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

Julian Matthes *

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 contained in 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 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 $K_{n,n}$ contains a rainbow copy of any tree T with $\frac{n}{k}(1-o(1))$ 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.

1 Introduction

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 $K_{n,n}$ with vertex classes $\{1_1, ..., n_1\}$ and $\{1_2, ..., n_2\}$, where the edge between i_1 and j_2 represents the cell of the Latin Square in the i-th row and j-th column. Conversely, each proper colouring of $K_{n,n}$ using exactly n colours defines a Latin Square.

In 1847, Kirkman showed that K_n can be decomposed into copies of a triangle if, and only if, $n \equiv 1, 3$ modulo 6 ([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, Gerhard Ringel posed the conjecture that K_{2n+1} can be decomposed into $2n+1$ copies of any tree with n edges ([10]). Early progress on this was done by Anton Kotzig

*julian.matthes@awi.uni-heidelberg.de, Heidelberg University, Department of Economics, Bergheimer Strasse 58, 69115 Heidelberg, Germany.

(according to [12]), who considered a specific colouring of the complete graph known as the "Nearest-Distance-colouring" (ND-colouring). This colouring derives its name from its construction, which is as follows: Order n 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 distance between the connected vertices are equal. Kotzig conjectured that the ND-coloured complete graph on $2n+1$ vertices contains a rainbow copy of any tree with n edges. This is stronger than 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 $2n+1$ distinct 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, diameter ≤ 5 trees, symmetrical trees and more, see [3]) and, in a breakthrough in 2018, for large bounded degree trees ([5]).

Also in 2018, Montgomery, Pokrovskiy and Sudakov proved an asymptotic version of the two conjectures ([8]). Building on their methods, they could finally prove Ringel's conjecture by way of proving Kotzig's conjecture in 2020 ([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 Renyi 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, Gyarfas and Pyber in [2] and Krivelevich in [7] and adapted e.g. by Rödl, Ruciński and Szemerédi ([11]).

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

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 $(1-\varepsilon)\frac{n}{k}$ vertices. Then any locally k-bounded colouring of $K_{n,n}$ contains a rainbow copy of T.*

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

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

This follows from an argument similar to that of Kotzig above. We introduce a proper colouring of $K_{n,n}$ with a special symmetry property, which we call "Difference-colouring" (or "D-colouring"): Enumerate the vertices of one class of $K_{n,n}$ by $\{1_1, 2_1, ..., n_1\}$ and those of the other class by $\{1_2, 2_2, ..., n_2\}$. Now we colour the edge between u_1 and v_2 in the colour $u - v$ modulo n. This yields a proper colouring of $K_{n,n}$ using n colours.

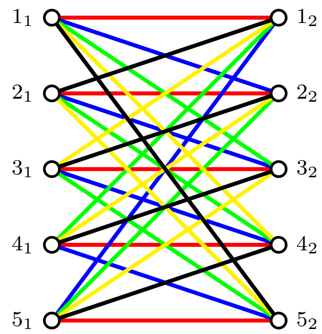

Figure 1: The D-Colouring for $n = 5$.

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 $K_{m,m}$. For $i \in \{1, ..., n-1\}$, set S_{i+1} the tree with vertex set $\{(v+i)_1 \,|\, v_1 \in V_1(S_1)\} \cup \{(v+i)_2 \,|\, v \in V_2(S_1)\}$ and edge set $\{\{(v+i)_1, (w+i)_2\} \,|\, \{v, w\} \in E(S_1)\}$. We have: $c(v_1 w_2) = c((v+i)_1 (w+i)_2)$ for each $i \in \{1, ..., n-1\}$, so each of the trees S_i, $i \in \{1, ..., n\}$ is rainbow and by construction, any two edges of the same colour are disjoint, so we have n disjoint copies of T. $\square$

2 Preliminaries

2.1 Definitions and Notations

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

2.1.1 Graph Theory

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 contained in 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 $c(e) \in C(G)$ the colour of the edge $e \in E(G)$.

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 $N_C(V) = N_C(v)$.

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

Lastly, we need the following definitions:

Definition 1. *A bare path in a tree T is a path whose interior vertices all have degree 2 in T.*

Definition 2. *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.*

2.1.2 Asymptotics

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

Lastly, 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 $x \leq f(y)$.

2.1.3 Probability Theory

We say that an event occurs *almost surely* or *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.

