Decompositions into spanning rainbow structures

Richard Montgomery, Alexey Pokrovskiy and Benjamin Sudakov

ABSTRACT

A subgraph of an edge-coloured graph is called rainbow if all its edges have distinct colours. The study of rainbow subgraphs goes back more than 200 years to the work of Euler on Latin squares and has been the focus of extensive research ever since. Euler posed a problem equivalent to finding properly n-edge-coloured complete bipartite graphs $K_{n,n}$ which can be decomposed into rainbow perfect matchings. While there are proper edge-colourings of $K_{n,n}$ without even a single rainbow perfect matching, the theme of this paper is to show that with some very weak additional constraints one can find many disjoint rainbow perfect matchings. In particular, we prove that if some fraction of the colour classes have at most $(1 - o(1))n$ edges, then one can nearly-decompose the edges of $K_{n,n}$ into edge-disjoint perfect rainbow matchings. As an application of this, we establish in a very strong form a conjecture of Akbari and Alipour and asymptotically prove a conjecture of Barat and Nagy. Both these conjectures concern rainbow perfect matchings in edge-colourings of $K_{n,n}$ with quadratically many colours. The above result also has implications to some conjectures of Snevily about subsquares of multiplication tables of groups.

Finally, using our techniques, we also prove a number of results on near-decompositions of graphs into other rainbow structures like Hamiltonian cycles and spanning trees. Most notably, we prove that any properly coloured complete graph can be nearly-decomposed into spanning rainbow trees. This asymptotically proves the Brualdi–Hollingsworth and Kaneko–Kano–Suzuki conjectures which predict that a perfect decomposition should exist under the same assumptions.

1. *Introduction*

A *Latin square* of order n is an $n \times n$ array filled with n symbols such that each symbol appears once in every row and column. A *partial transversal* is a collection of cells of the Latin square which do not share the same row, column or symbol. A *transversal* is a partial transversal of order n. Latin squares were introduced by Euler in the 18th century and are familiar to the layperson in the form of Sudoku puzzles, which, when completed, are Latin squares. Another well-known example of the Latin square is a multiplication table of any finite group. The study of Latin squares have applications both inside and outside mathematics, with connections to 2-dimensional permutations, design theory, finite projective planes and error correcting codes.

Euler was interested in *orthogonal Latin squares* — a pair of $n \times n$ Latin squares S and T with the property that every pair of symbols (i, j) occurs precisely once in the array. This is equivalent to Latin squares which can be decomposed into disjoint transversals (see [**24**, **38**]). He conjectured that there exist $n \times n$ Latin squares with a decomposition into disjoint transversals if, and only if, $n \not\equiv 2 \pmod 4$. When $n \not\equiv 2 \pmod 4$, Euler himself constructed such Latin squares. The '$n = 6$' case stood open for over 100 years until it was proved by Tarry in 1901. The remaining cases '$n \not\equiv 2 \pmod 4$, $n \geqslant 10$' were resolved in 1959 by Bose, Parker and Shrikande [**14**]. Surprisingly, they showed that Euler's Conjecture was false for these values of n by explicitly constructing Latin squares with a decomposition into disjoint transversals.

It is a hard problem to determine which Latin squares have transversals. This question is very difficult even in the case of multiplication tables of finite groups. In 1955, Hall and Paige [**34**]

conjectured that the multiplication table of a group G has a transversal exactly if the 2-Sylow subgroups of G are trivial or non-cyclic. It took 50 years to establish this conjecture and its proof is based on the classification of finite simple groups (see [**50**] and the references therein). The most famous open problem on transversals in general Latin squares is a conjecture of Ryser and Brualdi-Stein.

CONJECTURE 1.1 (Ryser [**45**], Brualdi-Stein [**17**, **49**]). Every $n \times n$ Latin square has a partial transversal of order $n - 1$ and a full transversal if n is odd.

The best results towards this conjecture are asymptotic and show that all Latin squares have partial transversals of size $n - o(n)$. Woolbright [**51**] and Brower, de Vries and Wieringa [**15**] independently proved this with $o(n) = \sqrt{n}$. The error term was further improved by Hatami and Shor [**35**], who showed that $o(n) = O(\log^2 n)$ suffices.

Generalized Latin squares are $n \times n$ arrays filled with an *arbitrary* number of symbols such that no symbol appears twice in the same row or column. They are natural extensions of Latin squares, and have also been extensively studied. A familiar example of a generalized Latin square is a multiplication table between elements of two subsets of equal size in some group. It is generally believed that extra symbols in a Latin square should help to find transversals. The goal of this paper is to confirm that this is indeed the case. Moreover we show that, under some very weak additional conditions, a generalized Latin square has not only one but many disjoint transversals.

THEOREM 1.2. *Let S be a generalized Latin square with at most $(1 - o(1))n$ symbols occurring more than $(1 - o(1))n$ times. Then, S has $(1 - o(1))n$ pairwise disjoint transversals.*

All previous results that guaranteed transversals studied arrays which were very far from Latin squares. For example, Erdős and Spencer [**23**] showed that a transversal exists in any $n \times n$ array in which each symbol appears at most $n/16$ times. Furthermore, Alon, Spencer and Tetali [**8**] found many disjoint transversals in the case when each symbol appears δn times, for some small but fixed $\delta > 0$. On the other hand, our result shows that the only generalized Latin squares without transversals are small perturbations of Latin squares.

Theorem 1.2 can be also used to attack several open problems on generalized Latin squares. For example, Akbari and Alipour conjectured the following.

CONJECTURE 1.3 (Akbari and Alipour [**1**]). Every generalized Latin square with at least $n^2/2$ symbols has a transversal.

More generally, Barat and Nagy [**13**] conjectured that under the same assumptions as above, any generalized Latin square should have a decomposition into disjoint transversals. Theorem 1.2 has implications for both of these conjectures. It is easy to show that in any generalized Latin square with at least εn^2 symbols at most $(1 - \varepsilon/2)n$ symbols occur more than $(1 - \varepsilon/2)n$ times (see Lemma 8.14). Thus, the following is a corollary of Theorem 1.2.

COROLLARY 1.4. *For all $\varepsilon > 0$ and sufficiently large n, every generalized Latin square with at least εn^2 symbols has $(1 - \varepsilon)n$ pairwise disjoint transversals.*

For large n, this establishes the conjecture of Akbari–Alipour in a very strong form, showing that the bound of $n^2/2$ can be reduced to εn^2. It also proves asymptotically the Barat–Nagy conjecture, giving a near-decomposition of the generalized Latin square into transversals.

Theorem 1.2 has also some interesting implications for transversals in actual Latin squares. Indeed, it is not hard to show that any Latin square contains many subsquares which satisfy

the assumptions of Theorem 1.2. In fact, a random $(1 - o(1))n \times (1 - o(1))n$ subsquare will have this property with high probability. Thus, we have the following corollary.

COROLLARY 1.5. *Let S be a random $(1 - o(1))n \times (1 - o(1))n$ subsquare of an $n \times n$ Latin square L. With high probability, S has a transversal.*

This corollary reproves the result that Latin squares have partial transversals of size $n - o(n)$. However, it proves much more, that is, partial transversals of size $n - o(n)$ must *be present almost everywhere* in the Latin square.

Our main theorem has additional applications to group theoretic problems and questions about rainbow structures in coloured graphs, which we discuss next.

Subsquares of multiplication tables

A natural way to obtain a generalized Latin square is to consider a subsquare S of a multiplication table of a group G. Snevily made the following general conjecture on transversals in subsquares of abelian groups.

CONJECTURE 1.6 (Snevily [**48**]). *Let $S = A \times B$ be a subsquare of the multiplication table of an abelian group G defined by two n-element sets $A, B \subseteq G$.*

  (i) *If G is an odd abelian group, then S has a transversal.*
 (ii) *If G is an even cyclic group, then S has no transversal only when both A and B are translates of the same even cyclic subgroup of G.*

Here a 'translate of A' means any set of the form gA for $g \in V(G)$. Part (i) of this conjecture has attracted a lot of attention. After work by Alon [**4**] and Dasgupta, Károlyi, Serra and Szegedy [**22**], it was solved by Arsovski [**10**]. Part (ii) of Conjecture 1.6 is still open.

Our work has implications for this conjecture, and for various generalizations for other groups and semigroups. Combining our Theorem 1.2 with the following lemma, one can find not just one but many transversals in certain subsquares of multiplication tables.

LEMMA 1.7. *Let $S = A \times B$ be a subsquare of the multiplication table of a group G defined by two n-element sets $A, B \subseteq G$. Then, either S has at most $(1 - o(1))n$ symbols occurring more than $(1 - o(1))n$ times or there is a subgroup H of G and elements $g, g' \in G$ such that $|A \Delta gH| = o(n)$ and $|B \Delta g'H| = o(n)$.*

In other words, this lemma says that either a subsquare S of a multiplication table is close to a translate of a subgroup, or it satisfies the condition of Theorem 1.2. In the latter case, we can use this theorem to nearly decompose S into disjoint transversals. Thus, we have the following corollary which works in any group, not just finite or abelian groups.

COROLLARY 1.8. *Let $S = A \times B$ be a subsquare of the multiplication table of a group G defined by two n-element sets $A, B \subseteq G$. Then, one of the following holds.*

  • *S has $(1 - o(1))n$ disjoint transversals.*
  • *There is a subgroup H of G and elements $g, g' \in G$ such that $|A \Delta gH| = o(n)$ and $|B \Delta g'H| = o(n)$.*

Lemma 1.7 is implicit in the work of Fournier [**25**] and appears as Theorem 1.3.3 in the lecture notes of Green [**29**]. It is formulated in terms of multiplicative energy, which for a subset A of group G is the number of quadruples $a_1, a_2, b_1, b_2 \in A$ such that $a_1 a_2^{-1} = b_1 b_2^{-1}$. It

follows easily from the definitions that if S has more than $(1 - o(1))n$ symbols occurring more than $(1 - o(1))n$ times, then both A and B have energy at least $(1 - o(1))n^3$ and therefore are very close to cosets of some subgroups, which can further be shown to be the same subgroup.

Rainbow matchings, Hamiltonian paths and cycles

Transversals in Latin squares are closely related to rainbow subgraphs of edge-coloured graphs. Recall that an edge-coloured graph is properly coloured if no two edges of the same colour share a vertex. A *matching* in a graph is a set of disjoint edges. We call a subgraph of a graph *rainbow* if all of its edges have different colours. There is a one-to-one correspondence between $n \times n$ generalized Latin squares and proper edge-colourings of the complete bipartite graph $K_{n,n}$. Indeed, given a generalized Latin square $S = (s_{ij})$ with m symbols in total, associate with it an m-edge-colouring of $K_{n,n}$ by setting $V(K_{n,n}) = \{x_1, \dots, x_n, y_1, \dots, y_n\}$ and letting the colour of the edge (x_i, y_j) be s_{ij}. Notee that this colouring is proper, that is, adjacent edges receive different colours. Therefore, the study of transversals in generalized Latin squares is equivalent to the study of perfect rainbow matchings in proper edge-colourings of $K_{n,n}$. Moreover, if S is symmetric, that is, $s_{ij} = s_{ji}$ for all i and j, it also defines the proper edge-colouring of the complete n-vertex graph K_n in which the edge ij is coloured by s_{ij}. Since S is symmetric, each edge has a well-defined colour. Under this second correspondence, transversals give rainbow maximum degree 2 subgraphs of K_n.

As explained above, partial transversals in the Latin square S correspond to rainbow matchings in the corresponding edge-coloured $K_{n,n}$. Thus, Conjecture 1.1 is equivalent to the statement that any proper n-edge-colouring of $K_{n,n}$ contains a rainbow matching of size $n - 1$. Theorem 1.2 then follows from the following statement.

THEOREM 1.9.   *There is an $\alpha > 0$ so that the following holds for all $1 > \varepsilon \geqslant n^{-\alpha}/\alpha$. Let $K_{n,n}$ be properly coloured with at most $(1 - \varepsilon)n$ colours having more than $(1 - \varepsilon)n$ edges. Then, $K_{n,n}$ has $(1 - \varepsilon)n$ edge-disjoint perfect rainbow matchings.*

We can also find perfect rainbow matchings in graphs that are more general than $K_{n,n}$. Our proof works for all suitably pseudo-random properly coloured balanced bipartite graphs. See Lemma 8.11 for an example of such a result, and see Lemma 8.13 for a restatement and proof of Theorem 1.9.

There is a lot of interest in studying rainbow structures in properly coloured complete graphs. Recall that transversals in symmetric generalized Latin squares correspond to rainbow maximum degree 2 subgraphs of properly coloured complete graphs. Since paths and cycles are a special type of maximum degree 2 subgraph, there has been a focus on finding nearly spanning rainbow paths/cycles in properly coloured complete graphs. For example, Andersen [9] in 1989 conjectured that all properly coloured K_n have a rainbow path of length $n - 2$. Hahn conjectured even more that such a path can be found in any (not necessarily properly) coloured complete graph with at most $n/2 - 1$ edges of each colour (see [33]). Hahn's conjecture was recently disproved by the second and third author [43], who showed that without the 'proper colouring' assumption the graph might not have rainbow paths longer than $n - \Omega(\log n)$. Thus, it makes sense to restrict ourselves to colourings which are proper. The progress on Andersen's conjecture was slow, despite efforts by various researchers, for example, see [3, 19, 27, 31, 32]. Until recently it was not even known how to find a rainbow path/cycle of length $(1 - o(1))n$. This was proved by Alon and the second and third author [6], who showed that any properly coloured K_n contains a rainbow cycle with $n - O(n^{3/4})$ vertices. Using our techniques, one can say much more, that is, we can nearly decompose such a complete graph into long rainbow cycles. This is a corollary of the following theorem.

THEOREM 1.10. *There is an $\alpha > 0$ so that the following holds for all $1 > \varepsilon \geqslant n^{-\alpha}/\alpha$. Let K_n be properly coloured with at most $(1 - \varepsilon)n$ colours having more than $(1 - \varepsilon)n/2$ edges. Then, K_n has $(1 - \varepsilon)n/2$ edge-disjoint rainbow Hamiltonian cycles.*

COROLLARY 1.11. *There is an $\alpha > 0$ so that the following holds for all $1 > \varepsilon \geqslant n^{-\alpha}/\alpha$. Given a properly coloured K_n, let U be a random subset of $(1 - \varepsilon)n$ vertices. Then, with high probability, the subgraph induced by U has $(1 - 2\varepsilon)n/2$ edge-disjoint rainbow Hamiltonian cycles.*

See Lemma 8.29 for a restatement and proof of Theorem 1.10, and Corollary 8.30 for a restatement and proof of Corollary 1.11.

Rainbow spanning trees

In this paper, we also study spanning rainbow trees in properly coloured complete graphs. Note that a rainbow Hamiltonian path is a very special case of a rainbow tree. Because of this one expects the results which hold for rainbow spanning trees to be stronger than ones for paths. For example, every properly coloured K_n contains a rainbow spanning tree (a star at any vertex is rainbow), whereas it is known that there are proper edge-colourings of K_n without rainbow Hamiltonian paths. In fact, much more is probably true. It was conjectured by a number of authors that properly coloured complete graphs should always have decompositions into spanning rainbow trees.

CONJECTURE 1.12 (Brualdi and Hollingsworth, [16]). *Every properly $(2n - 1)$-coloured K_{2n} can be decomposed into edge-disjoint rainbow spanning trees.*

CONJECTURE 1.13 (Kaneko, Kano, and Suzuki, [37]). *Every properly coloured K_n contains $\lfloor n/2 \rfloor$ edge-disjoint rainbow spanning trees.*

These conjectures attracted a lot of attention from various researchers (see, for example, [2, 18, 26]) who showed how to find several disjoint spanning rainbow trees. The best known results for these problem guarantee the existence of εn edge-disjoint rainbow trees (see [36] for Conjecture 1.12 and [12, 44] for Conjecture 1.13). Developing our results on Hamiltonian cycles, we are able to improve this and show that one can find $(1 - o(1))n$ disjoint spanning rainbow trees.

THEOREM 1.14. *There is an $\alpha > 0$ so that the following holds for all $1 > \varepsilon \geqslant n^{-\alpha}/\alpha$. Every properly coloured K_n has $(1 - \varepsilon)n/2$ edge-disjoint spanning rainbow trees.*

This theorem proves an asymptotic version of the Brualdi–Hollingsworth and Kaneko–Kano–Suzuki conjectures. Note that unlike our results about perfect matchings and Hamiltonian cycles, which require certain small additional conditions, this theorem is true for all proper edge-colourings.

2. *Proof overview*

Our various rainbow decomposition results build on each other. First, we find decompositions into rainbow perfect matchings, then into rainbow Hamiltonian cycles and then into rainbow spanning trees. There are other rainbow structures that we find in between these — the actual sequence of our proofs is the following:

(1) near-decompositions of nearly-regular balanced bipartite graphs into nearly-perfect rainbow matchings;
(2) near-decompositions of typical balanced bipartite graphs into perfect rainbow matchings;
(3) near-decompositions of typical graphs into rainbow 2-factors;
(4) near-decompositions of typical graphs into rainbow Hamiltonian cycles;
(5) near-decompositions of complete graphs into rainbow spanning trees.

The following definitions make precise various terms in this overview.

- *Near-decomposition:* A near-decomposition of a graph G is a set of edge-disjoint subgraphs $H_1, \ldots, H_t$ in G which cover almost all the edges of G, that is, which have $e(H_1 \cup \cdots \cup H_t) = (1 - o(1))e(G)$.
- *Average degree:* The average degree of a graph G is $\mathrm{d}(G) = 2e(G)/v(G)$.
- *Nearly-regular:* A graph is nearly regular if all the vertices $v \in V(G)$ have $d(v) = (1 \pm o(1))\frac{2e(G)}{v(G)}$, that is, if all its degrees are close to each other.
- *Typical:* A graph is typical if any pair of vertices $u, v \in V(G)$ has $d(u, v) = (1 \pm o(1))\frac{4e(G)^2}{v(G)^3}$, that is, if all its codegrees are close to each other. This is the main notion of pseudo-randomness that we use in this paper.
- *Global boundedness:* A coloured graph is globally b-bounded if it has $\leqslant b$ edges of each colour.
- *2-factor:* A 2-factor is a collection of vertex-disjoint cycles which span all the vertices of a graph.
- *Balanced bipartite:* A graph is balanced bipartite if its vertices can be partitioned into two sets of the same size, so that all the edges lie between the two sets.

2.1. Nearly-perfect rainbow matchings

There are two main results we prove about nearly-perfect rainbow matchings — one finds a single nearly-perfect rainbow matching in a graph, the other nearly-decomposes a graph into them. The following is an informal description of the first result:

A 1. Every properly coloured, nearly-regular, globally $\mathrm{d}(G)$-bounded, balanced bipartite graph G has a rainbow matching M of order $(1 - o(1))|V(G)|/2$. Additionally, M can be chosen probabilistically so that every edge of G is in M with roughly the same probability.

The precise statement of this is Lemma 4.6. The proof uses Rödl's semi-random method together with some extra ideas. The key point in A1 is that the matching it produces is randomized. Given a properly coloured, nearly-regular, globally $\mathrm{d}(G)$-bounded, balanced bipartite graph, we can repeatedly apply A1 in order to produce a sequence of disjoint nearly-perfect rainbow matchings $M_1, \ldots, M_t$. We can keep iterating this as long as the remaining graph satisfies the assumptions of A1 (near-regularity and global boundedness). Using the fact that the matching in A1 is randomized, we can show that with high probability we can iterate A1 until there are $o(|V(G)|^2)$ edges left in the graph, that is, until we have a near-decomposition into nearly-perfect rainbow matchings:

A 2. Every properly coloured, nearly-regular, globally $\mathrm{d}(G)$-bounded, balanced bipartite graph G can be nearly-decomposed into rainbow matchings of order $(1 - o(1))|V(G)|/2$.

