Random subgraphs of properly edge-coloured complete graphs and long rainbow cycles

Noga Alon* Alexey Pokrovskiy† Benny Sudakov ‡

Abstract

A subgraph of an edge-coloured complete graph is called rainbow if all its edges have different colours. In 1980 Hahn conjectured that every properly edge-coloured complete graph K_n has a rainbow Hamiltonian path. Although this conjecture turned out to be false, it was widely believed that such a colouring always contains a rainbow cycle of length almost n. In this paper, improving on several earlier results, we confirm this by proving that every properly edge-coloured K_n has a rainbow cycle of length $n - O(n^{3/4})$. One of the main ingredients of our proof, which is of independent interest, shows that the subgraph of a properly edge-coloured K_n formed by the edges a random set of colours has a similar edge distribution as a truly random graph with the same edge density. In particular it has very good expansion properties.

1 Introduction

In this paper we study properly edge-coloured complete graphs, i.e., in which edges which share a vertex have distinct colours. Properly edge-coloured complete graphs are important objects because they generalize 1-factorizations of complete graphs. A 1-factorization of K_{2n} is a proper edge-colouring of K_{2n} with $2n-1$ colours, or equivalently a decomposition of the edges of K_{2n} into perfect matchings. These factorizations were introduced by Kirkman more than 150 years ago and were extensively studied in the context of combinatorial designs (see, e.g., [9, 13] and the references therein.)

A *rainbow* subgraphs of a properly edge-coloured complete graph is a subgraph all of whose edges have different colours. One reason to study such subgraphs come from the Ramsey theory, more precisely the canonical version of Ramsey's theorem, proved by Erdős and Rado. Here the goal is to show that edge-colourings of K_n, in which each colour appears only few times contain rainbow copies of certain graphs (see, e.g., [12], Introduction for more details). Another motivation comes from problems in design theory. For example a special case of the Brualdi-Stein Conjecture about transversals in Latin squares is that every 1-factorization of K_{2n} has rainbow subgraph with $2n-1$ edges and maximum degree 2. A special kind of graph with maximum degree 2 is a Hamiltonian path i.e. a path which goes through every vertex of G exactly once. Since properly coloured complete graphs are believed to contail large rainbow maximum degree 2 subgraphs, it is natural to ask whether they have rainbow Hamiltonian paths as well. This was conjectured by Hahn [8] in 1980.

Conjecture 1.1. *For $n \geq 5$, every properly coloured K_n has a rainbow Hamiltonian path.*

*Sackler School of Mathematics and Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel. Email: nogaa@tau.ac.il. Research supported in part by a USA-Israeli BSF grant 2012/107, by an ISF grant 620/13 and by the Israeli I-Core program.

†Department of Mathematics, ETH, 8092 Zurich, Switzerland. dr.alexey.pokrovskiy@gmail.com. Research supported in part by SNSF grant 200021-149111.

‡Department of Mathematics, ETH, 8092 Zurich, Switzerland. benjamin.sudakov@math.ethz.ch. Research supported in part by SNSF grant 200021-149111.

It turns out that this conjecture is false and for $n = 2^k$, Maamoun and Meyniel [11] found 1-factorizations of K_n without rainbow Hamilton cycle. Nevertheless, it is widely believed (see e.g., [7]) that the intuition behind Hahn's Conjecture is correct and that various slight weakenings of this conjecture should be true. Moreover, they should hold not only for 1-factorizations but in general for proper edge-colourings. Hahn and Thomassen [10] suggested that every properly edge-coloured K_n, with $< n/2$ edges of each colour, has a rainbow Hamiltonian path. Akbari, Etesami, Mahini, Mahmoody [1] conjectured that every 1-factorization of K_n contains a Hamiltonian cycle which has at least $n - 2$ different colours on its edges. They also asked whether every 1-factorization has a rainbow cycle of length at least $n-2$. Andersen [3] conjectured that every proper edge-colouring of K_n has a rainbow path of length at least $n - 2$.

There have been many positive results supporting the above conjectures. By a trivial greedy argument every properly coloured K_n has a rainbow path of length $\geq n/2 - 1$. Indeed if the maximum rainbow path P in such a complete graph has length less than $n/2 - 1$, then an endpoint of P must have an edge going to $V(K_n) \setminus V(P)$ in a colour which is not present in P (contradicting the maximality of P.) Akbari, Etesami, Mahini, Mahmoody [1] showed that every properly coloured K_n has a rainbow cycle of length $\geq n/2 - 1$. Gyárfás and Mhalla [6] showed that every 1-factorization of K_n has a rainbow path of length $\geq (2n + 1)/3$. Gyárfás, Ruszinkó, Sárközy, and Schelp [7] showed that every properly coloured K_n has a rainbow cycle of length $\geq (4/7 - o(1))n$. Gebauer and Mousset [5], and independently Chen and Li [4] showed that every properly coloured K_n has a rainbow path of length $\geq (3/4 - o(1))n$. But despite all these results Gyárfás and Mhalla [6] remarked that "presently finding a rainbow path even with $n - o(n)$ vertices is out of reach."