Definition 3. *Let B be any set and $p \in [0, 1]$. A subset $A \subset B$ is called p-random if A is formed by including each vertex of B independently with probability p.*

Definition 4. *We say that a p-random set A is independent from a q-random set B 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.*

2.2 Concentration Inequalities and Union Bound

Lemma 1 (Chernoff's concentration inequality, [4])**.** *Let $X \sim \mathrm{Bin}(n, p)$ and $0 < \varepsilon < \frac{3}{2}$, then:*

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

Corollary 1. *For $\frac{1}{n} \ll \varepsilon$, if X is a p-random subset of $[n]$, then with high probability, $|X| = (1 \pm \varepsilon)pn$.*

Lemma 2 (Azuma's concentration inequality, [4])**.** *Let X be k-Lipschitz on $\prod_{i=1}^{n} \Omega_i$, then:*

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

Corollary 2. *For $\frac{1}{n} \ll \varepsilon$ and fixed k, if Y is a k-Lipschitz random variable influenced by at most n coordinates, then, with high probability, $Y = E(Y) \pm \varepsilon n$.*

Lemma 3 (Union Bound, [4])**.** *For fixed k, if $A_1, \ldots, A_k$ are events which hold with high probability, then they simultaneously occur with high probability.*

2.3 Tree Decomposition

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 $K_{n,n}$.

Lemma 4 ([8])**.** *Given integers D and $n, \mu > 0$ and a tree T with at most n vertices, there are integers $l \leq 10^4 D\mu^{-2}$ and $j \in \{2, ..., l\}$ and a sequence of subgraphs $T_0 \subset T_1 \subset ... \subset T_l = T$ such that:*

1. *For each $i \in [l] \backslash \{1, j\}$, 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 centered in T_0 with at least D leaves each and*
4. *$|T_0| \leq \mu n$*

2.4 Embedding Small Structures as Rainbows

Using the notation from Lemma 4, the following lemma gives us the means to find a rainbow copy of the "center piece" T_0 of a tree.

Lemma 5 ([8])**.** *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.*

2.5 Finding Rainbow Stars in Coloured Graphs

The following lemma is the technical result needed to add large stars to the center piece. It should be noted that the proof is purely deterministic, it uses a switching argument. The bound on the size of the union of all stars can not 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 1).

Lemma 6 ([8])**.** *Let $0 < \varepsilon < \frac{1}{100}$ and $l \leq \varepsilon^2 \frac{n}{2}$. 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 $v_1, ..., v_l$. Let $d_1, ..., d_l \geq 1$ be integers satisfying $\sum_{i \in [l]} d_i \leq (1-3\varepsilon)\frac{n}{k}$, and suppose G has a locally k-bounded edge-colouring. Then, G contains disjoint stars $S_1, ..., S_l$ so that, for each $i \in [l]$, S_i is a star rooted at v_i with d_i leaves, and $\bigcup_{i \in [l]} S_i$ is rainbow.*

2.6 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 lemmata collect some simple properties of such subgraphs.

Lemma 7. *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, we have $P(u \in N_C(v) \cap V) = pq$ by the independence of V and C. Because the set $N_C(v) \cap V$ is affected by $3n - 1$ coordinates (all the n colours and all the $2n$ vertices except v) and is k-Lipschitz as the colouring is locally k-bounded, we have by Azuma's inequality, that

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

This is a result for a single vertex v. Taking the union bound (Lemma 3) 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 8. *Let $\varepsilon > 0$, $k \in \mathbb{N}$ and let $p \geq n^{-1/100}$. Let $K_{n,n}$ have a locally k-bounded colouring and suppose G is a subgraph of $K_{n,n}$ chosen by including the edges of each colour 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 $|A|, |B| \geq n^{3/4}$:*

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

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

Proof. This goes along the lines of the proof of Lemma 5.1 in [8].
Let $l = \lceil \frac{\sqrt{2kn}}{\varepsilon^2 p^2} \rceil$, so that $l \leq n^{0.52+o(1)} \leq n^{0.6+o(1)}$ by the choice of the parameters. The key to the proof is to show that there exist partitions $A = A_1 \cup A_2 \cup ... \cup A_l$, $B = B_1 \cup B_2 \cup ... \cup B_l$ 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 pair (A_i, 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. $|A| \geq |B|$, then:

Claim 1. *There are partitions $A = A_1 \cup A_2 \cup ... \cup A_l$, $B = B_1 \cup B_2 \cup ... \cup B_l$ 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 $i, j \in [l]$.*

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 $\leq 2(k-1)/l$ (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 $\leq kn/2l^2$ (since that 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 $\frac{2(k-1)}{l} \leq \frac{2kl}{l^2} \leq \frac{kn}{l^2}$ (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

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

by the choice of l. 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 some partitions with the desired property. □

