Long directed rainbow cycles and rainbow spanning trees

Frederik Benzing [a], Alexey Pokrovskiy [b], Benny Sudakov [c,1]

[a] Department of Computer Science, ETH, 8092 Zurich, Switzerland
[b] Department of Economics, Mathematics, and Statistics, Birkbeck, London, United Kingdom of Great Britain and Northern Ireland
[c] Department of Mathematics, ETH, 8092 Zurich, Switzerland

ARTICLE INFO

ABSTRACT

A subgraph of an edge-coloured graph is called rainbow if all its edges have different colours. The problem of finding rainbow subgraphs goes back to the work of Euler on transversals in Latin squares and was extensively studied since then. In this paper we consider two related questions concerning rainbow subgraphs of complete, edge-coloured graphs and digraphs. In the first part, we show that every properly edge-coloured complete directed graph contains a directed rainbow cycle of length $n - O(n^{4/5})$. This is motivated by an old problem of Hahn and improves a result of Gyarfas and Sarkozy. In the second part, we show that any tree T on n vertices with maximum degree $\Delta_T \leq \beta n/\log n$ has a rainbow embedding into a properly edge-coloured K_n provided that every colour appears at most αn times and α, β are sufficiently small constants.

1. Introduction

In this paper we study rainbow subgraphs of properly edge-coloured complete graphs and digraphs. An edge-colouring of an undirected graph is *proper* if no two edges sharing a vertex have the same colour. In the directed setting, no pair of edges with a common start point and no pair of edges with a common end point may be monochromatic. In both cases a subgraph of the complete

E-mail addresses: beningf@inf.ethz.ch (F. Benzing), dr.alexey.pokrovskiy@gmail.com (A. Pokrovskiy), benjamin.sudakov@math.ethz.ch (B. Sudakov).

graph is *rainbow* if all its edges have distinct colours. We define the complete directed graph on n vertices, denoted $\overrightarrow{K_n}$, to be the graph on n vertices with an edge going in both directions between any two distinct vertices.

The search for rainbow structures can be traced back to the 18th century, when Euler initiated the study of transversals in Latin squares. In the meantime, a multitude of conjectures and results in this field has been published, see e.g. [31, Chapter 9] for a survey. Let us first define the notions involved and give their natural reformulation in terms of rainbow subgraphs. A *Latin square of order* n is an $n \times n$ array filled with symbols, so that each symbol appears precisely once in each row and column. A *partial transversal of length k* is a collection of k cells of the Latin square, so that no two cells share their row, column or symbol. To every $n \times n$ Latin square, one can assign a proper colouring of $\overleftrightarrow{K_n}$ with a loop added at each vertex as follows—label the vertices of $\overleftrightarrow{K_n}$ by the numbers $1, \ldots, n$ and colour each directed edge ij by the symbol in the cell (i, j). Identifying the cell (i, j) with the edge ij, a partial transversal now corresponds to a rainbow subgraph of $\overleftrightarrow{K_n}$ in which each vertex has in- and out-degree at most 1.

A long-standing conjecture attributed (in slightly different versions) to Ryser [28], Brualdi [10], and Stein [29] asks whether every Latin square of order n contains a partial transversal of length $n - 1$. The best known approximate version of this, due to Hatami and Shor [20], asserts that there always is a partial transversal of length $n - O\left(\log^2 n\right)$, improving several earlier results. A partial transversal is called *cycle-free* if the corresponding subgraph of $\overleftrightarrow{K_n}$ is cycle-free. Since partial transversals correspond to rainbow subgraphs of $\overleftrightarrow{K_n}$ with maximum in- and out-degree 1, cycle-free partial transversals correspond to rainbow subgraphs of $\overleftrightarrow{K_n}$ which are unions of vertex-disjoint directed paths. We shall call such subgraphs *path forests*. Gyárfás and Sárközy [17] conjectured that every Latin square contains a partial, cycle-free transversal of length $n - 2$ and showed that this would be best possible for $n = 4$. They proved that every Latin square contains a partial, cycle-free transversal of size $n - O(n \log \log n / \log n)$. One of our main results improves upon and generalises this result.

Theorem 1.1. *Every properly edge-coloured $\overleftrightarrow{K_n}$ contains a rainbow path forest of length $n - O(n^{2/3})$ and a rainbow cycle of length $n - O(n^{4/5})$.*

Note that this result is more general than that of Gyárfás and Sárközy [17]. On the one hand it shows the existence of long cycles rather than just path forests, on the other hand it only assumes $\overleftrightarrow{K_n}$ to be properly coloured, rather than each colour-class being a 1-factor (which is the case for colourings of $\overleftrightarrow{K_n}$ coming from Latin squares).

Theorem 1.1 is also interesting since it generalises a recent result by Alon and the second two authors [2]. They proved that every properly edge-coloured K_n contains a rainbow cycle of length $n - O(n^{3/4})$ and our proof of Theorem 1.1 is based on their key ideas. Both these results raise the question of how far the error terms, $O(n^{4/5})$ and $O(n^{3/4})$ respectively, can be pushed. In the undirected case, Balogh and Molla [6] reduced the $O(n^{3/4})$ error term to $O(n^{1/2} \log n)$. Hahn [18] conjectured that every properly coloured complete graph contains a rainbow Hamilton path. However, Maamoun and Meyniel [25] constructed an edge-colouring of K_n, for $n = 2^k$, which does not contain a rainbow Hamilton path, disproving Hahn's conjecture. Their construction generalises easily to the directed case: Given a colouring K_n which does not contain a Hamilton path, replace each undirected edge $\{u, v\}$ of colour f by two directed edges $(u, v), (v, u)$ giving both of them colour f. Then this colouring of $\overleftrightarrow{K_n}$ does not contain a directed (or, in fact, undirected) Hamilton path.

Beyond Latin squares, rainbow structures play an important role in Ramsey Theory, for they are one of the structures guaranteed by the canonical theorem of Ramsey Theory, proved by Erdős and Rado [13]. In this setting it is a natural question under which conditions a given graph G has a rainbow embedding into an edge-coloured K_n. If such an embedding exists, we call the colouring of K_n *G-rainbow*. We also say that an edge-colouring is *locally k-bounded* if every colour appears at most k times at each vertex. Analogously, the colouring is *globally k-bounded* if every colour appears at most k times in total.

In 1976 Bollobás and Erdős [8] raised the question which edge-colourings contain a properly coloured Hamilton cycle and proved that locally αn-bounded colourings do, for $\alpha = 1/69$. They

conjectured that their statement holds for any $\alpha < 1/2$. After several improvements, Lo [23] proved an asymptotic version of this conjecture.

Since then different generalisations have been made, asking for rainbow rather than properly coloured graphs. E.g., Hahn and Thomassen [19] formulated an analogue to the Bollobás–Erdős conjecture in 1986, suggesting that there exists an α so that any globally αn-bounded colouring contains a rainbow Hamilton cycle. They proved that any globally k-bounded colouring contains a rainbow Hamilton cycle for $k = \alpha n^{1/3}$. Other sublinear values for k were proved by Erdős, Nešetřil, and Rödl [12], and by Frieze and Reed [16]. Albert, Frieze and Reed [1] finally showed that any globally $n/64$-bounded colouring contains a rainbow Hamilton cycle.

Frieze and Krivelevich posed the question whether there is an $\alpha = \alpha(\Delta)$ so that any globally αn-bounded colouring contains a copy of every spanning tree with maximum degree Δ. Böttcher, Kohayakawa and Procacci [9] showed a stronger result, namely that any globally $\frac{n}{51\Delta^2}$-bounded colouring contains a rainbow copy of any graph (and not only tree) with maximum degree at most Δ. The third author and Volec [30] showed that the Δ-dependence in this theorem is best possible. More precisely, they constructed a locally 3-bounded and globally 9-bounded colouring of K_{3n} which is not T-rainbow for any tree T of radius 2. This especially holds for radius-2-trees with maximum degree $O\left(\sqrt{n}\right)$ showing that the Δ-dependence is indeed optimal.

In light of this construction, we consider the question whether properly coloured (i.e. locally 1-bounded) colourings behave differently than locally 3-bounded colourings. More precisely, we ask under which conditions a proper colouring of K_n is T-rainbow for a given spanning tree T. The intuition that we might be able to find a rainbow embedding of any spanning tree into any properly coloured K_n is false. Recall the colouring from Maamoun and Meyniel [25] of K_n, for $n = 2^k$, which does not contain a Hamilton path. We also observe that the colouring they present does not allow for a rainbow embedding of any spanning tree in which all but precisely 2 vertices have odd degree, giving a much wider class of counterexamples. The details of this and other colourings of K_n which do not contain rainbow copies of certain spanning trees are presented in Section 5.

The approach of Böttcher, Kohayakawa and Procacci [9], which uses a framework for the Local Lemma developed by Lu and Székely [24], can be modified straightforwardly to show that any globally $\frac{n}{C\Delta}$-bounded (where C is a sufficiently large constant), proper colouring of K_n contains a rainbow copy of every tree with maximum degree Δ. Combining this method with some additional ideas, we show that the condition of being properly coloured gives much stronger results, distinguishing them from locally 3-bounded colourings, for example.

Theorem 1.2. *There are constants $\alpha, \beta > 0$ so that the following holds for every integer n. Let T be a tree on n vertices with maximum degree at most $\beta n / \log n$ and let c be a proper colouring of K_n which is globally αn-bounded. Then c is T-rainbow.*

Recall that the condition of the colouring being globally αn-bounded cannot simply be dropped. However, we believe that the statement might be true for any $\alpha < 1/2$. It is also possible, that the condition on the maximum degree of the tree T can be dropped.

The proofs of Theorems 1.1 and 1.2 are given in Sections 3 and 4 respectively. Each section starts with a short outline of the main steps of the proof. In Section 5, we end with some remarks and open questions.

2. Notation

For any graph G, let $V(G)$ be the set of vertices of G, let $E(G)$ be the set of edges of G and write $v(G) = |V(G)|$ as well as $e(G) = |E(G)|$.

For two vertices v, v' let vv' denote the directed edge from v to v'. Suppose we are given a colouring c of the edges of G. For an edge e let $c(e)$ be its colour. For a graph G, let $C(G)$ denote all the colours appearing in G.

For a directed path $P = v_1 \to \ldots \to v_{|P|}$ denote by $f(P) := v_1$ the first vertex of P and by $l(P) := v_{|P|}$ the last vertex of P. If $v \in P$ with $v \neq l(P)$, then let v^+ be the successor of v on P and write $c(v) := c(vv^+)$. If $v \in P$ with $v \neq f(P)$, then let v^- be the predecessor of v on P. For subsets of vertices $X \subset V(P) \setminus l(P)$, define $X^+ = \{x^+ : x \in X\}$. Define X^- similarly.

We call a pairwise vertex-disjoint collection $\mathcal{P} = \{P_1, \ldots, P_r\}$ of directed paths a *path forest*. Given a path forest $\mathcal{P}$ as above, define $f(\mathcal{P}) = \{f(P_i) : 1 \leq i \leq r\}$ and $l(\mathcal{P}) = \{l(P_i) : 1 \leq i \leq r\}$. If $v \in V(\mathcal{P}) \setminus l(\mathcal{P})$, then let v^+ be the successor of v in $\mathcal{P}$ and also write $c(v) = c(vv^+)$. Also define v^- and X^+, X^- for suitable $v \in V(\mathcal{P}), X \subset V(\mathcal{P})$ in the expected way.

For two paths P, Q, which only intersect in the vertex $l(P) = f(Q)$, we write $P + Q$ for the concatenation of P and Q.

For sets of vertices $A, B \subset V(G)$ and a subset of colours $C \subset C(G)$, let $d(A, B, C)$ denote the number of edges from A to B whose colour lies in C. Moreover, we write $d^-(A, C) := d(V(G), A, C)$ and similarly $d^+(A, C) = d(A, V(G), C)$. If we drop C in this notation, we assume $C = C(G)$, e.g. $d(A, B) := d(A, B, C(G))$. Also, if $A = \{v\}$ consists of a single vertex, we will not resist the temptation of writing $d(v, B)$ rather than $d(\{v\}, B)$. For a vertex $v \in V(G)$ we also write $N^-(v) = \{w \in V(G) : \ wv \in G\}$.

3. Long directed rainbow cycles

In this section we prove Theorem 1.1. Before giving the details we outline the strategy: We will first put aside a few colours from $\overleftrightarrow{K_n}$ to an expander H. The remaining graph $\overleftrightarrow{K_n} \setminus H$ will have high minimum degree and colours disjoint form H. In this graph we will find a long rainbow path forest (a vertex-disjoint collection of paths). In the final step we will use edges from H to 'rotate' and 'glue together' the rainbow path forest to obtain a long cycle. This technique was also used in [2] for the undirected case and we show here how to transfer this technique to the directed case.