In this paper we improve on all the above mentioned results by showing that every properly edge-coloured K_n has an almost spanning rainbow cycle.

Theorem 1.2. *For all sufficiently large n, every properly edge-coloured K_n contains a rainbow cycle of length at least $n - 24n^{3/4}$.*

This theorem gives an approximate version of Hahn's and Andersen's conjectures, leaving as an open problem to pin down the correct order of the error term (currently between -1 and $-O(n^{3/4})$). The constant in front of $n^{3/4}$ can be further improved and we make no attempt to optimize it.

The proof of our main theorem is based on the following result, which has an independent interest. For a graph G and two sets $A, B \subseteq V(G)$, we use $e_G(A, B)$ to denote the number of edges of G with one vertex in A and one vertex in B. We show that the subgraph of a properly edge-coloured K_n formed by the edges in a random set of colours has a similar edge distribution as a truly random graph with the same edge density. Here we assume that n is sufficiently large and write $f \gg g$ if f/g tends to infinity with n.

Theorem 1.3. *Given a proper edge-colouring of K_n, let G be a subgraph obtained by choosing every colour class randomly and independently with probability $p \leq 1/2$. Then, with high probability, all vertices in G have degree $(1 - o(1))np$ and for every two disjoint subsets A, B with $|A|, |B| \gg (\log n/p)^2$, $e_G(A, B) \geq (1 - o(1))p|A||B|$.*

Our proof can be also used to show that this conclusion holds for not necessarily disjoint sets. For larger sets A, B of size $\gg \log^2 n/p^4$ we can also obtain a corresponding upper bound, showing that $e_G(A, B) \leq (1 - o(1))px^2$ (see remark in the next section). Note that the edges in the random subgraph G are highly correlated. Specifically, the edges of the same colour are either all appear or all do not appear in G. Yet we show that the edge distribution between the sufficiently large sets are not affected much by this dependence.

Notation

For two disjoint sets of vertices A and B, we use $E(A, B)$ to denote the set of edges between A and B. A *path forest* $\mathcal{P} = \{P_1, \dots, P_k\}$ is a collection of vertex-disjoint paths in a graph. For a path forest $\mathcal{P}$, let $V(\mathcal{P}) = V(P_1) \cup \dots \cup V(P_k)$ denote the vertices of the path forest, and let $E(\mathcal{P}) = E(P_1) \cup \dots \cup E(P_k)$ denote the edges of the path forest. We'll use additive notation for concatenating paths i.e. if $P = p_1p_2 \dots p_i$ and $Q = q_1q_2 \dots q_j$ are two vertex-disjoint paths and p_iq_1 is an edge, then we let $P + Q$ denote the path

$p_1p_2\dots p_iq_1q_2\dots q_j$. For a graph G and a vertex v, $d_G(v)$ denotes the number of edges in G containing v. The minimum and maximum degrees of G are denoted by $\delta(G)$ and $\Delta(G)$ respectively. For the sake of clarity, we omit floor and ceiling signs where they are not important.

2 Random subgraphs of properly coloured complete graphs

The goal of this section is to prove that the edges from a random collection of colours in a properly edge-coloured complete graph have distribution similar to truly random graph of the same density. Our main result here will be Theorem 1.3. Throughout this section we assume that the number of vertices n is sufficiently large and all error terms $o(1)$ tend to zero when n tends to infinity. We say that some probability event holds *almost surely* if its probability is $1-o(1)$. First we need to recall the well known Chernoff bound, see e.g., [2].

Lemma 2.1. *Let X be binomial random variable with parameters (n,p). Then for $\varepsilon \in (0,1)$ we have*

$$\mathbb{P}\big(|X-pn| > \varepsilon pn\big) \le 2e^{-\frac{pn\varepsilon^2}{3}}.$$

Given a proper edge-colouring of K_n, we call a pair of disjoint subsets A, B *nearly-rainbow* if the number of colours of edges between A and B is at least $(1-o(1))|A||B|$. The following lemma shows that we can easily control the number of random edges inside nearly-rainbow pairs.