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 **P**, if

P1 $|X| \geq |Y| \geq n^{1/10}$

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 **P**, we get:

Claim 2. *Let $X \subset V_1(K_{n,n})$ and $Y \subset V_2(K_{n,n})$. If both **P1** and **P2** hold, then with probability $1 - o(n^{-1})$, we have*

$$|e_G(A, B) - p|X||Y|| \leq \frac{\varepsilon p|X||Y|}{2}$$

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 $e(R \cap G) \sim \text{Bin}(\left(1 - \frac{\varepsilon p}{8}\right))|X||Y|, p)$. By Lemma 1 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

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

In combination with $e(K_{n,n}[X, Y] - R) \leq \frac{\varepsilon p|A||B|}{8}$, this implies that **P2** holds for X and Y with the above probability. To simplify the latter, note that we have $p^3|Y| \geq n^{7/100} = \omega(\log n)$, as $p \geq n^{-\frac{1}{100}}$. So, since $|X| \geq |Y| \geq n^{1/10}$, condition **P2** holds for X and Y with probability at least $1 - \exp(-|X| \cdot \omega(\log n))$. Summing over possible sizes of X and Y yields that **P2** holds with probability at least

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

□

Next, fix partitions of $A = A_1 \cup A_2 \cup ... A_l$ and $B = B_1 \cup B_2 \cup ... \cup B_l$ like in the above Claim 1.

Next, we will need to look at pairs not satisfying $\mathbf{P}$. 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 $\mathbf{P}$.

Claim 3. *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 $\mathbf{P1}$, we delete at most

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

edges.

Let I denote the set of indices of pairs that do not satisfy $\mathbf{P2}$, but do satisfy $\mathbf{P1}$. Note that we have

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

by Claim 1.

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

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

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

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

The first inequality is the triangle inequality, then we use Claim 2, 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

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

We can finally prove the claim of the lemma:

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

holds by the triangle inequality. By applying (1) to the last term 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 $|e_G(A,B) - e_{G\cap H}(A,B)| \leq |e_H(A,B) - |A||B||$, we get:

$$\begin{aligned}|e_G(A,B) - p|A||B|| &\leq 2|e_H(A,B) - |A||B|| + \frac{\varepsilon p|A||B|}{2} \\ &\leq \varepsilon p|A||B|\end{aligned}$$

as desired. □

2.7 Finding Rainbow Bare Paths

Next, we will look for rainbow paths between given end-points with internal vertices inside a fixed randomized set. This will be needed for the step in the proof of Theorem 1 where we connect the components of the forest T_{j-1} to get the tree T_j, using the definition of j from Lemma 4. To find some such paths, we employ the concentration result from the previous subsection.

Lemma 9. *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. Almost surely, 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 7, 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 $10k\mu^{1/3}n \geq n^{3/4}$, which means that we can apply Lemma 8 to get 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 $|A|, |B| \geq 10k\mu^{1/3}n$, there are at least $p|A||B|/2 \geq 4k\mu n^2$ 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. 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

$$|N_{C_1 \backslash C'}(u, X \backslash U)| \geq 100k^2\mu 1/3m - 2\mu n - 3k\mu n \geq 10k\mu^{1/3}n$$

Let $A \subset N_{C_1 \backslash (C')}(u, X \backslash U)$ such that $|A| = 10k\mu^{1/3}n$ and let C'' be the set of colous between u and A. Using **Q1** again, we have

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

Now let $B \subset N_{C_1 \backslash (C' \cup C'')}(v, X \backslash (U \cup A))$ satisfy $|B| = 10k\mu^{1/3}n$. By **Q2**, there are at least $4k\mu n^2$ colour-C_2 edges between A and B, at most $kn|C'| \leq 3\mu n^2$ 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 $X \backslash U$! □

Lemma 10 ([8])**.** *Suppose we have a graph G with a locally k-bounded colouring containing the disjoint vertex sets $X^1 = \{x_1, ..., x_m\} \subset V(G)$, $X^2 = \{x'_1, ..., x'_m\} \subset V(G)$ 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.*

2.8 Finding Almost-Covering 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. The larger first part of the section deals with covering almost all of a not-too-large such set A while a lemma at the end 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 Chapters 6 and 7 in [8].