3.1. Random subgraph

This section is concerned with 'random subgraphs' of a properly coloured $\overleftrightarrow{K_n}$. Here by 'random subgraph' we mean the subgraph that for each colour of $\overleftrightarrow{K_n}$ contains all edges of that colour with probability p and otherwise contains no edge of that colour, where the coin is thrown independently for each colour. The theorem below assures that such a subgraph looks almost like a truly random graph. The result was originally proved by Alon and the last two authors [2] for properly edge-coloured, complete, undirected graphs. It is also pointed out that the argument can be generalised to different settings. The version we need, stated below, follows from Theorem 5.1 in [2] and the remarks thereafter.

Theorem 3.1 (Alon, Pokrovskiy, and Sudakov [2])**.** *Given a proper edge-colouring of $\overleftrightarrow{K_n}$, let G be the random subgraph obtained by choosing each colour class independently with probability p satisfying $\log(n)/n \ll p \leq 1/2$. Then, with high probability, all vertices in G have in-degree and out-degree $(1 - o(1))np$ and for every two disjoint subsets A, B with*

$$|A|, |B| \gg (\log n/p)^2, \ e_G(A, B) \geq (1 - o(1))p|A \parallel B|.$$

3.2. Long rainbow path forest

The following lemma shows that we can find long rainbow path forests in properly coloured digraphs with high in-degree. It is based on a technique by Andersen [4], developed for undirected graphs. Its adaption to the directed case requires an additional idea relying on the following simple observation.

Observation 3.2. *Let $\mathcal{P}$ be a path forest and let $v \in l(\mathcal{P})$ and $f, f' \in f(\mathcal{P})$ be distinct. Then, at least one of $\mathcal{P} + vf$ and $\mathcal{P} + vf'$ is still a path forest.*

Proof. Observe that a path forest is precisely a graph which contains no cycle and has maximum in-degree and maximum out-degree at most 1. In $\mathcal{P} + vf$ every vertex still has in-degree/out-degree at most 1. Thus, if $\mathcal{P} + vf$ is not a path forest, then it must contain a cycle. This implies that v and f lie on the same path of $\mathcal{P}$. But then v and f' lie on different paths in $\mathcal{P}$, and hence $\mathcal{P} + vf'$ is a path forest. □

We record another simple observation (without proof, since it is immediate) that will help us later.

Observation 3.3. *Let $\mathcal{P}$ be a path forest, $v \in l(\mathcal{P})$ and $f \in f(\mathcal{P})$ as before and assume that $\mathcal{P}' = \mathcal{P} + vf$ is a path forest. We then have $|\mathcal{P}'| = |\mathcal{P}| - 1$, meaning that $\mathcal{P}'$ consists of one path less than $\mathcal{P}$. Also, deleting an edge (but not the vertices it is incident to) from any path forest increases its size (number of paths) by 1.*

We now state and prove the main lemma of this section, guaranteeing long rainbow path forests in properly coloured digraphs with high in-degree.

Lemma 3.4. *Let $n \in \mathbb{N}$ and $0 < \gamma \leq \delta < 1$ satisfy*

$$\left\lfloor \frac{\gamma n}{\lfloor 1/\delta \rfloor} \right\rfloor \frac{1}{\lfloor 1/\delta \rfloor} > 1.$$

Then any properly coloured digraph G on n vertices with minimum in-degree at least $(1 - \delta)n$ contains a directed rainbow path forest $\mathcal{P} = \{P_1, \ldots, P_r\}$ with $r \leq \gamma n$ and $|e(\mathcal{P})| \geq (1 - 3\delta)n$.

Remark. In our applications of Lemma 3.4 we will have $\delta \to 0$ and $\delta \gamma n \to \infty$, so that

$$\left\lfloor \frac{\gamma n}{\lfloor 1/\delta \rfloor} \right\rfloor \frac{1}{\lfloor 1/\delta \rfloor} = (1 + o(1)) \, \delta^2 \gamma n.$$

Proof of Lemma 3.4. For ease of readability, we shall assume that $1/\delta$, γn, $\delta \gamma n$ are all integers, so that the condition on γ, δ, n simplifies to

$$\delta^2 \gamma n > 1. \tag{1}$$

Let $\mathcal{P} = \{P_1, \ldots, P_r\}$ be a maximum (w.r.t. $e(\mathcal{P})$) directed rainbow path forest with $r \leq \gamma n$. Assume for contradiction that $|e(\mathcal{P})| < (1 - 3\delta)n$. We will show that $\mathcal{P}$ cannot be maximum. Note that we may assume that $r = \gamma n$, since if $r < \gamma n$, then we can simply add single vertices to $\mathcal{P}$ in order to increase the number of paths. This is possible because $|V(G) \setminus V(\mathcal{P})| = n - |V(P)| \geq n - |e(P)| - r \geq 3\delta n - \gamma n \geq \gamma n$.

We partition the set $f(\mathcal{P})$ into sets $Q_0, \ldots, Q_s$ of equal size $1/\delta$, where $s := \delta \gamma n - 1$, in other words:

$$f(\mathcal{P}) = \bigcup_{i=0}^{s} Q_i, \text{ where the union is disjoint and } |Q_i| = 1/\delta \text{ for } i = 0, \ldots, s.$$

Define $C_0 = C(G) \setminus C(\mathcal{P})$ to be the set of colours not appearing in $\mathcal{P}$. For $1 \leq i \leq s + 1$ we define subsets $V_i \subset V(G)$ and $C_i \subset C(G)$ recursively as follows:

$$V_i := \{v \in V(G) : d(v, Q_{i-1}, C_{i-1}) \geq 2\}$$
$$C_i := C_{i-1} \cup \{c(v) : v \in V_i \cap V(\mathcal{P}) \setminus l(\mathcal{P})\}.$$

With these definitions at hand, our plan is to show that $|C_i| \geq |C_{i-1}| + \delta n$. Once we establish this, we are almost done: We then have $|C_{s+1}| - |C_0| \geq (s + 1)\delta n = \delta^2 \gamma n^2 > n$ by (1). On the other hand $C_{s+1} \setminus C_0 \subset C(\mathcal{P})$ by construction, so that $|C_{s+1}| - |C_0| \leq n$. This contradiction then proves Lemma 3.4.

It thus only remains to show $|C_i| \geq |C_{i-1}| + \delta n$. We will soon establish the following claim which relies on the maximality of $\mathcal{P}$.

Claim 3.5. *We have $V_i \subset V(\mathcal{P}) \setminus l(\mathcal{P})$ for $i = 1, \ldots, s + 1$.*

Before proving the claim let us first show how this claim implies $|C_i| \geq |C_{i-1}| + \delta n$. We have $|C_i| \geq |C_0| + |V_i \cap V(\mathcal{P}) \setminus l(\mathcal{P})|$, since for each $v \in V_i \cap V(\mathcal{P}) \setminus l(\mathcal{P})$, we have $c(v) \in C_i$. Using the claim we get $|V_i \cap V(\mathcal{P}) \setminus l(\mathcal{P})| = |V_i|$. Combining these two inequalities we obtain
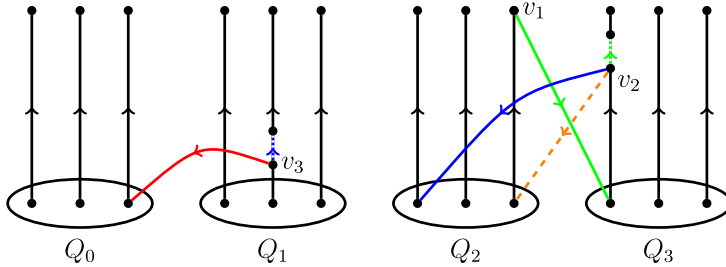
$$|C_i| \geq |C_0| + |V_i|. \tag{2}$$

Fig. 1. An example of how the algorithm in the proof of Claim 3.5 might work. Suppose $i_1 = 4$, i.e. there is a vertex $v_1 \in V_4$, so that there are two edges from v_1 to Q_3 using colours in C_3. In the first step, the algorithm adds one of these edges (the solid green edge in our case) to the rainbow path forest and then deletes an edge of the same colour (the dotted green edge). Now, i_2 is the smallest index, so that 'green' is an element of C_{i_2}, suppose $i_2 = 3$. Moreover, v_2 is the starting point of the deleted (green dotted) edge. There are two edges from v_2 to Q_2 (drawn as orange dashed and blue solid lines). In the second step, the algorithm will add one of them to the path forest. Adding the orange edge would create a cycle. Thus, the algorithm uses the other possibility and adds the blue, solid edge. It then deletes the dashed blue edge, to keep the path forest rainbow. Suppose we have $i_3 = 1$. In the last step, the algorithm adds the red edge and terminates, since there was no other red edge in the path forest. The resulting rainbow path forest contains one more edge than the original path forest contradicting maximality of $\mathcal{P}$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

We now use a double counting argument to bound $|V_i|$ in terms of $|C_{i-1}| - |C_0|$. Choose $A \in \mathbb{R}$ so that $|C_{i-1}| - |C_0| = A\delta n$. Observe that for any vertex v we have $d^-(v, C_{i-1}) \geq d^-(v) - (|C(G)| - |C_{i-1}|) \geq (1 - \delta)n - (|C(G)| - |C_0|) + (|C_{i-1}| - |C_0|) \geq 2\delta n + A\delta n$. Here, in the last inequality we used $|C(G)| - |C_0| = |C(\mathcal{P})| \leq (1 - 3\delta)n$. Using $d^-(v, C_{i-1}) \geq (A + 2)\delta n$ we obtain

$$d^-(Q_{i-1}, C_{i-1}) \geq (A + 2)\delta n \cdot |Q_{i-1}| = (A + 2)n. \tag{3}$$

On the other hand, writing $X = \{v \in V(G) : d(v, Q_{i-1}, C_{i-1}) \leq 1\}$ and $V_i = \{v \in V(G) : d(v, Q_{i-1}, C_{i-1}) > 1\}$, we also get

$$d^-(Q_{i-1}, C_{i-1}) \leq |X| + |V_i| \cdot |Q_{i-1}| \leq n + |V_i|/\delta. \tag{4}$$

Here, the first inequality uses $d(v, Q_{i-1}, C_{i-1}) \leq |Q_{i-1}|$ for $v \in V_i$. Putting (3) and (4) together, we finally obtain $|V_i| \geq (A + 1)\delta n$. Using this in (2) we get

$$|C_i| - |C_0| \geq (A + 1)\delta n.$$

Recalling the definition of A, this establishes $|C_i| \geq |C_{i-1}| + \delta n$.

It is only left to prove Claim 3.5 to complete the proof of Lemma 3.4. We first record the following simple, but helpful reformulation of the definitions of C_i and V_i without proof:

Observation 3.6. *If $c \in C_i \setminus C_{i-1}$ for some $i > 0$, then there is some $v \in V_i \cap V(\mathcal{P}) \setminus l(\mathcal{P})$ with $c(v) = c$ and such that there are at least two edges from v to Q_{i-1} whose colours lie in C_{i-1}.*

Proof of Claim 3.5. We prove the claim by contradiction, i.e. we assume that there exists an index i such that $V_i \not\subset V(\mathcal{P}) \setminus l(\mathcal{P})$ and deduce that $\mathcal{P}$ is not a maximum rainbow path forest. Let i_1 be the smallest index with $V_{i_1} \not\subset V(\mathcal{P}) \setminus l(\mathcal{P})$ and choose some $v_1 \in V_{i_1} \cap (l(\mathcal{P}) \cup (V(G) \setminus V(\mathcal{P})))$. For now, assume $v_1 \in l(\mathcal{P})$. The case $v_1 \in V(G) \setminus V(\mathcal{P})$ will be very similar and treated at the end of the proof of this claim.

We will apply an algorithm that transforms the original rainbow path forest $\mathcal{P}_1 = \mathcal{P}$ into a rainbow path forest $\mathcal{P}_{T+1}$ with one more edge than $\mathcal{P}$ in T steps. The notation needed for the formal description of this algorithm makes the proof a bit technical, we provide an instance of the algorithm in Fig. 1.

The algorithm will consist of T steps, where the kth step of the algorithm will have as input an index i_k, a vertex v_k, and a rainbow path forest $\mathcal{P}_k$ and it will generate an output $i_{k+1}, v_{k+1}, \mathcal{P}_{k+1}$.