Lemma 2.2. *Given a proper edge-colouring of K_n , let G be a subgraph of K_n obtained by choosing every colour class with probability p. Then, almost surely, all nearly-rainbow pairs A, B with $|A| = |B| = y \gg \log n/p$ satisfy $e_G(A,B) \ge (1-o(1))py^2$*

Proof. Since $y \gg \log n/p$, we can choose $\varepsilon = o(1)$ so that every nearly-rainbow pair A, B has $(1-\varepsilon/2)|A||B|$ colours, and $\varepsilon^2 y \ge 30 \log n/p$. Let (A, B) be a nearly-rainbow pair with $|A| = |B| = y$. Then the number of different colours in G between A and B is binomially distributed with parameters (m, p), where $m \ge (1-\varepsilon/2)y^2$ is the number of colours in between A and B in K_n. Since $e_G(A, B)$ is always at least the number of colours between A and B, Lemma 2.1 implies

$$\mathbb{P}\big(e_G(A,B) \le (1-\varepsilon)py^2\big) \le e^{-\varepsilon^2 py^2/13}.$$

Since $\varepsilon^2 y \ge 30 \log n/p$, the result follows by taking a union bound over all $\binom{n}{y}^2$ pairs of sets A, B of size y and all $y \le n/2$. □

Our next lemma shows that we can partition any pair of sets A, B into few parts such that almost all pairs of parts are nearly-rainbow.

Lemma 2.3. *Let A, B be two subsets of size x of properly edge-coloured K_n and let y satisfy $x \gg y^2$. Then there are partitions of A and B into sets $\{A_i\}$ and $\{B_j\}$ of size y such that all but an $o(1)$ fraction of pairs A_i, B_j are nearly-rainbow.*

Proof. Since $x \gg y^2$ for $\varepsilon = o(1)$ we can assume that $x \ge \varepsilon^{-2}y^2$ and x is divisible by y. To prove the lemma, we will show that there are partitions of A and B into sets $\{A_i\}$ and $\{B_j\}$ of size y such that all but an ε fraction of pairs A_i, B_j are nearly-rainbow.

Consider pair of random subsets $S \subset A$ and $T \subset B$ of size y chosen uniformly at random from all such subsets. For every colour c, let E_c be the set of edges of colour c between A and B. Given any two vertices $a, a' \in A$ notice that $\mathbb{P}(a \in S) = y/x$ and $\mathbb{P}(a, a' \in S) = \frac{y(y-1)}{x(x-1)}$. The same estimates hold for vertices in $T \subseteq B$. This implies that for two disjoint edges ab and $a'b'$ between A and B have $\mathbb{P}(ab \in E(S,T)) = y^2/x^2$ and $\mathbb{P}(ab, a'b' \in E(S,T)) = \frac{y^2(y-1)^2}{x^2(x-1)^2}$. Also note that $|E_c| \le x$, since the edge-colouring on K_n is proper.

Thus, by the inclusion-exclusion formula we can bound the probability that a colour c is present in $E(S,T)$ as follows

$$\begin{aligned}\mathbb{P}(c \text{ present in } E(S,T)) &\geq \sum_{e\in E_c}\mathbb{P}(e\in E(S,T)) - \sum_{e,f\in E_c}\mathbb{P}(e,f\in E(S,T)) \geq \frac{y^2}{x^2}|E_c| - \frac{y^2(y-1)^2}{x^2(x-1)^2}\binom{|E_c|}{2}\\ &= \frac{y^2}{x^2}|E_c|\Big(1-(y-1)^2/(x-1)\Big) \geq \frac{y^2}{x^2}|E_c|\Big(1-y^2/x\Big) \geq \frac{y^2}{x^2}|E_c|(1-\varepsilon^2).\end{aligned}$$

Let Z be the number of colours in $E(S,T)$. Note that $\sum_c |E_c| = x^2$. Hence, by linearity of expectation, $\mathbb{E}(Z) \geq \sum_c (1-\varepsilon^2)\frac{y^2}{x^2}|E_c| = (1-\varepsilon^2)y^2$. Since $Z \leq e(S,T) = y^2$ we have that $y^2 - Z$ is non-negative with $\mathbb{E}(y^2 - Z) \leq \varepsilon^2 y^2$. Therefore, by Markov's inequality we have $\mathbb{P}(y^2 - Z \geq \varepsilon y^2) \leq \varepsilon$. This implies that with probability at least $1-\varepsilon$ a pair S,T is nearly-rainbow.

Let $\{A_i\}$ and $\{B_j\}$ be random partitions of A and B into sets of size y. By the above discussion the expected fraction of pairs which are not nearly-rainbow A_i, B_j is at most ε. Therefore there exists some partition satisfying the assertion of the lemma. □

Combining the above two lemmas we can now complete the proof of Theorem 1.3.

Proof of Theorem 1.3. Given a proper edge-colouring of K_n , let G be a subgraph of K_n obtained by choosing every color class with probability p. Since all the edges incident to some vertex have distinct colours the degrees of G are binomially distributed with parameters$(n-1,p)$. Moreover from condition $x \gg (\log n/p)^2$ we have that $pn \gg \log n$. Therefore for every vertex v, by the Chernoff bound the probability that $|d_G(v) - np| \geq \epsilon np$ is at most $e^{\frac{-pn\varepsilon^2}{4}}$. By the union bound, all the degrees are almost surely $(1-o(1))np$.

