Finding Large Rainbow Trees in Colourings of $K_{n,n}$

Julian Matthes[a]

Submitted: Apr 3, 2022; Accepted: Oct 6, 2023; Published: Nov 17, 2023

Abstract

A subgraph of an edge-coloured graph is called rainbow if all of its edges have distinct colours. An edge-colouring is called locally k-bounded if each vertex is incident with at most k edges of the same colour. Recently, Montgomery, Pokrovskiy and Sudakov showed that for large n, a certain locally 2-bounded edge-colouring of the complete graph K_{2n+1} contains a rainbow copy of any tree with n edges, thereby resolving a long-standing conjecture by Ringel: For large n, K_{2n+1} can be decomposed into copies of any tree with n edges. In this paper, we employ their methods to show that any locally k-bounded edge-colouring of the complete bipartite graph $K_{n,n}$ contains a rainbow copy of any tree T with $(1 - o(1))n/k$ edges. We show that this implies that every tree with n edges packs at least n times into $K_{n+o(1),n+o(1)}$. We conjecture that for large n, $K_{n,n}$ can be decomposed into n copies of any tree with n edges.

Mathematics Subject Classifications: 05C70, 05C05, 05B40

1 Introduction

1.1 History and context of the paper

A subgraph of an edge-coloured graph is called rainbow if all of its edges have distinct colours. An edge-colouring is called locally k-bounded if each vertex is incident with at most k edges of the same colour. This paper deals with edge-colourings exclusively, which is why the term colouring always refers to edge-colouring from here on. If such a colouring is locally 1-bounded, it is also called a proper colouring.

The theories of rainbow substructures go back to Euler and his work on Latin Squares, arrays of $n \times n$ cells where each cell is coloured by one of n colours so that no colour occurs twice in any row or column. Latin Squares are in fact intimately related to coloured bipartite graphs: A Latin Square can be represented as a properly coloured instance of the complete bipartite graph $K_{n,n}$ with vertex classes $\{u_1, \ldots, u_n\}$ and $\{v_1, \ldots, v_n\}$, where the edge between u_i and v_j represents the cell of the Latin Square in the i-th row and

[a] Department of Economics, Heidelberg University, Heidelberg, Germany (julian.matthes@awi.uni-heidelberg.de).

j-th column. Conversely, each proper colouring of $K_{n,n}$ using exactly n colours defines a Latin Square.

In 1847, Kirkman [6] showed that the complete graph on n vertices K_n can be decomposed into copies of a triangle if and only if $n \equiv 1, 3$ modulo 6. Problems like these, on graph decompositions, sparked interest among recreational and professional mathematicians and eventually helped inspire the theories of Steiner triple systems and design theory (see e.g. [14]).

In 1963, Ringel [10] posed the conjecture that K_{2n+1} can be decomposed into $2n + 1$ copies of any tree with n edges. Early progress on this was done by Kotzig (according to [12]), who considered a specific colouring of the complete graph known as the *nearest-distance-colouring* (or *ND-colouring*). This colouring derives its name from its construction, which is as follows: Order $2n + 1$ vertices in the plane such that they form the vertices of regular $(2n+1)$-gon and connect any pair of vertices with an edge where two edges have the same colour if and only if the Euclidean distances between the connected vertices are equal. Combinatorically, this corresponds to labelling the edge between the i-th and j-th vertex with colour $k \in \{1, \dots, n\}$ if either $i = j + k$ or $j = i + k$ with addition modulo $2n + 1$. Kotzig conjectured that the ND-coloured complete graph on $2n + 1$ vertices contains a rainbow copy of any tree with n edges. This implies Ringel's conjecture: If one finds a rainbow copy of an n-edge tree T in the ND-coloured K_{2n+1}, then by rotating the tree $2n$ times, each time by the angle $2\pi/(2n + 1)$, one ends up with $2n + 1$ edge-disjoint rainbow copies of the tree.

In subsequent years, Ringel's conjecture could be proved for certain small classes of trees (e.g. caterpillars, trees with at most 4 leaves, firecrackers, and more, see [3]) and, in a breakthrough in 2018, for large bounded degree trees (see [5]).

Also in 2018, Montgomery, Pokrovskiy and Sudakov [8] proved an asymptotic version of the two conjectures. Building on their methods, they could finally prove Ringel's conjecture by way of proving Kotzig's conjecture in 2020 (see [9]).

The methods they used are mostly of probabilistic nature, related to the famous probabilistic method, which was popularized by the work of Erdős and Rényi in 1959 in [1] (although earlier examples dating back as far as 1943 exist, see [13]). They also employed different instances of absorption, a technique which was initiated by Erdős, Gyárfás and Pyber in [2] and Krivelevich in [7] and adapted e.g. by Rödl, Ruciński and Szemerédi (see [11]).

1.2 Presentation of main results

In this paper, we work with the methods developed in [8] to show a similar result in the setting of complete bipartite graphs. We obtain:

Theorem 1. *Let $\varepsilon > 0$, $k, n \in \mathbb{N}$ such that $0 < \frac{1}{n} \ll \frac{1}{k}, \varepsilon$. Let T be a tree on at most $(1-\varepsilon)n/k$ vertices. Then any locally k-bounded colouring of $K_{n,n}$ contains a rainbow copy of T.*

The precise definition of $\ll$ we use is given in the preliminaries section.

From this first theorem, it follows that asymptotically, $K_{n,n}$ can be decomposed into n copies of any tree with n edges:

Theorem 2. *Let* $0 < \frac{1}{m} \ll \varepsilon$ *and* $n = \lceil m(1+\varepsilon) \rceil$*. Let* T *be a tree with* m *edges, then* T *packs at least* n *times into* $K_{n,n}$*.*

In order to conclude Theorem 2 from Theorem 1, we use a shifting argument similar to that of Kotzig above, utilizing the symmetry properties of a particular proper colouring of $K_{n,n}$, which we call *difference-colouring* (or *D-colouring*). It is defined as follows: Enumerate the vertices of one class of $K_{n,n}$ by $u_1, u_2, \ldots, u_n$ and those of the other class by $v_1, v_2, \ldots, v_n$. Now colour the edge between u_i and v_j in the colour k, if $i - j \equiv k$ modulo n. This yields a proper colouring of $K_{n,n}$ using n colours.

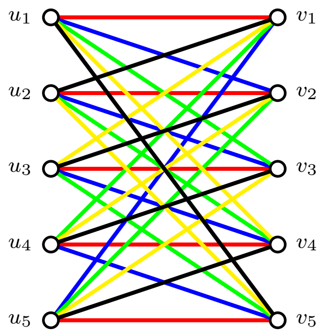

Figure 1: The D-Colouring for $n = 5$.

Note that in Theorem 1, the only restriction on the colouring of $K_{n,n}$ is that it is locally k-bounded, so we can apply the theorem to this colouring, setting $k = 1$.

Proof of Theorem 2. By Theorem 1, we can find a rainbow copy S_1 of T in a D-coloured instance of $K_{n,n}$. For $i \in \{1, \ldots, n-1\}$, let S_{i+1} be the tree that is the result of i applications of the colour-preserving graph endomorphism which maps u_j to u_{j+1} and v_j to v_{j+1} for $j + 1 \leqslant n$ and which maps u_n to u_1 and v_n to v_1. Now, $S_1, \ldots, S_n$ are n edge-disjoint rainbow copies of T. $\square$

This paper adapts the results from [8] from the setting of complete graphs to complete bipartite graphs, which works quite smoothly without much additional work. The authors of [8] went on to prove Ringel's conjecture in [9]: For large n, K_{2n+1} can be decomposed into n copies of any tree with n edges. It therefore seems reasonable to propose the following:

Conjecture 3. Let T be a tree with n edges. Then $K_{n,n}$ can be decomposed into n copies of T.