The path forest $\mathcal{P}_{k+1}$ will be obtained from $\mathcal{P}_k$ by adding one edge to $\mathcal{P}_k$ and deleting one from it. The only exception from this is the last step, i.e. the Tth step, which will only output $\mathcal{P}_{T+1}$, which will be obtained by adding one edge to $\mathcal{P}_T$ (without deleting one).

All in all, the algorithm will generate a sequence of indices $i_1, \ldots, i_T$, a sequence of vertices $v_1, \ldots, v_T$ and a sequence of path forests $\mathcal{P}_1, \ldots, \mathcal{P}_{T+1}$. These indices, vertices and path forests will satisfy the following properties.

(a) We have $i_k > i_{k+1} > 0$ for $1 \leq k < T$, so the indices i_k are decreasing and bounded below.

(b) The edge added to $\mathcal{P}_k$ and the edge deleted from $\mathcal{P}_k$ at the kth step to obtain $\mathcal{P}_{k+1}$ have the same colour for $k < T$. In particular, $C(\mathcal{P}_{k+1}) = C(\mathcal{P}_k)$.

(c) For $k = 1, \ldots, T$ we have $v_k \in l(\mathcal{P}_k)$.

(d) The edge e added at the kth step to $\mathcal{P}_k$ goes from v_k to Q_{i_k-1}.

(e) For $k \geq 2$ we have $v_k \in V_{i_k} \cap V(\mathcal{P}) \setminus l(\mathcal{P})$. For $k = 1$ we have $v_k \in V_{i_k}$.

(f) The edge e added at the kth step to $\mathcal{P}_k$ satisfies $c(e) \in C_{i_k-1}$ as well as $c(e) \in C_{i_{k+1}} \setminus C_{i_{k+1}-1}$.

It is easily checked that these properties hold for $i_1, v_1, \mathcal{P}_1$ as defined above and we will show that they still hold after the kth step for $i_{k+1}, v_{k+1}, \mathcal{P}_{k+1}$. We list three easy consequences of the properties (b)–(f).

Claim 3.7.

(i) *The algorithm terminates after a finite number of steps.*

(ii) *We have $Q_{i_k-1} \subset f(\mathcal{P}_k)$.*

(iii) *Before the kth step, no edge e of colour $c(e) \in C_{i_k-1}$ was deleted (from $\mathcal{P}_i$ for any i with $1 \leq i < k$).*

Proof of Claim 3.7. For (i), note that by property (a) the sequence $i_1, i_2, \ldots$ is strictly decreasing and bounded below by 0. Since all the indices are integers, the sequence must be finite.

For (ii), note that in the kth step of the algorithm, we add an edge incident to Q_{i_k-1} by (d). In particular, since $i_1 > i_2 > \cdots$, this implies that before the kth step, we added no edge directed to Q_{i_k-1}. This implies that each vertex in Q_{i_k-1} has in-degree 0 in the graph $\mathcal{P}_k$ or equivalently $Q_{i_k-1} \subset f(\mathcal{P}_k)$.

For (iii), note that the edge deleted at the kth step has a colour lying in $C_{i_{k+1}} \setminus C_{i_{k+1}-1}$ by . Since $i_1 > i_2 > \cdots$ and $C_0 \subset C_1 \subset C_2 \subset \ldots$, this implies that before the kth step, no edge e of colour $c(e) \in C_{i_k-1}$ was deleted. This proves Claim 3.7. □

We now describe the kth step of the algorithm. Recall that it has input $i_k, v_k, \mathcal{P}_k$ satisfying properties (b)–(f). We need to show that its output $i_{k+1}, v_{k+1}, \mathcal{P}_{k+1}$ also satisfies these properties (for $k < T$).

kth step of algorithm:

- We first explain which edge will be added to $\mathcal{P}_k$. By (e) we have $v_k \in V_{i_k}$, so there are at least two edges from v_k to Q_{i_k-1} using colours in C_{i_k-1} (by the definition of V_{i_k}). Call these edges $e = v_k f$ and $e' = v_k f'$. By (c) and (ii) we have $v_k \in l(\mathcal{P}_k)$ and $f, f' \in Q_{i_k-1} \subset f(\mathcal{P}_k)$. Thus, by Observation 3.2 either $\mathcal{P}_k + e$ or $\mathcal{P}_k + e'$ is a path forest. Assume without loss of generality that $\mathcal{P}_k + e$ is a path forest. e is the edge that we will add to $\mathcal{P}_k$. Hence, (d) will be true after the kth step. Note also that $c(e) \in C_{i_k-1}$, so that the first part of (f) is satisfied.

- Since we want to add e to $\mathcal{P}_k$, we might need to delete an edge from $\mathcal{P}_k$ which has colour $c(e)$ to make sure that $\mathcal{P}_{k+1}$ will be rainbow. We distinguish two cases.
  Case 1: If $c(e) \in C_0$, then define $\mathcal{P}_{k+1} = \mathcal{P}_k + e$. Observe that $\mathcal{P}_{k+1}$ is rainbow, since $C(\mathcal{P}_k) = C(\mathcal{P}_0)$ (by (b)) does not contain $c(e)$ before adding e. Terminate the algorithm.
  Case 2: If $c(e) \notin C_0$, then choose i_{k+1} so that $c(e) \in C_{i_{k+1}} \setminus C_{i_{k+1}-1}$. This is possible since $C_0 \subset C_1 \subset C_2 \subset \ldots$ by construction. Note that the second part of (f) is now also satisfied. Recall that $c(e) \in C_{i_k-1}$, so that $0 < i_{k+1} \leq i_k - 1$, assuring that (a) is true. By Observation 3.6, we can find $v_{k+1} \in V_{i_{k+1}} \cap V(\mathcal{P}) \setminus l(\mathcal{P})$ with $c(v_{k+1}) = c(e)$. Define $\mathcal{P}_{k+1} = \mathcal{P}_k + e - v_{k+1}v_{k+1}^+$, so that (b) holds.

This step implicitly assumes that $v_{k+1}v_{k+1}^+ \in \mathcal{P}_k$. To see that this is true, note that $v_{k+1}v_{k+1}^+ \in \mathcal{P}_1$ and that, by (iii), it has not been deleted from $\mathcal{P}_1$ before the kth step, so that it must still be in $\mathcal{P}_k$.

- We have already seen that properties (b), (a), (d), (f) hold for this choice of i_{k+1}, v_{k+1}, $\mathcal{P}_{k+1}$. (c) is satisfied since we deleted $v_{k+1}v_{k+1}^+$ from $\mathcal{P}_k$ to obtain $\mathcal{P}_{k+1}$. (e) holds by our choice of v_{k+1}.
- Continue with step $k+1$.

Using Observation 3.3, it is now easy to check that $|\mathcal{P}_{i+1}| = |\mathcal{P}_i|$ for $i = 1, \ldots, T-1$ and $|\mathcal{P}_{T+1}| = |\mathcal{P}_T| - 1$. So, all in all, we get $|\mathcal{P}_{T+1}| = |\mathcal{P}_1| - 1 \leq \gamma n$. It is also not hard to check that $e(\mathcal{P}_1) = \cdots = e(\mathcal{P}_T) = e(\mathcal{P}_{T+1}) - 1$ so that $\mathcal{P}_{T+1}$ is a rainbow path forest with at most γn paths and one more edge than $\mathcal{P}_1 = \mathcal{P}$, contradicting maximality of $\mathcal{P}$ and thus finishing the proof of Claim 3.5 in the case $v_1 \in l(\mathcal{P})$ (recall that we distinguished this case from the case $v_1 \in V(G) \setminus V(\mathcal{P})$ in the beginning of the proof of Claim 3.5).

It remains to consider $v_1 \in V(G) \setminus V(\mathcal{P})$. In this case, property (c) is violated in the first step of our algorithm. We can carry it out nevertheless, adding an edge from v_1 to Q_{i_1-1} and deleting an edge of the same colour to obtain $\mathcal{P}_2$ from $\mathcal{P}_1$. We then have $|\mathcal{P}_2| = |\mathcal{P}_1| + 1$. From now on, the algorithm works exactly as before and we have $|\mathcal{P}_{i+1}| = |\mathcal{P}_i|$ for $i = 2, \ldots, T-1$ and $|\mathcal{P}_{T+1}| = |\mathcal{P}_T| - 1$. So, all in all, we get $|\mathcal{P}_{T+1}| = |\mathcal{P}_1| \leq \gamma n$ as before, completing the proof of Claim 3.5. □

As outlined previously, Lemma 3.4 follows from the above claim. □

3.3. Rotating and gluing the path forest

We now give the last step of the proof of Theorem 1.1. We combine ideas from the undirected case [2, Lemma 4.1] and 'directed path rotations' [15, Section 4.3] with some new ideas needed to ensure a rainbow structure.

Lemma 3.8.   Let $b, m, r \in \mathbb{N}$ so that $mr \leq b$. Let $\mathcal{P} = \{P_1, \ldots, P_r\}$ be a rainbow path forest with $V(\mathcal{P}) \subset V(G)$ and let H be a properly coloured digraph, with $V(H) = V(G)$, so that $C(H)$ is disjoint from $C(\mathcal{P})$ and so that for any two disjoint sets of vertices $A, B \subset V(G)$ with $|A| = |B| = b$ we have $d_H(A, B) \geq 2b + 1$. Also assume that H has minimum in-degree $\delta_H^- \geq 5b$. Then either $|P_1| \geq v(\mathcal{P}) - 2b$ or there exist edges $e_1, e_2, e_3 \in H$ and a rainbow path forest $\mathcal{P}' = \{P'_1, \ldots, P'_r\}$ such that $E(P') \subset E(\mathcal{P}) + e_1 + e_2 + e_3$ and $V(\mathcal{P}') \supset V(\mathcal{P})$ and $|P'_1| \geq |P_1| + m$.

Proof.   Assume that $|P_1| < v(\mathcal{P}) - 2b$. For $i = 1, \ldots, r$ write $P_i = v_{i,1} \to \ldots \to v_{i,|P_i|}$ and define $v_k = v_{1,k}$ for the vertices on P_1. Set

$$A = \{v_j : 2 \leq j \leq 2b\} \qquad A' = \{v_j : j > 2b\}$$
$$B = \{v_{i,j} : i \geq 2, j < m\} \qquad B' = \{v_{i,j} : i \geq 2, j \geq m\}.$$

Note that $2b \leq v(\mathcal{P}) - |P_1| = |B| + |B'| \leq mr + |B'|$ which implies $|B'| \geq b$ since $mr \leq b$. In the following let $B_0 \subset B'$ be a subset of size $|B_0| = b$.

Observe that in H there are at least $5b$ edges directed to v_1. They start at $A \cup B$ or at A' or at B' or at $V(G) \setminus V(\mathcal{P})$. Observe that $|A \cup B| \leq 3b$ so that there are at least $2b$ edges from $A' \cup B' \cup (V(G) \setminus V(\mathcal{P}))$ to v_1. We distinguish three cases: Either there is an edge from B' to v_1 or there are $\geq b$ edges from $V(G) \setminus V(\mathcal{P})$ to v_1 or there are $> b$ edges from A' to v_1. Fig. 2 illustrates how the path forest $\mathcal{P}$ changes in each of the three cases as described below.

Case 1: If H contains an edge $e = v_{s,t}v_1$ from B' to v_1 (i.e. $s \geq 2$ and $t \geq m$), then divide P_s into two parts Q_1 from $v_{s,1}$ to $v_{s,t}$ and Q_2 from $v_{s,t+1}$ to $v_{s,|P_s|}$. Note that Q_2 might be empty. Now define $P'_1 = Q_1 + e + P_1$ and $P'_s = Q_2$ as well as $P'_i = P_i$ for $i \neq 1, s$. Then $\mathcal{P}' = \{P'_1, \ldots, P'_r\}$ satisfies the claim of the Lemma for $e_1 = e_2 = e_3 = e$.

Case 2: Assume now that there are $\geq b$ edges from $V(G) \setminus V(\mathcal{P})$ to v_1. Choose a subset $X \subset N_H^-(v_1) \cap V(G) \setminus V(\mathcal{P})$ of size $|X| = b$. Recall $B_0 \subset B'$ with $|B_0| = b$. Since $d_H(B_0, X) \geq 2b + 1$, there is a vertex $x \in X$ with $d_H(B_0, x) \geq 2$. Pick two edges $f, f' \in H$ from B_0 to x, they have distinct colours since H is properly coloured. So we may without loss of generality assume that $c(f) \neq c(xv_1)$. Write
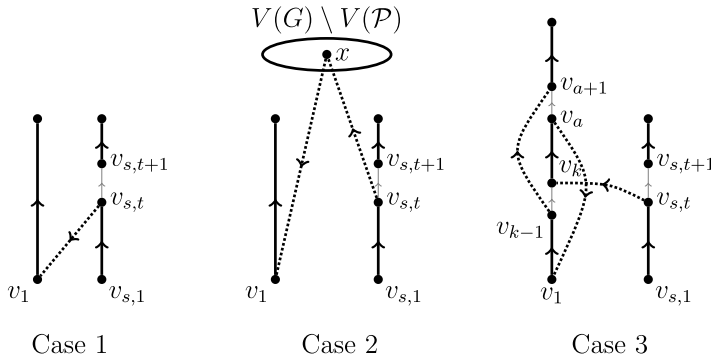
Fig. 2. The three different cases that can arise in the proof of Lemma 3.8. In each case, only the two paths that change, namely P_1 and P_s, are displayed. The edges that will be deleted from the path forest are thin and grey, the edges from the expander H that will be added, i.e. e_1, e_2, e_3, are dotted.