The precise statement of this is Lemma 8.2. The proof of A2 iterates A1 while ensuring that the assumptions of A1 are maintained. We show this using a martingale concentration inequality.

2.2. *Rainbow perfect matchings*

The basic result we prove about near-decompositions into perfect rainbow matchings is the following:

A 3.   Let G be a properly coloured, nearly-regular, globally $\mathrm{d}(G)$-bounded, balanced bipartite graph. Let H be a properly coloured, typical graph on $V(G)$ which is edge-disjoint and colour-disjoint from G. Then, $G \cup H$ has a near-decomposition into rainbow perfect matchings.

The precise statement of this is Lemma 8.12. The assumptions of this lemma (that we have two disjoint graphs, one of which is typical and the other nearly-regular and globally bounded) will reoccur several times in this paper. We pause now to explain why these are natural assumptions under which to seek spanning rainbow structures.

We look at a nearly-regular, globally $\mathrm{d}(G)$-bounded graph for two reasons. Firstly A2 shows that under this assumption one can find rainbow nearly-perfect matchings (so it is reasonable to try to strengthen A2 to get perfect matchings). Secondly, if one starts in any properly coloured $K_{n,n}$ and selects a random subgraph G by choosing every colour independently with probability p (and letting G be the edges of the chosen colours), then the resulting subgraph will be a nearly-regular, globally $(1 + o(1))\mathrm{d}(G)$-bounded graph with high probability. We prove results about nearly-regular, globally $\mathrm{d}(G)$-bounded graphs so that we can say things about random subgraphs of properly coloured complete graphs.

Unfortunately one cannot hope to find perfect rainbow matchings if one just considers a nearly-regular graph G. This is because nearly-regular graphs might have no perfect matchings at all (for example, a disjoint union of two copies of $K_{n,n+1}$ is nearly-regular, balanced bipartite and has no perfect matching). This is the motivation for the typical graph H disjoint from G in A3. The union of a nearly-regular graph G and a typical graph H has a perfect matching, making A3 more plausible.

To prove A3, we first apply A2 to G to get a near-decomposition of G into nearly-rainbow matchings. Then, we use edges of H to modify the matchings one-by-one to turn them into perfect matchings. The modifications we use are simple switchings where we exchange two edges of a matching M for three edges of H in order to get a larger matching M'. Using a sequence of switchings, we will obtain perfect matchings.

Proving Theorem 1.9.   A3 can be used to prove Theorem 1.9. To do this, we need two intermediate results. The first concerns choosing a random set of colours in a properly coloured graph.

A 4.   Let G be properly coloured and typical. Choose every colour independently with probability p, and let H be the subgraph formed by the edges of the chosen colours. Then, with high probability, H is typical.

This result says that the subgraph chosen by a random set of colours is pseudo-random. A result like this was first used by Alon and the second and third author when studying rainbow cycles in graphs [**6**].

Applying A4 to the complete bipartite graph $K_{n,n}$ from Theorem 1.9 gives a typical subgraph H which can be used in A3. The graph G formed by the colours from $K_{n,n}$ unused in H will be nearly-regular with high probability. However, we cannot yet apply A3 since the graph G might not be globally $\mathrm{d}(G)$-bounded. Indeed, G may have colour classes of size n, whereas the average degree of G will be $(1 \pm o(1))(n - p)$ (where p is the parameter from A4). To get around this, we have another intermediate result saying that there is a subgraph G' of G which is globally $\mathrm{d}(G')$-bounded.

A 5. Let G be a properly coloured balanced bipartite graph with $\leqslant (1 - \varepsilon)n$ colours having $\geqslant (1 - \varepsilon)n$ edges and $\delta(G) \geqslant (1 - \varepsilon^2)n$. Then, G has a spanning subgraph G' with $\mathrm{d}(G') \geqslant (1 - 2\varepsilon)n$ which is globally $\mathrm{d}(G')$-bounded and nearly-regular.

See Lemma 6.7 for a precise statement of A5. This is proved in two stages. First, for every colour c with $\geqslant (1 - \varepsilon)n$ edges, we randomly delete every colour c edge with a small probability q. The remaining graph G_1 will be globally $(1 - o(1))\mathrm{d}(G_1)$-bounded with high probability, but might no longer be nearly-regular. We then apply a 'regularization' lemma to G_1 which deletes a small number of edges from G_1 to make it nearly-regular, without overly affecting the global boundedness. The resulting graph G' is then globally $\mathrm{d}(G')$-bounded and nearly-regular. Plugging G' into A3 together with the graph H from A4, we obtain Theorem 1.9.

2.3. *Rainbow 2-factors*

Rainbow 2-factors are intermediate structures we use between finding perfect matchings and Hamiltonian cycles. The main result about 2-factors that we need is a direct analogue of A3.

A 6. Let G be a properly coloured, nearly-regular, globally $\frac{1}{2}\mathrm{d}(G)$-bounded graph. Let H be a properly coloured, typical graph on $V(G)$ which is edge-disjoint and colour-disjoint from G. Then, $G \cup H$ has a near-decomposition into rainbow 2-factors.

See Lemma 8.21 for a precise statement of this. The main difference between A3 and A6 is that the global boundedness in A6 is $\frac{1}{2}\mathrm{d}(G)$ (rather than $\mathrm{d}(G)$ as it was in A3). The reason for this is that to find a rainbow 2-factor we would need $|V(G)|$ colours in the graph, which is forced by global $\frac{1}{2}\mathrm{d}(G)$-boundedness (but not by $\mathrm{d}(G)$-boundedness). Thus, the global $\frac{1}{2}\mathrm{d}(G)$-boundedness condition is natural because it is the weakest global boundedness we can impose on the graph to guarantee enough colours for a rainbow 2-factor

The proof of A6 consists of using A3 to find matchings in the graph, which are then put together to get 2-factors. To see how we might do this, we randomly partition $V(G \cup H)$ and $C(G \cup H)$ into vertex sets $U_1, \dots, U_k$ and colour sets $C_1, \dots, C_k$ of the same size. Then, using variants of A4, we can show that the subgraphs $G_{C_i}[U_j, U_k]$ are nearly-regular, while the subgraphs $H_{C_i}[U_j, U_k]$ are typical. By A3, these subgraphs have near-decompositions into families $\mathcal{M}_{i,j,k}$ of perfect rainbow matchings for all distinct i, j, k. By taking unions of these matchings for suitable i, j, k we obtain rainbow 2-factors, that is, $\bigcup_{i=1}^{k} \mathcal{M}_{i,i+1 \pmod{k},i}$ is a family of rainbow 2-factors.

2.4. *Rainbow Hamiltonian cycles*

The main result about Hamiltonian cycles that we need is a direct analogue of A3 and A6.

A 7. Let G be a properly coloured, nearly-regular, globally $\frac{1}{2}\mathrm{d}(G)$-bounded graph. Let H be a properly coloured, typical graph on $V(G)$ which is edge-disjoint and colour-disjoint from G. Then, $G \cup H$ has a near-decomposition into rainbow Hamiltonian cycles.

See Lemma 8.27 for a precise statement of this. The proof of A7 consists of first splitting the colours of H at random into two subgraphs H_1 and H_2. Using a result like A4, we have that H_1 and H_2 are both typical. Applying A6 to G and H_1, we get a near-decomposition of $G \cup H_1$ into rainbow 2-factors. Then, we use the typical graph H_2 to modify the 2-factors one-by-one into Hamiltonian cycles. This modification is done by 'rotations' — switching a small number of edges on a 2-factor for edges of H_2 in order to decrease the number of cycles in the 2-factor. After a small number of rotations like this, we create a Hamiltonian cycle.

Theorem 1.10 is proved using A7. The proof is similar to the proof of Theorem 1.9 — starting with a properly coloured K_n, we use analogues of A4 and A5 to get the graphs G and H needed in A7.

2.5. *Rainbow spanning trees*

Here, we explain the proof of Theorem 1.14 — that the Brualdi–Hollingsworth and Kaneko–Kano–Suzuki conjectures hold asymptotically. The starting point of this is to observe that a near-decomposition into rainbow Hamiltonian cycles gives a near-decomposition into rainbow spanning trees. Because of this, our results about Hamiltonian cycles have implications for spanning tree decompositions. The first implication is that if we have a properly coloured K_n with $\leqslant (1 - \varepsilon)n$ colours having $\geqslant (1 - \varepsilon)n/2$ edges, then this K_n has a near-decomposition into rainbow spanning trees (by Theorem 1.10).

Thus, it remains to look at colourings of K_n with $\geqslant (1 - \varepsilon)n$ colours having $\geqslant (1 - \varepsilon)n/2$ edges. In this section, we will focus on the case when the colouring has exactly $n - 1$ colours each having exactly $n/2$ edges. This is the setting of the Brualdi–Hollingsworth Conjecture and is substantially easier to deal with. To deal with this case, we need the following result on how the colours in a random subset of vertices behave.

A 8. Let K_n be properly coloured and choose a subset of $(1 - \varepsilon)n$ vertices $U \subseteq V(K_n)$ at random. Then, $K_n[U]$ is globally $(1 - 2\varepsilon)n/2$-bounded.

See Lemma 5.2 (c) for a precise statement of this. Note that the subgraph $K_n[U]$ from A8 is globally $(1 - 2\varepsilon)n/2$-bounded and has $\mathrm{d}(K_n[U]) = (1 - \varepsilon)n$.

Randomly partition $K_n[U]$ into graphs G' and J, with every edge placed in J independently with probability $p \ll \varepsilon$. Then, randomly partition the colours of G' into sets C_G and C_H, with each colour ending up in C_H independently with probability p. Let G'' and H be the subgraphs of G' consisting of edges with colours in C_G and C_H, respectively. Using results like A8, it can be shown that G'', H and J are all nearly-regular and typical. Since $G'' \subseteq G$, we have that G'' is also globally $(1 - 2\varepsilon)n/2$-bounded. Since $p \ll \varepsilon$ and G had $\mathrm{d}(G) = (1 - \varepsilon)n$, we have that $\mathrm{d}(G'') \approx (1 - \varepsilon - 2p)n \geqslant (1 - 2\varepsilon)n$. Thus, G'' and H satisfy the assumptions of A7, which gives a near-decomposition of $G'' \cup H$ into rainbow Hamiltonian paths.

We now have a set of rainbow paths of length $(1 - \varepsilon)n$ and an edge-disjoint typical subgraph J. We turn the paths into spanning rainbow trees by extending each path one vertex at a time using edges of J. The operations we use to extend the trees are very simple: We always have a collection of rainbow trees $T_1, \dots, T_{(1-\varepsilon)n}$ which we want to enlarge. To enlarge a tree T_i, we find three edges e_1, e_2, e_3 outside $T_1, \dots, T_{(1-\varepsilon)n}$ and two edges f_1, f_2 on T_i so that $T_i' = T_i \cup \{e_1, e_2, e_3\} \setminus \{f_1, f_2\}$ is another rainbow tree. Replacing T_i by T_i' gives us a collection of larger rainbow trees, so by iterating this process we would eventually get rainbow spanning trees. The remaining question is then 'how can we find the edges e_1, e_2, e_3, f_1, f_2 which we use to enlarge T_i?' This is where the typicality of the graph J is used. The fact that J is pseudo-random means that its edges are suitably spread out around $V(K_n)$, and this allows us to find edges in J to switch with edges of T_i.

3. *Preliminaries*

Here we collect some useful notation and results which will be used later in the paper.

3.1. *Basic notation*

For a graph G, the set of edges of G is denoted by $E(G)$ and the set of vertices of G is denoted by $V(G)$. For a vertex v in a graph G, the set of edges in G through v is denoted by $E_G(v)$,

the set of colours of edges going through v is denoted by $C_G(v)$, the set of neighbours of v in G is denoted by $N_G(v)$ and $d_G(v) = |N_G(v)|$. For a coloured graph G and a colour c, the set of colour c edges in G is denoted by $E_G(c)$ and the set of vertices touching colour c edges in G is denoted by $V_G(c)$. In all of these, we omit the 'G' subscript when the graph G is clear from context. We will use additive notation for adding and deleting vertices and edges from graphs.

For a graph G and a set of vertices A, let $G[A]$ denote the induced subgraph of G on A. For disjoint sets of vertices A and B, we use $G[A, B]$ to denote the bipartite subgraph of G on $A \cup B$ consisting of all edges between A and B. For any event E, we let $\mathbf{1}_E$ be the indicator function for E, taking the value 1 when E occurs, and 0 otherwise.

For two functions $f(x_1, \ldots, x_t)$ and $g(y_1, \ldots, y_s)$, we use $f(\pm x_1, \ldots, \pm x_t) = g(\pm y_1, \ldots, \pm y_s)$ to mean that '$\max_{\sigma_i \in \{-1,+1\}} f(\sigma_1 x_1, \ldots, \sigma_t x_t) \leqslant \max_{\sigma_i \in \{-1,+1\}} g(\sigma_1 y_1, \ldots, \sigma_s y_s)$ and also that $\min_{\sigma_i \in \{-1,+1\}} f(\sigma_1 x_1, \ldots, \sigma_t x_t) \geqslant \min_{\sigma_i \in \{-1,+1\}} g(\sigma_1 y_1, \ldots, \sigma_s y_s)$'. The most frequently used case of this notation will to say $x = y \pm z$ for some $z \geqslant 0$, in which case the notation is equivalent to both '$y - z \leqslant x \leqslant y + z$' and '$|x - y| \leqslant z$'.

Note that $a = b \pm c$, $b = d \pm e \Rightarrow a = d \pm c \pm e$. Also note that for any a, b, b' with $|b'| \geqslant |b|$, we have $a \pm b = a \pm b'$. Finally note that the notation is transitive $f(\pm x_1, \ldots, \pm x_t) = g(\pm y_1, \ldots, \pm y_s)$ and $g(\pm y_1, \ldots, \pm y_s) = h(\pm z_1, \ldots, \pm z_r) \Rightarrow$ $f(\pm x_1, \ldots, \pm x_t) = h(\pm z_1, \ldots, \pm z_r)$.

We will often use the following which hold for any $0 \leqslant x < 0.5$.

$$\frac{1}{1-x} \leqslant 1 + 2x \text{ and } \frac{1}{1+x} \geqslant 1 - 2x \tag{1}$$

$$1 - x = (1 \pm x^2)e^{-x} \tag{2}$$

$$1 + x \leqslant e^x \tag{3}$$

$$(1 - x)^t \geqslant 1 - tx \tag{4}$$

$$\sum_{i=1}^{T} e^{-(i-1)x} = (1 \pm x^2 \pm 2e^{-xT})x^{-1}. \tag{5}$$

The last inequality comes from $\sum_{i=1}^{T} e^{-(i-1)x} = (1 \pm x^2) \sum_{i=1}^{T} (1 - x)^{i-1} =$ $(1 \pm x^2)\frac{1-(1-x)^T}{x} = (1 \pm x^2 \pm 2e^{-xT})x^{-1}$. We will also use that, for any $0 \leqslant x \leqslant 0.25$,

$$\frac{1+x}{1-x} \leqslant 1 + 3x. \tag{6}$$

Throughout the paper, most of our results will be either about balanced bipartite graphs or about general graphs. When dealing with balanced bipartite graphs, they will always come with a specific bipartition into two parts usually labelled by 'X' and 'Y' with $|X| = |Y| = n$. When dealing with general graphs, they will usually have $v(G) = n$. Whenever we define a graph G, if we do not specifically say that G is balanced bipartite, we implicitly mean that G is a general graph.

We make two definitions about graphs, which vary slightly depending on whether the graph they are talking about is balanced bipartite or not.

DEFINITION 3.1. • A balanced bipartite graph G with parts X and Y is (γ, δ, n)-regular if $|X| = |Y| = (1 \pm \gamma)n$ and $d_G(v) = (1 \pm \gamma)\delta n$ for every vertex $v \in V(G)$.
• A general graph G is (γ, δ, n)-regular if $|G| = n$ and $d_G(v) = (1 \pm \gamma)\delta n$ for every vertex $v \in V(G)$.

DEFINITION 3.2. • A balanced bipartite graph G with parts X and Y is (γ, δ, n)-typical if it is (γ, δ, n)-regular and we have $d(x, y) = (1 \pm \gamma)\delta^2 n$ for any pair of vertices $x, y \in X$ or $x, y \in Y$.

• A general graph is (γ, δ, n)-typical if it is (γ, δ, n)-regular and for any pair of vertices x, y we have $d(x, y) = (1 \pm \gamma)\delta^2 n$.

DEFINITION 3.3. A graph G is globally b-bounded if G has $\leqslant b$ edges of each colour, that is, if $|E_G(c)| \leqslant b$ for all colours c.

DEFINITION 3.4. A graph G is locally ℓ-bounded if G has $\leqslant \ell$ edges of each colour passing through any vertex $v \in V(G)$, that is, if $\Delta(E_G(c)) \leqslant \ell$ for all colours c.

3.2. Asymptotic notation

For a number $C \geqslant 1$ and $x, y \in (0, 1]$, we use '$x \ll_C y$' to mean '$x \leqslant \frac{y^C}{C}$'. We will write '$x \overset{\text{\tiny POLY}}{\ll} y$' to mean that there is some absolute constant C for which the proof works with '$x \overset{\text{\tiny POLY}}{\ll} y$' replaced by '$x \ll_C y$'. This notation parallels more standard notation '$x \ll y$' which means 'there is a fixed positive continuous function f on (0,1] so that if '$x \ll y$' is replaced by '$x \leqslant f(y)$', then the remainder of the proof works'. (Equivalently, '$x \ll y$' can be interpreted as 'for all $x \in (0, 1]$, there is some $y \in (0, 1]$ such that the remainder of the proof works with x and y'.) The two notations '$x \overset{\text{\tiny POLY}}{\ll} y$' and '$x \ll y$' are largely interchangeable — most of our proofs remain correct with all instances of '$\overset{\text{\tiny POLY}}{\ll}$' replaced by '$\ll$'. The advantage of using '$\overset{\text{\tiny POLY}}{\ll}$' is that it proves polynomial bounds on the parameters (rather than bounds of the form 'for all $\varepsilon > 0$ and sufficiently large n'). This is important towards the end of this paper, where the proofs need polynomial bounds on the parameters.

While the constants C will always be implicit in each instance of '$x \overset{\text{\tiny POLY}}{\ll} y$', it is possible to work them out explicitly. To do this, one should go through the lemmas in the paper in numerical order, choosing the constants C for earlier lemmas before later lemmas. This is because an inequality $x \ll_C y$ in a later lemma may be needed to imply an inequality $x \ll_{C'} y$ from an earlier lemma. Within an individual lemma, we will often have several inequalities of the form $x \overset{\text{\tiny POLY}}{\ll} y$. There the constants C need to be chosen in the reverse order of their occurrence in the text. The reason for this is the same — as we prove a lemma we may use an inequality $x \ll_C y$ to imply another inequality $x \ll_{C'} y$ (and so we should choose C' before choosing C).

Throughout the paper, there are four operations we perform with the '$x \overset{\text{\tiny POLY}}{\ll} y$' notation.

(a) We will use $x_1 \overset{\text{\tiny POLY}}{\ll} x_2 \overset{\text{\tiny POLY}}{\ll} \cdots \overset{\text{\tiny POLY}}{\ll} x_k$ to deduce finitely many inequalities of the form '$p(x_1, \ldots, x_k) \leqslant q(x_1, \ldots, x_k)$' where p and q are monomials with non-negative coefficients and $\min\{i : p(0, \ldots, 0, x_{i+1}, \ldots, x_k) = 0\} < \min\{j : q(0, \ldots, 0, x_{j+1}, \ldots, x_k) = 0\}$, for example, $1000x_1 \leqslant x_2^5 x_4^2 x_5^3$ is of this form.