Lemma 11 (Finding almost covering rainbow matchings)**.** *Let $k \in \mathbb{N}$ and $\varepsilon > 0$, and suppose $K_{n,n}$ has a locally k-bounded colouring and $p \geq n^{-1/10^4}$. Let $X \subset V(K_{n,n})$ and $C \subset C(K_{n,n})$ be independent and p-random. Then, with probability $1 - o(n^{-1})$, for each set $A \subset V_j(K_{n,n})$ with $|A| \leq \frac{pn}{k}$, there is a C-rainbow matching in $K_{n,n}$ of size at least $|A| - \varepsilon pn$ between A and $X \cap V_{3-j}(K_{n,n})$.*

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 1 ([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 1, we need to 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 2. *Let $k \in \mathbb{N}$, $\varepsilon > 0$, and $p \geq n^{-1/10^4}$. Let $K_{n,n}$ have a locally k-bounded colouring. Let $X \subset V(K_{n,n})$ 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 $\{c \in C\}$. Then, for $j \in [2]$, with probability $1 - o(n^{-1})$, for each $A \subset V_j(K_{n,n})\backslash X$ and $B \subset X \cap V_{3-j}(K_{n,n})$ 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.*

Most of this section will be dedicated to showing Proposition 2. Before proceeding to this technical part, we will give is the proof of Lemma 11 assuming the Propositions 1 and 2, as Lemma 11 is the result that we will want to use later on.

Proof of Lemma 11. Let η be a fixed constant not dependent on n which satisfies $0 < \eta \ll \varepsilon$. With probability $1 - o(n^{-1})$, by Proposition 2 applied with $\frac{\eta}{2k}$ for ε, we get:

Q For each $A \subset V_j(K_{n,n})\backslash X$ and $B \subset X \cap V_{3-j}(K_{n,n}9$ with $|A|, |B| \geq \frac{\eta pn}{2k} \geq n^{3/4}$, there are at least $\frac{(1-\eta)|B|}{k}$ colours in C between A and B.

Also, With probability $1 - o(n^{-1})$, by Chernoff's inequality, we have

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

We claim that the property in the lemma holds. Let $A \subset V_j(K_{n,n})\backslash X$ with $|A| \leq pn/k$. Add vertices to A from $V_j(K_{n,n})\backslash X$, or delete vertices from A, to get a set A' with $|A'| = \lfloor (1-\frac{\eta}{2})\frac{pn}{k} \rfloor =: m$ and $|A\backslash A'| \leq \eta\frac{pn}{k} \leq \varepsilon\frac{pn}{2}$. Let X' be a subset of X of size km. Since $\eta m \geq \eta\frac{pn}{2k}$, we can apply **Q** to see that for any subsets $A'' \subset A'$ and $B \subset X'$ with $|A''|, |B| \geq \eta m$, there are at least $(1-\eta)\frac{|B|}{k}$ colours in C between A'' and B.
Thus, by Proposition 1, there is a C-rainbow matching with at least $(1-\frac{\varepsilon}{4})m \geq |A'| - \varepsilon\frac{pn}{2}$ edges between A' and X'. As $|A\backslash A'| \leq \varepsilon\frac{pn}{2}$, at least $|A| - \varepsilon pn$ of the edges in this C-rainbow matching must lie between A and X. □

We will now proceed to prove Proposition 2. This will be done as follows: First, we establish that for $j \in [2]$, a set $A \subset V_j(K_{n,n})$ that is neither too large nor too small and a p-random subset X of $V(K_{n,n})$, most colours do not appear significantly more often than expected between A and X (Proposition 3). We will then be able to extend this result to any A that is not too small (Proposition 4). Once this has been proven, we can combine it with the Edge-Concentration Lemma (8) to get Proposition 2. Note that Proposition 3 and Proposition 4 are written in such a way that they do not require A to be contained in only one of the components of $K_{n,n}$.

Proposition 3. *Let k be constant and $\varepsilon, p \geq n^{-1/100}$. Let $K_{n,n}$ have a locally k-bounded colouring and let X be a p-random subset of $V(K_{n,n})$. Then, with probability $1-o(n^{-1})$, for each $A \subset V(K_{n,n})$, $j \in [2]$, with $n^{1/20} \leq |A| \leq n^{1/4}$, there are at most εn colours of which there are more than $(1+\varepsilon)pk|A|$ edges between A and X.*

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

Claim 4. *There is a partition $C' = C_1 \cup ... \cup C_l$ so that, for each $i \in [l]$ and $a, b \in C_i$, A_a and A_b are disjoint.*

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

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

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

and $B = \cup_{i \in [l]} B_i$. We will need to show that $|B| \leq \varepsilon n$ to prove the proposition.