$f = v_{s,t}x$ and define Q_1 and Q_2 as before. Set $P'_1 = Q_1 + f + xv_1 + P_1$ and $P'_s = Q_2$ as well as $P'_i = P_i$ for $i \neq 1, s$. Now $\mathcal{P}' = \{P'_1, \ldots, P'_r\}$ with $e_1 = e_2 = f$ and $e_3 = xv_1$ satisfies the claim.

Case 3: We now consider the case $d_H(A', v_1) > b$. We call a vertex $v_k \in A$ good if there is an index a with $2b < a < |P_1|$ so that the edges $v_a v_1, v_{k-1} v_{a+1}$ both lie in H and have different colours. In this case we call v_a a friend of v_k. If v_k is good, then there is a 'rotation' of P_1 that starts at v_k: Define $R_1 = v_1 \to v_2 \to \ldots \to v_{k-1}$ and $R_2 = v_k \to v_{k+1} \to \ldots \to v_a$ and $R_3 = v_{a+1} \to v_{a+2} \to \ldots \to v_{|P_1|}$ and observe that the path $P_1(v_k, v_a) := R_2 + v_a v_1 + R_1 + v_{k-1} v_{a+1} + R_3$ is a rainbow path covering exactly the vertices of P_1 and having v_k as a starting point.

We now claim that there are at least b good vertices. If not, pick a set $X \subset A$ of vertices that are not good and has size $|X| = b$. Also pick a subset $Y \subset N_H^-(v_1) \cap A' \setminus \{v_{|P_1|}\}$ of size $|Y| = b$ (this is possible since $d_H(A', v_1) > b$). We have $d_H(X^-, Y^+) \geq 2b + 1$, so we can pick a vertex $v_{a+1} \in Y^+$ so that there are two edges $e = v_{k-1} v_{a+1}, e' = v_{j-1} v_{a+1}$ in H with $v_{k-1}, v_{j-1} \in X^-$. Since $c(e) \neq c(e')$ we may without loss of generality assume that $c(e) \neq c(v_a v_1)$. But this shows that $v_k \in X$ is good, contradicting our choice of X. We have thus established that there are at least b good vertices in A.

Let $X \subset A$ be a set of good vertices of size $|X| = b$. Since $d_H(B_0, X) \geq 2b + 1$, there is a vertex $v_k \in X$ so that $d_H(B_0, v_k) \geq 3$. Pick three edges $f_1, f_2, f_3 \in H$ that go from B_0 to v_k. Since v_k is good, it has a friend v_a. Since the colours of f_1, f_2, f_3 are pairwise distinct, we may assume without loss of generality that $c(f_1) \neq c(v_{k-1} v_{a+1}), c(v_a v_1)$. Write $f_1 = v_{s,t} v_k$, define Q_1, Q_2 as before. Finally, set $P'_1 = Q_1 + f_1 + P_1(v_k, v_a)$ and $P'_s = Q_2$ as well as $P'_i = P_i$ for $i \neq 1, s$. Again, $\mathcal{P}' = \{P_1, \ldots, P_r\}$ satisfies the claim for $e_1 = v_{k-1} v_{a+1}, e_2 = v_a v_1, e_3 = f_1$. □

3.4. Proof of Theorem 1.1

We have now collected all the ingredients to prove Theorem 1.1 and show how to combine them in this section. We start by noting that the existence of long rainbow path forests follows directly from Lemma 3.4 applied to a properly edge-coloured $\overleftrightarrow{K_n}$. Take e.g. $\delta = 2n^{-1/3}$ and $\gamma = n^{-1/3}$. We now turn our attention to the existence of long rainbow cycles.

Consider a proper colouring of $\overleftrightarrow{K_n}$. Create a subgraph H of $\overleftrightarrow{K_n}$ as in Theorem 3.1 with $p = 6n^{-1/5}$. Set $G = K_n \setminus H$. Let $b = n^{4/5}$. With high probability every $v \in V(G)$ satisfies $5.5b \leq d_H^-(v) \leq 7b - 1$. Thus, any $v \in G$ satisfies $d_G^-(v) = n - 1 - d_H^-(v) \geq (1 - 7n^{-1/5})n$. By Lemma 3.4, for $\gamma = n^{-3/5}$ and $\delta = 7n^{-1/5}$, G contains a rainbow path forest $\mathcal{P} = \{P_1, \ldots, P_r\}$ with $r \leq \gamma n = n^{2/5}$ and $e(\mathcal{P}) \geq (1 - 21n^{-1/5})n$.

We now apply Lemma 3.8 repeatedly to $\mathcal{P}$ with the expander H constructed above, recall $b = n^{4/5}, r \leq n^{2/5}$ and set $m = n^{2/5}$. All in all, we apply the lemma $n^{3/5}$ times. After each application, we delete all the edges from H that have the same colour as e_1, e_2 or e_3. After $i \leq n^{3/5}$ applications of

[Lemma 3.8](#) each vertex $v \in H$ has in-degree at least $5.5n^{4/5} - 3i \geq 5.5n^{4/5} - 3n^{3/5} \geq 5b$. Moreover, for any two disjoint sets $A, B \subset V(G)$ of size $|A| = |B| = b$, we have, after the ith application, $d_H(A, B) \geq (1 - o(1))b^2p - 3bi \geq 2n^{7/5} \geq 2b + 1$. So we actually may apply [Lemma 3.8](#) $n^{3/5}$ times. At each application, we have either that already $|P_1| \geq V(\mathcal{P}) - 2b \geq (1 - 23n^{-1/5})n$ or that the length of P_1 increases by m. Since $m \cdot n^{3/5} = n$ we must have the first of these two cases before the last iteration. We have thus constructed a rainbow path P_1 with $|P_1| \geq (1 - 23n^{-1/5})n$. In order to turn P_1 into a cycle, observe that between the last b and the first b vertices on P_1 there is an edge that lies in H (or rather what is left of H). We thus obtain a cycle with at least $(1 - 25n^{-1/5})n$ vertices as desired.

4. Rainbow spanning trees

In this section we prove [Theorem 1.2](#). We start with a brief outline of the proof. Let T be a tree on n vertices with maximum degree $\Delta(T) \leq \beta n/\log n$ and let c be a globally αn-bounded, proper colouring of K_n. We will find a rainbow embedding of T into K_n in two steps. Denote by R the $n/4$ vertices of T with highest degrees and set $W = V(T) \setminus R$. In the first step we will find a rainbow embedding of R into K_n in a random-greedy fashion. More precisely, we will find a subset $B \subset V(K_n)$ and a bijection $\pi : R \to B$ so that $\pi(T[R]) =: T_0$ is rainbow. We also write $A = V(K_n) \setminus B$. We will make sure that this partial embedding π of T will satisfy sufficiently nice properties, so that we can carry out the second step of our embedding: We will show that a uniformly random bijection $\tau : W \to A$ gives rise to a rainbow embedding of T. More precisely, the embedded tree $(\pi \cup \tau)(T)$ will be rainbow with positive probability. For this part our main tool is a version of the Local Lemma due to Lu and Székely in [24].

Before giving precise descriptions of these two steps in the next sections we explain what the 'nice' properties of the partial embedding π are and give some intuition why they help us in the second step. To this end we need some notation. For an edge $e \in K_n$ denote by $c(e)$ its colour and for a subgraph H of K_n denote by $C(H)$ the set of all colours appearing in H. For a colour $f \in C(K_n)$ and a vertex $a \in K_n$ we say that f is present at a if there is an edge incident to a which has colour f.

Recall that we aim for a 'nice' rainbow embedding of $T[R]$, i.e. a bijection $\pi : R \to B$ so that $\pi(T[R]) =: T_0$ is rainbow. For such an embedding we define values w_f, w_g, m_g for each $f \in C(K_n)$ and each $g \in V(K_n)$ below. The embedding π will be 'nice' if all these quantities are small.

For $f \in C(K_n)$, we define the weight w_f of f to be

$$w_f = \sum_{\substack{b \in B \\ f \text{ present at } b}} d_T\left(\pi^{-1}(b)\right).$$

Intuitively, this is a rough measure of how likely the colour f is to be used by the embedding $(\pi \cup \tau)$ (where τ is a uniformly random bijection from W to A). If f is present at some vertex $b \in B$, so that $\pi^{-1}(b)$ has a high degree in T, then it is more likely to be used in the embedding $\pi \cup \tau$.

For $g \in V(K_n)$, set $w_g = 0$ if $g \in B$. Otherwise $g \in A$ and we let

$$w_g = \sum_{\substack{u \in R \\ c(g\pi(u)) \in C(T_0)}} d_T(u).$$

This quantity roughly measures (in fact it uses a union bound) how likely we are to embed an 'unsuitable' vertex $w \in W$ onto $\tau(w) = g$. w is 'unsuitable' for g, if there is a vertex $u \in R$ so that $uw \in T$ and so that the colour of the edge $\pi(u)g$ was already in $C(T_0)$.

Finally, for $g \in V(K_n)$ we define m_g to be

$$m_g = |\{a \in A : c(ag) \in C(T_0)\}|.$$

This value simply counts how many colours incident to g were already used in T_0 and therefore should not be used again when using τ to extend the embedding.

We now state our two main lemmas describing the results of the first and second steps of our embedding.

Lemma 4.1. *Let T be a tree on n vertices with maximum degree at most $\beta n/\log n$ and let c be a proper, globally αn-bounded edge colouring of K_n.*

Let $\alpha_1, \alpha_2, \alpha_3 > 0$ satisfy the following relations:

$$16\alpha \leq \alpha_1, \ 16\alpha_2 \leq \alpha_3, \ 32\alpha_1 \leq \alpha_3 \quad and \quad \alpha_1^2 > 32\beta, \ \alpha_2^2 > 64\beta, \ \alpha_3^2 > 64\beta. \tag{5}$$

Assume moreover, that $\gamma, \delta > 0$ satisfy

$$\gamma\delta \geq 2\alpha, \quad 16\delta \leq \alpha_2, \quad 2\gamma \leq \alpha_2. \tag{6}$$

Then there is a rainbow embedding π of the $n/4$ vertices of T which have highest degrees, so that the partial embedding π satisfies $w_f \leq \alpha_1 n$, $m_g \leq \alpha_2 n$ and $w_g \leq \alpha_3 n$ for all colours $f \in C(K_n)$ and all vertices $g \in V(K_n)$.

Lemma 4.2. *Let $0 < r, \chi, \alpha < 1 \leq D$ so that*

$$16(1 - r)^{-3} \cdot \Big(\chi(3D + 11) + \alpha(16D + 12)\Big) \leq 1. \tag{7}$$

Assume that c is a proper colouring of K_n which is globally αn-bounded. Let T be a tree on n vertices, which are partitioned into $V(T) = W \cup R$. Assume also that $\pi : R \to B \subset V(K_n)$ is a rainbow embedding of $T[R]$ and write $A = V(K_n) \setminus B$.

Suppose that

$$|R| \leq rn \quad and \quad \forall w \in W : d_T(w) \leq D.$$

Further suppose that for all $g \in V(K_n)$ and for all $f \in C(K_n)$, we have

$$w_g \leq \chi n, \quad m_g \leq \chi n, \quad w_f \leq \chi n.$$

Then, for n sufficiently large, there is a bijection $\sigma : W \to A$ extending π so that $(\pi \cup \sigma)(T)$ is rainbow.

In Section 4.1 we give the precise statements of our two main tools, namely the Azuma–Hoeffding concentration inequality and the Local Lemma in the framework of Lu and Székely. In Sections 4.2 and 4.3 we prove the Lemmas 4.1 and 4.2 respectively. We combine these two lemmas to give the proof of Theorem 1.2 in Section 4.4.