We did, however, not succeed in directly translating the methods from the proof of Ringel's conjecture in [9] to the bipartite case to prove this. The main difference between the two cases is in the structure of the colouring in question: The D-colouring is a proper colouring, as opposed to the ND-colouring of K_{2n+1} being locally 2-bounded. The fact that the ND-colouring is locally 2-bounded allows for more flexibility in finding a rainbow copy of some trees, namely those that consist mainly of large stars, roughly speaking. More precisely, the proof of Theorem 2.8 in Chapter 6 of [9] can seemingly not be translated to the case of complete bipartite graphs and the D-colouring.

The rest of this paper is structured as follows: The next section lists notations, definitions and preliminary results used for the proof. Section 3 then presents a result on the decomposition of a given tree, and explains how this result shapes the proof of Theorem 1. Section 4 collects a number of lemmas which are required to carry out the proof as lined out in Section 3. This includes many results from [8]: Some are cited in their original form, some have been adapted to the context of complete bipartite graphs and the adapted proofs are given. Section 5 then gives a proof of Theorem 1 using the preliminaries, Section 6 concludes.

2 Notation and Preliminaries

As usual, for $n \in \mathbb{N}$, let $[n]$ denote the discrete interval $\{1, \ldots, n\}$.

For the complete bipartite graph $K_{n,n}$, denote the two vertex classes by $V_1(K_{n,n})$ and $V_2(K_{n,n})$.

For lack of other colourings in this paper, an edge-colouring will also be referred to as a colouring. A colouring is called locally k-bounded if for any colour, each vertex in G is incident with at most k edges of that colour. A locally 1-bounded colouring is also referred to as a proper colouring.

Let G be a coloured graph. We will denote by $C(G)$ the set of colours of G and by $c(e)$ the colour of the edge $e \in E(G)$. $|G|$ denotes the number of vertices of G.

In a coloured graph G, for $C \subset C(G)$ and $V \subset V(G)$, denote by $N_C(V)$ the set of vertices that share an edge of a colour in C with a vertex in V. If $V = \{v\}$ is a singleton, we will also write $N_C(V) = N_C(v)$ and call the elements of this set the colour-C neighbours of v. For a vertex u, we use $N_C(u, V)$ to specify $N_C(u) \cap V$.

We refer to a subgraph of G as rainbow if no two of its edges have the same colour.

We will make use of the following asymptotic notation conventions: Let f and g be real-valued functions on $\mathbb{N}$. We say that $f = o(g)$ and $g = \omega(f)$, if $\lim_{n \mapsto \infty} \frac{f(n)}{g(n)} = 0$. We say $f = O(g)$ and $g = \Omega(f)$, if there is a constant $C > 0$ such that $f(n) \leqslant Cg(n)$ for all n.

We write $x \ll y$, if there is a positive continuous function f on $(0, 1]$ for which the remainder of the proof works with $x \ll y$ replaced by $x \leqslant f(y)$.

We say that an event occurs *with high probability*, if it occurs with probability $1 - o(1)$. For events A, B, denote $P(A)$ the probability of A, $P(A|B)$ the probability of A given B. For a real random variable X, denote by $E[X]$ the expected value of X.

Lemma 4. *[Chernoff's concentration inequality, [4]] Let $X \sim \mathrm{Bin}(n, p)$ and $0 < \varepsilon < \frac{3}{2}$, then:*

$$P(|X - E[X]| \geqslant \varepsilon E[X]) \leqslant 2 \exp\left(-\frac{\varepsilon^2 E[X]}{3}\right).$$

A random variable is called k-Lipschitz, if changing $\omega \in \prod_{i=1}^{n} \Omega_i$ in any one coordinate changes $X(\omega)$ by at most k.

Lemma 5. *[Azuma's concentration inequality, [4]] Let X be a k-Lipschitz random variable on $\prod_{i=1}^{n} \Omega_i$, then:*

$$P(|X - E[X]| > t) \leqslant 2 \exp\left(-\frac{t^2}{k^2 n}\right).$$

3 Tree Decomposition and Sketch of the Proof of Theorem 1

The following decomposition is at the heart of the proof of Theorem 1. The remaining sections will be dealing with embedding parts of the decomposition of a given tree in rainbow fashion in a coloured $K_{n,n}$.

Definition 6. A *bare path* in a tree T is a path whose interior vertices all have degree 2 in T.

Definition 7. We say that L is a *set of non-neighbouring leaves* of a tree T, if $L \subset V(T)$ is a set of leaves such that no two vertices in L share a neighbour.

Lemma 8. *[Lemma 4.2 from [8]] Given integers D and $n, \mu > 0$ and a tree T with at most n vertices, there are integers $\ell \leqslant 10^4 D \mu^{-2}$ and $j \in \{2, \dots, \ell\}$ and a sequence of subgraphs $T_0 \subset T_1 \subset \cdots \subset T_\ell = T$ such that:*

1. *For each $i \in [\ell] \backslash \{1, j\}$, T_i is formed from T_{i-1} by adding a set of non-neighbouring leaves,*
2. *T_j is formed from T_{j-1} by adding at most μn vertex-disjoint bare paths of length 3,*
3. *T_1 is formed from T_0 by adding vertex-disjoint stars with at least D leaves each, and*
4. *$|T_0| \leqslant \mu n$.*

At this point, it is appropriate to give a very brief sketch of the proof of Theorem 1 which will be presented in Section 4. Finding a rainbow copy of a tree T in a coloured instance of $K_{n,n}$ will be done by embedding the members of the sequence $T_0 \subset T_1 \subset \cdots \subset T_\ell = T$ existing for T by Lemma 8 sequentially as rainbows. Lemma 8 states that to extend a rainbow copy of T_i in $K_{n,n}$ to a rainbow copy of T_{i+1} for $i \in \{0, \dots, \ell - 1\}$, one will have to do one of three things: add large stars, add a matching of non-neighbouring leaves or add bare paths of length 3. In the next subsections, we will develop tools for each of these tasks and for embedding a rainbow copy of T_0, using random subsets of the vertex and colour set of $K_{n,n}$ with carefully chosen distributions for each step.

The proof of Theorem 1 will then consist of showing that we can carefully choose a partition of the vertex and colour set of a coloured instance of $K_{n,n}$ such that a rainbow copy of T_0 can be found with high probability and each of the extension steps can subsequently be carried out successfully with high probability, as well. By using vertices and colours from different sets of the partitions for each step, we are making sure that the copy of the subforest of T attained after each step is indeed rainbow. Knowing that an algorithm that succeeds with high probability for such a random partition exists, we can then deduce that there must be some initial partitions of the vertices and colours for which the algorithm indeed succeeds and hence we know for sure that there must be a rainbow copy of T.

4 Finding Rainbow Structures

4.1 Embedding Small Structures as Rainbows

Using the notation from Lemma 8, the following lemma gives us the means to find a rainbow copy of the center piece T_0 of a tree.