(b) We will use $x \overset{\text{\tiny POLY}}{\ll} y$ to deduce finitely many inequalities of the form '$x \ll_C y$' for a fixed constant C.

(c) For $x \overset{\text{\tiny POLY}}{\ll} y$ and fixed constants C_1, C_2, we can choose a variable z with $x \ll_{C_1} z \ll_{C_2} y$.

(d) For $n^{-1} \overset{\text{\tiny POLY}}{\ll} 1$ and any fixed constant C, we can deduce $n^{-1} \ll_C \log^{-1} n \ll_C 1$.

To see that (a) is possible, we need to show that for any finite collection $\mathcal{I}$ of inequalities of the given form, we can choose constants $C_1, \ldots, C_{k-1}$ so that $0 < x_1 \ll_{C_1} x_2 \ll_{C_2} \cdots \ll_{C_{k-1}} x_k < 1$ implies all the inequalities in $\mathcal{I}$. To see this, first consider a single inequality '$p(x_1, \ldots, x_k) \leqslant q(x_1, \ldots, x_k)$' of the form in (a). From the assumptions on p and q, we know that $p(x_1, \ldots, x_k) = D_p x_1^{\ell_1} \cdots x_k^{\ell_k}$ and $q(x_1, \ldots, x_k) = D_q x_1^{r_1} \cdots x_k^{r_k}$ for some $D_p, D_q > 0$ and

$\min\{i : \ell_i \neq 0\} < \min\{i : r_i \neq 0\}$. Now, it is easy to check that for $C = r_1 + \cdots + r_k + D_p/D_q$, we have $0 < x_1 \overset{\text{POLY}}{\ll}_C x_2 \overset{\text{POLY}}{\ll}_C \cdots \overset{\text{POLY}}{\ll}_C x_k < 1 \Rightarrow p(x_1, \ldots, x_k) \leqslant q(x_1, \ldots, x_k)$. Now given a finite collection $\mathcal{I}$ of inequalities of the given form, for each $I \in \mathcal{I}$, we can choose a constant C_I so that $0 < x_1 \overset{\text{POLY}}{\ll}_{C_I} x_2 \overset{\text{POLY}}{\ll}_{C_I} \cdots \overset{\text{POLY}}{\ll}_{C_I} x_k < 1 \Rightarrow I$. Letting $C = \max_{I \in \mathcal{I}} C_I$ gives a single constant for which $0 < x_1 \overset{\text{POLY}}{\ll}_C x_2 \overset{\text{POLY}}{\ll}_C \cdots \overset{\text{POLY}}{\ll}_C x_k < 1$ implies all the inequalities in $\mathcal{I}$.

We remark that occasionally we will use a slight strengthening of (a), when p and q are *multinomials* with non-negative coefficients and $\min\{i : p(0, \ldots, 0, x_{i+1}, \ldots, x_k) = 0\} < \min\{j : q(0, \ldots, 0, x_{j+1}, \ldots, x_j) = 0\}$, for example, $50x_1x_2 + 5x_2^2 \leqslant x_3^5x_4^2 + x_1^2x_5^3$ is of this form. This strengthening can be reduced to the monomial version. To do this, consider multinomials p and q with non-negative coefficients and an integer i for which $p(0, \ldots, 0, x_{i+1}, \ldots, x_k) = 0$ and $q(0, \ldots, 0, x_{i+1}, \ldots, x_k) \neq 0$. Let D_p be the sum of the coefficients of p and note that the monomial $\hat{p} = D_p x_i$ satisfies $\hat{p} \geqslant p$ (for $0 < x_1 \leqslant \cdots \leqslant x_k < 1$). Letting D_q be the smallest coefficient of q and d the degree of q, note that the monomial $\hat{q} = D_q x_{i+1}^d$ satisfies $\hat{q} \leqslant q$ (for $0 < x_1 \leqslant \cdots \leqslant x_k < 1$). Thus, we can use the monomial version of (a) to get constants $C_1, \ldots, C_{k-1}$ so that $0 < x_1 \ll_{C_1} x_2 \ll_{C_2} \cdots \ll_{C_{k-1}} x_k < 1$ implies $\hat{p} \leqslant \hat{q}$ and hence also $p \leqslant q$.

Note that (b) is just a special case of (a) since the inequality '$x \ll_C y$' is of the form of the inequalities in (a). Operation (b) is important because it allows us to plug one instance of the '$\overset{\text{POLY}}{\ll}$' notation into another one. As an example, suppose that we have proved a lemma which assumes '$a \overset{\text{POLY}}{\ll} b$'. This means that we have proved that there is some explicit constant C for which the lemma holds with '$a \overset{\text{POLY}}{\ll} b$' replaced by '$a \ll_C b$'. Now if we subsequently have variables x, y with $x \overset{\text{POLY}}{\ll} y$, then (b) guarantees that we can plug x and y into the earlier lemma with $a = x$ and $b = y$.

For operation (c), note that for $C = C_1 C_2^{C_1 C_2}$, if we have numbers x, y with $x \ll_C y$, then the number $z = y^{C_2}/C_2$ satisfies $x \ll_{C_1} z \ll_{C_2} y$. Operation (c) is important because it allows us to introduce new variables inside our proof. For example, if we have a lemma which assumes $x \overset{\text{POLY}}{\ll} y$, then in the proof of the lemma we can say 'choose z with $x \overset{\text{POLY}}{\ll} z \overset{\text{POLY}}{\ll} y$'. Here the constants C_1 and C_2 in '$x \ll_{C_1} z \ll_{C_2} y$' are chosen first, and operation (c) guarantees that we can later choose a constant for '$x \ll_C y$'.

For operation (d), note that '$n^{-1} \ll_C \log^{-1} n \ll_C 1$' is equivalent to '$\frac{1}{C^{1/c}} n^{1/c} \geqslant \log n \geqslant C$' which is true for sufficiently large n. Operation (d) is important because it allows us to use $n^{-1} \overset{\text{POLY}}{\ll} 1$ to deduce any instance of $n^{-1} \overset{\text{POLY}}{\ll} \log^{-1} n \overset{\text{POLY}}{\ll} 1$.

How does our '$\overset{\text{POLY}}{\ll}$' notation compare with the standard '$\ll$' notation? Versions of the operations (a), (b) and (c) work with the '$\ll$' notation as well. Particularly, (a) is more versatile with '$\ll$', because it is possible to show that $x_1 \ll x_2 \ll \cdots \ll x_k$ can be used to deduce finitely many inequalities of the form '$p(x_1, \ldots, x_k) \leqslant q(x_1, \ldots, x_k)$' where p and q are arbitrary positive continuous functions on $(0,1]$ satisfying $\min\{i : p(0, \ldots, 0, x_{i+1}, \ldots, x_k) = 0\} < \min\{j : q(0, \ldots, 0, x_{j+1}, \ldots, x_j) = 0\}$ (rather than multinomials). Operation (d) however has no analogue for the '$\ll$' notation (the natural analogue would be that 'for $n^{-1} \ll 1$ and any positive continuous f, g on $(0,1]$ we can deduce $n^{-1} \leqslant f(\log^{-1} n) \leqslant g(1)$'. However this is not true for $f(x) = 0.5e^{-1/x}$). Because of this, in our proofs, the '$\overset{\text{POLY}}{\ll}$' and '$\ll$' notations are interchangeable whenever operation (d) is not used (while when operation (d) is used, we need to use the '$\overset{\text{POLY}}{\ll}$' notation).

3.3. Probabilistic tools

We will use the following cases of the Bonferroni Inequalities.

LEMMA 3.5 (Bonferroni Inequalities). *Let $X_1, \ldots, X_n$ be events in a probability space. Then,*

$$\mathbb{P}(\cup_{i=1}^n X_i) \geqslant \sum_{i=1}^n \mathbb{P}(X_i) - \sum_{i=1}^n \sum_{j=1}^{i-1} \mathbb{P}(X_i \cap X_j).$$

Given a probability space $\Omega = \prod_{i=1}^n \Omega_i$ and a random variable $X : \Omega \to \mathbb{R}$ we make the following definitions.

- Suppose that there is a constant C such that changing $\omega \in \Omega$ in any one coordinate changes $X(\omega)$ by at most C. Then, we say that X is C-Lipschitz.
- For $i \in \{1, \ldots, n\}$, we say that X is uninfluenced by i if $\omega_j = \omega_j'$ for $j \neq i \Rightarrow X(\omega) = X(\omega')$. Otherwise we say that X is influenced by i.

We will use the following concentration inequalities

LEMMA 3.6 (Azuma's Inequality). *Suppose that X is C-Lipschitz and influenced by $\leqslant m$ coordinates in $\{1, \ldots, n\}$. Then, for any $t > 0$,*

$$\mathbb{P}(|X - \mathbb{E}(X)| > t) \leqslant 2e^{\frac{-t^2}{mC^2}}.$$

Note that the bound in the above inequality can be rewritten as $\mathbb{P}(X \neq \mathbb{E}(X) \pm t) \leqslant 2e^{\frac{-t^2}{mC^2}}$. A sequence of random variables $X_0, X_1, X_2, \ldots$ is a supermartingale if $\mathbb{E}(X_{t+1}|X_0, \ldots, X_t) \leqslant X_t$ for all t.

LEMMA 3.7 (Azuma's Inequality for Supermartingales). *Suppose that $Y_0, Y_1, \ldots, Y_n$ is a supermartingale with $|Y_i - Y_{i-1}| \leqslant C$ for each $i \in [n]$. Then, for any $t > 0$,*

$$\mathbb{P}(Y_n > Y_0 + t) \leqslant e^{\frac{-t^2}{2nC^2}}.$$

LEMMA 3.8 (Chernoff Bound). *Let X be the binomial random variable with parameters (n, p). Then, for $\varepsilon \in (0, 1)$, we have*

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

LEMMA 3.9 (Greenhill, Isaev, Kwan, McKay [**30**]). *Let $\binom{[N]}{r}$ be the set of r-subsets of $\{1, \ldots, N\}$ and let $h : \binom{[N]}{r} \to \mathbb{R}$ be given. Let C be a uniformly random element of $\binom{[N]}{r}$. Suppose that there exists $\alpha \geqslant 0$ such that $|h(A) - h(A')| \leqslant \alpha$ for any $A, A' \in \binom{[N]}{r}$ with $|A \cap A'| = r - 1$. Then, for any $t > 0$,*

$$\mathbb{P}(|h(C) - \mathbb{E}h(C)| \geqslant t) \leqslant 2e^{-\frac{2t^2}{\alpha^2 \min(r, N-r)}}.$$

4. *Finding one rainbow matching probabilistically*

The goal of this section is to prove that every properly coloured d-regular, globally $(1 + o(1))d$-bounded balanced bipartite graph has a nearly-spanning rainbow matching M. This matching is found using a randomized process, which allows us to prove that every edge ends up in M with at least the expected probability d^{-1}. It will be more convenient for us to prove the result for graphs which are approximately regular rather than regular. Thus, throughout this section, we will always deal with (γ, δ, n)-regular graphs for suitable parameters. See Lemma 4.6 for a precise statement of the result we prove.

The random process that we use to find a rainbow matching is a variation of the semi-random method introduced by Rödl. We remark that in the case when the graph G has exactly d edges of each colour, then our results follow directly from standard versions of the Rödl Nibble (this is done by first expressing the problem in terms of finding a matching in an uncoloured 3-uniform hypergraph, and then using, for example, [**7**, Theorem 4.7.1]). Thus, the difficult case of the result we aim to prove is when G is a graph in which some colour classes have size much smaller than d. We deal with this situation by using a balancing coin flips approach to keep our graphs nearly-regular.

Random process.   Let G be a coloured balanced bipartite graph which is (γ, δ, n)-regular and globally $(1 + o(1))d$-bounded. We describe a randomized process which will find a rainbow matching M of size $(1 - o(1))n$ in G with high probability. The process will last for T rounds. In each round, we will focus on some subgraph G_t of G and partition G_t into a rainbow matching M_t and a vertex-disjoint, colour-disjoint graph G_{t+1}. At the end of the process, we will have a collection of vertex-disjoint, colour-disjoint matchings $M_1, \ldots, M_T$, and so letting $M = M_1 \cup \cdots \cup M_T$, we get a rainbow matching. We will prove that with high probability $e(M) = (1 - o(1))n$.

Individual rounds.   To partition G_t into M_{t+1} and G_{t+1}, in each round, we use a random process which we call an (α, b)-*random edge-assignment*. Let the parts of the bipartition of G_t be called X and Y. The definition of the (α, b)-random edge-assignment is the following:

- first, we activate every vertex of X with probability α;
- for every activated vertex x, we choose a random neighbour y_x of x in Y, and say the the edges xy_x is *chosen*;
- let M_{t+1} be the matching formed by all the edges of the form xy_x whose colour is not the colour of any other chosen edge $x'y_{x'}$ and for which y_x does not appear on any other chosen edge $x'y_{x'}$;
- let H be the subgraph of G on $V(G) \setminus V(M_{t+1})$ consisting of all the edges whose colours do not occur on any chosen edge;
- delete every edge xy on H with probability $\frac{\alpha b}{d(x)} - \frac{\alpha |E(c(xy))|}{d(x)}$ to get G_{t+1}.

Suppose that G_t is $(\gamma_t, \delta_t, n_t)$-regular and globally $(1 + \gamma_t)\delta_t n_t$-bounded. We will run an $(\alpha, (1 + \gamma_t)\delta_t n_t)$-random edge-assignment on G_t and estimate the probabilities of edges and vertices of G_t ending up in M_t or G_{t+1}.

$$\mathbb{P}(v \in V(G_{t+1})) \approx 1 - \alpha \approx e^{-\alpha} \qquad \text{for any vertex } v \in V(G_t)$$

$$\mathbb{P}(\{x, y\} \subseteq V(G_{t+1})) \approx 1 - 2\alpha \approx e^{-2\alpha} \qquad \text{for any pair } \{x, y\} \subseteq V(G_t)$$

$$\mathbb{P}(y \in N_{G_{t+1}}(x) | x \in V(G_t)) \approx 1 - 2\alpha \approx e^{-2\alpha} \qquad \text{for any } y \in N_{G_t}(x)$$

$$\mathbb{P}(e \in E(G_{t+1})) \approx 1 - 3\alpha \approx e^{-3\alpha} \qquad \text{for any edge } e \in E(G_t) \qquad (7)$$

$$\mathbb{P}(e \in E(M_{t+1})) \approx \frac{\alpha}{\delta_t n_t} \qquad \text{for any edge } e \in E(G_t). \qquad (8)$$

Using linearity of expectation, we can estimate the expected number of vertices, degrees of vertices, and sizes of colour classes in G_{t+1}.

$$\mathbb{E}(|X \cap V(G_{t+1})|) = \mathbb{E}(|Y \cap V(G_{t+1})|) \approx e^{-\alpha} n_t$$

$$\mathbb{E}(|E_{G_{t+1}}(c)|) \lesssim e^{-2\alpha} \delta_t n_t \qquad \text{for any colour } c$$

$$\mathbb{E}(d_{G_{t+1}}(x)) \approx e^{-2\alpha} \delta_t n_t = (e^{-\alpha}\delta)(e^{-\alpha} n_t) \qquad \text{for any vertex } x \in V(H_\omega)$$

It can be shown that the quantities above are Lipschitz, and so by Azuma's Inequality they are concentrated around their expectation with high probability. This implies that with high probability G_{t+1} is $(\gamma_{t+1}, e^{-\alpha}\delta_t, e^{-\alpha}n_t)$-regular and globally $(1 + \gamma_{t+1})(e^{-\alpha}\delta_t)(e^{-\alpha}n_t)$-bounded for some suitable error γ_{t+1}.

Iterating. Let $G_0 = G$ be a coloured graph which is (γ, δ, n)-regular and globally $(1 + o(1))\delta n$-bounded. We iteratively construct graphs $G_1, \ldots, G_T$ and matchings $M_1, \ldots, M_T$ — at step t we run an $(\alpha, (1 + o(1))e^{-2\alpha t}\delta n)$-random edge-assignment on G_t in order to obtain M_{t+1} and G_{t+1}.

From the previous section, we have that, with suitable errors $\gamma_1, \ldots, \gamma_T$, the following hold for all t with high probability.

(i) G_t is $(\gamma_t, e^{-\alpha t}\delta, e^{-\alpha t}n)$-regular.
(ii) G_t is globally $(1 + \gamma_t)(e^{-\alpha t}\delta)(e^{-\alpha t}n)$-bounded.

In particular, if $T = \omega(\alpha^{-1})$, then (i) implies that $|V(G_T)| \lesssim (1 + \gamma_T)e^{-\alpha T}n = o(n)$. Since $M_1, \ldots, M_T$ are vertex-disjoint, colour-disjoint rainbow matchings with $|V(G)| = |V(G_T)| + |\bigcup_{i=1}^T V(M_i)|$, we get that $M = \bigcup_{i=1}^T V(M_i)$ is a rainbow matching of size $(1 - o(1))n$ in G.

Showing that the matching is random. It remains to show that for any edge $e \in E(G)$, the probability that e is in M is (approximately) at least $(\delta n)^{-1}$. First note that (7) implies $\mathbb{P}(e \in E(G_t)) = \prod_{i=0}^t \mathbb{P}(e \in G_i | e \in G_{i-1}) \gtrsim e^{-3t\alpha}$. Combining this with (8), and (5), we get

$$\mathbb{P}(e \in E(M_1 \cup \cdots \cup M_T)) = \sum_{t=0}^T \mathbb{P}(e \in G_t)\mathbb{P}(e \in M_{t+1} | e \in G_t)$$

$$\gtrsim \sum_{t=0}^T \left(e^{-3t\alpha}\right) \left(\frac{\alpha}{(e^{-\alpha t}\delta)(e^{-\alpha t}n)}\right) = \frac{\alpha}{\delta n} \sum_{t=0}^T e^{-\alpha t} \gtrsim \frac{1}{\delta n}.$$

This concludes the proof sketch in this section. The main thing we need to do in the full proof is to keep track of the errors γ_t and make sure that they do not get too big.

4.1. Formal definition of the random edge assignment

Here we formally define the probability space of the (α, b)-random edge-assignment which runs on a graph G. The process will depend on two parameters α and b. The graph G will be a globally b-bounded balanced bipartite graph with parts X and Y. The process has a coordinate for every vertex in X, and a coordinate for every edge $e \in E(G)$ (the balancing coin flips).

• *Vertex choices:* For $x \in X$, the vertex x is activated with probability α. Every activated vertex *chooses* a neighbour y_x of x uniformly at random from its neighbours.
• *Balancing coin flips:* For $xy \in E(G)$, the edge xy is *killed* with probability $\frac{\alpha b}{d(x)} - \frac{\alpha|E(c(xy))|}{d(x)}$.

We say that an edge $xy \in E(G)$ is *chosen* if x is activated and chooses y. We say that a colour c is *chosen* if some colour c edge is chosen. We construct a matching M and graphs Γ, H depending on the process as follows.

$$M = \{xy \in E(G) : xy \text{ is chosen, and no } x'y' \in E(G) \setminus \{xy\} \text{ is chosen}$$

$$\text{with } y' = y \text{ or } c(x'y') = c(xy)\}$$

$$V(\Gamma) = V(G)$$

$$E(\Gamma) = \{e \in E(G) : c(e) \text{ is not chosen and } e \text{ is not killed}\}$$

$$H = \Gamma[V(G) \setminus V(M)].$$

We say that the M, Γ and H are *produced* by the process. Note that by the definitions of M and H, we always have that M is a rainbow matching, that $V(M)$ and $V(H)$ partition $V(G)$, and that M and H share no colours.