4.1. Tools

In Section 4.2 we will be concerned with certain random variables and showing that they are concentrated around their expected value. To this end we will use the well known Azuma–Hoeffding inequality due to Azuma [5] and Hoeffding [21]. We will use the following special case:

Theorem 4.3 (*Azuma–Hoeffding Inequality*)*. Suppose $Y_1, \ldots, Y_n$ are random variables, so that Y_i takes the value 0 or d_i and so that Y_i is mutually independent of $Y_1, \ldots, Y_{i-1}$. Then, writing $Y = Y_1 + \cdots + Y_n$ and $\sigma = \sum_{i=1}^{n} d_i^2$, we have for any $t > 0$ that*

$$\Pr\big[Y - \mathrm{E}[Y] > t\big] \leq e^{-t^2/2\sigma}.$$

In Section 4.3 we will make use of the Lovász Local Lemma. It is a powerful tool for probabilistic existence proofs established by Erdős and Lovász [11]. Erdős and Spencer observed that the statement of the Local Lemma could be generalised and formulated the so called *Lopsided Local Lemma* [14], see e.g. [26, Chapter 19.3] for a precise statement and proof. The work of Lu and Székely [24] makes use of the Lopsided Local Lemma by finding a *negative dependency digraph* of a certain set of events in the probability space of random injections. A combination of their main result in [24] and the Lopsided Local Lemma gives Theorem 4.4. To make its statement precise, we start with some definitions.

Suppose we are given finite sets X and Y of the same cardinality and consider the probability space Ω given by picking a bijection $\sigma : X \to Y$ uniformly at random from all such bijections, denote the set of these bijections by $\mathcal{S}$. Let $\tau : T \to U$ be a given bijection between two sets $T \subset X$

and $U \subset Y$. The corresponding *canonical event* $\Omega(T, U, \tau)$ consists of all bijections $\sigma : X \rightarrow Y$ extending τ, i.e.

$$\Omega(T, U, \tau) = \{\sigma \in \mathcal{S} : \sigma(x) = \tau(x) \text{ for all } x \in T\}.$$

We say that two events $\Omega(T_1, U_1, \tau_1)$ and $\Omega(T_2, U_2, \tau_2)$ $\mathcal{S}$-*intersect* if T_1 and T_2 intersect or U_1 and U_2 intersect and write $\Omega(T_1, U_1, \tau_1) \sim \Omega(T_2, U_2, \tau_2)$.

Theorem 4.4 (*Asymmetric Lopsided Lovász Local Lemma in the Framework of Lu And Székely [24]*). *Let $\mathcal{B}$ be a collection of canonical events. Then, with positive probability none of the events $B \in \mathcal{B}$ occurs, provided that for all $B \in \mathcal{B}$ it holds that*

$$\sum_{\substack{B' \in \mathcal{B} \\ B' \sim B}} \Pr[B'] \leq \frac{1}{4}.$$

4.2. Proof of Lemma 4.1

The overall strategy of our proof is simple. We will embed the $n/4$ high degree vertices in a random-greedy fashion and show that the desired properties hold with high probability. We first describe the random-greedy embedding, which we will refer to as RGE from now on.

Denote the vertices of T by $v_1, \ldots, v_n$ so that (i) $v_1, \ldots, v_{n/4}$ are those vertices of T with the highest degrees (splitting ties arbitrarily) and so that (ii) for each $i \in [n/4]$ there is at most one $1 \leq j < i$ so that $v_j v_i$ is an edge of T. Condition (ii) can be satisfied since the $n/4$ vertices with highest degrees of T span a forest, which is 1-degenerate. Sticking with the notation introduced previously, we set $R = \{v_1, \ldots, v_{n/4}\}$, $W = \{v_{n/4+1}, \ldots, v_n\}$.

Moreover, write $d_i = d_T(v_i)$ for $i = 1, \ldots, n$. For distinct vertices $b_1, \ldots, b_k \in K_n$ we denote by $T[b_1, \ldots, b_k]$ the subgraph of K_n in which $b_i b_j$ is an edge precisely if $v_i v_j$ is an edge of T.

We now describe the random-greedy embedding: We will perform $n/4$ steps of the following form. For the kth step, suppose that we have already chosen distinct $b_1, \ldots, b_{k-1} \in K_n$ so that $T[b_1, \ldots, b_{k-1}]$ is rainbow. We say that a vertex $b \in K_n$ is *feasible* if (i) $b \notin \{b_1, \ldots, b_{k-1}\}$ and (ii) the subgraph $T[b_1, \ldots, b_{k-1}, b]$ is rainbow. Out of the feasible vertices we choose one vertex b_k uniformly at random (and independently of previous choices) and then continue with the $(k+1)$th step. It is clear that, if successful, RGE produces a rainbow subgraph $T[b_1, \ldots, b_{n/4}]$. We will not only show that RGE is successful, but that in each step, there are at least $n/2$ feasible vertices, making the embedding sufficiently random for our purposes.

Claim 4.5. *At each of its $n/4$ steps, RGE has at least $n/2$ feasible choices.*

Proof. At the kth step of RGE, there are at most $k - 1 \leq n/4$ vertices b that violate (i) $b \notin \{b_1, \ldots, b_{k-1}\}$. To bound how many vertices b violate (ii) $T[b_1, \ldots, b_{k-1}, b]$ being rainbow, observe the following: v_k has at most one neighbour v_j in $\{v_1, \ldots, v_{k-1}\}$. If v_k has no such neighbour, then any vertex $b \in K_n \setminus \{b_1, \ldots, b_{k-1}\}$ will make $T[b_1, \ldots, b_{k-1}, b]$ rainbow since $T[b_1, \ldots, b_{k-1}, b]$ consists of the same set of edges as $T[b_1, \ldots, b_{k-1}]$. Otherwise, let v_j be the unique neighbour of v_k with $j < k$. Then the only vertices b that will not lead to a rainbow copy of $T[b_1, \ldots, b_{k-1}, b]$ are those with $c(b_j b) \in C(T[b_1, \ldots, b_{k-1}])$. Since the colouring c is proper, all colours present at b_j are distinct and there are at most $|c(T[b_1, \ldots, b_{k-1}])| \leq e(T[b_1, \ldots, b_{k-1}]) \leq n/4$ vertices b with $c(b_j b) \in C(T[b_1, \ldots, b_{k-1}])$. So all in all, there are at most $n/4 + n/4 = n/2$ vertices in K_n which are not feasible, leaving at least $n/2$ feasible vertices. This finishes the proof of Claim 4.5. □

This finishes the description of RGE and we are now ready to prove Lemma 4.1. We need to prove the three inequalities $w_f \leq \alpha_1 n$, $m_a \leq \alpha_2 n$ and $w_a \leq \alpha_3 n$. They will be derived in a standard way: We first bound the expectation of the given random variable, prove tight concentration using the Azuma–Hoeffding inequality and finally apply a union bound over all colours/vertices. This is where we will use our conditions that the colouring c is proper and globally αn-bounded as well as the assumption that $\Delta(T) \leq \beta n/\log n$. Observe that the latter means $d_i \leq \beta n/\log n$ for $i = 1, \ldots, n$

and that we have $\sum_{i=1}^{n} d_i < 2n$. We use this to bound $\sum_{i=1}^{n} d_i^2$. By standard convexity arguments, this sum is maximised if $2 \log n / \beta$ of the d_i take the largest possible value $\beta n / \log n$, so that

$$\sum_{i=1}^{n} d_i^2 < \frac{2 \log n}{\beta} \left(\frac{\beta n}{\log n} \right)^2 = \frac{2 \beta n^2}{\log n}. \tag{8}$$

The three following claims are concerned with bounding the expectation and showing concentration of w_f, m_a, w_a.

Claim 4.6. *For every colour $f \in C(K_n)$ we have $\Pr[w_f \geq \alpha_1 n] = o\left(n^{-2}\right)$.*

Proof. Fix some $f \in C(K_n)$. Write $w_f = X_1 + \cdots + X_{n/4}$, where X_i is a random variable that takes the value d_i if f is present at b_i and zero otherwise. We start by bounding the expectation of w_f. Observe that f is present at at most $2\alpha n$ vertices of K_n since c is globally αn bounded. Recall, that when choosing b_i in the ith step of RGE, there are at least $n/2$ feasible choices. Thus, the probability that we choose a vertex at which f is present is at most $2\alpha n/(n/2) = 4\alpha$. Thus, $\Pr[X_i = d_i] \leq 4\alpha$. It is crucial that this bound on $\Pr[X_i = d_i]$ holds independently of the previous 'history' of RGE. By linearity of expectation, we obtain $E[w_f] \leq 4\alpha \sum_{i=1}^{n/4} d_i \leq 8\alpha n \leq \alpha_1 n/2$, where we used (5) in the last inequality. Since the bound on $\Pr[X_i = d_i]$ holds independently of previous choices of b_j and since changing the outcome of X_i changes the random variable w_f by at most d_i, we can apply Azuma-Hoeffding. Recalling (5) and (8), we find

$$\Pr\left[w_f \geq \alpha_1 n\right] \leq \Pr\left[w_f - E[w_f] \geq \alpha_1 n/2\right] \leq \exp\left(-\frac{(\alpha_1 n/2)^2}{2 \sum_{i=1}^{n/4} d_i^2} \right)$$

$$\leq \exp\left(-\frac{\alpha_1^2}{16\beta} \log n \right) = o\left(n^{-2}\right).$$

This finishes the proof of Claim 4.6. □

Claim 4.7. *For each vertex $g \in K_n$ we have $\Pr[m_g \geq \alpha_2 n] = o\left(n^{-1}\right)$.*

Proof. Fix some vertex $g \in K_n$. We start by observing that not 'many' vertices of K_n can share 'a lot of' colours with g. More precisely, for a vertex $b \in K_n$ let $ov(b) = |\{f \in K_n | f \text{ present at } g \text{ and } b\}|$. Since there are less than n colours present at g and since every colour is present at at most $2\alpha n$ vertices, we can bound

$$\sum_{b \in K_n} ov(b) \leq n \cdot 2\alpha n = 2\alpha n^2.$$

Call a vertex $b \in K_n$ 'bad' if $ov(b) > \gamma n$. Observe that by the bound given above, we can have at most δn bad vertices since $\delta n \cdot \gamma n \geq 2\alpha n^2$ by (6).

We now define random variables $X_1, \ldots, X_{n/4}$. Set

$$X_i = |\{j > i \mid b_i b_j \in T_0 \text{ and } c(b_i b_j) \text{ present at } g\}|.$$

The condition $j > i$ avoids double counting edges, so that we have $m_g \leq X_1 + \cdots + X_{n/4}$. To bound each X_i we distinguish whether or not b_i is bad. To this end, introduce random variables $Y_1, \ldots, Y_{n/4}$ where $Y_i = d_i$ if b_i is bad and $Y_i = 0$ else. Moreover, let $Z_1, \ldots, Z_{n/4}$ be random variables such that $Z_i = 0$ if b_i is bad and $Z_i = X_i$ otherwise. It is easy to see that $X_i \leq Y_i + Z_i$. Writing $Y = Y_1 + \cdots + Y_{n/4}$ and $Z = Z_1 + \cdots + Z_{n/4}$ for convenience, we thus have $m_g \leq Y + Z$. We first bound Y. Observe that, since there are at most δn bad vertices and at least $n/2$ feasible choices for RGE at the kth step, we have $\Pr[Y_k = d_k] \leq 2\delta$, equivalently $E[Y_k] \leq 2\delta d_k$. We can now proceed analogously to the proof of Claim 4.6: By linearity of expectation $E[Y] \leq 4\delta n \leq \alpha_2 n/4$ (using (6)) and thus, using

Azuma-Hoeffding as well as (5) and (8) again,

$$\Pr[Y \geq \alpha_2 n/2] \quad \leq \Pr\left[Y - E[Y] \geq \alpha_2 n/4\right] \tag{9}$$

$$\leq \exp\left(-\frac{\alpha_2^2}{64\beta} \log n\right) = o\left(n^{-1}\right). \tag{10}$$

To bound Z consider each summand Z_k individually: Fix some k and assume that b_k is not bad, otherwise $Z_k = 0$. Denote $I = \{i > k | b_k b_i \in T_0\}$. For any $i \in I$, let W_i be the indicator variable of the event that $c(b_k b_i)$ is present at g. We then have $Z_k = \sum_{i \in I} W_i$. We now bound the probability that W_i is 1. Note that once b_k is chosen, the event $W_i = 1$ depends on the choice of b_i only. Since b_k is not bad, there are at most γn colours present at b_k which are also present at g. Moreover, there are at least $n/2$ feasible choices for b_i. Thus, $\Pr[W_i = 1] \leq \gamma n/(n/2) = 2\gamma$. It is now easy to see that, since T_0 is a forest and has at most $n/4$ edges, we can bound $Z_1 + \cdots + Z_{n/4}$ by the sum of $n/4$ independent Bernoulli variables with success probability 2γ. By this coupling argument and the classic Chernoff bound [see e.g. 3] we can use $\alpha_2 \geq 2\gamma$ from (6) to obtain:

$$\Pr[Z \geq \alpha_2 n/2] = \Pr\left[Z \geq 2E[Z]\right] \leq \exp\left(-2E[Z]^2/n\right) = o\left(n^{-1}\right). \tag{11}$$

Combining (9) and (11), we finally obtain the desired result:

$$\Pr[m_g \geq \alpha_2 n] \leq \Pr[Y + Z \geq \alpha_2 n]$$
$$\leq \Pr[Y \geq \alpha_2 n/2] + \Pr[Z \geq \alpha_2 n/2] = o\left(n^{-1}\right).$$

This concludes the proof of Claim 4.7. □

Claim 4.8. *For every vertex $g \in K_n$ we have $\Pr[w_g \geq \alpha_3 n] = o\left(n^{-1}\right)$.*

Proof. Fix some vertex $g \in K_n$. For the remainder of the proof we will assume that $m_g \leq \alpha_2 n$ and that for each colour f which is present at g we have $w_f \leq \alpha_1 n$. This is fine, since by Claims 4.6 and 4.7 and a union bound these inequalities hold with probability $1 - o\left(n^{-1}\right)$.