Lemma 9. *[Proposition 10.1 from [8]] Suppose we have an m-vertex tree T and a graph G with a locally k-bounded colouring in which the minimum degree of G, $\delta(G)$, satisfies $\delta(G) \geqslant 3km$. Then, there is a rainbow copy of T in G.*

4.2 Finding Rainbow Stars in Coloured Graphs

The following lemma is the technical result needed to add large stars to the center piece. It is Corollary 8.4 from [8] with the small change that they define G to be an n-vertex graph. The proof they give is also valid if we relax this condition to n being an upper bound on the maximum degree. The proof is purely deterministic, it uses a switching argument. It should be noted that the bound on the size of the union of all stars cannot be dropped using this technique, which provides a serious obstacle when trying to extend the result to the non-asymptotic case of Theorem 1 (Conjecture 3).

Lemma 10. *Let $0 < \varepsilon < \frac{1}{100}$ and $\ell \leqslant \varepsilon^2 \frac{n}{2}$. Let G be a graph with minimum degree at least $(1-\varepsilon)n$ and maximum degree at most n which contains an independent set on the distinct vertices $v_1, \ldots, v_\ell$. Let $d_1, \ldots, d_\ell \geqslant 1$ be integers satisfying $\sum_{i \in [\ell]} d_i \leqslant (1-3\varepsilon)\frac{n}{k}$, 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 centered at v_i with d_i leaves, and $\bigcup_{i \in [\ell]} S_i$ is rainbow.*

4.3 Edge Concentration in Randomized Rainbow Subgraphs

In the next subsections, we will consider graphs that arise as subgraphs of a coloured $K_{n,n}$ by including each vertex at random with some probability p and all edges of a given colour with some probability q. The following lemmas collect some simple properties of such subgraphs.

Definition 11. Let B be any set and $p \in [0,1]$. A random subset $A \subset B$ is called p-random if A is formed by including each element of B independently with probability p.

Definition 12. Let A be a p-random sub of some finite set and B be a q-random subset of some finite set. We say that A and B are independent, if the choices for A and B are made independently, that is, if $P(A = A_0 \wedge B = B_0) = P(A = A_0)P(B = B_0)$ for any outcomes A_0 and B_0 of A and B.

Lemma 13. *Let $0 < \frac{1}{n} \ll \frac{1}{k}, p, q$ and suppose $K_{n,n}$ has a locally k-bounded colouring. Let $V \subset V(K_{n,n})$ and $C \subset C(K_{n,n})$ be independent and p-random and q-random, respectively. With probability $1 - o(n^{-1})$, each vertex has at least $\frac{pqn}{2}$ colour-C neighbours in V.*

Proof. This goes along the lines of the proof of Proposition 9.1 in [8].

Fix $v \in V(K_{n,n})$, then for any other vertex u not in the same class of the bipartition as v, we have $P(u \in N_C(v) \cap V) = pq$ by the independence of V and C. Because $|N_C(v) \cap V|$ is a function of $3n-1$ random variables (all the n colours that could be present in the class of the bipartition not containing v, and all the $2n$ vertices except v) and is k-Lipschitz as the colouring is locally k-bounded (changing a vertex changes the integer in question by at most one, changing a colour by at most k), we have by Lemma 5, that

$$P\left(|N_C(v) \cap V| \leqslant \frac{pqn}{2}\right) \leqslant 2\exp\left(\frac{-p^2q^2n}{1000k^2}\right) = o(n^{-2}).$$

This is a result for a single vertex v. Taking the union bound over all of the $2n$ vertices then shows that the statement of the lemma does not hold with probability $o(n \cdot n^{-2})$, hence it holds with probability $1 - o(n^{-1})$ as claimed. □

Lemma 14. *Let $0 < \frac{1}{n} \ll \varepsilon, k$ and let $p \geqslant n^{-1/100}$, and $\varepsilon \leqslant 1$. Let $K_{n,n}$ have a locally k-bounded colouring and suppose G is a subgraph of $K_{n,n}$ chosen by including all of the edges of a random subset of the colours, where each colour is included independently at random with probability p. Then, with probability $1-o(n^{-1})$, for any two sets $A \subset V_1(K_{n,n}) \cap V(G)$ and $B \subset V_2(K_{n,n}) \cap V(G)$ with $|A|, |B| \geqslant n^{3/4}$:*

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

Note that $p|A||B|$ is the expected number of such edges, so we are essentially asking about the about the deviation of the random variable that counts the edges of G between A and B from its expected value.

Proof. This goes along the lines of the proof of Lemma 5.1 in [8].

Let $\ell = \lceil \frac{\sqrt{2kn}}{\varepsilon^2 p^2} \rceil$, so that $\ell \leqslant n^{0.52+o(1)} \leqslant n^{0.6+o(1)}$ by the choice of the parameters. The key to the proof is to show that there exist partitions $A = A_1 \cup A_2 \cup \cdots \cup A_\ell$, $B = B_1 \cup B_2 \cup \cdots \cup B_\ell$, such that for fixed i and j, there are few edges between A_i and B_j that share their colour with another edge between the same A_i and B_j. This is the first claim in this proof. Once this is established, we will then assign each pair (A_i, B_j) to one of two classes, depending on whether there are many different colours between A_i and B_j or not. The result then follows by counting edges between pairs in both classes.

Let us specify the assertion about the existence of a partition as mentioned above. Let w.l.o.g. $|A| \geqslant |B|$, then:

Claim 15. *There are partitions $A = A_1 \cup A_2 \cup \cdots \cup A_\ell$, $B = B_1 \cup B_2 \cup \cdots \cup B_\ell$, such that there are at most $\frac{\varepsilon^2 p^2 |A||B|}{100}$ edges ab between A and B for which there is another edge $a'b'$ with $c(ab) = c(a'b')$ and $ab, a'b' \in E(K_{n,n}[A_i, B_j])$ for some $i, j \in [\ell]$.*

Proof. Pick such partitions at random by choosing the class of each element of A and B, respectively, independently and uniformly at random. Observe the following: Fix any colour-c edge e, and let A_i, B_j be the classes it goes between. Then the probability that there is another colour-c edge between A_i and B_j sharing a vertex with e is at most $2(k-1)/\ell$ (by independence and as by local k-boundedness, e touches at most $2(k-1)$ other colour-c edges). The probability that there is a colour-c edge that is vertex-disjoint from e between A_i and B_j is at most kn/ℓ^2 (since there are at most kn colour-c edges in total, each edge is incident to two vertices and i and j are both fixed). By adding up and noting $\frac{2(k-1)}{\ell} \leqslant \frac{2kl}{\ell^2} \leqslant \frac{kn}{\ell^2}$ (for n large enough), we see that the probability that there is another colour-c edge between A_i and B_j is at most

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

by the choice of ℓ, and since $\varepsilon \leqslant 1$. Thus, as there are $|A||B|$ edges between A and B in $K_{n,n}$, the expected number of edges which have non-unique colour across their classes is at most $\varepsilon^2 p^2 |A||B|/100$. By applying the usual reasoning for the probabilistic method, there must exist a partition with the desired property. $\square$

We now introduce the following property:
Two sets $X \subset V_1(K_{n,n})$ and $Y \subset V_2(K_{n,n})$ satisfy the property $\mathbf{P}$, if

P1 $|X|, |Y| \geqslant n^{1/10}$, and

P2 There are at least $\left(1 - \frac{\varepsilon p}{8}\right)|X||Y|$ different colours between X and Y in $K_{n,n}$

This essentially tells us whether or not there are many different colours present between X and Y, which we will use to classify the pairs (A_i, B_j) from above. For pairs satisfying $\mathbf{P}$, we get:

Claim 16. *With probability $1 - o(n^{-1})$, the following holds. Whenever $X \subset V_1(K_{n,n})$ and $Y \subset V_2(K_{n,n})$ satisfy $\boldsymbol{P}$, then*

$$|e_G(X, Y) - p|X||Y|| \leqslant \frac{\varepsilon p |X||Y|}{2}.$$

Proof. For any such sets X and Y, we can select a rainbow subgraph R of $K_{n,n}[X, Y]$ with $\left(1 - \frac{\varepsilon p}{8}\right)|X||Y|$ edges. Notice that $e(R \cap G) \sim \text{Bin}\left((1 - \frac{\varepsilon p}{8})|X||Y|, p\right)$. By Lemma 4 applied with $\frac{\varepsilon p}{8}$ for ε, with probability at least $1 - \exp\left(-\frac{\varepsilon^2 p^3 |X||Y|}{10^3}\right)$, we have

$$\left(1 - \frac{\varepsilon p}{4}\right) p|X||Y| \leqslant e(R \cap G) \leqslant \left(1 + \frac{\varepsilon p}{4}\right) p|X||Y|.$$

In combination with $e(K_{n,n}[X,Y]-R)\leqslant \frac{\varepsilon p|X||Y|}{8}$, this implies that the equation from the claim holds for X and Y with the above probability. To simplify the latter, note that by $|X|,|Y|\geqslant n^{1/10}$ and $p\geqslant n^{-\frac{1}{100}}$, we have $p^3|Y|\geqslant n^{7/100}=\omega(\log n)$. Hence, the equation from the claim holds for X and Y with probability at least $1-\exp(-\max\{|X|,|Y|\}\cdot\omega(\log n))$. Summing over possible sizes of X and Y yields that the equation from the claim holds with probability at least

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

Fix partitions of $A=A_1\cup A_2\cup\cdots\cup A_\ell$ and $B=B_1\cup B_2\cup\cdots\cup B_\ell$ like in the above Claim 15.

Next, we will need to look at pairs not satisfying **P**. Let the subgraph H of $K_{n,n}[A,B]$ be defined as follows: Start with $K_{n,n}$ and remove all the edges between the pair A_i and B_j if the pair (A_i,B_j) does not satisfy **P**.

Claim 17. *To obtain H from $K_{n,n}[A,B]$, one has to delete at most $\frac{\varepsilon p|A||B|}{4}$ edges.*

Proof. By deleting edges between pairs that don't satisfy **P1**, we delete at most

$$\ell n^{1/10}(|A|+|B|)\leqslant n^{0.7+o(1)}(|A|+|B|)\leqslant\frac{\varepsilon p|A||B|}{8}$$

edges.

Let I denote the set of pairs of indices of pairs that do not satisfy **P2**, but do satisfy **P1**. Note that we have

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

by Claim 15.

By multiplication with $\frac{8}{\varepsilon p}$, we see that

$$\sum_{(i,j)\in I}|A_i||B_j|\leqslant\frac{\varepsilon p|A||B|}{8}.$$

In total, to obtain H from $K_{n,n}[A,B]$, one has to delete at most

$$\frac{\varepsilon p|A||B|}{8}+\frac{\varepsilon p|A||B|}{8}=\frac{\varepsilon p|A||B|}{4}$$

edges. □

Lastly, we put Claim 16 to work. Denote by I' the set of pairs of indices of the pairs that satisfy **P**, hence between which all of the initial edges are also in H. Then, with probability $1 - o(n^{-1})$, we have

$$\begin{aligned}|e_{G\cap H}(A,B) - pe_H(A,B)| &= \left|\sum_{(i,j)\in I'} e_G(A_i,B_j) - p|A_i||B_j|\right| \\ &\leqslant \sum_{(i,j)\in I'} |e_G(A_i,B_j) - p|A_i||B_j|| \\ &\leqslant \sum_{(i,j)\in I'} \varepsilon p|A_i||B_j|/2.\end{aligned}$$

The first inequality is the triangle inequality, then we use Claim 16, which we can as $\mathbf{P}$ is true for all pairs in the sum (this is how we constructed H).

Thus, as A and B are partitioned by the A_i, B_j respectively, with probability $1-o(n^{-1})$ we get

$$|e_{G\cap H}(A,B) - pe_H(A,B)| \leqslant \varepsilon p|A||B|/2. \tag{1}$$

We can finally prove the claim of the lemma:

$$\begin{aligned}|e_G(A,B) - p|A||B|| \leqslant |e_G(A,B) - e_{G\cap H}(A,B)| &+ p|e_H(A,B) - |A||B|| \\ &+ |e_{G\cap H}(A,B) - pe_H(A,B)|\end{aligned}$$

holds by the triangle inequality. By applying (1) to the last term, Claim 16 to the second term and noting that
$E_G(A,B)\backslash E_{G\cap H}(A,B) \subset E_{K_{n,n}}(A,B)\backslash E_H(A,B)$ implies the inequality
$|e_G(A,B) - e_{G\cap H}(A,B)| \leqslant |e_H(A,B) - |A||B||$, we get:

$$|e_G(A,B) - p|A||B|| \leqslant 2|e_H(A,B) - |A||B|| + \frac{\varepsilon p|A||B|}{2},$$

and hence

$$|e_G(A,B) - p|A||B|| \leqslant \varepsilon p|A||B|,$$

as desired. □

4.4 Finding Rainbow Bare Paths

Next, we will look for rainbow paths between given end-points with internal vertices inside a fixed random set. This will be needed for the step in the proof of Theorem 1 where we add bare paths to T_{j-1} to obtain T_j, where T_{j-1} and T_j are the corresponding elements in the sequence of subgraphs provided by Lemma 8. To find such paths, we employ Lemma 14 from the previous subsection. We will be looking for collections of internally vertex disjoint collectively C-rainbow u, v-paths for given vertices u and v, i.e. a set S of paths between u and v with the property that S is C-rainbow and any two paths in S do not share an internal vertex.

Lemma 18. *Let $0 < \frac{1}{n} \ll \mu \ll \frac{1}{k}, p$ and suppose $K_{n,n}$ has a locally k-bounded colouring. Let $X \subset V(K_{n,n})$ and $C \subset C(K_{n,n})$ be independent and p-random subsets. With high probability, for each pair of distinct vertices $u \in V_1(K_{n,n})\backslash X$, $v \in V_2(K_{n,n})\backslash X$ there are at least μn internally vertex-disjoint collectively C-rainbow u, v-paths with length 3 and internal vertices in X.*

Proof. This goes along the lines of the proof of Lemma 9.2 in [8].

Create a random partition $C = C_1 \cup C_2$ by assigning each colour in C uniformly at random to either C_1 or C_2. By Lemma 13, we get that with high probability,

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

By the choice of k and μ, we have that $20k\mu^{1/3}n \geqslant n^{3/4}$, which means that we can apply Lemma 14 to get the following with high probability:

Q2 Between every pair of disjoint subsets $A \subset V_1(K_{n,n})$, $B \subset V_2(K_{n,n})$ with $|A|, |B| \geqslant 20k\mu^{1/3}n$, there are at least $p|A||B|/3 \geqslant 4k\mu n^2$ colour-C_2 edges.