Fix some $x \gg (\log n/p)^2$. Notice that for any pair of disjoint sets A,B with $|A|,|B| \gg x$, $E_{K_n}(A,B)$ contains $(1-o(1))|A||B|/x^2$ edge-disjoint pairs $A_i \subseteq A, B_j \subseteq B$ with $|A_i| = |B_j| = x$. Using this, it is sufficient to prove the theorem just for pairs of sets A,B with $|A|,|B| = x$

Let y be some integer satisfying $y \gg \log n/p$ and $y^2 \ll x$, which exists since $x \gg (\log n/p)^2$. Then, by Lemma 2.2, we have that for every nearly-rainbow pair S,T of sets of size y there are at least $(1-o(1))py^2$ edges of G between S and T. Let A and B be two arbitrary subsets of G size x. By Lemma 2.3, there are partitions $\{A_i\},\{B_j\}$ of A and B into subsets of size y such that all but $o(1)$ fraction of the pairs A_i, B_j are nearly-rainbow in K_n. Then, almost surely,

$$e_G(A,B) \geq \sum_{\text{nearly-regular } A_i,B_j} e_G(A_i,B_j) \geq (1-o(1))\frac{x^2}{y^2}\cdot(1-o(1))py^2 \geq (1-o(1))px^2,$$

completing the proof. □ □

Remark. If $x \gg \log^2 n/p^4$ then one can use our proof to also bound $e_G(A,B)$ from above by $(1-o(1))px^2$. Indeed in this case we can first change the definition of nearly-rainbow pair to have at least $1-o(p)$ fraction of the edges with distinct colours and then adjust Lemma 2.3 to show that for $y^2 \ll p^2 x$ there exist partitions with at most a $o(p)$ fraction of non nearly-rainbow pairs. Then in the above proof even if all non nearly-rainbow pairs are complete and all the edges with non-distinct colours in nearly-rainbow pairs are present it can only contribute at most $o(px^2)$ edges.

3 Rainbow path forest

The following lemma is the second main ingredient which we will need to prove Theorem 1.2. It says that every properly coloured graph with very high minimum degree has a nearly-spanning rainbow path forest. The proof is a version of a technique of Andersen [3] who proved the same result for complete graphs.

Lemma 3.1. *For all γ, δ, n with $\delta \geq \gamma$ and $3\gamma\delta - \gamma^2/2 > n^{-1}$ the following holds. Let G be a properly coloured graph with $|G| = n$ and $\delta(G) \geq (1-\delta)n$. Then G contains a rainbow path forest with $\leq \gamma n$ paths and $|E(\mathcal{P})| \geq (1-4\delta)n$.*

Proof. Let $\mathcal{P} = \{P_1, \dots, P_{\gamma n}\}$ be a rainbow path forest with $\leq \gamma n$ paths and $|E(\mathcal{P})|$ as large as possible. Suppose for the sake of contradiction that $|E(\mathcal{P})| < (1-4\delta)n$. We claim that without loss of generality we may suppose that all the paths $P_1, \dots, P_{\gamma n}$ are nonempty. Indeed notice that we have $|V(\mathcal{P})| \leq |E(\mathcal{P})| + \gamma n < (1-4\delta)n + \gamma n \leq n - \gamma n$. Therefore if any of the paths in $\mathcal{P}$ are empty, then we can replace them by single-vertex paths outside $V(\mathcal{P})$ to get a new path forest with the same number of edges as $\mathcal{P}$. For each i, let the path P_i have vertex sequence $v_{i,1}, v_{i,2}, \dots, v_{i,|P_i|}$. For a vertex $v_{i,j}$ for $j > 1$, let $e(v_{i,j})$ denote the edge $v_{i,j}v_{i,j-1}$ going from $v_{i,j}$ to its predecessor on P_j, and let $c(v_{i,j})$ denote the colour of $e(v_{i,j})$.

We define sets of colours $C_0, C_1, \dots, C_{\gamma n}$ recursively as follows. Let C_0 be the set of colours not on paths in $\mathcal{P}$. For $i = 1, \dots, \gamma n$, let

$$C_i = \left\{c(x) : x \in N_{C_{i-1}}(v_{i,1}) \cap V(\mathcal{P}) \setminus \{v_{1,1}, \dots, v_{\gamma n,1}\}\right\} \cup C_{i-1}.$$

Notice that for any colour $c \in C_i \setminus C_{i-1}$, there is an edge from $v_{i,1}$ to the vertex $x \in V(\bigcup \mathcal{P})$ with $c(x) = c$.

Claim 3.2. *$N_{C_{i-1}}(v_{i,1}) \subseteq V(\mathcal{P}) \setminus \{v_{i+1,1}, \dots, v_{\gamma n,1}\}$ for $i = 1 \dots, \gamma n$.*

Proof. First we'll deal with the case when for $j > i$ there is an edge $v_{i,1}v_{j,1}$ by something in C_{i-1}. Define integers, s, $i_0, \dots, i_s$, colours $c_1, \dots, c_s$, and vertices $x_0, \dots, x_{s-1}$ as follows.