4.2. *Probabilities*

To analyse various features of (α, b)-random edge-assignments, we need estimates of the probability of various events. The following lemma computes all the probability estimates required.

LEMMA 4.1. *Suppose that we have d, b, α, ℓ and γ with $(1 + \gamma)d \geqslant b$ and $\ell d^{-1} \leqslant \alpha \leqslant \gamma \leqslant 0.01$.*
Let G be a coloured balanced bipartite graph which is $(\gamma, d/n, n)$-regular, globally b-bounded, and locally ℓ-bounded. Let M, Γ, H be produced by an (α, b)-random edge-assignment on G. Then, the following probability bounds (10)–(23) hold.

Proof. Let the bipartition classes of G be X and Y. We will often use the following

$$\frac{1}{d(v)} = (1 \pm 2\gamma)\frac{1}{d} \quad \text{for any } v \in V(G). \tag{9}$$

This comes from the $(\gamma, d/n, n)$-regularity of G and (1).

$$\mathbb{P}(xy \text{ chosen}) = \frac{\alpha}{d}(1 \pm 2\gamma) \quad \text{for any } xy \in E(G). \tag{10}$$

This comes from $\mathbb{P}(xy \text{ chosen}) = \frac{\alpha}{d(x)}$ and (9).

$$\mathbb{P}(xy \text{ killed}) = \left(\frac{\alpha b}{d} - \frac{\alpha|E(c(e))|}{d}\right)(1 \pm 2\gamma) \quad \text{for any } xy \in E(G). \tag{11}$$

This comes from $\mathbb{P}(xy \text{ killed}) = \frac{\alpha b}{d(x)} - \frac{\alpha|E(c(e))|}{d(x)}$ and (9).

$$\mathbb{P}(e \text{ and } e' \text{ chosen}) \leqslant \frac{\alpha^2}{d^2}(1 + 5\gamma) \quad \text{for } e \neq e' \in E(G). \tag{12}$$

If $e \cap X = e' \cap X$, then both e and e' cannot be chosen, so we may assume that $e = xy$ and $e' = x'y'$ for $x \neq x'$. The events that xy and $x'y'$ are chosen are independent which gives $\mathbb{P}(\text{edges } xy \text{ and } x'y' \text{ chosen}) = \frac{\alpha^2}{d(x)d(x')}$. Now (12) comes from (9) and $\gamma \leqslant 0.01$.

$$\mathbb{P}(c \text{ chosen}) = \frac{\alpha|E(c)|}{d}(1 \pm 4\gamma) \quad \text{for any } c \in C(G). \tag{13}$$

By the union bound and (10), we have that c is chosen with probability $\leqslant \sum_{e \in E(c)} \mathbb{P}(e \text{ chosen}) \leqslant |E(c)|(1 + 2\gamma)\frac{\alpha}{d}$. By the Bonferroni inequalities (see Lemma 3.5), (10) and (12), we have the bound $\mathbb{P}(\text{colour } c \text{ chosen}) \geqslant \sum_{e \in E(c)} \mathbb{P}(e \text{ chosen}) - \sum_{\substack{e, e' \in E(c) \\ e \neq e'}} \mathbb{P}(e \text{ and } e' \text{ chosen}) \geqslant |E(c)|(1 - 2\gamma)\frac{\alpha}{d} - \binom{|E(c)|}{2}(1 + 5\gamma)\frac{\alpha^2}{d^2}$. The lower bound in (13) then comes from $|E(c)| \leqslant b \leqslant (1 + \gamma)d$ and $\alpha \leqslant \gamma \leqslant 0.01$.

$$\mathbb{P}(xy \text{ chosen and } xy \notin E(M)) \leqslant \frac{3\alpha^2}{d} \quad \text{for any } xy \in E(G). \tag{14}$$

From the definition of M, the only way $xy \notin E(M)$ can hold for a chosen edge xy is if another edge $x'y'$ is chosen with either $y' = y$ or $c(x'y') = c(xy)$. By the union

bound, we have $\mathbb{P}(xy \text{ chosen and } xy \notin E(M)) \leqslant \sum_{x' \in N(y) \setminus \{x\}} \mathbb{P}(xy \text{ chosen and } x'y \text{ chosen}) + \sum_{x'y' \in C(xy) \setminus \{xy\}} \mathbb{P}(xy \text{ chosen and } x'y' \text{ chosen})$. Using (12), $|E(c)| \leqslant b \leqslant (1 + \gamma)d$, $\Delta(G) \leqslant (1 + \gamma)d$, and $\gamma \leqslant 0.01$, this is at most $(d(y) + |E(c)|)(1 + 5\gamma)\frac{\alpha^2}{d^2} \leqslant \frac{3\alpha^2}{d}$, as required.

$$\mathbb{P}(e \notin \Gamma) = (1 \pm 9\gamma)\frac{\alpha b}{d} \quad \text{for any } e \in E(G). \tag{15}$$

Since e is killed independently of any colour being chosen, we have $\mathbb{P}(e \notin \Gamma) = \mathbb{P}(c(e) \text{ chosen}) + \mathbb{P}(e \text{ killed}) - \mathbb{P}(c(e) \text{ chosen})\mathbb{P}(e \text{ killed})$. Combining this with (13), (11), $|E(c(e))| \leqslant b \leqslant (1 + \gamma)d$ and $\alpha \leqslant \gamma \leqslant 0.01$, we get $\mathbb{P}(e \notin \Gamma) = \frac{\alpha|E(c(e))|}{d}(1 \pm 4\gamma) + (\frac{\alpha b}{d} - \frac{\alpha|E(c(e))|}{d})(1 \pm 2\gamma) - (1 \pm 7\gamma)\frac{\alpha|E(c(e))|}{d}(\frac{\alpha b}{d} - \frac{\alpha|E(c(e))|}{d}) = (1 \pm 9\gamma)\frac{b\alpha}{d}$.

$$\mathbb{P}(xy \in E(M)) = \frac{\alpha}{d}(1 \pm 5\gamma) \quad \text{for any } xy \in E(G). \tag{16}$$

Recall that M contains only chosen edges. Using this, the upper bound comes from (10), while the lower bound comes from (10), (14) and $\alpha \leqslant \gamma$.

$$\mathbb{P}(v \in V(M)) = \alpha(1 \pm 7\gamma) \quad \text{for any } v \in V(G). \tag{17}$$

Recall that M is a matching, which implies that the events '$vu \in E(M)$' are disjoint for $u \in N(v)$. Using (16) and that $d(v) = (1 \pm \gamma)d$, this gives $\mathbb{P}(v \in V(M)) = \sum_{u \in N(v)} \mathbb{P}(vu \in E(M)) = d(v) \cdot \frac{\alpha}{d}(1 \pm 5\gamma) = \alpha(1 \pm 7\gamma)$.

$$\mathbb{P}(u, v \in V(M)) \leqslant 3\alpha^2 \quad \text{for any vertices } u \neq v \in V(G). \tag{18}$$

Note that

$$\mathbb{P}(u, v \in V(M)) \leqslant \sum_{\substack{z \in N(u), \\ w \in N(v)}} \mathbb{P}(uz, vw \text{ chosen}) = \mathbb{P}(uv \text{ chosen}) + \sum_{\substack{z \in N(u), \\ w \in N(v), \\ uz \neq vw}} \mathbb{P}(uz, vw \text{ chosen}).$$

Here the first term is defined to be zero if there is no edge uv in G. Using (10), (12) and $\Delta(G) \leqslant (1 + \gamma)d$, we get that this is at most $(1 + 2\gamma)\frac{\alpha}{d} + (1 + \gamma)^2 d^2 \cdot (1 + 5\gamma)\frac{\alpha^2}{d^2}$ which, combined with $d^{-1} \leqslant \alpha \leqslant \gamma \leqslant 0.01$, implies the result.

$$\mathbb{P}(\{u, v\} \cap V(M) \neq \emptyset) = 2\alpha(1 \pm 11\gamma) \quad \text{for any vertices } u \neq v \in V(G). \tag{19}$$

This comes from the Bonferroni inequalities together with (17), (18) and $\alpha \leqslant \gamma$.

$$\mathbb{P}(xy \text{ chosen and } v \in V(M)) \leqslant \frac{\alpha^2}{d}(1 + 7\gamma) \quad \text{for } xy \in E(G) \text{ and } v \notin \{x, y\}. \tag{20}$$

By the union bound, (12) and $\Delta(G) \leqslant (1 + \gamma)d$, this probability is $\leqslant \sum_{u \in N(v)} \mathbb{P}(xy \text{ and } vu \text{ chosen}) \leqslant d(v)(1 + 5\gamma)\frac{\alpha^2}{d^2} \leqslant (1 + 7\gamma)\frac{\alpha^2}{d}$.

$$\mathbb{P}(e \notin \Gamma \text{ and } v \in V(M)) \leqslant 6\alpha^2 \quad \text{for } e \in E(G) \text{ and } v \in V(G). \tag{21}$$

By the union bound, $\mathbb{P}(e \notin \Gamma \text{ and } v \in V(M)) \leqslant \mathbb{P}(e \text{ killed and } v \in V(M)) + \mathbb{P}(c(e) \text{ chosen and } v \in V(M))$. Using (11), (17) and $b \leqslant (1 + \gamma)d$, the first term can be bounded above by $\mathbb{P}(e \text{ killed and } v \in V(M)) = \mathbb{P}(e \text{ killed})\mathbb{P}(v \in V(M)) \leqslant (\frac{\alpha b}{d} - \frac{\alpha|E(c(e))|}{d})(1 + 2\gamma)\alpha(1 + 7\gamma) \leqslant 3\alpha^2$. Let $E_{c(e),v}$ be the set of $\leqslant \ell$ colour $c(e)$ edges through v. The second term can be bounded by

$$\mathbb{P}(c(e) \text{ chosen and } v \in V(M)) \leqslant \sum_{e' \in E_{c(e),v}} \mathbb{P}(e' \text{ chosen}) + \sum_{\substack{e' \notin E_{c(e),v}, \\ c(e') = c(e)}} \mathbb{P}(e' \text{ chosen and } v \in V(M)).$$

Using (10), (20), $|E_{c(e),v}| \leqslant \ell \leqslant \alpha d$, $d^{-1} \leqslant \alpha \leqslant \gamma \leqslant 0.01$ and $b \leqslant (1+\gamma)d$, this is at most $(1 + 2\gamma)(\alpha d)\frac{\alpha}{d} + |E(c)|\frac{\alpha^2}{d}(1+7\gamma) \leqslant 3\alpha^2$.

$$\mathbb{P}(e \in \Gamma \text{ and } v \notin V(M)) = 1 - \alpha - \frac{\alpha b}{d} \pm 22.5\alpha\gamma$$

$$= (1 \pm 23\alpha\gamma)\left(1 - \alpha - \frac{\alpha b}{d}\right) \quad \text{for } e \in E(G) \text{ and } v \in V(G). \quad (22)$$

This comes from '$\mathbb{P}(A \text{ and } B) = 1 - \mathbb{P}(\overline{A}) - \mathbb{P}(\overline{B}) + \mathbb{P}(\overline{A} \text{ and } \overline{B})$' together with (15), (17), (21), $\alpha \leqslant \gamma \leqslant 0.01$ and $b \leqslant (1+\gamma)d$.

$$\mathbb{P}(xy \notin E(H)) = \mathbb{P}(xy \notin E(\Gamma) \text{ or } x \in V(M) \text{ or } y \in V(M))$$

$$= \left(2\alpha + \frac{b\alpha}{d}\right)(1 \pm 40\gamma) \quad \text{for } xy \in E(G). \quad (23)$$

This comes from the Bonferroni inequalities together with (15), (17), (18), (21) and $\alpha \leqslant \gamma$. □

4.3. *Expectations*

Using the probabilities in the previous section, it is immediate to compute the expectations of relevant quantities.

LEMMA 4.2. *Suppose that we have* d, ℓ, α, γ *with* $\ell d^{-1} \leqslant \alpha \leqslant \gamma \leqslant 0.01$.
Let G *be a coloured bipartite graph which is* $(\gamma, d/n, n)$-*regular, globally* $(1+\gamma)d$-*bounded and locally* ℓ-*bounded. Let* M, Γ *and* H *be the graphs produced by an* $(\alpha, (1+\gamma)d)$-*random edge-assignment on* G. *The following hold:*

- $\mathbb{E}(|X \cap V(H)|) = (1 \pm (1+10\alpha)\gamma)(1-\alpha)n$
- $\mathbb{E}(|Y \cap V(H)|) = (1 \pm (1+10\alpha)\gamma)(1-\alpha)n$
- $\mathbb{E}(|E_H(c)|) \leqslant \mathbb{E}(|\{uv \in E_G(c) : \{u,v\} \cap V(M) = \emptyset\}|) \leqslant (1 + 24\alpha\gamma)(1-2\alpha)|E_G(c)|$ for any colour c;
- $\mathbb{E}(d_H(x)) = \mathbb{E}(|N_\Gamma(x) \setminus V(M)|) = (1 \pm (1+26\alpha)\gamma)(1-2\alpha)d$ for any vertex $x \in V(H)$.

Proof. These are immediate from linearity of expectation, (17), (19), (22) and the $(\gamma, d/n, n)$-regularity of G. □

4.4. *Concentration*

By Azuma's Inequality, the random variables considered in the previous section are concentrated around their expectations.

LEMMA 4.3. *Suppose that we have* $n, \delta, \gamma, \alpha, \ell$ *with* $n^{-0.001} \leqslant \alpha \leqslant \gamma \leqslant 0.00001$, $\gamma \leqslant \delta \leqslant 1$ *and* $\ell \leqslant n^{0.001}$.
Let G *be a coloured bipartite graph with bipartition classes* X *and* Y *which is* (γ, δ, n)-*regular, locally* ℓ-*bounded and globally* $(1+\gamma)\delta n$-*bounded. Let* M, Γ *and* H *be the graphs produced by an* $(\alpha, (1+\gamma)\delta n)$-*random edge-assignment on* G. *The following hold with probability* $\geqslant 1 - n^{-2}$.

(i) $|X \cap V(H)| = |Y \cap V(H)| = (1 \pm (1+12\alpha)\gamma)(1-\alpha)n$.
(ii) $|E_H(c)| \leqslant |E(c) \setminus V(M)| \leqslant (1 + 26\alpha\gamma)(1-2\alpha)\delta n$ for every colour c.
(iii) $d_H(v) = |N_\Gamma(v) \setminus V(M)| = (1 \pm (1+30\alpha)\gamma)(1-2\alpha)\delta n$ for every vertex $v \in V(H)$.

Proof. First, we prove the Lipschitzness of the relevant random variables.

CLAIM 4.4. $|X \cap V(H)|$, $|Y \cap V(H)|$, $|E(c) \setminus V(M)|$ and $|N_\Gamma(v) \setminus V(M)|$ are all 26ℓ-Lipschitz for any colour c and vertex v.

Proof. Consider two $(\alpha, (1+\gamma)d)$-random edge-assignments which differ on one coordinate — Edge-Assignment 1 which produces graphs M_1, Γ_1, H_1 and Edge-Assignment 2 which produces graphs M_2, Γ_2, H_2. Furthermore, let C_1 and C_2 be the colours chosen, respectively, by the two edge-assignments, and let K_1 and K_2 be the edges killed, respectively, by the two edge-assignments. We will show that $|V(M_1)\Delta V(M_2)| \leqslant 20$, $|E(K_1)\Delta E(K_2)| \leqslant 1$ and $|C_1\Delta C_2| \leqslant 2$.

First, note that C_1 and C_2 only differ in the colour of some edge xy if xy is chosen by one assignment and not the other.

Suppose that the coordinate on which the two edge-assignments differ is a balancing coin flip on an edge xy. Note that $M_1 = M_2$, $C_1 = C_2$ and K_1 and K_2 can differ only on the edge xy, so that, as required, $|V(M_1)\Delta V(M_2)| \leqslant 20$, $|K_1\Delta K_2| \leqslant 1$ and $|C_1\Delta C_2| \leqslant 2$.

Suppose that the coordinate on which the two edge-assignments differ is a vertex-activation choice for a vertex $x \in X$, which is, say, activated in Edge-Assignment 1 but not in Edge-Assignment 2. Say that y is chosen by x in Edge-Assignment 1. Either $M_1 = M_2$, or $M_1 = M_2 + xy$, or M_1 is M_2 with up to two edges removed — edges $x'y'$ with $y' = y$ or $c(xy) = c(x'y')$. Thus, we have $|V(M_1)\Delta V(M_2)| \leqslant 4$. As $C_1 = C_2 \cup \{c(xy)\}$ and $K_1 = K_2$, we have $|E(K_1)\Delta E(K_2)| \leqslant 1$ and $|C_1\Delta C_2| \leqslant 2$.

Suppose finally that the coordinate on which the two edge-assignments differ is a vertex-choice for a vertex $x \in X$. Note that if x is not activated, then the outcome of the edge-assignments is the same and $C_1 = C_2$, so we can assume that x is activated. Let y_x^1 and y_x^2 be the vertices chosen by x in Edge-Assignments 1 and 2, respectively.

Note that xy_x^1 and xy_x^2 are the only edges which may be chosen by one, but not by both assignments. Hence $c(xy_x^1)$ and $c(xy_x^2)$ are the only colours which may be chosen by one, but not by both assignments, so that $|C_1\Delta C_2| \leqslant 2$. The two rainbow matchings M_1 and M_2 can only differ on edges sharing a vertex or a colour with one of the edges xy_x^1 or xy_x^2. Note that M_1 has at most one edge touching each of the vertices x, y_x^1 and y_x^2 (since M_1 is a matching), and has at most one edge of each of the colours $c(xy_x^1)$ and $c(xy_x^2)$ (since M_1 is rainbow). Thus, $e(M_1 \setminus M_2) \leqslant 5$. Similarly, $e(M_2 \setminus M_1) \leqslant 5$. This implies that $|V(M_1)\Delta V(M_2)| \leqslant 20$. Furthermore, $K_1 = K_2$, so certainly $|K_1\Delta K_2| \leqslant 1$

Thus, we always have that $|V(M_1)\Delta V(M_2)| \leqslant 20$, $|K_1\Delta K_2| \leqslant 1$ and $|C_1\Delta C_2| \leqslant 2$. By the definition of H_1 and H_2, we then have $V(H_1)\Delta V(H_2) = V(M_1)\Delta V(M_2)$ which implies $|(X \cap V(H_1))\Delta(X \cap V(H_2))|, |(Y \cap V(H_1))\Delta(Y \cap V(H_2))| \leqslant 20$. For a colour c, $E(c) \setminus V(M_1)$ and $E(c) \setminus V(M_2)$ can only differ on colour c edges passing through $V(M_1)\Delta V(M_2)$. Combined with local ℓ-boundedness, this gives $|(E(c) \setminus V(M_1))\Delta(E(c) \setminus V(M_2))| \leqslant \ell|V(M_1)\Delta V(M_2)| \leqslant 20\ell$. For a vertex v, $N_{\Gamma_1}(v) \setminus V(M_1)$ and $N_{\Gamma_2}(v) \setminus V(M_2)$ can differ only on vertices of $V(M_1)\Delta V(M_2)$, on vertices of $K_1\Delta K_2$, or on vertices $z \in N(v)$ with vz having colour in $C_1\Delta C_2$. Combined with local ℓ-boundedness, this gives $|(N_{\Gamma_1}(v) \setminus V(M_1))\Delta(N_{\Gamma_2}(v) \setminus V(M_2))| \leqslant 2\ell + 22 \leqslant 26\ell$. □

Note that $|X \cap V(H)|$, $|Y \cap V(H)|$ and $|E(c) \setminus V(M)|$ are influenced only by the choices of the vertices $x \in X$ and which vertices in X are activated, but not which edges are killed. Furthermore, $|N_\Gamma(v) \setminus V(M)|$ is influenced only by the choices of the vertices $x \in X$ and which vertices in X are activated and which edges between v and $N_G(v)$ are killed. There are at most $(1+\gamma)n$ vertices in X, and $d_G(v) \leqslant (1+\gamma)n$ neighbours of v. Overall, we have that the quantities $|X \cap V(H)|$, $|Y \cap V(H)|$, $|E(c) \setminus V(M)|$ and $|N_\Gamma(v) \setminus V(M)|$ are each influenced by at most $3n$ coordinates.