Claim 5. *For each $i \in [l]$, if $|C_i| \geq \frac{\varepsilon n}{2l}$, then $P(|B_i| \geq \varepsilon p|C_i|/2) \leq \exp(\frac{-\varepsilon^2 pn^{1/2}}{200k})$.*

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 $|A_c \cap X| \sim \text{Bin}(|A_c|, p)$. By Lemma 1, and as $|A_c| \leq k|A|$, for each $c \in C_i$,

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

for sufficiently large n. Furthermore, for each i, 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 $\text{Bin}(|C_i|, \varepsilon p/4)$. By Lemma 1 applied with $1/2$ for ε, if $|C_i| \geq \frac{\varepsilon n}{2l}$, we obtain that the claim is true:

$$P\left(|B_i| \geq \frac{\varepsilon p|C_i|}{2}\right) \leq 2\exp\left(-\frac{\varepsilon p|C_i|/4}{12}\right) \leq \exp\left(-\frac{\varepsilon^2 pn}{100l}\right) \leq \exp\left(-\frac{\varepsilon^2 pn^{1/2}}{200k}\right)$$

□

Now, note that we have that $\frac{\varepsilon^2 pn^{1/2}}{200k} > n^{2/5}$, as $\varepsilon, p \geq n^{-\frac{1}{100}}$. Therefore the probability that for some subset A of $V(K_{n,n})$ with $n^{1/20} \leq |A| \leq n^{1/4}$, there is an index $i \in [l]$ such that $|C_i| \geq \frac{\varepsilon n}{2l}$ and $|B_i| \geq \frac{\varepsilon p|C_i|}{2}$ is at most

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

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

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

where we have used that $\sum_{i\in[l]} |C_i| = |C'| \leq n/pk$. $\square$

We can now show that the property in Proposition 3 is likely also to hold for any subset A which is not too small.

Proposition 4. *Let $k \in \mathbb{N}$ and $\varepsilon, p \geq n^{-1/10^3}$. Let $K_{n,n}$ have a locally k-bounded colouring and let X be a random subset of $V(K_{n,n})$ with each vertex included independently at random with probability p. Then, with probability $1-o(n^{-1})$, for each $A \subset V(K_{n,n})\backslash X$, $j \in [2]$, 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 Proposition 3 applied with $\frac{\varepsilon^2 p}{4}$ for ε, with probability $1-o(n^{-1})$, for each subset $B \subset V(K_{n,n})$ with $n^{1/20} \leq |B| \leq n^{1/4}$, for all but at most $\frac{\varepsilon^2 p}{4}$ colours there are at most $(1+\varepsilon^2)pk|B|$ edges of that colour between B and X. Let $A \subset V(K_{n,n]})$ satisfy $|A| \geq n^{1/4}$ and choose $A = A_1 \cup ... \cup A_l$ with $l = \lfloor |A|/n^{1/20} \rfloor$ and $n^{1/20} \leq |A_i| \leq 2n^{1/20}$ for each $i \in [l]$. For each $i \in [l]$, 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 $|C_i| \leq \frac{\varepsilon^2 pn}{4}$. Let C' be the set of colours which there are more than $(1+\varepsilon)pk|A|$ edges of between A and X. We need to show that $|C'| \leq \varepsilon n$. Note that, if $c \in C'$, then

$$\begin{aligned}(1+\varepsilon)pk|A| \leq \sum_{a\in A} |N_c(a) \cap X| &\leq \sum_{i:c\in C_i} k|A_i| + \sum_{i:c\notin C_i} (1+\varepsilon^2)pk|A_i| \\ &\leq |\{i : c \in C_i\}|2kn^{1/20} + (1+\varepsilon^2)pk|A|\end{aligned}$$

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 $K_{n,n}$). The third inequality comes from $|A_i| \leq 2n^{1/20}$ and $\sum_{i\in[l]} |A_i| = |A|$. This implies that

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

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

$$|C'|\varepsilon pl/4 \leq \sum c \in C'|\{i : c \in C_i\}| = \sum_{i\in[l]} |C' \cap C_i| \leq \sum_{i\in[l]} |C_i|$$

Therefore,

$$|C'| \leq \frac{\sum_{i\in[l]} |C_i|}{\varepsilon pl/4} \leq \frac{l\varepsilon^2 pn/4}{\varepsilon pl/4} = \varepsilon n$$

as required. $\square$

Now we can finally prove Proposition 2.