For this proof it will be convenient to have the following piece of notation: Recall that for each $i \leq n/4$, there is at most one index $j < i$ so that $b_j b_i \in T_0$. If such an index j exists, define $pre(i) = j$. Otherwise let $pre(i) = \emptyset$, say.

Define random variables $X_1, \ldots, X_{n/4}$ as follows: If at the kth step of RGE b_k is chosen so that $c(g b_k) \in C(T[b_1 \ldots, b_{k-1}])$, then we set $X_k = d_k$, otherwise we set $X_k = 0$.

We also define random variables $Y_1, \ldots, Y_{n/4}$: If there is a colour $f = c(b_{pre(k)} b_k)$ chosen in the kth step[2] and some index $j < k$ so that $c(g b_j) = f$, then we set $Y_k = d_j$. Otherwise we set $Y_k = 0$. We will slightly abuse notation and denote the event that "$c(g b_j) = c(b_k b_{pre(k)})$ for $j < k$" by $\{Y_k = d_j\}$, even though this is ambiguous if there is some $i \neq j$ with $d_i = d_j$.

We briefly explain why, with these definitions, we have $w_g \leq X + Y$, where $X = X_1 + \cdots + X_{n/4}$ and $Y = Y_1 + \cdots + Y_{n/4}$. Recall that, by definition, $w_g = \sum_{c(g b_i) \in C(T_0)} d_i$ unless $g \in \{b_1, \ldots, b_{n/4}\}$ in which case $w_g = 0$. For some fixed k, d_k is a summand of w_g if the colour $c(g b_k)$ lies in $C(T_0)$ — this can happen in one of two ways: If $C(g b_k) \in T[b_1, \ldots, b_{k-1}]$, then $X_k = d_k$, otherwise the colour $c(g b_k)$ is chosen at the ith step of RGE for some $i > k$ and then $Y_i = d_k$. This shows that d_k is a summand of $X + Y$ if it is a summand of w_g implying $w_g \leq X + Y$. We remark that, crucially, the colour $c(g b_k)$ cannot be chosen in the kth step of RGE itself, unless $g = b_{pre(k)}$ in which case $w_g = 0$ by definition. This is the only step that fails in our proof of Theorem 1.2 if we replace the condition that our colouring is proper (i.e. locally 1-bounded) by the condition that it is locally 3-bounded, say.

We will now derive concentration results for X and Y implying the desired result. Consider X first, we use the same line of reasoning as in Claims 4.6 and 4.7. Recall that we conditioned on the event $m_g \leq \alpha_2 n$. Thus, at the kth step of RGE there are at most $m_g \leq \alpha_2 n$ vertices

---

[2] This of course only makes sense if $pre(k) \neq \emptyset$. In case $pre(k) = \emptyset$ then no colour is chosen at the kth step and we set $Y_k = 0$.

$b \in K_n$ so that $c(gb) \in C(T[b_1, \ldots, b_{k-1}])$. Hence, $\Pr[X_k = d_k] \le \alpha_2 n/(n/2) = 2\alpha_2$. We deduce $E[X] \le 4\alpha_2 n \le \alpha_3 n/4$ (using (5)) and, by Azuma-Hoeffding and Eqs. (5) and (8),

$$\Pr[X \ge \alpha_3 n/2] \le \Pr\left[X - E[X] \ge \alpha_3 n/4\right] \tag{12}$$

$$= \exp\left(-\frac{\alpha_3^2}{64\beta}\log n\right) = o(n^{-1}). \tag{13}$$

It remains to bound Y. We start with the observation that, if $Y_k = d_i$, then there is no index $k' \ne k$ with $Y_{k'} = d_i$, since otherwise we would have $c(b_{pre(k')}b_{k'}) = c(b_{pre(k)}b_k) = c(gb_i)$ contradicting the fact that T_0 is rainbow.

For a more detailed analysis of Y, for each $1 \le k \le n/4$ we record which values Y_k can take given the choices of $b_1, \ldots, b_{k-1}$. We record these values in a list L_k. We add i to L_k if $i < k$ and the colour $c(gb_i)$ is present at $b_{pre(k)}$. Observe that, if $c(gb_i)$ is not present at $b_{pre(k)}$, then it is impossible that $c(b_{pre(k)}b_k) = c(gb_i)$ and therefore it is impossible that $Y_k = d_i$. Hence, Y_k can only take the value d_i if $i \in L_k$.

If it is the case that $i \in L_k$, then we have $\Pr[Y_k = d_i] \le 2/n$, since in the kth step RGE has at least $n/2$ feasible choices for b_k and at most one of them leads to $Y_k = d_i$. Moreover, since there are $n/2$ feasible choices and at most $k - 1 \le n/4$ values that Y_k can take, we know that for any given $i \in L_k$ the probability that $Y_k = d_i$ is at most $4/n$ even if we condition on the event $\{Y_k \ne d_j \text{ for all } j \in L_k \setminus \{i\}\}$. Thus, the bound $\Pr[Y_k = d_i] \le 4/n$ holds, independently of whether or not we know that $Y_k \ne d_j$ for some values of j (note that this is not necessarily true for the bound $\Pr[Y_k = d_i] \le 2/n$: If we know that $Y_k \ne d_j$, this means that RGE does not choose the edge of colour $c(gb_j)$ in the kth step, so that it becomes more likely that RGE chooses the edge of colour $c(gb_i)$). Hence, using a coupling argument, we can bound $Y_k \le \sum_{i \in L_k} W_{k,i}$ where the $W_{k,i}$ are mutually independent random variables with $\Pr[W_{k,i} = d_i] = 4/n$ and $\Pr[W_{k,i} = 0] = 1 - 4/n$.

We now slightly adapt our choice of $W_{k,i}$ to reflect the fact that for each i there is at most one k with $Y_k = d_i$. This will allow us to get a better bound on Y. For each $i \in L_k$ we introduce a random variable $W'_{k,i}$. These random variables are sampled in increasing order of k. If $W'_{k',i} = 0$ for all $k' < k$ (with $i \in L_{k'}$, of course), then we set $W'_{k,i} = W_{k,i}$. If, on the other hand, there is some $k' < k$ so that $W'_{k',i} = d_i$, then we set $W'_{k,i} = 0$. Recalling that for each i there is at most one k with $Y_k = d_i$, we can thus use a coupling argument to bound

$$Y \le \sum_{k=1}^{n/4} \sum_{i \in L_k} W'_{k,i}.$$

Changing the order of summation gives

$$Y \le \sum_{i=1}^{n/4} \sum_{\substack{k: \\ L_k \ni i}} W'_{k,i}.$$

Now, write $W'_i = \sum_{k:L_k \ni i} W'_{k,i}$ and $W' = \sum_{i=1}^{n/4} W'_i$. By construction, the W'_i are mutually independent and each W'_i either takes the value d_i or 0. To conclude, we will bound $\Pr[W'_i = d_i]$ and then apply Azuma-Hoeffding to bound W'. To achieve the former, note that $\Pr[W'_i = d_i] = \Pr[\exists k : W_{k,i} = d_i] \le \frac{4}{n} \cdot |\{k : L_k \ni i\}|$, where the last inequality is a union bound. It follows from the definitions of L_k and $w_{c(gb_i)}$ that for every i

$$|\{k : L_k \ni i\}| = \sum_{\substack{j : \\ c(gb_i) \text{ present at } b_j}} |\{k : pre(k) = j\}|$$

$$\le \sum_{\substack{j \\ c(gb_i) \text{ present at } b_j}} d_j = w_{c(gb_i)}.$$

Since we conditioned on the event that $w_{c(gb_i)} \le \alpha_1 n$, we obtain $\Pr[W'_i = d_i] \le 4\alpha_1$.

This implies (together with (5)) that $E[W'] \leq 8\alpha_1 n \leq \alpha_3 n/4$. Applying Azuma-Hoeffding as well as (5) and (8) one last time, we get

$$\Pr\left[W' \geq \alpha_3 n/2\right] \leq \Pr\left[W' - E[W'] \geq \alpha_3 n/4\right] \leq \exp\left(-\frac{\alpha_3^2}{64\beta}\log n\right)$$
$$= o(n^{-1}).$$

Combining this with $Y \leq W'$, $w_g \leq X + Y$ and (12) finishes the proof of Claim 4.8. $\quad\square$

Lemma 4.1 now follows from Claims 4.6–4.8 and a union bound over all n vertices and all at most $\binom{n}{2}$ colours.

4.3. Proof of Lemma 4.2

We start with the simple observation that $|A| \geq (1 - r)n$. By the statement 'n sufficiently large' in Lemma 4.2 we simply mean that $\Pi_{i=0}^3 (|A| - i) \geq ((1 - r)n)^4 /2$.

To prove Lemma 4.2 we will apply the Lopsided Lovász Local Lemma as stated in Theorem 4.4. To this end, we consider a random bijection $\sigma : W \rightarrow A$ chosen uniformly from all such bijections. We want to find σ so that $\pi \cup \sigma$ gives a rainbow embedding of T. The only obstruction to the embedding being rainbow is if two embedded edges have the same colour. We distinguish cases depending on where these edges lie — each edge either lies in T_0 or goes from B to A or lies in A. We therefore have the following kinds of 'bad events'.

- (One edge in T_0, one going from B to A) For a vertex $v \in W$ and $v_0 \in R$ with $vv_0 \in T$ and a vertex $a \in A$ with $c(\pi(v_0)a) \in C(T_0)$, it is a bad event if $\sigma(v) = a$, since then $\pi \cup \sigma$ will not be a rainbow embedding of T. We refer to this event as FR_{v,v_0}^a. Here in 'FR' the F stands for 'forbidden' indicating that the edge $\pi(v_0)a$, uses a 'forbidden' colour from $C(T_0)$. The R stands for 'root' indicating that the edge v_0v is incident to the 'roots' R.
- (One edge in T_0, one in A) For distinct vertices $v, w \in W$ and distinct vertices $a, b \in A$ with $vw \in T$ and $c(ab) \in C(T_0)$ it is a bad event if $\sigma(v) = a, \sigma(w) = b$. We refer to this event as $F_{v,w}^{a,b}$.
- (Both edges from B to A) If there are distinct vertices $v, w \in W$ and $v_0, w_0 \in R$ so that $v_0v, w_0w \in T$ and if there are distinct vertices $a, b \in A$ with $c(\pi(v_0)a) = c(\pi(w_0)b)$, then it is a bad event if $\sigma(v) = a, \sigma(w) = b$. We refer to this event as $SRR_{v,v_0,w,w_0}^{a,b}$. Here the 'S' indicates that two edges which have the 'same' colour are chosen and the 'RR' indicates that both these edges are incident to R.
- (One edge from B to A and one in A) If there are distinct vertices $v, w, x \in W$ and $v_0 \in R$ so that $v_0v, wx \in T$ and if there are distinct vertices, $a, b, d \in A$ with $c(\pi(v_0)a) = c(bd)$, then it is a bad event if $\sigma(v) = a, \sigma(w) = b, \sigma(x) = d$. We refer to this event as $SR_{v,v_0,w,x}^{a,b,d}$.
- (Both edges in A) If there are distinct vertices $v, w, x, y \in W$ so that $vw, xy \in T$ and if there are vertices $a, b, d, e \in A$ so that $c(ab) = c(de)$, then it is a bad event if $\sigma(v) = a, \sigma(w) = b, \sigma(x) = d, \sigma(y) = e$. We refer to this event as $S_{v,w,x,y}^{a,b,d,e}$.