(1) Let $i_0 = i$ and $x_0 = v_{j,1}$.

(2) We will maintain that if $i_t \geq 1$, then the colour of $v_{i_t,1}x_t$ is in $C_{(i_t)-1}$. Notice that this does hold for i_0 and x_0.

(3) For $t \geq 1$, let c_t be the colour of $v_{i_{t-1},1}x_{t-1}$. By (2), we have $c_t \in C_{(i_{t-1})-1}$.

(4) For $t \geq 1$, let i_t be the smallest number for which $c_t \in C_{i_t}$. Notice that this ensures $c_t \in C_{i_t} \setminus C_{(i_t)-1}$

(5) For $t \geq 1$, if $i_t > 0$ then let x_t be the vertex of $V(\mathcal{P})$ with $c(x_t) = c_t$. Such a vertex must exist since from (4) we have $c_t \in C_{i_t} \setminus C_0$. Notice that by the definition of C_{i_t} and $c_t \in C_{i_t} \setminus C_{(i_t)-1}$, the edge $v_{i_t,1}x_t$ must be present and coloured by something in $C_{(i_t)-1}$ as required by (2).

(6) We stop at the first number s for which $i_s = 0$.

See Figure 1 for a concrete example of these integers, colours, and vertices being chosen. Notice that from the choice of c_t and i_t in (3) and (4) we have $i_0 > i_1 > \dots > i_s$. We also have $x_t \neq x_{t'}$ for $t \neq t'$. To see this notice that from (4) and (5) we have $c(x_t) = c_t \in C_{i_t} \setminus C_{(i_t)-1}$ and $c(x_{t'}) = c_{t'} \in C_{i_{t'}} \setminus C_{(i_{t'})-1}$. Since the sets $C_0, C_1, \dots$ are nested, the only way $c(x_t) = c(x_{t'})$ could occur is if $i_t = i_{t'}$ (which would imply $t = t'$.) The following claim will let us find a larger rainbow path forest than $\mathcal{P}$.

Claim 3.3. *$\mathcal{P}' = P_1 \cup \dots \cup P_{\gamma n} \cup \{(v_{i_0,1}x_0), (v_{i_1,1}x_1), (v_{i_2,1}x_2), \dots, (v_{i_{s-1},1}x_{s-1})\} \setminus \{e(x_1), e(x_2), \dots, e(x_{s-1})\}$ is a rainbow path forest.*

Proof. To see that $\mathcal{P}'$ is rainbow, notice that for $0 \leq t < s-1$, the edges $v_{i_t,1}x_t$ and $e(x_{t+1})$ both have the same colour, namely c_{t+1}. This shows that $\mathcal{P}' - v_{i_{s-1},1}x_{s-1} = P_1 \cup \dots \cup P_{\gamma n} \cup \{(v_{i_0,1}x_0), (v_{i_1,1}x_1), (v_{i_2,1}x_2), \dots, (v_{i_{s-2},1}x_{s-2})\} \setminus \{e(x_1), e(x_2), \dots, e(x_{s-1})\}$ has exactly the same colours that $\mathcal{P}$ had. By the definition of s, we have that the colour c_s of $v_{i_{s-1},1}x_{s-1}$ is in C_0 and hence not in $\mathcal{P}$, proving that $\mathcal{P}'$ is rainbow.

To see that $\mathcal{P}'$ is a forest, notice that since $\mathcal{P}$ is a forest any cycle in $\mathcal{P}'$ must use an edge $v_{i_t,1}x_t$ for some t. Let the vertex sequence of such a cycle be $v_{i_t,1}, x_t, u_1, u_2, \dots, u_\ell, v_{i_t,1}$. Since $x_t \in V(\bigcup \mathcal{P})$ we have that $x_t = v_{k,j}$ for some j and k. Notice that since the edge $e(x_t) = v_{k,j}v_{k,j-1}$ is absent in $\mathcal{P}'$, we have that that

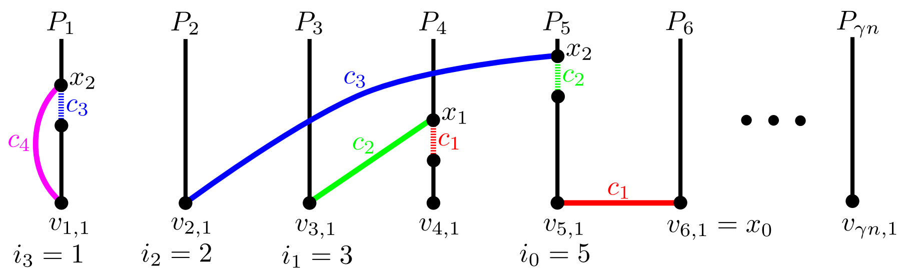

Figure 1: An example of the proof of Claim 3.2. In this example $s = 3$. The dashed coloured edges get deleted from the path forest and are replaced by the solid coloured ones. This gives a bigger path forest, contradicting the maximality of $\mathcal{P}$.