Proof of Proposition 2. The desired property in the lemma strengthens as ε decreases, so we may assume that $\varepsilon \leq 1/2$. Let $G = K_{n,n}[C]$. By Proposition 4 applied with $\frac{\varepsilon^2 p^2}{4k}$ for ε, with probability $1 - o(n^{-1})$, for each $A \subset V_j(G)$ 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 $K_{n,n}$. With probability $1-o(n^{-1})$, by Lemma 8 applied with ε^2 for ε, for every two subsets $A \subset V_j(K_{n,n}) \cap V(G)$, $B \subset V_{3-j}(K_{n,n}) \cap V(G)$ with $|A|, |B| \geq n^{3/4}$, we have $e_G(A, B) \geq (1-\varepsilon^2)p|A||B|$. Now, for each $A \subset V_j(K_{n,n})\backslash X$ and $B \subset X \cap V_{3-j}(K_{n,n})$ with $|A| \geq n^{3/4}$ and $|B| \geq \varepsilon pn \geq n^{3/4}$, there are at leas $(1-\varepsilon^2)p|A||B|$ 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 $K_{n,n}$ at most $k|A|$ times, this removes at most $k|A|\frac{\varepsilon^2 p^2 n}{4k}$ 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

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

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

2.9 Finishing Almost-Covering Rainbow Matchings

Lemma 12 ([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.*

3 Proof of 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 4. Find a sequence of forests $T_0 \subset T_1 \subset ... \subset T_l$ such that

1. T_0, which we call the base forest T_0 has at most $\lfloor \mu n \rfloor$ vertices.

2. T_1 can be obtained from T_0 by adding a collection of large stars with centers in T_0, large meaning that each of them has at least $D = \lceil \log^{10} n \rceil$ leaves

3. For $i \in \{2, ..., j-1\}$, T_i is obtained from T_{i-1} by adding a set of not 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 the tree T_j. Small means that there are at most μn such paths.

5. For $i \in \{j+1, ..., l\}$, again T_i is obtained from T_{i-1} by adding a set of not neighbouring leaves.

6. $T_l = T$.

Lemma 4 guarantees that this is possible and that $l \leq 10^4 D\mu^{-2}$.

Step 2: Providing templates 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 a 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 12.

Collection of Properties:

Lemma 7 guarantees that we almost surely have

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

By Lemma 9, we almost surely 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 1, almost surely we get $|X_0|, |C_0| \leq 2\frac{\varepsilon |V(K_{n,n})|}{400k} = \frac{\varepsilon n}{100k}$, and hence any vertex is contained in at most $\frac{\varepsilon n}{100}$ C_0-edges (as there are at most k edges of each color incident to any vertex). Thus, the following almost surely holds:

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

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 5 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. Add an edge between two such root vertices in T_0 if the unique path between these vertices' pre-images in T is of of odd length to get a tree T_0' with $|T_0'| = |T_0|$. **R1** guarantees that we can apply Lemma 5 to find a rainbow copy of tree 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 of T_0 in X_0 with all the components in the right place to be able to extend T_0.

Step 4: Find large rainbow stars. In a next step, applying Lemma 6, for the appropriate integers $m \leq \frac{n}{D}$ and $d_1, ..., d_m \geq D$, let $v_1, ..., v_m \in V(S_0)$ 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 $d = \sum_{i=1}^{m} d_i = |T_1| - |T_0| \leq |T| = (1-\varepsilon)\frac{n}{k}$. For each $i \in [m]$, let $n_i = \left\lceil 1 - \frac{\varepsilon}{8}\frac{nd_i}{kd} \right\rceil$. Note that

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

Using **R3** and Lemma 6, 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 $n_i \geq d_i$, so we have found stars that are larger than required. We will later "randomize" this result by taking the section 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 colors C_1 which is used exclusively for the stars, but this will have to depend on X_1 to make sure we have enough colors in C_1 that actually appear between S_0 and 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 1 pair. We will later (Step 6 and 7) choose X_1 and C_1 ins such a way that, for such a pairing, $x \in X_1$ 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 colors, 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 a template of vertices and colours for each of the extension steps of our forest. We will choose these templates as p_i-random subsets of $V(K_{n,n})$ and $C(K_{n,n})$ respectively.

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

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

where the inequality follows as $l \leq 10^4 D\mu^{-2} = o(\log^{10} n)$. For p_l, we have the residual term

$$
\begin{aligned}
p_l &= 1 - p_0 - \sum_{i\in[l-1]} p_i \\
&= 1 - \frac{\varepsilon}{400} - (1+\frac{\varepsilon}{4})k\frac{|T| - m_l - |T_0|}{n} - \frac{\varepsilon(l-1)}{4l} \\
&\geq 1 - (1+\frac{\varepsilon}{4})k\frac{(1-\varepsilon)n/k - m_l}{n} - \frac{\varepsilon}{4} \\
&\geq n^{-1/10^4}
\end{aligned}
$$

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 $m_1 = d$. By (6), we have, for each $i \in [m]$, that

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

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

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

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 $C^{unpaired} = C(V(K_{n,n})\backslash C^{paired}$ 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 $c \in C^{paired}$, 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 $c \in C^{unpaired}\backslash C_0$, 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 uniformly at random. Thus, each colour appears in C_1 with the same probability: p_1. Whether a color is included depend on any other colour's in- or exclusion.

