pokrovskiy, and B. Sudakov. Long directed rainbow cycles and rainbow spanning trees. *arXiv:1711.03772*, 2017.

[8] B. Bollobás. **Modern graph theory**, volume 184. Springer, 2002.

[9] J. Böttcher, J. Hladký, D. Piguet, and A. Taraz. An approximate version of the tree packing conjecture. *Israel Journal of Mathematics*, 211(1):391–446, 2016.

[10] A. Brouwer, A. de Vries, , and R. Wieringa. A lower bound for the length of partial transversals in a latin square. *Nieuw Archief Voor Wiskunde*, 26(2):330, 1978.

[11] R. A. Brualdi and H. J. Ryser. **Combinatorial matrix theory**. Cambridge University Press, 1991.

[12] H. Chen and X. Li. Long rainbow path in properly edge-colored complete graphs. *arXiv preprint arXiv:1503.04516*, 2015.

[13] J. H. Dinitz and D. R. Stinson. **Contemporary design theory: A collection of surveys**, volume 26. John Wiley & Sons, 1992.

[14] L. Euler. Recherches sur une nouvelle espéce de quarrés magiques. *Verh. Zeeuwsch. Gennot. Weten. Vliss.*, 9:85–239, 1782.

[15] A. Ferber, C. Lee, and F. Mousset. Packing spanning graphs from separable families. *Israel Journal of Mathematics*, 219(2):959–982, 2017.

[16] A. Ferber and W. Samotij. Packing trees of unbounded degrees in random graphs. *J. London Math. Soc.*, to appear.

[17] J. A. Gallian. A dynamic survey of graph labeling. *The electronic journal of combinatorics*, 16(6):1–219, 2009.

[18] H. Gebauer and F. Mousset. On rainbow cycles and paths. *arXiv preprint arXiv:1207.0840*, 2012.

[19] R. Graham and N. Sloane. On additive bases and harmonious graphs. *SIAM Journal on Algebraic Discrete Methods*, 1(4):382–404, 1980.

[20] H.-D. Gronau, R. C. Mullin, and A. Rosa. Orthogonal double covers of complete graphs by trees. *Graphs and Combinatorics*, 13(3):251–262, 1997.

[21] A. Gyárfás and J. Lehel. Packing trees of different order into K_n. In *Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976)*, volume 1, pages 463–469, 1978.

[22] A. Gyárfás and M. Mhalla. Rainbow and orthogonal paths in factorizations of K_n. *Journal of Combinatorial Designs*, 18(3):167–176, 2010.

[23] A. Gyárfás, M. Ruszinkó, G. Sárközy, and R. Schelp. Long rainbow cycles in proper edge-colorings of complete graphs. *Australasian Journal of Combinatorics*, 50:45–53, 2011.

[24] G. Hahn. Un jeu de coloration. In *Regards sur la theorie des graphes, Actes du Colloque de Cerisy*, volume 12, page 18, 1980.

[25] D. Hughes. Biplanes and semi-biplanes. *Combinatorial Mathematics*, pages 55–58, 1978.

[26] S. Janson, T. Łuczak, and A. Rucinski. **Random graphs**. Wiley, 2011.

[27] F. Joos, J. Kim, D. Kühn, and D. Osthus. Optimal packings of bounded degree trees. *J. European Math. Soc., to appear*, 2018.

[28] A. Keedwell and J. Dénes. **Latin squares and their applications**. Elsevier Science, 2015.

[29] J. Kim, D. Kühn, D. Osthus, and M. Tyomkyn. A blow-up lemma for approximate decompositions. *Trans. Amer. Math. Soc., to appear*, 2018.

[30] U. Leck and V. Leck. On orthogonal double covers by trees. *Journal of Combinatorial Designs*, 5(6):433–441, 1997.

[31] M. Maamoun and H. Meyniel. On a problem of G. Hahn about coloured Hamiltonian paths in K_{2^t}. *Discrete mathematics*, 51(2):213–214, 1984.

[32] S. Messuti, V. Rödl, and M. Schacht. Packing minor-closed families of graphs into complete graphs. *Journal of Combinatorial Theory, Series B*, 119:245–265, 2016.

[33] G. Ringel. Theory of graphs and its applications. In *Proceedings of the Symposium Smolenice*, 1963.

[34] A. Rósa. On certain valuations of the vertices of a graph. In *Theory of Graphs (Internat. Symposium, Rome)*, pages 349–355, 1966.

[35] S. K. Stein. A combinatorial theorem. *J. London Math. Soc.*, 25:249–255, 1950.

[36] S. K. Stein. Transversals of Latin squares and their generalizations. *Pacific J. Math.*, 59:567–575, 1975.

[37] I. M. Wanless. Transversals in Latin squares: A survey. *Surveys in Combinatorics*, 2011.

[38] D. Woolbright. An $n \times n$ latin square has a transversal with at least $n - \sqrt{n}$ distinct symbols. *Journal of Combinatorial Theory, Series A*, 24(2):235–237, 1978.

[39] M. Woźniak. Packing of graphs and permutations—a survey. *Discrete Mathematics*, 276:379–391, 2004.

[40] H. Yap. Packing of graphs—a survey. *Annals of Discrete Mathematics*, 38:395–404, 1988.

[41] A. Żak. Harmonious order of graphs. *Discrete Mathematics*, 309(20):6055–6064, 2009.