Now, for a contradiction, suppose that we have $u \in V_1(K_{n,n}), v \in V_2(K_{n,n})$ and less than μn internally vertex-disjoint paths of length 3 with internal vertices in X that are collectively C-rainbow. Let $\mathcal{P}$ denote a maximal set of such paths and let U denote the set of their internal vertices, so that by assumption, $U \subset X$. Further, denote by C' the set of edge colours in $\mathcal{P}$. We have $|U| < 2\mu n$ and $|C'| < 3\mu n$. Now, by applying **Q1**, we find that

$$|N_{C_1\backslash C'}(u, X\backslash U)| \geqslant 100k^2\mu^{1/3}m - 2\mu n - 3k\mu n \geqslant 20k\mu^{1/3}n.$$

Let $A \subset N_{C_1\backslash C'}(u, X\backslash U)$ such that $|A| = 20k\mu^{1/3}n$ and let C'' be the set of colours between u and A. Using **Q1** again, we have

$$|N_{C_1\backslash(C'\cup C'')}(v, X\backslash(U \cup A))| \geqslant 100k^2\mu^{1/3}m - 2\mu n - 3k\mu n - |A| - k|A| \geqslant 20k\mu^{1/3}n.$$

Now let $B \subset N_{C_1\backslash(C'\cup C'')}(v, X\backslash(U \cup A))$ satisfy $|B| = 20k\mu^{1/3}n$. By **Q2**, there are at least $4k\mu n^2$ colour-C_2 edges between A and B, at most $kn|C'| \leqslant 3k\mu n^2$ of which have their colour in C'. This contradicts the maximality of $\mathcal{P}$, as there must then be some $x \in A$, $y \in B$ such that $uxyv$ is a $(C\backslash C')$-rainbow path with internal vertices in $X\backslash U$! $\square$

Lemma 19. *[Proposition 10.3 from [8]] Suppose we have a graph G with a locally k-bounded colouring containing the disjoint vertex sets $X_1 = \{x_1, \ldots, x_m\} \subset V(G)$, $X_2 = \{x'_1, \ldots, x'_m\} \subset V(G)$ 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.*

4.5 Finding Rainbow Matchings

We will lastly need to address the problem of finding rainbow matchings from a set $A \subset V_j(K_{n,n})$ for $j \in [2]$ into a random set $X \in V(K_{n,n})$ such that A is covered by the matching. Lemma 20 deals with covering almost all of a not-too-large such set A while Lemma 24 specifies a simple condition under which such almost-covering matchings can be finished to cover all of A using some set-aside colours and vertices. This is preparing for an absorption argument in the proof of Theorem 1. The arguments in this subsection are similar to those in Sections 6 and 7 in [8].

Lemma 20. *Let $0 < \frac{1}{n} \ll \varepsilon, k$, and suppose $K_{n,n}$ has a locally k-bounded colouring and $p \geqslant n^{-1/10^4}$. Let $j \in [2]$ and let $X \subset V_j(K_{n,n})$ and $C \subset C(K_{n,n})$ be p-random, allowing for dependence between colours and vertices. Then, with probability $1 - o(n^{-1})$, for each set $A \subset V_{3-j}(K_{n,n})$ with $|A| \leqslant \frac{pn}{k}$ and $j \in [2]$, there is a C-rainbow matching in $K_{n,n}$ of size at least $|A| - \varepsilon pn$ between A and X.*

The following proposition establishes the existence of almost-covering matchings in certain bipartite graphs where not too few colours appear between any two subsets of sufficient size.

Proposition 21 (Lemma 7.2 from [8])**.** *Let $0 < \frac{1}{n} \ll \eta \ll \varepsilon$. Let G be a bipartite graph with classes X and Y, $|X| = n$ 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.*

In order to leverage Proposition 21, we construct a bipartite graph of the form given in the proposition inside our coloured $K_{n,n}$, which is essentially achieved by the following result:

Proposition 22. *Let $k \in \mathbb{N}$, $\varepsilon > 0$, and $p \geqslant n^{-1/10^4}$. Let $K_{n,n}$ have a locally k-bounded colouring. Let $j \in [2]$, $X \subset V_j(K_{n,n})$ and $C \subset C(K_{n,n})$ be p- and q-random subsets respectively, where the events $\{x \in X\}$ might depend on the events $\{c \in C\}$, i.e. are not necessarily independent. Then, with probability $1 - o(n^{-1})$, for each $A \subset V_{3-j}(K_{n,n})$ and $B \subset X$ with $|A| \geqslant (2n)^{3/4}$ and $|B| \geqslant 2\varepsilon pn$, there are at least $(1 - \varepsilon)|B|/k$ colours in C which appear between A and B.*

In order to prove Proposition 22, we need the following result:

Proposition 23. *[Lemma 6.1 from [8]] Let $k \in \mathbb{N}$ and $\varepsilon > 0$ be constant, and $p \geqslant n^{-1/10^4}$. 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 $c \in C(K_n)$, $P(x \in X) = P(c \in C) = p$, 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 $|A| \geqslant n^{3/4}$ and $|B| \geqslant \varepsilon pn$, there are at least $(1 - \varepsilon)|B|/k$ colours in C which appear between A and B in K_n.*

Proof of Proposition 22. For a given locally k-bounded colouring $K_{n,n}$, add a rainbow set of edges with new colours to it to obtain a coloured instance of K_{2n}. Since the used colours are distinct and new, this colouring of K_{2n} is also locally k-bounded. Since $p \geqslant n^{-1/10^4} \geqslant (2n)^{-1/10^4}$, we can apply Proposition 23 to this colouring of K_{2n} and derive that with probability $1 - o(n^{-1})$, for each $A \subset V_j(K_{n,n}) \setminus X$ and $B \subset X \cap V_{3-j}(K_{n,n})$ with $|A| \geqslant (2n)^{3/4}$ and $|B| \geqslant 2\varepsilon pn$, there are at least $(1-\varepsilon)|B|/k$ colours in C which appear between A and B in K_{2n}. By the definition of this colouring K_{2n}, these edges between A and B in K_{2n} are also edges between A and B in $K_{n,n}$: Only edges connecting vertices in the same component were added, but A and B lie in different components! $\square$

Proof of Lemma 20. Let η be a fixed constant not dependent on n which satisfies $0 < \eta \ll \varepsilon$. With probability $1 - o(n^{-1})$, by Proposition 22 applied with $\frac{\eta}{4k}$ for ε, we get:

Q For each $A \subset V_{3-j}(K_{n,n})$ and $B \subset X$ with $|A|, |B| \geqslant \frac{\eta pn}{2k} \geqslant (2n)^{3/4}$, there are at least $\frac{(1-\eta)|B|}{k}$ colours in C between A and B.

Also, With probability $1 - o(n^{-1})$, by Lemma 4, we have

$$(1-\eta/2)pn \leqslant |X| \leqslant (1+\eta/2)pn.$$

We claim that the property in the lemma holds. Let $A \subset V_j(K_{n,n})$ with $|A| \leqslant pn/k$. Add vertices to A from $V_j(K_{n,n})$, or delete vertices from A, to get a set A' with $|A'| = \lfloor (1-\frac{\eta}{2})\frac{pn}{k} \rfloor =: m$ and $|A \backslash A'| \leqslant \eta \frac{pn}{k} \leqslant \varepsilon \frac{pn}{2}$. Let X' be a subset of X of size km. Since $\eta m \geqslant \eta \frac{pn}{2k}$, we can apply **Q** to see that for any subsets $A'' \subset A'$ and $B \subset X'$ with $|A''|$, $|B| \geqslant \eta m$, there are at least $(1-\eta)\frac{|B|}{k}$ colours in C between A'' and B. Thus, by Proposition 21, there is a C-rainbow matching with at least $(1-\frac{\varepsilon}{2})m \geqslant |A'| - \varepsilon \frac{pn}{2}$ edges between A' and X'. As $|A \backslash A'| \leqslant \varepsilon \frac{pn}{2}$, at least $|A| - \varepsilon pn$ of the edges in this C-rainbow matching must lie between A and X. $\square$

Lemma 24. *[Proposition 10.2 in [8]] 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 C-rainbow 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.*

5 Proving Theorem 1

The following proof goes along the lines of the proof of Theorem 1.1 in [8].