Step 8a: Choose X_2, ..., X_l. Randomly partition $V(K_{n,n})\backslash(X_0 \cup X_1)$ as $X_2 \cup ... \cup X_l$ so that, for each $x \in V(K_{n,n})(X_0 \cup X_1)$, the class of x is chosen independently at random with $P(x \in X_i) = \frac{p_1}{1-p_0-p_1}$ for each $2 \leq i \leq l$. Note that, for each $i \in \{0, 1, ..., l\}$, each $x \in V(K_{n,n})$ appears in X_i independently at random 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, ..., C_l. Randomly partition $C(K_{n,n})\backslash(C_0 \cup C_1)$ as $C_2 \cup ... \cup C_l$ 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 $2 \leq i \leq l$. Note that, for each $0 \leq i \leq l$, each colour $c \in C(K_{n,n})$ appears in C_i independently at random 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 $m_i \leq p_i \frac{n}{k}$ for all $i \in [l]$. For $A \subset V(K_{n,n}) \backslash X_i$, set $q_{i,j} = \frac{|A \cap V_j(K_{n,n})|}{|A|}$ ($i \in [l]$, $j \in [2]$). Then pick $X_{i,j} \subset V_j(K_{n,n})$, $C_{i,j} \subset C(K_{n,n})$ q_i^{3-j}-random ($i \in [l]$, $j \in [2]$) and note that for $|A| = m_i$:
$|A \cap V_j(K_{n,n})| = q_{i,j}|A| = q_{i,j}m_i \leq q_{i,j}\frac{p_i n}{k}$. By Lemma 11, we find a $C_{i,1}$-rainbow matching from $A \cap V_1(K_{n,n}$ to $X_{i,2}$ of size at least $|A \cap V_1(K_{n,n})| - \mu p_i q_{i,1} n = m_i q_{i,1} - \mu p_i q_{i,1} n$ and a $C_{i,2}$-rainbow matching from $A \cap V_2(K_{n,n})$ to $X_{i,1}$ of size at least $|A \cap V_2(K_{n,n})| - \mu p_i q_{i,2} n = m_i q_{i,2} - \mu p_i q_{i,2} n$, so that by combining these matchings, using $q_{i,1} + q_{i,2} = 1$, we get with probability $1 - o(n^{-1})$

R5 For each $i \in [l]$ and subset $A \subset V(K_{n,n}) \backslash X_i$ with $|A| = m_i \leq p_i \frac{n}{k}$ there is a C_i-rainbow matching with at least $m_i - \mu p_i n$ edges from A into 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 $|T_0| \leq \mu n$ colours in C_0 and at most μn vertices in X_0.

Step 11: Extend to a copy of $T_2, ..., T_{j-1}$ by adding non-neighbouring leaves. Iteratively, for each $2 \leq i \leq j-1$, extend S_{i-1} to $S_i \subset K_{n,n}[X_0 \cup ... \cup X_i]$, a $(C_0 \cup ... \cup C_i)$-rainbow copy of T_i with $|C(S_i) \cap C_0| \leq \mu n + p_i \sum_{i'=2}^{i} \mu p'_i n$ and $|V(S_i) \cap X_0| \leq \mu n + p_i \sum_{i'=2}^{i} \mu p'_i n$ (so that, certainly, $|C(S_i) \cap C_0| \leq 2\mu n$ and $|V(S_i) \cap X_0| \leq 2\mu n$). 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 then finish using Lemma 12 with G as the subgraph of $K_{n,n}$ of colour-$((C_0 \backslash C(S_{i-1})) \cup C_i)$ edges, $C = C_i, C' = C_0 \backslash C(S_{i-1}), X^1 = A_i \cap V_1(K_{n,n}), X^2 = A_i \cap V_2(K_{n,n})), Y = X_i$ and $Z = X_0 \backslash (S_{i-1})$ to find a rainbow matching covering the whole set A_i and thus completing the extension step. The estimations for the colours and vertices from X_0 and C_0 respectively hold because we have used at most $\mu p'_i n$ colours and vertices to "finish" the i'-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 to $S_j \subset K_{n,n}[X_0 \cup ... \cup X_j]$, a $(C_0 \cup ... \cup C_j)$-rainbow copy of T_j with $|C(S_j) \cap C_0| \leq 4\mu n + p_j \sum_{i'=1}^{i} \mu p'_i n$ and $|V(S_j) \cap X_0| \leq 3\mu n + p_j \sum_{i'=1}^{i} \mu p'_i n$. 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 10 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 $T_{j+1}, ..., T_l$. Finally, for each $i \in j+1, ..., l$, use **R1**, **R5** and 12 as before to extend S_{i-1} to $S_i \subset K_{n,n}[X_0 \cup ... \cup X_i]$, a $(C_0 \cup ... \cup C_i)$-rainbow copy of T_i with at most $4\mu n + p_i \sum_{i'=2}^{i} \mu p'_i n$ colours in C_0 and at most $3\mu n + p_i \sum_{i'=2}^{i} \mu p'_i n$ vertices in X_0. When this is finished, we have a rainbow copy of $T_l = T$, as required.