Note that $n^{-0.001} \leqslant \alpha \leqslant \gamma \leqslant 0.00001, \delta \leqslant 1$ and $\ell \leqslant n^{0.001}$ implies that $\ell(\delta n)^{-1} \leqslant \alpha \leqslant \gamma \leqslant 0.01$. Fix $t = \alpha\gamma\delta n/10$. By Lemma 4.2, we have

$$\mathbb{E}(|X \cap V(H)|) \pm t = (1 \pm (1 + 12\alpha)\gamma)(1 - \alpha)n$$

$$\mathbb{E}(|Y \cap V(H)|) \pm t = (1 \pm (1 + 12\alpha)\gamma)(1 - \alpha)n$$

$$\mathbb{E}(|E(c) \setminus V(M)|) + t \leqslant (1 + (1 + 26\alpha)\gamma)(1 - 2\alpha)\delta n$$

$$\mathbb{E}(|N_\Gamma(v) \setminus V(M)|) \pm t = (1 \pm (1 + 30\alpha)\gamma)(1 - 2\alpha)\delta n$$

By Azuma's Inequality, we have that for any given c, v any of (i)–(iii) fail to hold with probability $\leqslant 2e^{-\frac{t^2}{3n(26\ell)^2}} \leqslant 2e^{-\frac{\alpha^2\gamma^2\delta^2 n^{0.9}}{300000}} \leqslant 2e^{-n^{0.8}}$ (using $n^{-0.001} \leqslant \ell^{-1} \leqslant \alpha \leqslant \gamma \leqslant 0.00001$). Taking a union bound over all c, v, we have that all of (i)–(iii) hold with probability $> 1 - 8n^2 e^{-n^{0.8}} \geqslant 1 - n^{-2}$ (using $n^{-0.001} \leqslant 0.001$). $\qquad\square$

The following version of the above lemma will be more convenient to apply.

COROLLARY 4.5. *Suppose that we have $n, \delta, \gamma, \alpha, \ell$ with $n^{-1} \overset{\text{POLY}}{\ll} \alpha \leqslant \gamma \overset{\text{POLY}}{\ll} \delta \leqslant 1$ and $\ell \overset{\text{POLY}}{\ll} n$. Let G be a coloured balanced bipartite graph which is (γ, δ, n)-regular, locally ℓ-bounded and globally $(1 + \gamma)\delta n$-bounded. Let H be produced by an $(\alpha, (1 + \gamma)\delta n)$-random edge-assignment on G.*

With probability $\geqslant 1 - n^{-2}$, the graph H is $(e^{35\alpha}\gamma, e^{-\alpha}\delta, e^{-\alpha}n)$-regular and globally $(1 + e^{35\alpha}\gamma)(e^{-\alpha}\delta)(e^{-\alpha}n)$-bounded.

Proof. Note that $n^{-1} \overset{\text{POLY}}{\ll} \alpha \leqslant \gamma \overset{\text{POLY}}{\ll} \delta \leqslant 1$ and $\ell \overset{\text{POLY}}{\ll} n$ implies $n^{-0.001} \leqslant \alpha \leqslant \gamma \leqslant 0.00001, \delta \leqslant 1$ and $\ell \leqslant n^{0.001}$. Let X, Y be the bipartition classes of G. By Lemma 4.3, we have that with probability $\geqslant 1 - n^{-2}$ all of (i), (ii) and (iii) hold. Note that, from (3) and $\alpha \leqslant \gamma \overset{\text{POLY}}{\ll} 1$, we have

$$(1 \pm (1 + 30\alpha)\gamma)(1 \pm 4\alpha^2) = (1 \pm (1 + 30\alpha)\gamma \pm 5\alpha^2) = (1 \pm e^{35\alpha}\gamma). \tag{24}$$

From (i), (2) and (24), we have $|X \cap V(H)| = |Y \cap V(H)| = (1 \pm (1 + 12\alpha)\gamma)(1 - \alpha)n = (1 \pm (1 + 12\alpha)\gamma)(1 \pm \alpha^2)e^{-\alpha}n = (1 \pm e^{35\alpha}\gamma)(e^{-\alpha})n$.

From (iii), (2) and (24), we have that for all vertices $v \in V(G)$ we have $d_H(v) = (1 \pm (1 + 30\alpha)\gamma)(1 - 2\alpha)\delta n = (1 \pm (1 + 30\alpha)\gamma)(1 \pm 4\alpha^2)e^{-2\alpha}\delta n = (1 \pm e^{35\alpha}\gamma)(e^{-\alpha}\delta)(e^{-\alpha}n)$. These show that H is $(e^{35\alpha}\gamma, e^{-\alpha}\delta, e^{-\alpha}n)$-regular.

From (ii), (2) and (24), we have that for every colour c, we have $|E_H(c)| \leqslant (1 + (1 + 26\alpha)\gamma)(1 - 2\alpha)\delta n \leqslant (1 + (1 + 26\alpha)\gamma)(1 + 4\alpha^2)e^{-2\alpha}\delta n \leqslant (1 + e^{35\alpha}\gamma)e^{-2\alpha}\delta n$. This shows that H is globally $(1 + e^{35\alpha}\gamma)(e^{-\alpha}\delta)(e^{-\alpha}n)$-bounded. $\qquad\square$

4.5. *Finding a nearly-perfect matching*

Here we prove the main result of this section. By iterating the (α, b)-random edge-assignment process on a properly coloured graph G, we can find a nearly spanning rainbow matching M in G. The following lemma does this and shows that the resulting rainbow matching is random-like in a sense that every edge is in M with at least (approximately) the right probability.

LEMMA 4.6. *Suppose that we have $n, \delta, \gamma, p, \ell$ with $1 \geqslant \delta \overset{\text{POLY}}{\gg} p \overset{\text{POLY}}{\gg} \gamma \overset{\text{POLY}}{\gg} n^{-1}$ and $n \overset{\text{POLY}}{\gg} \ell$.*

Let G be a locally ℓ-bounded, (γ, δ, n)-regular, globally $(1 + \gamma)\delta n$-bounded, coloured, balanced bipartite graph. Then G has a random rainbow matching M which has size $\geqslant (1 - 2p)n$ and

$$\mathbb{P}(e \in E(M)) \geqslant (1 - 9p)\frac{1}{\delta n} \quad \text{for each } e \in E(G). \tag{25}$$

Proof. Fix $H_0 = G$, $\alpha = \gamma$, and $T = \alpha^{-1} \ln(p^{-1})$. Without loss of generality, we may suppose that γ and α are chosen so that T is an integer (to see this, replace γ by $\gamma' = \frac{\ln(p^{-1})}{\lfloor \gamma^{-1} \ln(p^{-1}) \rfloor}$. This ensures that $T' = \gamma'^{-1} \ln(p^{-1})$ is an integer. Note that $p \stackrel{\text{\tiny POLY}}{\gg} 2\gamma \geqslant \gamma' \geqslant \gamma$ holds, so we could perform the proof of the lemma with γ replaced by γ'). Note that this gives $p = e^{-\alpha T}$. Fix the following constants:

$$\gamma_t = e^{35\alpha t}\gamma \quad \delta_t = e^{-\alpha t}\delta \quad n_t = e^{-\alpha t}n.$$

Using $p = e^{-\alpha T}$, we have $n_T = pn$, $\gamma_T = p^{-35}\gamma \leqslant p$, and $\delta_T = p\delta$.

We construct graphs $H_1, \ldots, H_T$ and matchings $M_1, \ldots, M_T$ recursively as follows.

- For $t \geqslant 0$, if H_t is not $(\gamma_t, \delta_t, n_t)$-regular or globally $(1 + \gamma_t)\delta_t n_t$-bounded, then stop the process at step t.
- Otherwise, if H_t is $(\gamma_t, \delta_t, n_t)$-regular and globally $(1 + \gamma_t)\delta_t n_t$-bounded, then we run an $(\alpha, (1 + \gamma_t)\delta_t n_t)$-random edge-assignment on H_t to get a graph H_{t+1} and a matching M_{t+1}.

Note that for all t, H_t is locally ℓ-bounded and we have $n_t^{-1} \leqslant n_T^{-1} = p^{-1}n^{-1} \stackrel{\text{\tiny POLY}}{\ll} \alpha = \gamma \leqslant \gamma_t \leqslant p^{-35}\gamma \stackrel{\text{\tiny POLY}}{\ll} p\delta \leqslant \delta_t \leqslant 1$ and $\ell \stackrel{\text{\tiny POLY}}{\ll} n$. Let A_t be the event that the process has not stopped at any of the steps $1, \ldots, t$. The events A_t are clearly decreasing. Since $\gamma_0 = \gamma, \delta_0 = \delta$ and $n_0 = n$, the assumptions of the lemma imply that $\mathbb{P}(A_0) = 1$. From Corollary 4.5, we have $\mathbb{P}(A_t | A_{t-1}) \geqslant 1 - n_{t-1}^{-2}$ (in this application, we have $\gamma = \gamma_{t-1}, \delta = \delta_{t-1}, n = n_{t-1}, \alpha = \alpha, \ell = \ell$). This implies $\mathbb{P}(A_0 \cap A_1 \cap \cdots \cap A_T) = \mathbb{P}(A_0)\mathbb{P}(A_1 | A_0)\mathbb{P}(A_2 | A_1) \cdots \mathbb{P}(A_T | A_{T-1}) \geqslant \prod_{i=1}^{T}(1 - n_{t-1}^{-2}) \geqslant (1 - n_T^{-2})^T = (1 - p^{-2}n^{-2})^T \geqslant 1 - \frac{1}{\gamma^2 p^4 n^2} > 0$ (using $p, \gamma \stackrel{\text{\tiny POLY}}{\gg} n^{-1}$).

Define M to be the rainbow matching $M_1 \cup \cdots \cup M_T$ conditional on the events $A_0, \ldots, A_T$ occurring (to see that M is a rainbow matching, recall that H_i and M_i were vertex-disjoint and colour-disjoint). As A_T holds, H_T has $(1 \pm \gamma_T)n_T$ vertices, so that M is a matching of size $\geqslant n - (1 + \gamma_T)n_T \geqslant (1 - 2p)n$.

CLAIM 4.7. *The following hold for each $t = 1, \ldots, T$ and $e \in E(G)$.*

$$\mathbb{P}(e \in E(H_t), A_t | e \in E(H_{t-1}), A_{t-1}) \geqslant (1 - T^{-1}p)(1 - 3\alpha) \tag{26}$$

$$\mathbb{P}(e \in E(M_{t+1}) | e \in E(H_t), A_t) = (1 \pm p)e^{2t\alpha}\frac{\alpha}{\delta n}. \tag{27}$$

Proof. Using Corollary 4.5, note that $\mathbb{P}(\overline{A_t} | e \in E(H_{t-1}), A_{t-1}) \leqslant n_{t-1}^{-2} \leqslant n_T^{-2} \leqslant p^{-2}n^{-2}$ (this application of Corollary 4.5 is the same as our previous one). Using (23), (16) and $p^{-2}n^{-2} \leqslant \gamma\alpha$ (which comes from $1 \stackrel{\text{\tiny POLY}}{\gg} p \stackrel{\text{\tiny POLY}}{\gg} \gamma \stackrel{\text{\tiny POLY}}{\gg} n^{-1}$) gives:

$$\mathbb{P}(e \in E(H_t), A_t | e \in E(H_{t-1}), A_{t-1}) \geqslant \mathbb{P}(e \in E(H_t) | e \in E(H_{t-1}), A_{t-1})$$

$$- \mathbb{P}(\overline{A_t} | e \in E(H_{t-1}), A_{t-1})$$

$$\geqslant \left(1 - 2\alpha - \frac{(1 + \gamma_t)\delta_t}{\delta_t}\alpha\right) - 150\alpha\gamma_t - p^{-2}n^{-2}$$

$$\geqslant (1 - 160p^{-45}\alpha\gamma)(1 - 3\alpha)$$

$$\mathbb{P}(e \in E(M_{t+1}) | e \in E(H_t), A_t) = (1 \pm 5\gamma_t)\frac{\alpha}{\delta_t n_t} = (1 \pm 6p^{-45}\gamma)e^{2t\alpha}\frac{\alpha}{\delta n}.$$

Now the claim follows from $160p^{-45}\alpha\gamma \leqslant T^{-1}p$ and $6p^{-45}\gamma \leqslant p$ (which both come from $1 \stackrel{\text{\tiny POLY}}{\gg} p \stackrel{\text{\tiny POLY}}{\gg} \gamma$). $\qquad \square$

Let $t \leqslant T$. Note that the events '$e \in H_t$ and A_t holds' are decreasing with t. Using (2), (4), (26) and $p \geqslant 10\alpha^2 T$, we have

$$\mathbb{P}(e \in E(H_t), A_t) = \prod_{i=1}^{t} \mathbb{P}(e \in H_i, A_i | e \in H_{i-1}, A_{i-1}) \geqslant \left(1 - T^{-1}p\right)^t (1 - 3\alpha)^t \geqslant (1 - 3p)e^{-3t\alpha}.$$

Combining the above with (5), (27), $p = e^{-\alpha T}$ and $p \geqslant \alpha$, we get:

$$\mathbb{P}(e \in E(M_1 \cup \cdots \cup M_T)) = \sum_{t=1}^{T} \mathbb{P}(e \in H_{t-1}, A_{t-1}) \mathbb{P}(e \in M_t | e \in H_{t-1}, A_{t-1})$$

$$\geqslant (1 - p)(1 - 3p)\frac{\alpha}{\delta n} \sum_{t=1}^{T} e^{-\alpha(t-1)}$$

$$\geqslant (1 - 4p)(1 - \alpha^2 - 2e^{-\alpha T})\frac{1}{\delta n}$$

$$\geqslant (1 - 8p)\frac{1}{\delta n}.$$

Now, (25) comes from $\mathbb{P}(e \in E(M_1 \cup \cdots \cup M_T) | A_0 \cap \cdots \cap A_T) \geqslant \mathbb{P}(e \in E(M_1 \cup \cdots \cup M_T)) - \mathbb{P}(\overline{A_0 \cap \cdots \cap A_T}) \geqslant \mathbb{P}(e \in E(M_1 \cup \cdots \cup M_T)) - 2\gamma^2 p^4 n^{-2} \geqslant (1 - 9p)\frac{1}{\delta n}.$ $\square$

5. Random and pseudo-random subgraphs

In this section, we collect intermediate lemmas which we will need concerning random and pseudo-random subgraphs. We will often use the typicality of complete graphs.

LEMMA 5.1. *For $\gamma \overset{\text{\tiny POLY}}{\gg} n^{-1}$, K_n is $(\gamma, 1, n)$-typical and $K_{n,n}$ is $(\gamma, 1, n)$-typical.*

Proof. Note that K_n is $(\frac{2}{n}, 1, n)$-typical while $K_{n,n}$ is $(0, 1, n)$-typical. Combined with $\gamma \overset{\text{\tiny POLY}}{\gg} n^{-1}$, this implies the lemma. $\square$

5.1. *Random subgraphs*

We will need a number of results of the form 'for a nice graph G, a random subgraph H of G is still nice'. Here 'nice' can mean that G is (γ, δ, n)-regular, (γ, δ, n)-typical or globally bounded. We will look at four different kinds of 'random subgraphs' H.

LEMMA 5.2 (Random subgraphs of a general graph). *Let $1 \geqslant \gamma, \delta, p, \mu \overset{\text{\tiny POLY}}{\gg} n^{-1}$. Let G be a properly coloured, globally μn-bounded (γ, δ, n)-regular/(γ, δ, n)-typical general graph.*

(a) *Random set of colours: Let H_1 be a subgraph of G formed by choosing each colour with probability p. Then, H_1 is $(2\gamma, p\delta, n)$-regular/$(2\gamma, p\delta, n)$-typical with probability $1 - o(n^{-1})$.*

(b) *Random set of edges: Let H_2 be a subgraph of G formed by choosing each edge with probability p. Then, H_2 is $(2\gamma, p\delta, n)$-regular/$(2\gamma, p\delta, n)$-typical and globally $(1 + \gamma)p\mu n$-bounded with probability $1 - o(n^{-1})$.*

(c) *Random set of vertices: For $pn \in \mathbb{Z}$ with $p < 1$, let $A \subseteq V(G)$ be a subset of order pn chosen uniformly at random out of all such subsets. Then, $G[A]$ is globally $(1 + \gamma)(\mu p^2)n$-bounded and $(2\gamma, \delta, pn)$-regular/$(2\gamma, \delta, pn)$-typical with probability $1 - o(n^{-1})$.*

(d) *Two disjoint random sets of vertices: For $pn \in \mathbb{Z}$ with $p < 1/2$, let $A, B \subseteq V(G)$ be two disjoint subsets of order pn chosen uniformly at random out of all pairs of such subsets. Then,*

$G[A, B]$ is a globally $(1 + \gamma)(2\mu p^2)n$-bounded, $(2\gamma, \delta, pn)$-regular/$(2\gamma, \delta, pn)$-typical balanced bipartite graph with probability $1 - o(n^{-1})$.

Proof. Note that the following bounds on expectations are true by linearity of expectation for all vertices $u \neq v$ and colours c.

$$\mathbb{E}(d_{H_1}(v)) = \mathbb{E}(d_{H_2}(v)) = pd_G(v) = (1 \pm \gamma)p\delta n \qquad \text{when } G \text{ is } (\gamma, \delta, n)\text{-regular.}$$

$$\mathbb{E}(d_{H_1}(u,v)) = \mathbb{E}(d_{H_2}(u,v)) = p^2 d_G(u,v) = (1 \pm \gamma)p^2\delta^2 n \quad \text{when } G \text{ is } (\gamma, \delta, n)\text{-typical.}$$

$$\mathbb{E}(|N(v) \cap A|) = \mathbb{E}(|N(v) \cap B|) = pd_G(v) = (1 \pm \gamma)\delta(pn) \qquad \text{when } G \text{ is } (\gamma, \delta, n)\text{-regular.}$$

$$\mathbb{E}(|N(u,v) \cap A|) = \mathbb{E}(|N(u,v) \cap B|) = pd_G(u,v) = (1 \pm \gamma)\delta^2(pn) \quad \text{when } G \text{ is } (\gamma, \delta, n)\text{-typical.}$$

$$\mathbb{E}(|E_{H_2}(c)|) = p|E_G(c)| \leqslant p\mu n$$

$$\mathbb{E}(|E_G(c) \cap A|) = \mathbb{E}(|E_G(c) \cap B|) = \frac{pn(pn-1)}{n(n-1)}|E_G(c)| = \left(p^2 - \frac{p(1-p)}{n-1}\right)|E_G(c)| \leqslant p^2\mu n$$

$$\mathbb{E}(|E_G(c) \cap (A \cup B)|) = \frac{2pn(2pn-1)}{n(n-1)}|E_G(c)| \leqslant 4p^2|E_G(c)| - \frac{2p(1-2p)}{n-1}$$

$$\leqslant 4p^2\mu n.$$

First we prove (a) and (b). Note that the random variables $d_{H_1}(v)$, $d_{H_2}(v)$, $d_{H_1}(u,v)$, $d_{H_2}(u,v)$ and $|E_{H_2}(c)|$ are all 2-Lipshitz (using the fact that the colouring is proper), and are all influenced by $\leqslant 2n$ coordinates. By Azuma's Inequality (Lemma 3.6), we have that the probability that any of these deviate from their expectation by more than $\gamma p^2\delta^2\mu n$ is $\leqslant 2e^{\frac{-(\gamma p^2\delta^2\mu n)^2}{8n}} = o(n^{-3})$ (using $1 \geqslant \gamma, \delta, p, \mu \overset{\text{POLY}}{\gg} n^{-1}$ which implies $\gamma^2 p^4\delta^4\mu^2 n \geqslant 40 \log n$). Taking a union bound over all pairs of vertices and colours, we obtain (a) and (b).