Proof of Theorem 1. Let $0 < 1/n \ll \mu \ll \varepsilon, 1/k$.

Step 1: Splitting up T using Lemma 8.

With the parameters we just chose, Lemma 8 guarantees that there exists some $\ell \leqslant 10^4 D\mu^{-2}$, such that we can find a sequence of forests $T_0 \subset T_1 \subset \cdots \subset T_\ell$ satisfying the following properties:

1. T_0 has at most μn vertices.
2. T_1 can be obtained from T_0 by adding a collection of large vertex-disjoint stars with centers in T_0, large meaning that each of them has at least $D = \lceil \log^{10} n \rceil$ leaves.
3. For $i \in \{2, \dots, j-1\}$, T_i is obtained from T_{i-1} by adding a set of non-neighbouring leaves.
4. We can add a small set of vertex disjoint bare paths of length 3 to connect the components of T_{j-1} to obtain a tree, T_j. Small means that there are at most μn such paths.
5. For $i \in \{j+1, \dots, \ell\}$, again T_i is obtained from T_{i-1} by adding a set of non-neighbouring leaves.
6. $T_\ell = T$.

Step 2: Providing sets X_0, C_0 for later absorption.

Set $p_0 := \frac{\varepsilon}{400k}$ and choose $X_0 \subset V(K_{n,n})$ and $C_0 \subset C(K_{n,n})$ p_0-randomly. These are our vertex and colour reserves for the more delicate tasks of the proof. We will later embed T_0 as C_0-rainbow in X_0. We will also use X_0 and C_0 to find the connecting bare paths of length 3 to get T_j from T_{j-1} and we will use them to finish off some the matchings we are adding, using Lemma 24.

Collection of Properties:

Lemma 13 guarantees that with high probability, we have

R1 Each vertex in $V(K_{n,n})$ has at least $\frac{p_0^2 n}{2} \geqslant 10k\mu n$ colour-C_0 neighbours in X_0.

By Lemma 18, with high probability, we have

R2 For each pair of vertices $u \in V_1(K_{n,n})$, $v \in V_2(K_{n,n})$, 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 4, with high probability, we get $|X_0|, |C_0| \leqslant 2\frac{\varepsilon|V(K_{n,n})|}{400k} = \frac{\varepsilon n}{100k}$, and hence any vertex is contained in at most $\frac{\varepsilon n}{100}$ edges with colour in C_0 (as there are at most k edges of each colour incident to any vertex). Thus, the following holds with high probability:

R3 If G is the subgraph of $K_{n,n}$ of the edges with colour in $C(K_{n,n}) \setminus C_0$, with any edges inside X_0 removed, then $\delta(G) \geqslant (1 - \frac{\varepsilon}{100} - \frac{\varepsilon}{100k})n \geqslant (1 - \frac{\varepsilon}{50})n$.

Step 3: Embed T_0.

Now we embed T_0 into X_0 in a rainbow fashion, using colours from C_0. This is achieved by Lemma 9 in essence, but we need to make sure that we embed the different components of T_0 in such a way that we do not end up with two vertices that need to be connected in the same vertex class of $K_{n,n}$ in the bare-path-adding-step. So, pick a C_0-rainbow copy, S_0 say, of T_0 in X_0 in such a way that T_0 can be extended to T within $K_{n,n}$. Practically, that this works can be seen e.g. as follows: Pick a root vertex in each component of T_0 and fix the embedding $\iota : T_0 \to T$. Add an edge between two such root vertices in T_0 if the unique path between these vertices' images in T under ι is of of odd length to get a graph (not necessarily a tree) T_0' with $|T_0'| = |T_0|$. **R1** guarantees that we can apply Lemma 9 to find a C_0-rainbow copy of T_0' in X_0 as a C_0-rainbow - then we just delete the images of the added edges to get a C_0-rainbow copy S_0 of T_0 in X_0 with all the components in the right place to be able to extend S_0.

Step 4: Find large rainbow stars.

In a next step, for the appropriate integers $m \leqslant \frac{n}{D}$ and $d_1, \dots, d_m \geqslant D$, let $v_1, \dots, v_m \in V(S_0)$ 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 $d = \sum_{i=1}^m d_i = |T_1| - |T_0| \leqslant |T| = (1-\varepsilon)\frac{n}{k}$. For each $i \in [m]$, let $n_i = \left\lceil \left(1 - \frac{\varepsilon}{8}\right) \frac{nd_i}{kd} \right\rceil$. Note that

$$\sum_{i=1}^m n_i \leqslant \left(1 - \frac{\varepsilon}{8}\right)\frac{n}{k} + m \leqslant \left(1 - \frac{\varepsilon}{10}\right)\frac{n}{k}.$$

Using **R3** and Lemma 10, find disjoint subsets $Y_i \subset V(K_{n,n}) \backslash X_0$, $i \in [m]$, so that $|Y_i| = n_i$ and $\{v_i y : i \in [m], y \in Y_i\}$ is $(C(K_{n,n}) \backslash C_0)$-rainbow.

Note that all of this was a purely deterministic process and also that $n_i \geqslant d_i$, so we have found stars that are larger than required. We will later randomize this result by taking the intersection of these large stars with a random subset of $V(K_{n,n})$ such that stars of the correct sizes remain with high probability. After that, we also want a random set of colours C_1 which is used exclusively for the stars, but this will have to depend on the vertices put aside for the stars to make sure we have enough colours in C_1 that actually appear as colours of edges connecting S_0 with the stars. We achieve this by establishing the following correspondence: 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 one pair. We will later (Step 6 and 7) choose sets X_1 and C_1 in such a way that, for a vertex x paired with a colour c, we have $x \in X_1$ if and only if $c \in C_1$. In order to have a nice distribution of the set C_1, we will also add some more random colours, see step 7.

As a side remark: This step is a severe obstruction when trying to apply the developed methods to a tree of size $\frac{n}{k}$ in a balanced k-bounded colouring as would be sufficient to prove Ringel's conjecture (which has been achieved in [9] by finding a completely new, deterministic method to embed trees that consist largely of large stars). In that case, one would need to use all of the colours and, if e.g. T is a union of large stars, it wouldn't be possible to find collectively rainbow stars larger than these. Also, this step is the reason why in some of the above lemmas, it is necessary to allow the vertices and colours to depend on each other.