$u_1 = v_{k,j+1}$. Let r be the smallest index for which $u_r \neq v_{k,j+r}$. By the definition of $\mathcal{P}'$, we have that the edge $u_{r-1}u_r$ must be of the form $v_{i'_t,1}x_{t'}$ for some $t' \neq t$ with $u_{r-1} = x_{t'}$ and $u_r = v_{i'_t,1}$. However, then the edge $u_{r-2}u_{r-1} = e(x_{t'})$ would be absent, contradicting C being a cycle.

To see that $\mathcal{P}'$ is a path forest, notice that it has maximum degree 2—indeed the only vertices whose degrees increased are $x_0, v_{i_0,1}, \ldots, v_{i_{s-1},1}$. Their degrees increased from 1 to 2 when going from $\mathcal{P}$ to $\mathcal{P}'$, which implies that $\Delta(\mathcal{P}') \leq 2$ is maintained. □

Now $\mathcal{P}'$ is a path forest with $\leq \gamma n$ paths and one more edge than $\mathcal{P}$ had, contradicting the maximality of $\mathcal{P}$.

The case when $p_{i,1}v$ is an edge for some $v \notin V(\bigcup \mathcal{P})$ is identical using $x_0 = v$. □

For $i = 1, \ldots, s$, let $m_i = |C_i| - |C_0|$. Let C be the set of all the colours which occur in G. Notice that, by the definition of C_0, we have $|C| = |C_0| + e(\bigcup \mathcal{P})$. Notice that for any vertex v we have

$$|N_{C_i}(v)| \geq |N(v)| - \big(|C| - |C_i|\big) \geq (1-\delta)n - (|C_0| + e(\bigcup \mathcal{P})) + (|C_0| + m_i) \geq 3\delta n + m_i. \qquad (1)$$

The first inequality comes from the fact that G is properly coloured and there are $|C| - |C_i|$ colours which are not in C_i. The second inequality comes from $\delta(G) \geq (1-\delta)n$ and the definitions of m_i and C_0. The third inequality comes from $e(\bigcup \mathcal{P}) \leq (1 - 4\delta)n$.

From the definition of C_i, we have $|C_i| \geq |C_0| + |N_{C_{i-1}}(v_{i,1}) \cap \{v_{k,j} : k \geq 2, j = 1, \ldots, \gamma n\}|$. From Claim 3.2, we have $|N_{C_{i-1}}(v_{i,1}) \cap \{v_{k,j} : k \geq 2, j = 1, \ldots, \gamma n\}| \geq |N_{C_{i-1}}(v_{i,1})| - i$. Combining these with (1) we get $|C_i| \geq |C_0| + 3\delta n + m_{i-1} - i$ which implies $m_i \geq m_{i-1} + 3\delta n - i$ always holds. Iterating this gives $m_i \geq 3i\delta n - \binom{i}{2}$. Setting $i = \gamma n$, gives $n \geq m_{\gamma n} \geq 3\gamma\delta n^2 - (\gamma n)^2/2$, which contradicts $3\gamma\delta - \gamma^2/2 > n^{-1}$. □

4 Long rainbow cycle

In this section we prove Theorem 1.2, using the following strategy. First we apply Theorem 1.3 to find a good expander H in K_n whose maximum degree is small and whose edges use only few colours . Then we apply Lemma 3.1 to find a nearly spanning path forest with few paths, which is shares no colours with H. Then we use the expander H to "rotate" the path forest in order to successively extend one of the paths in it until we have a nearly spanning rainbow path P. Finally we again use the expander H to close P into a rainbow cycle.

We start with the lemma which shows how to use an expander to enlarge one of the paths in a path forest.

Lemma 4.1. *For $b, m, r > 0$ with $2mr \leq b$, the following holds. Let $\mathcal{P} = \{P_1, \ldots, P_r\}$ be a rainbow path forest in a properly coloured graph G. Let H be a subgraph of G sharing no colours with $\mathcal{P}$ with $\delta(H) \geq 3b$ and $E_H(A, B) \geq b + 1$ for any two sets of vertices A and B of size b. Then either $|P_1| \geq |V(\mathcal{P})| - 2b$ or there are two edges $e_1, e_2 \in H$ and a rainbow path forest $\mathcal{P}' = \{P'_1, \ldots, P'_r\}$ such that $E(\mathcal{P}') \subseteq E(\mathcal{P}') + e_1 + e_2$ and $|P'_1| \geq |P_1| + m$.*

Proof. Suppose that $|P_1| < |V(\mathcal{P})| - 2b$. Let $P_1 = v_1, v_2, \dots, v_k$ and let T be the union of vertices on those path among $P_2, \dots, P_r$ which have length at least $2m$. Notice that there are at most $2mr$ vertices on paths P_i of length $\leq 2m$. Since $|P_1| < |V(\mathcal{P})| - 2b$, the set T has size at least $2b - 2mr \geq b$.