It remains to prove (c) and (d). Note that the functions $|N(v) \cap A|$, $|N(v) \cap B|$, $|N(u,v) \cap A|$, $|N(u,v) \cap B|$, $|E_G(c) \cap A|$ and $|E_G(c) \cap B|$ each satisfy the assumptions of Lemma 3.9 with $\alpha = 1$, $r = pn$, $N = n$. Also $|E_G(c) \cap (A \cup B)|$ satisfies the assumptions of Lemma 3.9 with $\alpha = 1$, $r = 2pn$ and $N = n$. Finally, note that we have $0 < \min(r, N - r) < n$ for all of these. By Lemma 3.9, we have that the probability that any of these functions deviate from their expectation by more than $\gamma p^2\delta^2\mu n/4$ is $\leqslant 2e^{-\frac{2(\gamma p^2\delta^2\mu n/4)^2}{n}} = o(n^{-3})$ (using $1 \geqslant \gamma, \delta, p, \mu \overset{\text{POLY}}{\gg} n^{-1}$ which implies $\gamma^2 p^4\delta^4\mu^2 n \geqslant 40 \log n$). Taking a union bound over all pairs of vertices and all colours, we obtain (c) and the '$(2\gamma, \delta, pn)$-regular/$(2\gamma, \delta, pn)$-typical' part of (d). We also get that with probability $1 - o(n^{-1})$, we have $|E_G(c) \cap A|, |E_G(c) \cap B| = p^2|E_G(c)| - \frac{p(1-p)}{n-1}|E_G(c)| \pm \gamma p^2\delta^2\mu n/4 = p^2|E_G(c)| \pm \gamma p^2\mu n/3$ for all colours c (using $\gamma, p, \mu \overset{\text{POLY}}{\gg} n^{-1}$). Similarly, we have $|E_G(c) \cap (A \cup B)| = 4p^2|E_G(c)| \pm \gamma p^2\mu n/3$. These give $e(G[A, B] \cap E_G(c)) = |E_G(c) \cap (A \cup B)| - |E_G(c) \cap A| - |E_G(c) \cap B| = 4p^2|E_G(c)| - 2 \cdot p^2|E_G(c)| \pm \gamma p^2\mu n \leqslant (1 + \gamma)2p^2\mu n$ (the last inequality coming from global μn-boundedness). This implies the global boundedness part of (d). □

We will need a balanced bipartite version of part of the above lemma.

LEMMA 5.3 (Random subgraphs of a balanced bipartite graph). *Let $1 \geqslant \gamma, \delta, p, \mu \overset{\text{POLY}}{\gg} n^{-1}$. Let G be a properly coloured, globally μn-bounded (γ, δ, n)-regular/(γ, δ, n)-typical balanced bipartite graph.*

(a) *Random set of colours: Let H_1 be a subgraph of G formed by choosing each colour with probability p. Then, H_1 is $(2\gamma, p\delta, n)$-regular/$(2\gamma, p\delta, n)$-typical with probability $1 - o(n^{-1})$.*

(b) *Random set of edges: Let H_2 be a subgraph of G formed by choosing each edge with probability p. Then, H_2 is $(2\gamma, p\delta, n)$-regular/$(2\gamma, p\delta, n)$-typical and globally $(1+\gamma)p\mu n$-bounded with probability $1 - o(n^{-1})$.*

Proof. Let u, v be vertices, and c a colour. Note that the following bounds on expectations are true by linearity of expectation.

$$\mathbb{E}(d_{H_1}(v)) = \mathbb{E}(d_{H_2}(v)) = pd_G(v) = (1 \pm \gamma)p\delta n \qquad \text{when } G \text{ is } (\gamma, \delta, n)\text{-regular}$$

$$\mathbb{E}(d_{H_1}(u,v)) = \mathbb{E}(d_{H_2}(u,v)) = p^2 d_G(u,v) = (1 \pm \gamma)p^2\delta^2 n \qquad \text{when } G \text{ is } (\gamma, \delta, n)\text{-typical}$$

$$\mathbb{E}(|E_{H_2}(c)|) = p|E_G(c)| \leqslant p\mu n.$$

Note that the random variables $d_{H_1}(v)$, $d_{H_2}(v)$, $d_{H_1}(u,v)$, $d_{H_2}(u,v)$ and $|E_{H_2}(c)|$ and are all 2-Lipshitz (using the fact that the colouring is proper), and are all influenced by $\leqslant 2n$ coordinates. By Azuma's Inequality (Lemma 3.6), we have that the probability that any of these deviate from their expectation by more than $\gamma p^2 \delta^2 \mu n$ is $\leqslant 2e^{\frac{-(\gamma p^2 \delta^2 \mu n)^2}{8n}} = o(n^{-3})$ (using $1 \geqslant \gamma, \delta, p, \mu \overset{\text{POLY}}{\gg} n^{-1}$ which implies $\gamma^2 p^4 \delta^4 \mu^2 n \geqslant 40 \log n$). Taking a union bound over all pairs of vertices and colours, we obtain (a) and (b). $\square$

The following lemma gives another property of the random subgraph formed by choosing every edge independently with probability p. This time we are concerned with how many vertices a small set of colours covers.

LEMMA 5.4. *Let $1 \geqslant p, \varepsilon \overset{\text{POLY}}{\gg} k^{-1} \overset{\text{POLY}}{\gg} \nu \overset{\text{POLY}}{\gg} n^{-1}$. Let G be a properly coloured graph with all colours covering $\geqslant (1-\nu)n$ vertices. Let H be a random subgraph formed by choosing every edge with probability p. Then, with high probability, any set of k colours of H covers $\geqslant (1-\varepsilon)n$ vertices.*

Proof. Let S be a set of k colours and G_S, H_S the subgraphs of G and H consisting of colour S edges. Note that $e(G_S) \geqslant k(1-\nu)n/2$. By the Handshaking Lemma, we have $\sum_{v \in V(G_S)} d_{G_S}(v) = 2e(G_S) \geqslant k(1-\nu)n$. Let L be the set of vertices in G_S of degree $\geqslant k/2$. Using $\Delta(G_S) \leqslant k$, we have $|L|k + (n-|L|)k/2 \geqslant \sum_{v \in V(G_S)} d_{G_S}(v) \geqslant k(1-\nu)n$, which is equivalent to $|L| \geqslant (1-2\nu)n$.

For a vertex $v \in L$, we have $\mathbb{P}(d_{H_S}(v) = 0) = (1-p)^{d_{G_S}(v)} \leqslant (1-p)^{k/2} \leqslant e^{-pk/2} \leqslant \varepsilon/4$, as $p, \varepsilon \overset{\text{POLY}}{\gg} k^{-1}$. Let X be the number of isolated vertices in L. By linearity of expectation $\mathbb{E}(X) \leqslant \varepsilon n/4$. Note that X is 2-Lipschitz and is influenced by $\leqslant e(G_S) \leqslant kn/2$ edges. By Azuma's Inequality (Lemma 3.6) applied with $t = \varepsilon n/4$, we have $\mathbb{P}(X \geqslant \varepsilon n/2) \leqslant 2e^{-\frac{(\varepsilon n/4)^2}{4(kn/2)}} = 2e^{-\frac{\varepsilon^2 n}{32k}} \leqslant n^{-3k}$ (as $\varepsilon, k^{-1} \overset{\text{POLY}}{\gg} n^{-1}$ implies that $\varepsilon^2/k^2 \geqslant \frac{400 \log n}{n}$). Thus, with probability $\geqslant 1 - n^{-3k}$, H_S has $\leqslant X + (n - |L|) \leqslant \varepsilon n$ isolated vertices. Taking a union bound over all sets S of k colours gives the result. $\square$

5.2. (e, m)-dense graphs

In this paper, it is convenient to use two different notions of pseudo-randomness. The first of these is (γ, δ, n)-typicality (See Definition 3.2). The second is the following.

DEFINITION 5.5. • A general graph G is (e, m)-dense if for any $\lambda \geqslant 1$ and disjoint sets A, B with $|A| = |B| = \lambda m$, we have $e(A, B) \geqslant \lambda^2 e$.
• A balanced bipartite graph G with parts X and Y is (e, m)-dense if for any $\lambda \geqslant 1$ and sets $A \subseteq X$, $B \subseteq Y$ with $|A| = |B| = \lambda m$, we have $e(A, B) \geqslant \lambda^2 e$.

We remark that most of the time we will use the above definition with $\lambda = 1$. Thus, the definition should be thought of as saying that there are e edges between any two sets of vertices of size m. Note that if G is (e, m)-dense, then it is also (e', m')-dense for any $e' \leqslant e$ and $m' \geqslant m$.

How is the above definition related to (γ, δ, n)-typicality? In fact, (γ, δ, n)-typicality is a stronger concept. We prove two lemmas relating typicality and density. The following is a variation of a lemma proved by the third author together with Alon and Krivelevich in [**5**].

LEMMA 5.6. *Every (γ, p, n)-typical graph H has the following for every pair of subsets A, B with $|B| \geqslant \gamma^{-1} p^{-2}$:*

$$|e(A, B) - p|A||B|| \leqslant 2|A|^{\frac{1}{2}} |B| \gamma^{\frac{1}{2}} n^{\frac{1}{2}} p.$$

Proof. Let Adj_H be the adjacency matrix of H, and let $M = \mathrm{Adj}_H - pJ$ where J is the appropriately sized all-ones matrix. Note that for every pair of distinct vertices x, x', we have

$$\sum_{v \in V(H)} M_{x,v} M_{x',v} = d_H(x, x') - p(d(x) + d(x')) + p^2 n \leqslant (1 + \gamma)p^2 n - 2(1 - \gamma)p^2 n + p^2 n$$

$$\leqslant 3\gamma p^2 n. \tag{28}$$

Next note that we have

$$|e(A, B) - p|A||B||^2 = \left(\sum_{x \in A} \sum_{y \in B} M_{x,y} \right)^2 \leqslant |A| \sum_{x \in A} \left(\sum_{y \in B} M_{x,y} \right)^2 \leqslant |A| \sum_{x \in V(H)} \left(\sum_{y \in B} M_{x,y} \right)^2$$

$$= |A| \sum_{x \in V(H)} \left(\sum_{y \in B} M_{x,y}^2 \right) + |A| \sum_{x \in V(H)} \left(\sum_{y \neq y' \in B} M_{x,y} M_{x,y'} \right)$$

$$\leqslant n|A||B| + |A| \sum_{y \neq y' \in B} \left(\sum_{x \in V(H)} M_{x,y} M_{x,y'} \right)$$

$$\overset{(28)}{\leqslant} n|A||B| + |A| \sum_{y \neq y' \in B} 3\gamma np^2 \leqslant n|A||B| + |A||B|^2 3\gamma np^2 \leqslant 4|A||B|^2 \gamma np^2.$$

Here the first inequality comes from the Cauchy–Schwarz inequality and the last inequality comes from $|B| \geqslant \gamma^{-1} p^{-2}$. Taking square roots gives the result. $\qquad\square$

The following version of the above is more convenient to apply.

LEMMA 5.7. *Let $1 \geqslant p, \mu \overset{\text{\tiny POLY}}{\gg} \gamma \overset{\text{\tiny POLY}}{\gg} n^{-1}$. Every (γ, p, n)-typical graph a(which is either balanced bipartite or general) is $(0.99p(\mu n)^2, \mu n)$-dense.*

Proof. First we deal with the case when G is a general graph. Note that $p, \mu \overset{\text{\tiny POLY}}{\gg} \gamma \overset{\text{\tiny POLY}}{\gg} n^{-1}$ implies $\mu n \geqslant \gamma^{-1} p^{-2}$. By Lemma 5.6, we have that for any $\lambda \geqslant 1$ and pair of subsets A, B with $|A| = |B| = \lambda \mu n$, we have $|e(A, B) - p(\lambda \mu n)^2| \leqslant 2(\lambda \mu n)^{\frac{1}{2}} (\lambda \mu n) \gamma^{\frac{1}{2}} n^{\frac{1}{2}} p \leqslant 0.01 p(\lambda \mu n)^2$ (the last inequality is $2\gamma^{\frac{1}{2}} \leqslant 0.01 \lambda^{\frac{1}{2}} \mu^{\frac{1}{2}}$ which comes from $\gamma \overset{\text{\tiny POLY}}{\ll} \mu$ and $\lambda \geqslant 1$).

Now suppose that G is a balanced bipartite (γ, p, n)-typical graph with parts X, Y. Add a copy of the Erdős–Rényi random graph $G(n, p)$ to both X and Y to get a graph H. Note that for any vertex v, we have $\mathbb{E}(d_H(v)) = (1 \pm \gamma)2pn$ and that any pair of vertices u, v have $\mathbb{E}(d_H(u, v)) = (1 \pm \gamma)2p^2n$. Note that these quantities are each 1-Lipschitz affected by $\leqslant 2n$ coordinates. By Azuma's Inequality (Lemma 3.6) and the union bound, we get that with high probability H is a $(2\gamma, p, 2n)$-typical general graph. By the general graph version of this lemma with $\mu' = \mu/2$, H is $(0.99p(\mu n)^2, \mu n)$-dense. This implies that between any sets $A \subseteq X$, $B \subseteq Y$ with $|A| = |B| = \lambda\mu n$, we have $e_G(A, B) = e_H(A, B) \geqslant \lambda^2 0.99p(\mu n)^2$, that is, that G is $(0.99p(\mu n)^2, \mu n)$-dense (as a balanced bipartite graph). $\qquad\square$

The following lemma shows that it is possible to delete a small number of edges from any graph so that its complement is pseudo-random. Here $\overline{H}$ denotes the set of edges on $V(H)$ not present in H.

LEMMA 5.8.   Let $1 \geqslant p, \mu \overset{\text{\tiny POLY}}{\gg} n^{-1}$. Every d-regular balanced bipartite graph G on $2n$ vertices has a $(d - \lfloor pn \rfloor)$-regular spanning subgraph H such that $\overline{H}$ is $(0.48p(\mu n)^2, \mu n)$-dense.

Proof.   Choose γ with $1 \geqslant p, \mu \overset{\text{\tiny POLY}}{\gg} \gamma \overset{\text{\tiny POLY}}{\gg} n^{-1}$. Consider an arbitrary 1-factorization of $K_{n,n}$ in which every colour either only occurs on G or only occurs outside G (this exists since every regular bipartite graph has a 1-factorization). By Lemma 5.1, $K_{n,n}$ is $(\gamma, 1, n)$-typical. Let E be a subgraph of $K_{n,n}$ formed by choosing every colour with probability $0.5p$. By Lemma 5.3 (a), E is $(2\gamma, 0.5p, n)$-typical with high probability. By Lemma 5.7, applied with $\gamma' = 2\gamma$, $p' = 0.5p$, $\mu = \mu$, E is $(0.48p(\mu n)^2, \mu n)$-dense.

Since E and $G \setminus E$ are unions of perfect matchings, they are regular. Since E is $(2\gamma, 0.5p, n)$-typical, the graph $G \setminus E$ is d'-regular for some $d' \geqslant d - (1 + 2\gamma)0.5pn \geqslant d - \lfloor pn \rfloor$ (using $\gamma \overset{\text{\tiny POLY}}{\ll} 1$). Therefore, we can find some $(d - \lfloor pn \rfloor)$-regular subgraph H of G which is edge-disjoint from E (using that G is bipartite). Since H^c contains E, we have that $\overline{H}$ is $(0.48p(\mu n)^2, \mu n)$-dense as required. $\qquad\square$

We will need two lemmas showing that deleting a small number of edges from an (e, m)-dense graph does not change the pseudo-randomness too much.

LEMMA 5.9.   Let G be (e, m)-dense and H a subgraph of G. Then, $G \setminus H$ is $(e - e(H), m)$-dense.

Proof.   For any $\lambda \geqslant 1$ and sets $A \subseteq X$, $B \subseteq Y$ with $|A| = |B| = \lambda m$, we have $e_{G \setminus H}(A, B) \geqslant e_G(A, B) - e(H) \geqslant \lambda^2 e - e(H) \geqslant \lambda^2(e - e(H))$. $\qquad\square$

LEMMA 5.10.   Let G be (e, m)-dense and M a matching in G. Then $G \setminus E(M)$ is $(e - m, m)$-dense.

Proof.   For $\lambda > 1$, let A and B be sets with $|A| = |B| = \lambda m$. Since G is (e, m)-dense, we have $e_G(A, B) \geqslant \lambda^2 e$. Since M is a matching there can be at most λm edges of M between A and B. Therefore, $e_{G \setminus M}(A, B) \geqslant \lambda^2 e - \lambda m \geqslant \lambda^2(e - m)$. $\qquad\square$

6. Regularization lemmas

In our proofs, we will need a number of intermediate lemmas saying that a graph G can be modified into a regular graph. Broadly speaking, there are three types of modifications that we will need: deleting a small number of edges, adding edges from a disjoint dense graph or adding a small number of vertices.

6.1. *Regularization by deleting edges*

Here we will prove results about finding a regular subgraph by deleting edges from a graph with very high minimum degree. The goal of this section is to prove Lemmas 6.6 and 6.7. The following theorem of Cristofides, Kühn and Osthus is a result of the type we want in this section (see [**20**, Theorem 12]).

THEOREM 6.1 (Cristofides, Kühn, Osthus). *Let G be a graph with minimum degree $\delta \geqslant n/2$ and r an even number with $r \leqslant \frac{1}{2}(\delta + \sqrt{n(2\delta - n)})$. Then, G has a spanning r-regular subgraph.*

The following version of this will be a bit easier to apply.

LEMMA 6.2. *Let $n^{-1} \overset{\text{POLY}}{\ll} \varepsilon \overset{\text{POLY}}{\ll} 1$. Let G be an n-vertex graph with $\delta(G) \geqslant (1 - \varepsilon)n$. Then, G has a spanning $2\lceil (1 - \varepsilon - 8\varepsilon^2)n/2\rceil$-regular subgraph.*

Proof. Set $\delta = (1 - \varepsilon)n \geqslant n/2$ and $r = 2\lceil (1 - \varepsilon - 8\varepsilon^2)n/2\rceil$. Note that r is even and has $r \leqslant \frac{1}{2}(\delta + \sqrt{n(2\delta - n)})$ (using $\sqrt{n(2\delta - n)} = n\sqrt{1 - 2\varepsilon} \geqslant n(1 - \varepsilon - 2\varepsilon^2)$, which holds for $\varepsilon \leqslant 1/2$). Apply Theorem 6.1 to get the lemma. $\square$

We will also need a balanced bipartite version of this lemma. To prove it we use the following theorem of Ore and Ryser (see [**41**]).

THEOREM 6.3 (Ore, Ryser). *A balanced bipartite graph with parts X, Y has no spanning d-regular subgraph if and only if there is a set $T \subseteq Y$ with $d|T| > \sum_{x \in X} \min(|N(x) \cap T|, d)$.*

Using this, we can prove a bipartite version of Lemma 6.2.