Notation. For an element $x \in A \cup W$ we write $x \in FR_{v,v_0}^a$ if $x \in \{v, a\}$. For the other bad events, we will use an analogous notation. For two bad events C, D we write $C \sim D$ if there is an $x \in A \cup W$ with $x \in C$ and $x \in D$.

Remark. Observe that, for example, the events $F_{v,w}^{a,b}$ and $F_{w,v}^{b,a}$ are identical. In the arguments below, we will count events $F_{v,w}^{a,b}$ that have $x \in F_{v,w}^{a,b}$ for some fixed $x \in W \cup A$. In these counting arguments we will count each event only once. If for example $x \in W$, we will count the events with labels $F_{x,w}^{a,b}$, but not those with labels $F_{w,x}^{a,b}$. The case $x \in A$ is treated similarly. This remark of counting each event only once even if it has multiple 'labels' also applies to the other kinds of bad events, that is to the events $SRR_{v,v_0,w,w_0}^{a,b}$, $SR_{v,v_0,w,x}^{a,b,d}$ and $S_{v,w,x,y}^{a,b,d,e}$.

Observe that all these events are indeed canonical and that our notation $C \sim D$ for bad events matches the notation from Theorem 4.4. Finally note that, if none of the bad events occurs, then the embedding $(\pi \cup \sigma)(T)$ is indeed rainbow. Denote the union of all these bad events by $\mathcal{B}$. Our aim is to apply Theorem 4.4 in order to show that the embedding $(\pi \cup \sigma)(T)$ is rainbow with positive probability. Thus, to prove Lemma 4.2 it suffices to show that for all $C \in \mathcal{B}$ we have

$$\sum_{\substack{C' \in \mathcal{B} \\ C' \sim C}} \Pr[C'] \leq \frac{1}{4}. \tag{14}$$

Proving (14) is a bit cumbersome but not hard, as long as we count carefully. We introduce some notation and outline the straightforward approach to prove (14) before giving the details. We will define functions $fr, f, srr, sr, s : A \cup W \to \mathbb{R}$. For example $fr(x)$ is defined as

$$fr(x) = \sum_{x \in FR^a_{v,v_0}} \Pr[FR^a_{v,v_0}].$$

The other functions are defined analogously so that $f(x)$ features bad events of type F rather than of type FR, so that $srr(x)$ features bad event of type SRR and so on. We will bound each of these functions separately, this is done in Claims 4.9–4.13. This is where we will use our assumptions on w_f, m_g, w_g and $d(w) \leq D$ for $w \in W$.

Having bounded $f(x), fr(x), \ldots$, we will use the following simple bound for each $C \in \mathcal{B}$ to conclude.

$$\sum_{\substack{C' \in \mathcal{B} \\ C' \sim C}} \Pr[C'] \leq \sum_{x \in C} \Big(fr(x) + f(x) + srr(x) + sr(x) + s(x) \Big)$$

$$\leq 4 \Big(\max_{v \in W} \big(fr(v) + f(v) + srr(v) + sr(v) + s(v) \big) \tag{15}$$

$$+ \max_{a \in A} \big(fr(a) + f(a) + srr(a) + sr(a) + s(a) \big) \Big).$$

We now give the details of bounding $fr(x), f(x), srr(x), sr(x), s(x)$.

Claim 4.9. *We have*

$$fr(v) \leq D\chi(1-r)^{-1} \quad \text{for all } v \in W \text{ and}$$
$$fr(a) \leq \chi(1-r)^{-1} \quad \text{for all } a \in A.$$

Proof. Note that events FR^a_{v,v_0} occur with probability at most $((1-r)n)^{-1}$.

First, we count the number of events FR^a_{v,v_0} where v is fixed: Since $d(v) \leq D$ there are at most D ways to choose $v_0 \in R$ so that $v_0 v \in T$. For a fixed choice of v_0 there are at most $|\{a \in A : c(\pi(v_0)a) \in C(T_0)\}| = m_{\pi(v_0)} \leq \chi n$ choices of $a \in A$ so that $c(\pi(v_0)a) \in C(T_0)$. Thus, all in all there are at most $D \cdot \chi n$ events of the form FR^a_{v,v_0}. Hence $fr(v) \leq D\chi(1-r)^{-1}$.

Next, we count the number of events FR^a_{v,v_0} where a is fixed: We first choose v_0 so that $c(\pi(v_0)a) \in C(T_0)$. Given the choice of v_0, we can choose v to be any one of the neighbours of v_0 that lie in W. There are at most $d(v_0)$ such neighbours. Summing over all v_0 with $c(\pi(v_0)a) \in C(T_0)$ we obtain a bound of $\sum_{\substack{v_0 \in R \\ c(a\pi(v_0)) \in C(T_0)}} d_T(v_0) = w_a \leq \chi n$ possible choices . Hence, $fr(a) \leq \chi(1-r)^{-1}$. $\square$

Claim 4.10. *We have*

$$f(v) \leq 4D\alpha r(1-r)^{-2} \quad \text{for all } v \in W \text{ and}$$
$$f(a) \leq 4\chi(1-r)^{-2} \quad \text{for all } a \in A.$$

Proof. The probability of an event $F^{a,b}_{v,w}$ is at most $\Pi^1_{i=0}((1-r)n - i)^{-1} \leq 2((1-r)n)^{-2}$, where we use that n is sufficiently large.

Fix some $v \in W$. We first count the number of events $F_{v',w}^{a,b} \ni v$. Without loss of generality assume $v = v'$. Since $d(v) \le D$, there are at most D choices of w. There are at most $\alpha n \cdot rn$ edges in K_n whose colour lies in $C(T_0)$, so that there are at most $2\alpha rn^2$ choices for the ordered pair a, b so that $c(ab) \in C(T_0)$. All in all, there are at most $D \cdot 2\alpha rn^2$ events of the form $F_{v,w}^{a,b}$ implying $f(v) \le 4D\alpha r(1-r)^{-2}$.

Now fix $a \in A$. We count the number of events $F_{v,w}^{a',b} \ni a$. Without loss of generality assume $a = a'$. There are at most $|\{b \in A : c(ab) \in C(T_0)\}| = m_a \le \chi n$ ways to choose b and, since there are at most n edges in T, there are at most $2n$ ways to choose the (ordered) pair v, w implying $f(a) \le 4\chi(1-r)^{-2}$.   □

Claim 4.11.   *We have*

$$srr(v) \le 2D\chi(1-r)^{-1} \quad \textit{for all } v \in W \text{ and}$$
$$srr(a) \le 2\chi(1-r)^{-2} \quad \textit{for all } a \in A.$$

Proof. The probability of an event $SRR_{v,v_0,w,w_0}^{a,b}$ is at most $\Pi_{i=0}^1 ((1-r)n - i)^{-1} \le 2((1-r)n)^{-2}$.

Fix $v \in W$. We count the number of events $SRR_{v',v_0,w,w_0}^{a,b} \ni v$. Without loss of generality assume $v = v'$. Since $d(v) \le D$, there are at most D choices of v_0. Now, there are at most $(1-r)n$ choices for a. Given the choices of v_0 and a, write $e = \pi(v_0)a$. We know that the edge $\pi(w_0)b$ has to have colour $c(e)$. We must choose w_0 so that $c(e)$ is present at w_0. Given such a w_0 there is at most one choice of b so that $c(\pi(w_0)b) = c(e)$ and there are at most $d(w_0)$ choices for w. Summing over the possible choices of w_0 we see that there are at most $\sum_{\substack{w_0 \in R \\ c(e) \text{ present at } \pi(w_0)}} d(w_0) = w_{c(e)} \le \chi n$ possible choices for the triple w_0, w, b. All in all, we can bound the number of events $SRR_{v,v_0,w,w_0}^{a,b} \ni v$ by $D(1-r)n \cdot \chi n$ implying $f(v) \le 2D\chi(1-r)^{-1}$.

Now fix $a \in A$ and count the number of events $SRR_{v,v_0,w,w_0}^{a',b} \ni x$. Without loss of generality assume $a = a'$. There are at most n ways to choose a pair v, v_0 since each such pair corresponds to an edge of T. Then, just as before there are at most χn ways to choose the triple w_0, w, b. Thus, $srr(a) \le 2\chi(1-r)^{-2}$.   □

Claim 4.12.   *We have*

$$sr(v) \le 8D\alpha(1-r)^{-2} \quad \textit{for all } v \in W \text{ and}$$
$$sr(a) \le 4\alpha(1-r)^{-3} + 4\chi(1-r)^{-2} \quad \textit{for all } a \in A.$$

Proof. The probability of an event $SR_{v,v_0,w,x}^{a,b,d}$ is at most $\Pi_{i=0}^2 ((1-r)n - i)^{-1} \le 2((1-r)n)^{-3}$.

As usual we count the number of $SR_{v',v_0,w,x}^{a,b,d} \ni v$ for a fixed $v \in W$. Here we need to distinguish two cases, either $v = v'$ or $v \in \{w, x\}$. Consider the case $v = v'$ first. As before, there are at most D ways to choose v_0, then there are at most $(1-r)n$ ways to choose a. We now know that b, d have to be chosen in a way so that $c(bd) = c(\pi(v_0)a)$. Since there are at most αn edges of colour $c(\pi(v_0)a)$ this can be done in at most $2\alpha n$ ways, here we count the number of ordered pairs b, d. Finally, there are at most n ways to choose the unordered pair w, x. All in all, we have at most $D \cdot (1-r)n \cdot 2\alpha n \cdot n = 2D\alpha(1-r)n^3$ choices. We now consider the case where $v \in \{w, x\}$. Assume $v = w$ without loss of generality. We have at most D choices for x and at most n choices for the pair v_0, v' and another at most $(1-r)n$ choices for a. Just as before, we are left with at most $2\alpha n$ choices for the ordered pair b, d so that $c(bd) = c(\pi(v_0)a)$. In total, we have $D \cdot n \cdot (1-r)n \cdot 2\alpha n = 2D\alpha(1-r)n^3$ choices. Combining the two cases $v = v'$ and $v = w$, we obtain $sr(v_r) \le 8D\alpha(1-r)^{-2}$.

We now count $SR_{v,v_0,w,x}^{a',b,d} \ni a$ for a fixed $a \in A$. We first consider the case where $a = a'$ and then the case where $a \in \{b, d\}$. So let $a = a'$. There are at most n choices for the pair v, v_0 and by the same argument as before, there are at most $n \cdot 2\alpha n$ ways to choose w, x and b, d. In total, we have at most $2\alpha n^3$ choices. We now consider the case $a \in \{b, d\}$. Assume $a = b$ without loss of generality. There at most $2n$ ways to choose ordered w, x and at most $(1-r)n$ ways to choose $d \ne b$. Once these choices have been made, we have to pick v_0, v, a' so that $c(\pi(v_0)a') = c(bd)$. By

a familiar argument (see proof of Claim 4.11), this can be done in at most $w_{c(bd)} \leq \chi n$ ways. In total we have at most $2n \cdot (1-r)n \cdot \chi n$ possible choices. Combining the cases $a = a'$ and $a = b$ we obtain $sr(x) \leq 4\alpha(1-r)^{-3} + 4\chi(1-r)^{-2}$. $\quad \square$

Claim 4.13. *We have*

$$s(v) \leq 4D\alpha(1-r)^{-2} \quad \text{for all } v \in W \text{ and}$$
$$s(a) \leq 8\alpha(1-r)^{-3} \quad \text{for all } a \in A.$$

Proof. The probability of an event $S_{v,w,x,y}^{a,b,d,e}$ is at most $\Pi_{i=0}^{3}((1-r)n - i)^{-1} \leq 2((1-r)n)^{-4}$.

We start by counting the number of events $S_{v',w,x,y}^{a,b,d,e} \ni v$ for a fixed $v \in W$. Without loss of generality assume $v = v'$. There are at most D choices for w and at most n choices for the unordered pair x, y. There are at most $((1-r)n)^2$ choices for the ordered pair a, b. Given the choice of a, b we have at most $2\alpha n$ choices for the ordered pair d, e. In total we have at most $D \cdot n \cdot ((1-r)n)^2 \cdot 2\alpha n$ so that $s(v) \leq 4D\alpha(1-r)^{-2}$.