First suppose that there is an edge of H from p_1 to a vertex $x \in T$ on some path P_i of length at least $2m$. We can partition P_i into two subpaths P^+ and P^- such that P^+ starts with x and has $|P^+| \geq |P_i|/2 \geq m$. Then we can take $e_1 = e_2 = v_1x$, $P_1 = P_1 + e_1 + P^+$, $P_i = P^-$ and $P'_j = P_j$ for all other j to obtain paths satisfying the assertion of the lemma.

Next suppose that $|N_H(v_1) \cap P_1| \geq b$. Let $S \subseteq N_H(v_1) \cap P_1$ be a subset of size b. Let S^+ be the set of predecessors on P_1 of vertices in S, i.e., $S^+ = \{v_{i-1} : v_i \in S\}$. Since $|S^+|, |T| \geq b$, there are at least $b+1$ edges between S^+ and T in H. In particular this means that there is some $v_\ell \in S^+$ which has $|N_H(v_\ell) \cap T| \geq 2$. Since H is properly coloured, there is some $x \in N_H(v_\ell) \cap T$ such that $v_\ell x$ has a different colour to $v_1 v_{\ell+1}$. By the definition of T, this x belongs to path $P_i, i \geq 2$ with $|P_i| \geq 2m$. Again we can partition P_i into two subpaths P^+ and P^- such that P^+ starts with x and has $|P^+| \geq m$. Then, taking $e_1 = v_1 v_{\ell+1}$ $e_2 = v_\ell x$, $P_1 = (v_k, v_{k-1}, \dots, v_{\ell+1} v_1, v_2, \dots, v_\ell) + e_2 + P^+$, $P'_i = P^-$ and $P'_j = P_j$ for all other j, we obtain paths satisfying the assertion of the lemma.

Finally we have that $|N_H(v_1) \cap P_1| < b$ and there are no edges from v_1 to T. Note that there are also at most $2mr \leq b$ edges from v_1 to vertices on paths P_i of length $\leq 2m$. Since $|N_H(v_1)| \geq 3b$, there is a set $S \subseteq N_H(v_1) \setminus V(\mathcal{P})$ with $|S| = b$. Since $|S|, |T| \geq b$, there is an edge sx in H from some $s \in S$ to some $x \in T$. Since H is properly coloured, sx has a different colour from $v_1 s$. By the definition of T, the vertex x is on some path $P_i, i \geq 2$ of length at least $2m$. Partition P_i into two subpaths P^+ and P^- such that P^+ starts with x and has $|P^+| \geq m$. Let $e_1 = v_1 s$ $e_2 = sx$. Then $P'_1 = P_1 + e_1 + e_2 + P^+$, $P'_i = P^-$ and $P'_j = P_j$ for all other j satisfy the assertion of the lemma, completing the proof □

Having finished all the necessary preparations we are now ready to show that every properly edge-coloured K_n has a nearly-spanning rainbow cycle.

Proof of Theorem 1.2. Given a properly edge-coloured complete graph K_n, we first use Theorem 1.3 to construct its subgraph H, satisfying conditions of Lemma 4.1. Let $b = n^{3/4}$. Let H be a subgraph obtained by choosing every colour class randomly and independently with probability $p = 4.5b/n$. Since $p = 4.5n^{-1/4}$ we have that $b = n^{3/4} \gg n^{1/2} \log^2 n > (\log n/p)^2$. Therefore, we can apply Theorem 1.3 to get that almost surely every vertex in H has degree $4b \leq d_H(v) = (1 - o(1))np \leq 5b - 1$ and $e_H(A, B) \geq (1 - o(1))pb^2 > 4.3n^{1/2}b$ for any disjoint sets A and B of size b.

Let $G = K_n \setminus H$ be the subgraph of K_n consisting of edges whose colours are not in $E(H)$. We have $\delta(G) \geq n - 1 - \Delta(H) \geq (1 - 5n^{-1/4})n$. Applying Lemma 3.1 with $\delta = 5n^{-1/4}$ and $\gamma = n^{-3/4}$ we get a rainbow path forest $\mathcal{P}$ with $n^{1/4}$ paths and $|E(\mathcal{P})| \geq n - 20n^{3/4}$. Moreover the colours of edges in $\mathcal{P}$ and H are disjoint.