LEMMA 6.4. *Let $n^{-1} \overset{\text{POLY}}{\ll} \varepsilon \overset{\text{POLY}}{\ll} 1$. Let G be a balanced bipartite graph with vertex classes X and Y with $|X| = |Y| = n$ and $\delta(G) \geqslant (1 - \varepsilon)n$. Then, G has a spanning $\lfloor (1 - \varepsilon - 8\varepsilon^2)n\rfloor$-regular subgraph.*

Proof. Note that there is an $\hat{\varepsilon} \in \mathbb{R}$ with $(1 - \hat{\varepsilon} - 8\hat{\varepsilon}^2)n = \lfloor (1 - \varepsilon - 8\varepsilon^2)n\rfloor$ and $1 \overset{\text{POLY}}{\gg} \varepsilon + n^{-1} \geqslant \hat{\varepsilon} \geqslant \varepsilon \overset{\text{POLY}}{\gg} n^{-1}$. Note that $\delta(G) \geqslant (1 - \hat{\varepsilon})n$. Fix $d = (1 - \hat{\varepsilon} - 8\hat{\varepsilon}^2)n = \lfloor (1 - \varepsilon - 8\varepsilon^2)n\rfloor$.

If the lemma does not hold then by Theorem 6.3, there is a set $T \subseteq Y$ with $d|T| > \sum_{x \in X} \min(|N(x) \cap T|, d)$. Fix $|T| = (1 - \tau)n$ and $\alpha n = |\{x \in X : \min(|N(x) \cap T|, d) = d\}|$. Note that $\delta(G) \geqslant (1 - \hat{\varepsilon})n$ implies $|N(x) \cap T| \geqslant |T| - \hat{\varepsilon}n$ for all $x \in X$. Using this $d|T| > \sum_{x \in X} \min(|N(x) \cap T|, d)$ implies $d(1 - \tau)n > (1 - \tau - \hat{\varepsilon})(1 - \alpha)n^2 + d\alpha n$. Plugging in the value of d gives $(1 - \hat{\varepsilon} - 8\hat{\varepsilon}^2)(1 - \tau) > (1 - \tau - \hat{\varepsilon})(1 - \alpha) + (1 - \hat{\varepsilon} - 8\hat{\varepsilon}^2)\alpha$, and therefore $(\hat{\varepsilon} + 8\hat{\varepsilon}^2 - \alpha)\tau > 8\hat{\varepsilon}^2(1 - \alpha)$. As $\alpha \leqslant 1$ and $\tau \geqslant 0$, we must have that $\alpha < \hat{\varepsilon} + 8\hat{\varepsilon}^2$. Therefore, $\tau > \frac{8\hat{\varepsilon}^2(1 - \alpha)}{\hat{\varepsilon} + 8\hat{\varepsilon}^2 - \alpha}$ and $\alpha < \hat{\varepsilon} + 8\hat{\varepsilon}^2 < 1.1\hat{\varepsilon} \leqslant 0.002$ give $\tau > 4\hat{\varepsilon}$.

Note that $|N(x) \cap T| \leqslant (1 - \tau)n \leqslant (1 - 4\hat{\varepsilon})n \leqslant d$ which implies that $(1 - \hat{\varepsilon} - 8\hat{\varepsilon}^2)(1 - \tau)n^2 = d|T| > \sum_{x \in X} \min(|N(x) \cap T|, d) = \sum_{x \in X} |N(x) \cap T| = e(X, T) \geqslant |T|\delta(G) \geqslant (1 - \tau)(1 - \hat{\varepsilon})n^2$, a contradiction to '$8\hat{\varepsilon}^2 > 0$'. $\square$

We want versions of the above lemmas for coloured graphs. Our lemmas will furthermore provide regular subgraphs with a better global boundedness than the starting graph. To do this, we use the following lemma which shows that any properly coloured graph has a subgraph with better global boundedness. It is proved by selecting the subgraph randomly.

LEMMA 6.5. *Let $1 \overset{\text{POLY}}{\gg} \varepsilon \overset{\text{POLY}}{\gg} n^{-1}$ and $k \in \{1, 2\}$. Let G be a properly coloured, globally n/k-bounded graph on $\leqslant 2n$ vertices with $\leqslant (1 - 20\varepsilon)n$ colours having $\geqslant (1 - 20\varepsilon)n/k$ edges and*

$\delta(G) \geqslant (1 - \varepsilon^2)n$. Then G has a spanning subgraph H with $\delta(H) \geqslant (1 - \varepsilon + 18\varepsilon^2)n$ which is globally $(1 - \varepsilon)n/k$-bounded.

Proof. We say that a colour is *large* if it has $\geqslant (1 - 20\varepsilon)n/k$ edges in G. Other colours are called *small*. For a vertex v, let $\ell_H(v)$ and $\ell_G(v)$ be the numbers of large colours through v in H and G, respectively. Similarly let $s_G(v)$ be the number of small colours through v. Note that $\ell_G(v) \leqslant (1 - 20\varepsilon)n$ and $s_G(v) = d_G(v) - \ell_G(v) \geqslant 20\varepsilon n - \varepsilon^2 n$ always hold. Let H be the subgraph of G formed by deleting every edge having a large colour independently with probability $p = \varepsilon + \varepsilon^2$. The following hold for all vertices v and large colours c by linearity of expectation.

$$\mathbb{E}(|E_H(c)|) = (1 - p)|E_G(c)| \leqslant (1 - \varepsilon - \varepsilon^2)n/k,$$

$$\mathbb{E}(d_H(v)) = (1 - p)\ell_G(v) + s_G(v) = (1 - p)d_G(v) + ps_G(v)$$

$$\geqslant (1 - \varepsilon - \varepsilon^2)(1 - \varepsilon^2)n + (\varepsilon + \varepsilon^2)(20\varepsilon - \varepsilon^2)n = (1 - \varepsilon + 18\varepsilon^2 + 20\varepsilon^3)n.$$

Note that $|E_H(c)|$ and $d_H(v)$ are both 1-Lipschitz and affected by $\leqslant n$ edges. By Azuma's Inequality (Lemma 3.6), the probability that either of these deviates from its expectation by more than $\varepsilon^3 n/4$ is $\leqslant 2e^{\frac{-(\varepsilon^3 n/4)^2}{n}} = o(n^{-2})$ (using $n \overset{\text{POLY}}{\gg} \varepsilon^{-1}$). Taking a union bound, we have that with high probability all large colours have $|E_H(c)| \leqslant (1 - \varepsilon)n/k$ and all vertices have $d_H(v) \geqslant (1 - \varepsilon + 18\varepsilon^2)n$. Also, small colours c always have $|E_H(c)| \leqslant (1 - \varepsilon)n/k$. Thus, with high probability, H is globally $(1 - \varepsilon)n/k$-bounded and has $\delta(H) \geqslant (1 - \varepsilon + 18\varepsilon^2)n$, as required. $\qquad\square$

By combining this with Lemmas 6.2 and 6.4, we prove the main results of this section.

LEMMA 6.6 (Regularization lemma for high degree general graphs). *Let $n^{-1} \overset{\text{POLY}}{\ll} \gamma \overset{\text{POLY}}{\ll} \varepsilon \overset{\text{POLY}}{\ll} 1$, and let G be a properly coloured n-vertex graph with $\leqslant (1 - 20\varepsilon)n$ colours having $\geqslant (1 - 20\varepsilon)n/2$ edges and $\delta(G) \geqslant (1 - \varepsilon^2)n$. Then, G has a spanning subgraph H which is globally $(1 - \varepsilon)n/2$-bounded and (γ, δ, n)-regular for some $\delta \geqslant 1 - \varepsilon + 9\varepsilon^2$.*

Proof. First apply Lemma 6.5 with $k = 2$ in order to get a subgraph G' with $\delta(G') \geqslant (1 - \varepsilon + 18\varepsilon^2)n$ which is globally $(1 - \varepsilon)n/2$-bounded. Then, apply Lemma 6.2 to G' with $\varepsilon' = \varepsilon - 18\varepsilon^2$ to get a subgraph H which is r-regular for $r \geqslant (1 - (\varepsilon - 18\varepsilon^2) - 8(\varepsilon - 18\varepsilon^2)^2)n \geqslant (1 - \varepsilon + 9\varepsilon^2)n$ (using $\varepsilon \overset{\text{POLY}}{\ll} 1$). $\qquad\square$

LEMMA 6.7 (Regularization lemma for high degree bipartite graphs). *Let $n^{-1} \overset{\text{POLY}}{\ll} \gamma \overset{\text{POLY}}{\ll} \varepsilon \overset{\text{POLY}}{\ll} 1$, and let G be a properly coloured balanced bipartite graph on $2n$ vertices with $\leqslant (1 - 20\varepsilon)n$ colours having $\geqslant (1 - 20\varepsilon)n$ edges and $\delta(G) \geqslant (1 - \varepsilon^2)n$. Then, G has a spanning subgraph H which is globally $(1 - \varepsilon)n$-bounded and (γ, δ, n)-regular for some $\delta \geqslant 1 - \varepsilon + 9\varepsilon^2$.*

Proof. First apply Lemma 6.5 with $k = 1$ in order to get a subgraph G' with $\delta(G') \geqslant (1 - \varepsilon + 18\varepsilon^2)n$ which is globally $(1 - \varepsilon)n$-bounded. Then, apply Lemma 6.4 to G' with $\varepsilon' = \varepsilon - 18\varepsilon^2$ to get a subgraph H which is r-regular for $r \geqslant (1 - (\varepsilon - 18\varepsilon^2) - 8(\varepsilon - 18\varepsilon^2)^2)n - 1 \geqslant (1 - \varepsilon + 9\varepsilon^2)n$ (using $\varepsilon \overset{\text{POLY}}{\ll} 1$). $\qquad\square$

6.2. *Regularization using a disjoint dense graph*

The following two lemmas take a graph G which is close to being regular and a disjoint dense graph E, and modify G slightly using edges of E in order to produce a truly regular graph.

Lemma 6.9 will be applied later in the paper, while Lemma 6.8 is a technical lemma to facilitate its proof.

LEMMA 6.8. *For $d > m$, let G be a balanced bipartite graph on $2n$ vertices with $\delta(G) \geqslant d - 1$ and $2k = \sum_{v \in V(G)} \max(0, d(v) - d)$, let E be an edge-disjoint $(1, d - m)$-dense graph, and let M be a matching in E of size $m - k$ such that $d_G(v) = d - 1 \iff v \in V(M)$. Then, there is a subgraph $H \subseteq G$ and a matching N in E of size m so that $H \cup N$ is d-regular.*

Proof. The proof is by induction on k. The initial case is when $k = 0$. In this case, note that every $v \in V(G)$ must have $\max(0, d(v) - d) = 0$ which implies $\Delta(G) \leqslant d$. Since $\delta(G) \geqslant d - 1$ and $d_G(v) = d - 1 \iff v \in V(M)$ we have that $G \cup M$ is d-regular

Suppose that $k \geqslant 1$ and that the lemma holds for all $k' < k$. Let X and Y be the parts of G. Note that since $\delta(G) \geqslant d - 1$ and both X and Y have exactly $e(M)$ degree $d - 1$ vertices, we must have $k = \sum_{x \in X} \max(0, d(x) - d) = e(G) - dn + e(M) = \sum_{y \in Y} \max(0, d(y) - d)$. In particular, this implies that X and Y each have $\leqslant k$ vertices of degree $\geqslant d + 1$. Let $x \in X, y \in Y$ be vertices with $\max(0, d(x) - d), \max(0, d(y) - d) \geqslant 1$. We have $d(x) \geqslant d + 1$, $d(y) \geqslant d + 1$. Since there are $m - k$ vertices in Y of degree $d - 1$ and $\leqslant k$ vertices in Y of degree $\geqslant d + 1$, Y has at least $n - m$ vertices of degree d. Similarly, X has at least $n - m$ vertices of degree d. This implies that $N(x)$ and $N(y)$ have subsets of size $d - m$ consisting of vertices of degree d. Since E is $(1, d - m)$-dense, there is an edge $uv \in E$ with $u \in N(x)$, $v \in N(y)$, and $d(v) = d(w) = d$.

Let $G' = G - xu - yv$ and $M' = M \cup uv$. We have that M' is a matching because $d_G(u) = d_G(v) = d \neq d - 1$ which implies $u, v \notin V(M)$. Note that $\delta(G') \geqslant d - 1$, $\sum_{w \in V(G)} \max(0, d_{G'}(w) - d) = 2k - 2$, $e(M') = m - k + 1$ and $d_{G'}(v) = d - 1 \iff v \in V(M')$. By induction, there is a subgraph H of G' and a matching M in E with the required properties. $\qquad\square$

The following version of the above lemma will be easier to apply.

LEMMA 6.9 (Regularization using a disjoint dense graph). *For $d > 2m$, let G be a balanced bipartite graph on $2n$ vertices with $e(G) \leqslant dn + m$, $\delta(G) \geqslant d$ and E an edge-disjoint $(1, d/2)$-dense graph. Then, there is a subgraph $H \subseteq G$ and a matching $M \subseteq E$ of size $\leqslant m$ so that $H \cup M$ is d-regular.*

Proof. Since $\delta(G) \geqslant d$, we have that $\sum_{v \in V(G)} \min(0, d(v) - d) = 2e(G) - 2dn \leqslant 2m$. The result follows by applying Lemma 6.8 with $m' = k = \sum_{v \in V(G)} \min(0, d(v) - d)/2 \leqslant m$ and $M = \emptyset$. To see that E is $(1, d - m')$-dense for this application, note that $d > 2m$ is equivalent to $d - m > d/2$. $\qquad\square$

6.3. *Regularization by adding vertices*

Here we show that a nearly-regular graph can be made regular by adding a small number of vertices (and edges adjacent to those vertices). We will need the Gale–Ryser Theorem concerning which degree sequences are realizable by bipartite graphs (see [**46**]).

THEOREM 6.10 (Gale, Ryser). *Let $x_1 \geqslant \cdots \geqslant x_m$, and $y_1 \geqslant \cdots \geqslant y_n$ be non-negative numbers. There exists a bipartite graph with parts X, Y with degree sequence $x_1, \ldots, x_m$ in X and $y_1, \ldots, y_n$ in Y if, and only if, we have $\sum_{i=1}^{n} y_i = \sum_{i=1}^{m} x_i$ and $\sum_{i=1}^{t} y_i \leqslant \sum_{i=1}^{m} \min(t, x_i)$ for $t = 1, \ldots, n$.*

We use the Gale–Ryser Theorem to prove the regularization lemma of this section.

LEMMA 6.11 (Regularization by adding vertices). *Suppose that* $0.01 \geqslant \delta \overset{\text{POLY}}{\gg} \gamma \overset{\text{POLY}}{\gg} n^{-1}$. *Every* (γ, δ, n)-*regular balanced bipartite graph* G *on* $2n$ *vertices is contained in a* d-*regular balanced bipartite graph* G' *with parts of size* $\leqslant (1 + 9\gamma)n$ *for* $d = \lceil (1 + 5\gamma)\delta n \rceil$, *where, additionally, all edges of* $G' \setminus G$ *touch* $V(G') \setminus V(G)$.

Proof. Let $m = dn - e(G)$. Since $e(G) = \sum_{x \in X} d(x) = (1 \pm \gamma)\delta n^2$, we have $2\gamma \delta n^2 \leqslant m \leqslant 8\gamma \delta n^2$. Let X' and Y' be two sets of new vertices with $|X'| = |Y'| = \lceil \frac{m}{d} \rceil$. Note that $\gamma n \leqslant |X'|, |Y'| \leqslant 9\gamma n$. Choose a graph H between X' and Y' with exactly $d\lceil \frac{m}{d} \rceil - m = d|X'| - m$ edges (a graph with this many edges exists since $\gamma \overset{\text{POLY}}{\gg} n^{-1}$ implies $|X'||Y'| \geqslant \gamma^2 n^2 \geqslant d$). Moreover, choose H so that $\Delta(H)$ is as small as possible. This ensures that $\delta(H) \geqslant \Delta(H) - 1$. For each vertex $v \in V(G) \cup V(H)$, let $k_v = d - d_{G \cup H}(v)$. Note that $\delta(H) \geqslant \Delta(H) - 1$ implies that any integer t has either '$t \leqslant d - \Delta(H)$' or '$t \geqslant d - \delta(H)$', and so

$$\sum_{x' \in X'} \min(t, k_{x'}) = \begin{cases} \displaystyle\sum_{x' \in X'} t & \text{if } t \leqslant d - \Delta(H) \\ \displaystyle\sum_{x' \in X'} k_{x'} & \text{if } t \geqslant d - \delta(H). \end{cases} \tag{29}$$

Since G is (γ, δ, n)-regular, for $v \in X \cup Y$ we have $k_v \leqslant 9\gamma d$. Note that $\sum_{x' \in X'} k_{x'} = d|X'| - e_H(X', Y') = m = d|Y| - e_G(X, Y) = \sum_{y \in Y} k_y$. Order the vertices of Y as $y_1, \ldots, y_n$ such that $k_{y_1} \geqslant k_{y_2} \geqslant \cdots \geqslant k_{y_n}$. Note that, for all $t = 1, \ldots, n$, $\sum_{i=1}^{t} k_{y_i} \leqslant 9\gamma dt \leqslant \gamma n t \leqslant |X'| t = \sum_{x' \in X'} t$. We also have $\sum_{i=1}^{t} k_{y_i} \leqslant \sum_{y \in Y} k_y = \sum_{x' \in X'} k_{x'}$. Combining these with (29), we get that for all t, we have $\sum_{i=1}^{t} k_{y_i} \leqslant \sum_{x' \in X'} \min(t, k_{x'})$. By the Gale–Ryser Theorem, there is a graph J_1 between X' and Y such that $d_{J_1}(v) = k_v$. By symmetry, there is a graph J_2 between X and Y' with $d_{J_2}(v) = k_v$ for all $v \in X \cup Y'$. Now $G' = G \cup H \cup J_1 \cup J_2$ is d-regular. Note that all edges of $G' \setminus G$ touch $X' \cup Y' = V(G') \setminus V(G)$, completing the proof of the lemma. $\square$

7. Completion

Our strategy for finding rainbow perfect matchings and Hamiltonian cycles is to first find nearly-perfect matchings or near-Hamiltonian cycles and then modify them. In this section, we collect the 'modification lemmas' which we use to complete nearly-spanning structures into truly spanning ones.

In all lemmas of this section, we will have a dense graph which is disjoint from the matchings/cycles which we are trying to complete. The matchings/cycles are turned into what we want by modifying them gradually using edges of the dense graph.

7.1. Perfect matchings

The following lemma extends a matching by one edge.

LEMMA 7.1. *Suppose that we have the following edge-disjoint subgraphs in a balanced bipartite graph* G *with parts* X, Y *of size* n.

- *A matching* M *with* $e(M) \geqslant (1 - \theta)n$.
- *A* $(1, \theta n)$-*dense graph* E.
- *Graphs* D_X, D_Y *with* $d_{D_X}(x') \geqslant 2\theta n$ *for each* $x' \in X$ *and* $d_{D_Y}(y') \geqslant 2\theta n$ *for each* $y' \in Y$.