Step 5: Choose probabilities p_i

. We want to set aside sets of vertices and colours for each of the extension steps of our forest. We will choose these sets as p_i-random subsets of $V(K_{n,n})$ and $C(K_{n,n})$ respectively. 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

$$p_i = \left(1 + \frac{\varepsilon}{4}\right) \frac{km_i}{n} + \frac{\varepsilon}{4\ell} \geqslant n^{-1/10^4}.$$

The inequality follows as $\ell \leqslant 10^4 D\mu^{-2} = O(\log^{10} n)$. For p_ℓ, we have the residual term

$$\begin{aligned}
p_\ell &= 1 - p_0 - \sum_{i \in [\ell-1]} p_i \\
&= 1 - \frac{\varepsilon}{400} - \left(1 + \frac{\varepsilon}{4}\right) k \frac{|T| - m_\ell - |T_0|}{n} - \frac{\varepsilon(\ell - 1)}{4\ell} \\
&\geqslant 1 - \left(1 + \frac{\varepsilon}{4}\right) k \frac{(1-\varepsilon)n/k - m_\ell}{n} - \frac{\varepsilon}{4} \\
&\geqslant n^{-1/10^4}.
\end{aligned}$$

Step 6: Choose X_1.

Pick $X_1 \subset V(K_{n,n}) \backslash X_0$ by including each vertex independently at random with probability $p_1/(1 - p_0)$. Recall that $m_1 = d$. We have, for each $i \in [m]$, that

$$\begin{aligned}
p_1|Y_i| = p_1 n_i &\geqslant \left(1 + \frac{\varepsilon}{4}\right) k \frac{m_1}{n} \left(1 - \frac{\varepsilon}{8}\right) \frac{nd_i}{kd} \\
&= \left(1 + \frac{\varepsilon}{4}\right)\left(1 - \frac{\varepsilon}{8}\right) d_i \geqslant \left(1 + \frac{\varepsilon}{16}\right) d_i \geqslant \log^{10} n.
\end{aligned}$$

Thus, by Lemma 4, for each $i \in [m]$, $P(|X_1 \cap Y_i| \geqslant d_i) = \exp(-\Omega(\varepsilon^2 \log^{10} n)) = o(n^{-1})$. This means that with high probability, the following property holds:

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

This was the previously described randomization process for the stars added to obtain T_1 from T_0! Note, for later, that each vertex $x \in V(K_{n,n})$ appears in X_1 independently at random with probability $(1 - p_0) \cdot \frac{p_1}{1-p_0} = p_1$.

Step 7: 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 $C^{\text{unpaired}} = C(K_{n,n}) \backslash C^{\text{paired}}$ 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 $c \in C^{\text{unpaired}} \backslash C_0$, c is included in C_1 independently at random with probability $\frac{p_1}{1-p_0}$. Thus, C_1 contains each colour paired with a vertex in X_1 and each unpaired colour outside C_0 is included independently at random. Thus, each colour appears in C_1 with the same probability: p_1. Whether a colour is included is independent of any other colour's in- or exclusion.

Step 8a: Choose $X_2, \ldots, X_\ell$.

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

Step 8b: Choose $C_2, \ldots, C_\ell$.

Randomly partition $C(K_{n,n}) \backslash (C_0 \cup C_1)$ as $C_2 \cup \cdots \cup C_\ell$ so that, for each $c \in C \backslash (C_0 \cup C_1)$, the class of c is chosen independently at random with $P(c \in C_i) = \frac{p_i}{1-p_1-p_0}$ for each $2 \leqslant i \leqslant \ell$. Note that, for each $0 \leqslant i \leqslant \ell$, each colour $c \in C(K_{n,n})$ appears in C_i with probability p_i, and the location of each colour in $C(K_{n,n})$ is independent of the location of all the other colours.

Step 9: Establish rainbow matching properties.

By the definitions of the probabilities, it is straightforward to see that we have $m_i \leqslant p_i \frac{n}{k}$ for all $i \in [\ell]$. For $A \subset V(K_{n,n}) \backslash X_i$, set $q_{i,j}^A = \frac{|A \cap V_j(K_{n,n})|}{|A|}$ ($i \in [\ell]$, $j \in [2]$). Then pick $X_{i,j} \subset X_i$, $C_{i,j} \subset C_i$ $q_{i,3-j}^A$-random ($i \in [\ell]$, $j \in [2]$) and note that if $|A| = m_i$:
$|A \cap V_j(K_{n,n})| = q_{i,j}^A |A| = q_{i,j}^A m_i \leqslant q_{i,j}^A \frac{p_i n}{k}$. By Lemma 20, we find a $C_{i,1}$-rainbow matching from $A \cap V_1(K_{n,n})$ to $X_{i,2}$ of size at least $|A \cap V_1(K_{n,n})| - \mu p_i q_{i,1}^A n = m_i q_{i,1}^A - \mu p_i q_{i,1}^A n$ and a $C_{i,2}$-rainbow matching from $A \cap V_2(K_{n,n})$ to $X_{i,1}$ of size at least $|A \cap V_2(K_{n,n})| - \mu p_i q_{i,2}^A n = m_i q_{i,2}^A - \mu p_i q_{i,2}^A n$, so that by combining these matchings, using $q_{i,1}^A + q_{i,2}^A = 1$, we get with probability $1 - o(n^{-1})$:

R5 For each $i \in [\ell]$ and subset $A \subset V(K_{n,n}) \backslash X_i$ with $|A| = m_i \leqslant p_i \frac{n}{k}$ there is a C_i-rainbow matching with at least $m_i - \mu p_i n$ edges from A to X_i.

Step 10: Extend S_0 to a copy of T_1 by adding stars.

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, $S_1 \subset K_{n,n}[X_0 \cup X_1]$ is a $(C_0 \cup C_1)$-rainbow copy of T_1 with at most $|T_0| \leqslant \mu n$ colours in C_0 and at most μn vertices in X_0.

Step 11: Extend to a copy of $T_2, \dots, T_{j-1}$ by adding non-neighbouring leaves.

Iteratively, for each $2 \leqslant i \leqslant j-1$, we now extend S_{i-1} to $S_i \subset K_{n,n}[X_0 \cup \dots \cup X_i]$. To carry out this step, 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 Lemma 24 twice to finish the matching. We apply Lemma 24 with G as the subgraph of $K_{n,n}$ of colour-$((C_0 \backslash C(S_{i-1})) \cup C_i)$ edges, $C = C_i$, $C' = C_0 \backslash C(S_{i-1})$, $Y = X_i$, $Z = X_0 \backslash (V(S_{i-1}))$ and with $X = A_i \cap V_1(K_{n,n})$ in the first instance, then in the second with the same arguments except X now set to $X = A_i \cap V_2(K_{n,n})$. This finds a rainbow matching covering the whole set A_i and thus completing the extension step. S_i is a $(C_0 \cup \dots \cup C_i)$-rainbow copy of T_i with $|C(S_i) \cap C_0| \leqslant \mu n + \sum_{i'=2}^{i} \mu p_{i'} n$ and $|V(S_i) \cap X_0| \leqslant \mu n + \sum_{i'=2}^{i} \mu p_{i'} n$. Note that $|C(S_i) \cap C_0| \leqslant 2\mu n$ and $|V(S_i) \cap X_0| \leqslant 2\mu n$, because we have used at most $\mu p_i n$ colours and vertices to finish the matching, as is guaranteed by **R5**.

Step 12: Extend to a copy of T_j by adding bare paths of length 3.

Using new vertices in X_0 and new colours in C_0, extend this copy of T_{j-1} to $S_j \subset K_{n,n}[X_0 \cup \dots \cup X_j]$, a $(C_0 \cup \dots \cup C_j)$-rainbow copy of T_j with $|C(S_j) \cap C_0| \leqslant 4\mu n + \sum_{i'=1}^{j} \mu p_{i'} n$ and $|V(S_j) \cap X_0| \leqslant 3\mu n + \sum_{i'=1}^{j} \mu p_{i'} n$. 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 Lemma 19 applied with G as the subgraph of $K_{n,n}$ of colour-$(C_0 \backslash C(S_{j-1}))$ edges and $Y = X_0 \backslash V(S_{j-1})$.