We now count the number of events $S_{v,w,x,y}^{a',b,d,e} \ni a$ for a fixed $a \in A$. Without loss of generality assume $a = a'$. We have at most $4n^2$ possibilities to choose the ordered pairs v, w and x, y, at most $(1-r)n$ choices for b and at most αn choices for the unordered pair d, e giving a total of $4n^2 \cdot (1-r)n \cdot \alpha n$ choices so that $s(a) \leq 8\alpha(1-r)^{-3}$. $\quad \square$

We now plug in the bounds from Claims 4.9–4.13 into (15). To simplify, we estimate $(1-r)^{-1} \leq (1-r)^{-2} \leq (1-r)^{-3}$ as well as $r \leq 1$. Altogether, this turns into (also recalling (7)):

$$\sum_{\substack{C' \in \mathcal{B} \\ C' \sim C}} \Pr[C'] \leq 4(1-r)^{-3}\big(\chi(3D + 11) + \alpha(16D + 12)\big) \leq 1/4$$

finishing the proof of Lemma 4.2.

4.4. Proof of Theorem 1.2

We now combine the previous results to prove Theorem 1.2.

Recall that $d_1, \ldots, d_{n/4}$ are the $n/4$ highest degrees of T, and $d_{n/4+1}, \ldots, d_n$ are the $3n/4$ lowest degrees of T. We start by observing that for every $i > n/4$ we have $d_i \leq 4$. To see this, observe $\sum_{i=1}^{n} d_i < 2n$, moreover $d_i \geq 1$ for all i so that $\sum_{i=1}^{n/4} d_i < 5n/4$. Thus the average degree of the vertices $v_1, \ldots, v_{n/4}$ is less than 5, which implies that $d_i \leq 4$ for all $i > n/4$.

Next, we choose $\alpha = \beta = 2^{-38}, \alpha_1 = 2^{-16}, \alpha_2 = 2^{-15}, \alpha_3 = 2^{-11}, \gamma = 2^{-16}, \delta = 2^{-19}$ and observe that the conditions of Lemma 4.1 are satisfied.

We now apply Lemma 4.1 to obtain a partial embedding of the $n/4$ vertices $v_1, \ldots, v_{n/4}$ of T with highest degrees. Now take $\chi = \alpha_3$. It then follows from Lemma 4.1 that the partial embedding satisfies $m_a \leq \chi n, w_a \leq \chi n$ for each $a \in V(K_n)$ and $w_f \leq \chi n$ for every $f \in C(K_n)$. Set $r = 1/4$ and $D = 4$. It is easily checked that all the conditions of Lemma 4.2 are satisfied so that we can find a rainbow embedding of T into K_n. This finishes the proof of Theorem 1.2.

5. Concluding remarks

In this section we make some remarks and present open questions, some of which are directly related to our results and some of which are inspired by them and might lead to further research.

Let us begin by describing some constructions of colourings of K_n and n-vertex trees T, so that K_n does not contain a rainbow copy of T. These constructions show that in order to find rainbow embeddings of n-vertex trees into K_n, we need stronger conditions than the colouring to just be proper. Specifically, we cannot simply drop the condition of Theorem 1.2, that the colouring be globally αn-bounded for some $\alpha < 1/2$.

We start with the observation that a 1-factorisation of K_{2n} contains precisely $2n - 1$ colours, which equals the number of edges of a spanning tree. Thus, any rainbow spanning tree within a 1-factorisation has to use each colour precisely once.

The first construction comes from Maamoun and Meyniel [25], who give a colouring of K_n for $n = 2^k$ which does not contain a rainbow Hamilton path. We adapt their argument to prove the following result.

Proposition 5.1. *For $n = 2^k$, there is a 1-factorisation of K_n which does not contain a rainbow spanning tree in which all but precisely two degrees are odd. The same 1-factorisation does not contain a rainbow Hamilton path.*

Proof. Suppose $n = 2^k$ and label the vertices of K_n by elements of the finite group $\mathbb{Z}_2^k$. For two distinct vertices $a, b \in V(K_n)$ colour the edge ab by the colour $a + b$, so that the set of colours used is precisely $\mathbb{Z}_2^k \setminus \{0\}$. It is not hard to see that this colouring is a 1-factorisation. Observe also that the sum over all colours is 0. Thus, when we sum over the colours of the edges of a rainbow spanning rainbow tree we obtain 0. Now suppose T is a spanning tree, embedded into K_n. Summing over all colours of the edges of T and reordering, we obtain

$$\sum_{ab \in E(T)} c(ab) = \sum_{ab \in E(T)} (a + b) = \sum_{a \in V(T)} d_T(a)a.$$

If T is a tree in which all but precisely two degrees are odd, say $x, y \in V(K_n)$ are the vertices with even degree, then

$$\sum_{ab \in E(T)} c(ab) = \left(\sum_{a \in V(T)} a \right) - x - y = x + y.$$

This expression never equals zero, no matter how we choose x and y. This contradicts the observation made above, that the sum should be 0. The same argument works if we replace the condition of all but two vertices having odd degree by the condition of all but two vertices having even degree. The only spanning tree fulfilling the latter is a Hamilton path and in this case, our argument is that of Maamoun and Meyniel [25]. ☐

We also show that the colouring from the proof above is not the only colouring which does not allow for a rainbow embedding of certain trees.

Proposition 5.2. *For every $n = 2k$, there is a tree on n vertices which has no rainbow embedding into any 1-factorised K_n.*

Proof. Let S be a tree on $2k$ vertices which has two distinct vertices x, y with the following properties. Firstly, x is not adjacent to y in S. Secondly every edge in S is incident to x or y.

There are many trees S satisfying these two properties. Any such tree consists of two stars rooted and x and y overlapping in one vertex.

Let c be a 1-factorisation of K_n and suppose we are given an embedding of S into K_n. We abuse notation to denote by x, y not only the vertices of S, but also their images in K_n. In our embedding, no edge has colour $c(xy)$, since the edge xy is absent in S and since every edge of S is adjacent to x or y and therefore has a colour different from xy. Thus, the embedding does not use each colour of K_n and cannot be rainbow. ☐

The above discussion shows that the "global αn-boundedness" condition in Theorem 1.2 cannot be completely removed. In contrast to this we expect that the assumption on the maximum degree of T in Theorem 1.2 is unnecessary—we believe that there is a constant $\alpha > 0$ so that any proper, globally αn-bounded colouring of K_n contains a rainbow copy of *every* n-vertex tree T. As evidence for this, we can prove that there is a small constant α so that any globally $\alpha n / \log^3 n$-bounded colouring contains a rainbow copy of every n-vertex tree [see 7].

Another direction one can take is to see how close we can come to embedding spanning trees into a properly coloured K_n without further conditions on the colouring. More precisely, one can ask whether there is a function $f(n) = o(n)$ so that any tree on $n - f(n)$ vertices has a rainbow

embedding into any properly coloured K_n. This would for example generalise the result by Alon, Pokrovskiy and Sudakov [2] about almost spanning paths in properly coloured K_n.

Finally, Theorem 1.1 invites to thinking about whether proper colourings of certain tournaments (e.g. regular tournaments) contain long rainbow cycles/paths. There is a great amount of conjectures and results in the area of Hamilton cycles in regular tournaments, see e.g. [22] and more recent results by the same authors, but so far little effort has gone into investigating rainbow structures in this setting.

Acknowledgement

References

[1] M. Albert, A. Frieze, B. Reed, Multicolored Hamilton cycles, Electron. J. Combin. 1 (62) (1995) Research paper R10.

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

[3] N. Alon, J.H. Spencer, The Probabilistic Method, in: Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, 2008, xv, 352.

[4] L. Andersen, Hamilton circuits with many colours in properly edge-coloured complete graphs, Math. Scand. 64 (1989) 5–14.

[5] K. Azuma, Weighted sums of certain dependent random variables, Tôhoku Math. J. 3 (19) (1967) 357–367.

[6] J. Balogh, T. Molla, Long rainbow cycles and Hamiltonian cycles using many colors in properly edge-colored complete graphs, European J. Combin. 79 (2017) http://dx.doi.org/10.1016/j.ejc.2019.02.008.

[7] F. Benzing, Long Directed Rainbow Cycles and Rainbow Spanning Trees (Masters thesis), ETH Zürich, 2017.

[8] B. Bollobás, P. Erdős, Alternating Hamiltonian cycles, Israel J. Math. 23 (1976) 126–131.

[9] J. Böttcher, Y. Kohayakawa, A. Procacci, Properly coloured copies and rainbow copies of large graphs with small maximum degree, Random Struct. Algorithms 40 (4) (2012) 425–436, http://dx.doi.org/10.1002/rsa.20383.

[10] R.A. Brualdi, H.J. Ryser, Combinatorial Matrix Theory, Cambridge University Press, Cambridge, UK, 1991, p. 255.

[11] P. Erdős, L. Lovász, Problems and results on 3-chromatic hypergraphs and some related questions, Infin. Finite Sets 2 (10) (1975) 609–627.

[12] P. Erdős, V. Nešetřil, V. Rödl, On Some Problems Related to Partitions of Edges of a Graph, Graphs and Other Combinatorial Topics, in: Teubner-Texte Math., vol. 59, Teubner, Leipzig, 1983, pp. 54–63.

[13] P. Erdős, R. Rado, A combinatorial theorem, J. Lond. Math. Soc. 25 (1950) 249–255.

[14] P. Erdős, J. Spencer, Lopsided Lovász Local Lemma and Latin transversals, Discrete Appl. Math. 30 (2–3) (1991) 151–154, http://dx.doi.org/10.1016/0166-218X(91)90040-4.

[15] A. Frieze, M. Krivelevich, On packing Hamilton cycles in ϵ-regular graphs, J. Combin. Theory Ser. B 94 (1) (2005) 159–172, http://dx.doi.org/10.1016/j.jctb.2004.12.003.

[16] A. Frieze, B. Reed, Polychromatic Hamilton cycles, Discrete Math. 118 (1–3) (1993) 69–74, http://dx.doi.org/10.1016/0012-365X(93)90054-W.

[17] A. Gyárfás, G.N. Sárközy, Rainbow matchings and cycle-free partial transversals of Latin squares, Discrete Math. 327 (1) (2014) 96–102, http://dx.doi.org/10.1016/j.disc.2014.03.010.

[18] G. Hahn, Un jeu de colouration, Actes Colloq. Cerisy 12 (1980) 18.

[19] G. Hahn, C. Thomassen, Path and cycle sub-ramsey numbers and an edge-colouring conjecture, Discrete Math. 62 (1) (1986) 29–33, http://dx.doi.org/10.1016/0012-365X(86)90038-5.

[20] P. Hatami, P.W. Shor, A lower bound for the length of a partial transversal in a Latin square, J. Combin. Theory Ser. A 115 (7) (2008) 1103–1113, http://dx.doi.org/10.1016/j.jcta.2008.01.002.

[21] W. Hoeffding, Probability inequalities for sums of bounded random variables, J. Amer. Statist. Assoc. 58 (301) (1963) 13–30, http://dx.doi.org/10.1080/01621459.1963.10500830, arXiv:0603030.

[22] D. Kühn, D. Osthus, A survey on Hamilton cycles in directed graphs, European J. Combin. 33 (5) (2012) 750–766, http://dx.doi.org/10.1016/j.ejc.2011.09.030, arXiv:1006.0590.

[23] A. Lo, Properly colored Hamiltonian cycles in edge-colored complete graphs, Combinatorica (2012) 1–22.

[24] L. Lu, L. Székely, Using Lovasz Local Lemma in the space of random injections, Electron. J. Combin. 1 (14) (2007) Research Paper R63.

[25] M. Maamoun, H. Meyniel, On a problem of G. Hahn about coloured Hamiltonian paths in K2t, Discrete Math. 51 (2) (1984) 213–214, http://dx.doi.org/10.1016/0012-365X(84)90073-6.

[26] M. Molloy, B. Reed, Graph Coloring and the Probabilistic Method, Springer-Verlag, Berlin, 2002.

[27] R. Montgomery, personal communication, 2017.

[28] H.J. Ryser, Neuere Probleme der Kombinatorik, in: Vorträge Über Kombinatorik Oberwolfach, 1967, pp. 69–91.

[29] S.K. Stein, Transversals of Latin squares and their generalizations, Pacific J. Math. 59 (1975) 567–575.

[30] B. Sudakov, J. Volec, Properly colored and rainbow copies of graphs with few cherries, J. Combin. Theory Ser. B 122 (2017) 391–416, http://dx.doi.org/10.1016/j.jctb.2016.07.001, arXiv:1504.06176.

[31] I. Wanless, Transversals in latin squares: a survey, in: R. Chapman (Ed.), Surveys in Combinatorics 2011, in: London Mathematical Society Lecture Note Series, Cambridge University Press, 2011, pp. 403–437, http://dx.doi.org/10.1017/CBO9781139004114.010.