Next, we repeatedly apply Lemma 4.1 $2n^{1/2}$ times with $b = n^{3/4}$, $r = n^{1/4}$, and $m = 0.5n^{1/2}$. At each iteration we delete from H all edges sharing a colour with e_1 or e_2 to get a subgraph H'. Notice that after i iterations, H has lost at most $2i$ colours, and so $\delta(H') \geq \delta(H) - 2i \geq 4b - 2i > 3b$ and for any $A, B \subseteq V(H)$ with $|A|, |B| \geq b$ we have $e_{H'}(A, B) \geq 4.3n^{1/2}b - 2ib \geq 0.3n^{1/2}b > b + 1$. This shows that we indeed can continue the process for $2n^{1/2}$ steps without violating the conditions of Lemma 4.1. At each iteration we either increase the length of P_1 by m, or we establish that $|P_1| \geq |V(\mathcal{P})| - 2b$. Since $2n^{1/2}m = n > n - 2b$ we have that the second option must occur at some point during the $2n^{1/2}$ iterations of Lemma 4.1. This gives a rainbow path P of length at least $|V(\mathcal{P})| - 2b \geq n - 22n^{3/4}$. As was mentioned above there must still be at least $0.3n^{1/2}b$ edges of H' left between any two disjoint sets A, B of size b. Let S be the set of first b vertices and T be the set of last b vertices of the rainbow path P. Then there is and edge of H' between S and T whose colour is not on P. Adding this edge we get a rainbow cycle of length at least $|P| - 2b \geq n - 24n^{3/4}$, completing the proof. □

5 Concluding remarks

Versions of Theorem 1.3 can be proved in settings other than properly coloured complete graphs. One particularly interesting variation is to look at properly coloured balanced complete bipartite graphs $K_{n,n}$.

Theorem 5.1. *Given a proper edge-colouring of $K_{n,n}$ with bipartition classes X and Y, let G be a subgraph obtained by choosing every colour class randomly and independently with probability p. Then, with high probability, all vertices in G have degree $(1-o(1))np$ and for every two subsets $A \subseteq X, B \subseteq Y$ of size $x \gg (\log n/p)^2$, $e_G(A,B) \geq (1-o(1))px^2$.*

The above theorem is proved by essentially the same argument as Theorem 1.3. Properly coloured balanced complete bipartite graphs are interesting because they generalize Latin squares. Indeed given any $n \times n$ Latin square, one can associate a proper colouring of $K_{n,n}$ with $V(K_{n,n}) = \{x_1, \dots, x_n, y_1, \dots, y_n\}$ to it by placing a colour i edge between the x_j and y_k whenever the (j,k)th entry in the Latin square is i. Thus Theorem 5.1 implies that every Latin square has a small set of symbols which have random-like behavior.

It would be interesting to find the correct value of second order term in Theorem 1.2. So far, the best lower bound on this is "-1" which comes from Maamoun and Meyniel's construction in [11]. It is quite possible that their construction is tight and "-1" should be the correct value. However this would likely be very hard to prove since, at present, we do not even know how to get a rainbow maximum degree 2 subgraph of a properly coloured K_n with $n - o(\sqrt{n})$ edges (a subgraph with $n - O(\sqrt{n})$ edges can be obtained by Lemma 3.1, or by a result from [3].)

Finally it would be interesting to know what is the smallest size of the sets for which Theorem 1.3 holds. In particular, is it true that for all $|A|, |B| \gg (\log n/p)^{1+\varepsilon}$ we have $e_G(A,B) \geq (1-o(1))p|A||B|$, where G is the graph in Theorem 1.3?

References

[1] S. Akbari, O. Etesami, H. Mahini, and M. Mahmoody. On rainbow cycles in edge coloured complete graphs. *Australas. J. Combin.* **37** (2007), 33–42.

[2] N. Alon and J. H. Spencer, **The Probabilistic Method**, 4th ed., Wiley, New York, 2016.

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

[4] H. Chen and X. Li. Long rainbow path in properly edge-coloured complete graphs. arXiv:1503.04516 (2015).

[5] H. Gebauer and F. Mousset. On Rainbow Cycles and Paths. *arXiv:1207.0840* (2012).

[6] A. Gyárfás and M. Mhalla. Rainbow and orthogonal paths in factorizations of K_n. *J. Combin. Des.* **18** (2010), 167–176.

[7] A. Gyárfás, M. Ruszinkó, G. Sárközy, and R. Schelp. Long rainbow cycles in proper edge-colourings of complete graphs. *Australas. J. Combin.* **50** (2011), 45–53.

[8] G. Hahn. Un jeu de colouration. *Actes du Colloque de Cerisy* **12** (1980), 18–18.

[9] E. Mendelsohn and A. Rosa, One-factorizations of the complete graph - a survey, *J. Graph Theory* **9** (1985), 43–65.

[10] G. Hahn and C. Thomassen, Path and cycle sub-ramsey numbers and an edge-colouring conjecture. *Discrete Math.* **62** (1986), 29–33.

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

[12] B. Sudakov and J. Volec, Properly colored and rainbow copies of graphs with few cherries, *J. Combinatorial Theory Ser. B*, to appear.

[13] W. Wallis, One-factorization of complete graphs, in: **Contemporary design theory**, Willey, New York, 1992, 593–631.