Step 13: Extend to a copy of $T_{j+1}, \dots, T_\ell$.

Finally, for each $i \in j+1, \dots, \ell$, use **R1**, **R5** and two applications of Lemma 24 as before to extend S_{i-1} to $S_i \subset K_{n,n}[X_0 \cup \dots \cup X_i]$, 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.

Step 14: Concluding the proof.

For the random partitions of $V(K_{n,n}) = X_0 \cup \dots \cup X_\ell$ and $C(K_{n,n}) = C_0 \cup \dots \cup C_\ell$ we chose, we have seen after Step 2 that **R1**-**R3** hold with high probability. Furthermore, if **R1**-**R3** hold, there exist vertex sets $Y_1, \dots, Y_m$ such that **R4** holds with high probability, as was remarked after Step 6. Finally, if **R1**-**R4** hold, we were able to show after Step 9 that **R5** holds for each $2 \leqslant i \leqslant \ell$ with probability $1 - o(\ell \cdot n^{-1}) = 1 - o(1)$.

All in all, this shows that **R1**-**R5** all collectively hold with high probability for partitions of the vertex and colours sets of our coloured $K_{n,n}$ chosen from the distributions given in the proof. Since that is the case, there must be some such partitions that **R1**-**R5** do in fact hold, which proves that a rainbow copy of T can be found.

It is important to note that this is a non-standard application of the probabilistic method, because some of the sets in the partition are not independent of each other and because the properties **R1-R5** are interlinked: **R4** may only hold if **R1-R3** hold and **R5** only holds if all other properties hold. For a detailed analysis of this particular way of reasoning, we refer to Section 10.2 in [8]. □

6 Conclusion

We have shown that for any given tree T with $\frac{n}{k}\cdot(1-o(1))$ edges, we can find a rainbow copy of T in any locally k-bounded colouring of $K_{n,n}$. Going through the steps one by one, the reader will notice that the mathematical machinery to prove this is kept relatively simple and that most of the work was done by cleverly combining concentration inequalities. The proof rests mainly on two ideas: Finding a decomposition of T into large stars, a few bare paths of length 3 and a lot of matchings and finding a partition of the vertex and colour set of $K_{n,n}$ that allows us to find disjoint rainbow copies of all the parts of the decomposition of T with vertices in different vertex classes and colours from different colour classes to then connect them to get a rainbow copy of the entire tree T. We did not find the vertex and colour decomposition needed explicitly, but used a randomized decomposition of the colours and vertices of $K_{n,n}$ and showed that for any given such setup, we could find a rainbow copy of T with high probability. This probabilistic argument was enriched by an instance of the absorption technique. By introducing the locally 1-bounded D-colouring of $K_{n,n}$, we could show that our main theorem (Theorem 1) implies that $K_{n,n}$ can asymptotically be decomposed into n copies of any tree with n edges. Overall, the methods used differ only in details from those used in [8] to show that one can find a rainbow copy of any tree with $n\cdot(1-o(1))$ edges in the locally 2-bounded ND-colouring of K_{2n+1}, from which follows that K_{2n+1} can asymptotically be decomposed into copies of any tree with n edges. In the latter case, the authors were able to follow up on their arguments and show a much stronger result in [9]: For large n, K_{2n+1} can be decomposed into n copies of any tree with n edges. As elaborated in the introduction, a similar result for complete bipartite graphs cannot be proved by using the methods of [9]. The problems can be summarized as follows: Using the probabilistic arguments expanded upon above hinge on the fact that for most trees T, the decomposition algorithm for the tree in question (cf. Lemma 8) has a lot of steps that consist of deleting non-neighbouring leaves. This is favourable for arguments of the type we use here because, roughly speaking, there is a lot of freedom when re-adding sets of non-neighbouring leaves via edges of unused colours to a rainbow subforest of $K_{n,n}$. For some trees, however, the decomposition does not have such a structure. This is the case if the tree in question contains of a few very large stars. In that case, one needs an additional argument to find a rainbow copy of such a tree. In the asymptotic case, a simple switching argument is sufficient to deal with the problematic class of trees, but this fails in the exact case. For the ND-colouring of a complete graph, Montgomery et al. found a way to work around this issue by giving a completely unrelated deterministic algorithm to deal with the problematic class of trees. Their algorithm depends heavily on the fact that each vertex in an ND-coloured K_n has exactly two neighbours connected via an edge of a given colour. In the here-introduced D-colouring, each vertex has only one neighbour connected via an edge of a given colour, which presents an additional obstacle that we were not able to overcome. We do, however, believe that other methods may succeed in proving Conjecture 3.

Acknowledgements

I would like to thank Felix Joos and António Girão for suggesting and supervising this topic, and many valuable discussions. I am also grateful to the anonymous referees for their suggestions.

References

[1] P. Erdős and A. Rényi. On random graphs. I. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.

[2] P. Erdős, A. Gyárfás, and L. Pyber. Vertex coverings by monochromatic cycles and trees. *Journal of Combinatorial Theory. Series B*, 51(1):90–95, 1991.

[3] J. A. Gallian. A dynamic survey of graph labeling. *The Electronic Journal of Combinatorics*, #DS6, 1998.

[4] S. Janson, T. Łuczak, and Andrzej Ruciński. *Random graphs.* John Wiley and Sons, 2000.

[5] F. Joos, J. Kim, D. Kühn, and D. Osthus. Optimal packings of bounded degree trees. *Journal of the European Mathematical Society (JEMS)*, 21(12):3573–3647, 2019.

[6] T. P. Kirkman. On a problem in combinations. *Cambridge and Dublin Mathematical Journal*, 2:191–204, 1847.

[7] M. Krivelevich. Triangle factors in random graphs. *Combinatorics, Probability and Computing*, 6(3):337–347, 1997.

[8] R. Montgomery, A. Pokrovskiy, and B. Sudakov. Embedding rainbow trees with applications to graph labelling and decomposition. *Journal of the European Mathematical Society (JEMS)*, 22(10):3101–3132, 2020.

[9] R. Montgomery, A. Pokrovskiy, and B. Sudakov. A proof of Ringel's conjecture. *Geometric and Functional Analysis*, 31(3):663–720, 2021.

[10] G. Ringel. Extremal problems in the theory of graphs. In *Theory of Graphs and its Applications (Proc. Sympos. Smolenice, 1963)*, pages 85–90. Publ. House Czech. Acad. Sci., Prague, 1964.

[11] V. Rödl, A. Ruciński, and E. Szemerédi. A Dirac-type theorem for 3-uniform hypergraphs. *Combinatorics, Probability and Computing*, 15(1-2):229–251, 2006.

[12] A. Rosa. On certain valuations of the vertices of a graph. In *Theory of Graphs (Internat. Sympos., Rome, 1966)*, pages 349–355. Gordon and Breach, New York; Dunod, Paris, 1967.

[13] T. Szele. Kombinatorische Untersuchungen über gerichtete vollständige Graphen. *Publicationes Mathematicae Debrecen*, 13:145–168, 1966.

[14] R. M. Wilson. Decompositions of complete graphs into subgraphs isomorphic to a

given graph. In *Proceedings of the Fifth British Combinatorial Conference (Univ. Aberdeen, Aberdeen, 1975)*, pages 647–659, 1976.