Closing Remark.

What the arguments above are achieving is that

1. We have a proof of Theorem 1 if if **R1**-**R5** *collectively* hold for the (random) partitions $V(K_{n,n}) = X_0 \cup ... \cup X_l$ and $C(K_{n,n}) = C_0 \cup ... \cup C_l$
2. **R1**-**R3** hold with high probability.
3. **R4** holds with high probability, if **R1**-**R3** hold, else we say it does not hold.
4. **R5** holds for each $2 \leq i \leq l$ with probability $1 - o(l \cdot n^{-1}) = 1 - o(1)$ if R1-**R4** hold, else we say it does not hold.

All in all, this shows that **R1**-**R5** all collectively hold with high probability for given partitions $V(K_{n,n}) = X_0 \cup ... \cup X_l$ and $C(K_{n,n}) = C_0 \cup ... \cup C_l$. This means that there must be some such partitions, which implies that a rainbow copy of T exists. □

4 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 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. It therefore seems reasonable to propose that $K_{n,n}$ can be decomposed into copies of any tree with n edges. We did, however, not succeed in translating the methods from their proof of Ringel's conjecture to the bipartite case to prove this. 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 4) 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, given the asymptotic result, believe that other methods may succeed in showing the following:

Conjecture 1. *Let T be a tree with n edges. Then $K_{n,n}$ can be decomposed into n copies of T.*

References

[1] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, 6:290–297, 1959. ISSN 0033-3883.

[2] P. Erdős, A. Gyárfás, and L. Pyber. Vertex coverings by monochromatic cycles and trees. *J. Combin. Theory Ser. B*, 51(1):90–95, 1991. ISSN 0095-8956. doi: 10.1016/0095-8956(91)90007-7. URL https://doi.org/10.1016/0095-8956(91)90007-7.

[3] J. A. Gallian. A dynamic survey of graph labeling. *Electron. J. Combin.*, 5:Dynamic Survey 6, 43, 1998. URL http://www.combinatorics.org/Surveys/index.html.

[4] S. Janson, T. Luczak, and A. Rucinski. *Theory of random graphs.* John Wiley and Sons, 2000.

[5] F. Joos, J. Kim, D. Kühn, and D. Osthus. Optimal packings of bounded degree trees. *J. Eur. Math. Soc. (JEMS)*, 21(12):3573–3647, 2019. ISSN 1435-9855. doi: 10.4171/JEMS/909. URL https://doi.org/10.4171/JEMS/909.

[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. *Combin. Probab. Comput.*, 6(3):337–347, 1997. ISSN 0963-5483. doi: 10.1017/S0963548397003106. URL https://doi.org/10.1017/S0963548397003106.

[8] R. Montgomery, A. Pokrovskiy, and B. Sudakov. Embedding rainbow trees with applications to graph labelling and decomposition. *J. Eur. Math. Soc. (JEMS)*, 22(10):3101–3132, 2020. ISSN 1435-9855. doi: 10.4171/jems/982. URL https://doi.org/10.4171/jems/982.

[9] R. Montgomery, A. Pokrovskiy, and B. Sudakov. A proof of Ringel's conjecture. *Geom. Funct. Anal.*, 31(3):663–720, 2021. ISSN 1016-443X. doi: 10.1007/s00039-021-00576-2. URL `https://doi.org/10.1007/s00039-021-00576-2`.

[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. *Combin. Probab. Comput.*, 15(1-2):229–251, 2006. ISSN 0963-5483. doi: 10.1017/S0963548305007042. URL `https://doi.org/10.1017/S0963548305007042`.

[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. *Publ. Math. Debrecen*, 13:145–168, 1966. ISSN 0033-3883.

[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.