Let $x \in X$ *and* $y \in Y$ *be vertices outside* M. *Then, there are vertices* u, v, m_u, m_v *with* $uv \in E$, $xm_u \in D_X$, $ym_v \in D_Y$, $um_u, vm_v \in M$ *such that* $M' = (M \cup \{xm_u, uv, ym_v\}) \setminus \{um_u, vm_v\}$ *is a matching.*

Proof. Let σ be the permutation which exchanges vertices of M and fixes all other vertices. Note that since $d_{D_X}(x) \geqslant 2\theta n$, $d_{D_Y}(y) \geqslant 2\theta n$ and $e(M) \geqslant (1-\theta)n$, we have $|N_{D_X}(x) \cap V(M)|, |N_{D_Y}(y) \cap V(M)| \geqslant \theta n$. Since E is $(1, \theta n)$-dense, it has an edge uv from $\sigma(N_{D_X}(x) \cap V(M))$ to $\sigma(N_{D_Y}(y) \cap V(M))$. Now taking $m_u = \sigma(u)$ and $m_v = \sigma(v)$ gives vertices satisfying the lemma. $\square$

By iterating the above lemma, we can turn nearly-perfect rainbow matchings into perfect ones.

LEMMA 7.2 (Completing a matching). *Let* $1 \overset{\text{\tiny POLY}}{\gg} \theta, p \overset{\text{\tiny POLY}}{\gg} \varepsilon \overset{\text{\tiny POLY}}{\gg} n^{-1}$. *Suppose that we have the following colour-disjoint subgraphs in a properly coloured, balanced bipartite graph G with parts X, Y of size n.*

- *A rainbow matching M_0 with $e(M_0) \geqslant (1-\varepsilon)n$.*
- *A $(p(\theta n)^2, \theta n)$-dense graph E.*
- *Graphs D_X, D_Y with $d_{D_X}(x) \geqslant 3\theta n$ for each $x \in X$ and $d_{D_Y}(y) \geqslant 3\theta n$ for each $y \in Y$.*

Then, there is a perfect rainbow matching N in $M_0 \cup E \cup D_X \cup D_Y$ using $\leqslant \varepsilon n$ edges in each of E, D_X and D_Y, where, additionally edges of D_X in N pass through $X \setminus V(M_0)$ and edges of D_Y in N pass through $Y \setminus V(M_0)$.

Proof. We will repeatedly apply Lemma 7.1 to produce rainbow matchings $M_1, \ldots, M_{n-e(M_0)}$ with $e(M_i) = e(M_0) + i$ and $V(M_0) \subseteq V(M_i)$. We will maintain that M_i always has at most i edges of each of E, D_X, D_Y, with the edges of $M_i \cap D_X$ passing through $X \setminus V(M_0)$ and the edges of $M_i \cap D_Y$ passing through $Y \setminus V(M_0)$. When finished, $N = M_{n-e(M_0)}$ will then satisfy all the requirements of the lemma.

At the $(i+1)$st application, we apply Lemma 7.1 with $M = M_i$, vertices $x_i \in X \setminus V(M_i), y_i \in Y \setminus V(M_i)$ and $E^i = E \setminus C(M_i)$, $D_X^i = D_X \setminus C(M_i), D_Y^i = D_Y \setminus C(M_i)$ — the graphs E, D_X, D_Y with all the edges with colour in $C(M_i)$ removed. Using Lemma 5.10 and $e(E \cap M_i) \leqslant i$, the graph E^i is $(p(\theta n)^2 - i\theta n, \theta n)$-dense (and so $(1, \theta n)$-dense since $\theta, p \overset{\text{\tiny POLY}}{\gg} \varepsilon$ implies $p(\theta n)^2 \geqslant \varepsilon n \theta n$). Also $e(D_X \cap M_i), e(D_Y \cap M_i) \leqslant i$ gives $d_{D_X^i}(x), d_{D_Y^i}(y) \geqslant 3\theta n - i \geqslant 2\theta n$ for each $x \in X$ and $y \in Y$. These show that the assumptions of Lemma 7.1 hold for M_i, $E^i, D_X^i, D_Y^i, x_i, y_i$ and so we can apply it to obtain a matching M_{i+1} containing one more edge than M_i. The matching M_{i+1} is necessarily rainbow since it is a union of a submatching of M_i (which is rainbow), and one edge from each of E^i, D_X^i and D_Y^i (which are all colour-disjoint from each other and from M_i). From Lemma 7.1, the new edges of M_{i+1} in D_X and D_Y pass through $x_i \in X \setminus V(M_0)$ and $y_i \in Y \setminus V(M_0)$, respectively, as required. $\square$

7.2. Hamiltonian cycles

The following lemma joins two long cycles together using edges from some disjoint, dense graphs.

LEMMA 7.3. *Let* $1 \geqslant \lambda \overset{\text{\tiny POLY}}{\gg} p \overset{\text{\tiny POLY}}{\gg} \theta \overset{\text{\tiny POLY}}{\gg} n^{-1}$. *Let C_1, C_2 be two vertex-disjoint cycles of length $\geqslant \lambda n$. Let E, F, G be $(p(\theta n)^2, \theta n)$-dense graphs. Suppose that C_1, C_2, E, F and G are all edge-disjoint. Let $xy \in E(C_1) \cup E(C_2)$.*

Then, there is a cycle C in $C_1 \cup C_2 \cup E \cup F \cup G$ with vertex set $V(C_1) \cup V(C_2)$ containing 1 edge of each of E, F, G, and, additionally, $xy \in E(C)$.

Proof. Without loss of generality, assume that $|C_1| \leqslant |C_2|$. Choose arbitrary orientations of C_1 and C_2. Let σ be the permutation mapping v to its successor (in the cycle containing v).

Let X_1 be the set of vertices $v \in V(C_1)$ with $|N_E(v) \cap V(C_2)| < 2\theta n + 2$ and let X_2 be the set of vertices $v \in V(C_1)$ with $|N_F(\sigma(v)) \cap V(C_2)| < 2\theta n + 2$.

Suppose $|X_1| \geqslant \lambda n/3$. Noting that $|C_2| \geqslant |C_1| \geqslant |X_1|$, pick a set $Y_1 \subset V(C_2)$ with $|Y_1| = |X_1|$ and note that $e_E(X_1, Y_1) < |X_1|(2\theta n + 2) < p|X_1|^2$ (as $\lambda \overset{\text{POLY}}{\gg} p, \theta$), contradicting that E is $(p(\theta n)^2, \theta n)$-dense. Thus, $|X_1| < \lambda n/3$. Similarly, $|X_2| < \lambda n/3$, and thus we may pick $x_0 \in V(C_1) \setminus (X_1 \cup X_2 \cup \{x, y\})$.

Let $y_0 = \sigma(x_0)$. We have $|(N_E(x_0) \cap V(C_2)) \setminus \{x, y\}| \geqslant 2\theta n$ and $|(N_F(y_0) \cap V(C_2)) \setminus \{x, y\}| \geqslant 2\theta n$. Since G is $(p(\theta n)^2, \theta n)$-dense, by considering appropriate disjoint subsets with size θn of $(N_E(x_0) \cap V(C_2)) \setminus \{x, y\}$ and $(N_F(y_0) \cap V(C_2)) \setminus \{x, y\}$, respectively, we have the following. There is an edge $zw \in E(G)$ with $\sigma(z) \in (N_E(x_0) \cap V(C_2)) \setminus \{x, y\}$ and $\sigma(w) \in (N_F(y_0) \cap V(C_2)) \setminus \{x, y\}$. Now $C = C_1 \cup C_2 - x_0 y_0 - z\sigma(z) - w\sigma(w) + x_0\sigma(z) + y_0\sigma(w) + zw$ is a cycle with the required properties (see Figure 1). □

Next we prove a similar lemma which joins cycles together. In the next two lemmas, we will have both directed and undirected graphs on the same vertex set. The cycles we obtain in both lemmas will contain a mixture of edges from the directed and undirected graphs. In such a situation, 'cycle' just means a graph which is a cycle after turning all directed edges into undirected edges, that is, we do not care about directions of edges in cycles at all (the actual purpose of the directed edges in the lemmas is to control degrees through certain vertices).

LEMMA 7.4. Let $1 \geqslant \delta \overset{\text{POLY}}{\gg} \theta \overset{\text{POLY}}{\gg} n^{-1}$ and $m \leqslant \theta n/2$. Let $\mathcal{C} = \{C_0, \ldots, C_m\}$ be a 2-factor. For each $i = 0, \ldots, m$, let $x_i y_i \in E(C_i)$. Let E be $(1, \theta n)$-dense and D_X, D_Y digraphs with $d^+(D_X), d^+(D_Y) \geqslant \delta n$. Suppose that $\mathcal{C}, E, D_X$ and D_Y are all edge-disjoint. Suppose $|C_0| \leqslant \delta n/2$.

Then there is a 2-factor $\mathcal{C}'$ with $\leqslant m$ cycles in which each cycle contains an edge $x_i y_i$ for some $i \geqslant 1$. Additionally, $\mathcal{C}'$ contains 1 edge from E, 1 edge from F which starts at x_0 and 1 edge from G which starts from y_0.

Proof. Choose arbitrary orientations of $C_1, \ldots, C_m$. Let σ be the permutation mapping v to its successor (in the cycle containing v). Let $U = \{x_i, y_i : i \geqslant 1\}$ and note that $|U| \leqslant 2m \leqslant \theta n$. We have $|N^+_{D_X}(x_0) \setminus (U \cup C_0)| \geqslant \delta n - \theta n - \delta n/2 \geqslant \delta n/3$ and $|N^+_{D_Y}(y_0) \setminus (U \cup C_0)| \geqslant \delta n - \theta n - \delta n/2 \geqslant \delta n/3$. Since E is $(1, \theta n)$-dense and $\theta n \leqslant \delta n/6$, by choosing appropriate disjoint subsets of size θn, we have the following. There is an edge $zw \in E$ with $\sigma(z) \in N^+_{D_X}(x_0) \setminus (U \cup C_0)$ and $\sigma(w) \in N^+_{D_Y}(y_0) \setminus (U \cup C_0)$. Now $\mathcal{C}' = \mathcal{C} - x_0 y_0 - z\sigma(z) - w\sigma(w) + x_0\sigma(z) + y_0\sigma(w) + zw$ is a 2-factor with the required properties (see Figure 1). □

By iterating the above lemmas we obtain the following lemma which turns a rainbow 2-factor into a rainbow Hamiltonian cycle. Once again we have have a combination of directed and undirected graphs. When we say a directed graph is properly coloured, we mean that the underlying undirected graph is properly coloured.

LEMMA 7.5 (Completing a Hamiltonian cycle). Let $1 \overset{\text{POLY}}{\gg} \delta \overset{\text{POLY}}{\gg} p \overset{\text{POLY}}{\gg} \theta \overset{\text{POLY}}{\gg} n^{-1}$ and $m \leqslant p\theta n$. Suppose that we have the following colour-disjoint, edge-disjoint, properly coloured graphs on a set of n vertices.

- A rainbow 2-factor $\mathcal{C} = \{C_0, \ldots, C_m\}$.
- $(p(\theta n)^2 + m\theta n, \theta n)$-dense graphs E, F, G.
- Digraphs D_X, D_Y with $d^+(D_X), d^+(D_Y) \geqslant \delta n + m$.
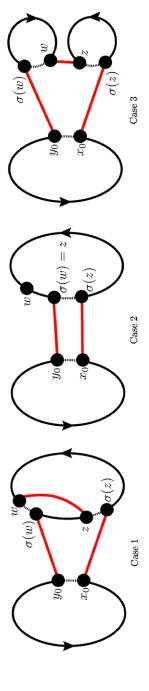
FIGURE 1 (colour online). *The three kinds of rotations we use in Lemmas 7.3 and Lemma 7.4. The dashed edges are removed from the cycles, while the solid red edges are added. In each case, the resulting graph is a single cycle. In Lemma 7.3, Cases 1 and 2 are relevant, while in Lemma 7.4, all three cases are relevant.*

Case 2 should be thought of as a degenerate version of Case 1 when we have $z = \sigma(w)$ (or symmetrically $w = \sigma(z)$). Note that this case never actually occurs because Lemmas 7.3 and Lemma 7.4 assume that E, G are disjoint from all C_i. We include Case 2 in the figure for purposes of exposition.

For each $i = 0, \ldots, m$, let x_i, y_i be a pair of adjacent vertices in C_i. Then, there is a rainbow Hamiltonian cycle C in $E \cup F \cup G \cup D_X \cup D_Y \cup C_0 \cup \cdots \cup C_m$ so that, additionally, $C \setminus (C_0 \cup \cdots \cup C_m)$ has $\leqslant m$ edges of each of E, F, G, D_X, D_Y, the edges of C in D_X all start in $\{x_0, \ldots, x_m\}$, and the edges of C in D_Y all start in $\{y_0, \ldots, y_m\}$.

Proof. Choose λ with $\delta \overset{\text{POLY}}{\gg} \lambda \overset{\text{POLY}}{\gg} p$. The proof is by induction on m. The initial case is when $m = 0$ in which case the lemma is trivial. Suppose that $m \geqslant 1$. If $|C_0|, |C_1| \geqslant \lambda n$, then apply Lemma 7.3 to C_0, C_1, E, F, G. If $|C_0|$ or $|C_1| \leqslant \lambda n$, then apply Lemma 7.4 to $\mathcal{C}, E, D_X, D_Y$. In either case, we get a new 2-factor $\mathcal{C}'$ with $\leqslant m$ cycles, where each cycle contains an edge $x_i y_i$ for some i. Additionally $\mathcal{C}'$ contains at most one edge from each of E, F, G, D_X, D_Y, with remaining edges from $\mathcal{C}$. This implies that $\mathcal{C}'$ is rainbow since $\mathcal{C}$ was rainbow and E, F, G, D_X, D_Y were all colour-disjoint from each other and from $\mathcal{C}$. From Lemma 7.4, we also have that if $\mathcal{C}'$ has edges in D_X and D_Y, then they start at x_0 and y_0, respectively.

Let E', F', G', D'_x, D'_Y be E, F, G, D_X, D_Y with colours of $\mathcal{C}' \setminus \mathcal{C}$ deleted to get $(p(\theta n)^2 + m\theta n - \theta n, \theta n)$-dense graphs E', F', G' (using Lemma 5.10), and digraphs D'_X, D'_Y with $d^+(D'_X), d^+(D'_Y) \geqslant \delta n + m - 1$. The lemma holds by induction. $\square$

8. Near-decompositions into rainbow structures

In this section, we prove our main results on matchings and Hamiltonian cycles. Most of the results here are of the form 'every properly coloured graph with certain properties can be nearly-decomposed into rainbow matchings/2-factors/Hamiltonian cycles'. These results build on one another. First, we find near-decompositions into nearly-perfect matchings. Then, we use completion results from the previous section to find near-decompositions into perfect matchings. We use these to find near-decompositions into 2-factors. Then, we again use completion results from the previous section to find near-decompositions into Hamiltonian cycles.

8.1. Nearly-perfect matchings

In this section, we show that every properly coloured d-regular, globally d-bounded bipartite graph has a near-decomposition into nearly-perfect rainbow matchings. This is proved by iteratively finding such matchings individually using Lemma 4.6. For a d-regular, globally d-bounded bipartite graph G_d, consider the following recursive process producing matchings $M_d, \ldots, M_{\varepsilon d}$.

P1: For $t = d, \ldots, \varepsilon d$, apply Lemma 4.6 to G_t in order to partition its edges into a randomized rainbow matching M_t and a graph G_{t-1}.

We emphasize that this process is run in *decreasing* order with t. The reason for this is that if we define the process like this, then the graphs G_t turn out to be approximately t-regular for every t.

If we were able to run this process for $(1 - o(1))d$ many steps, then we would obviously produce the desired $(1 - o(1))d$ edge-disjoint nearly-perfect rainbow matchings in G_0. To show that we can run it for that long, we need to show that with high probability G_t satisfies the assumptions of Lemma 4.6. There are two assumptions of that lemma which need to be maintained: $(\gamma, t/n, n)$-regularity and global $(1 + \gamma)t$-boundedness.

Maintaining global boundedness. Recall that in the matching M_t produced in G_t by Lemma 4.6, every edge ends up in M_t with probability roughly t^{-1}. This means that at step $t - 1$ of P1, for any colour c, we have

$$\mathbb{E}(|E_{G_{t-1}}(c)|) = \mathbb{P}(c \notin C(M_t))|E_{G_t}(c)| + \mathbb{P}(c \in C(M_t))(|E_{G_t}(c)| - 1)$$

$$\approx \left(1 - \frac{|E_{G_t}(c)|}{t}\right)|E_{G_t}(c)| + \frac{|E_{G_t}(c)|}{t}(|E_{G_t}(c)| - 1)$$

$$\leqslant t - 1.$$

Here, the last inequality is equivalent to '$|E_{G_t}(c)| \leqslant t$'. Thus, one would expect the global t-boundedness of G_t to be preserved throughout the entire process. By using Azuma's Inequality, we can show that this happens with high probability.

Maintaining regularity. Here we explain how to preserve $(\gamma, t/n, n)$-regularity between the applications of Lemma 4.6. First note that if Lemma 4.6 produced *perfect* matchings, then there would be nothing to check — then G_t would always be t-regular (and hence $(0, t/n, n)$-regular). But the matchings produced by Lemma 4.6 have size $(1 - o(1))n$, and so over time, one would expect the maximum degree of the graph to become bigger than t after a large number of steps. One thing that we will never lose is the minimum degree — the graphs will always have $\delta(G_t) \geqslant t$, since there can be at most one edge from each matching M_i present at any vertex.

To preserve regularity, we introduce another step to our process in addition to P1. Fix some large constant k, and do the following:

P2: whenever $t \equiv 0 \pmod{k}$, modify G_t slightly to make it into a t-regular graph.

This step ensures that G_t is $(\gamma, t/n, n)$-regular for all t and suitable γ. Indeed, for any t, there is some $k' \leqslant k$ with $G_{t+k'}$ $(t + k')$-regular and $\Delta(G_t) \leqslant \Delta(G_{t+k})$. Thus, we have $t \leqslant \delta(G_t) \leqslant \Delta(G_t) \leqslant t + k$, which implies that G_t is $(k/t, t/n, n)$-regular.

Step P2 is performed using Lemma 6.9. This lemma turns a graph G_t with $\delta(G_t) \geqslant t$ into a t-regular graph G'_t by deleting some edges and adding a small matching N disjoint from G_t. Edges in this matching N are given a new 'dummy colour' which was previously unused in G_t. While these dummy colours can end up in our matchings M_t, the total number of dummy colours is small (at most n/k), and so after deleting the dummy colours, we still have nearly-perfect matchings in G.

A concentration lemma

The following lemma will be used to show that the global boundedness of a graph decreases suitably after repeated applications of Lemma 4.6.

LEMMA 8.1. *For* $C\varepsilon \leqslant 0.1$ *and* $m \leqslant D$, *suppose that we have random variables* $X_0, X_1, X_2, \ldots, X_m$, *with* $D - m - 1 \geqslant \frac{D}{C}$ *and* $\frac{D}{C} \leqslant X_0 \leqslant D$, *such that, for every* $t = 0, \ldots, m - 1$, *and for any values of* $X_0, \ldots, X_t$, *we have,*

$$X_{t+1} = \begin{cases} X_t - 1 & \text{with probability } (1 - \varepsilon)\dfrac{1}{D-t}X_t \\[2mm] X_t & \text{with probability } 1 - (1 - \varepsilon)\dfrac{1}{D-t}X_t. \end{cases}$$

Then, we have,

$$\mathbb{P}\left(X_m \geqslant (1 + 3C\varepsilon)\left(1 - \frac{m}{D}\right)X_0\right) \leqslant e^{\frac{-\varepsilon^2 D}{18C^4}}.$$

Proof. Let $q = \mathbb{P}(X_{t+1} = X_t - 1 | X_0, \ldots, X_t) = (1 - \varepsilon)\frac{1}{D-t}X_t$, and note that $\mathbb{E}(X_{t+1}|X_0, \ldots, X_t) = X_t(1 - q) + (X_t - 1)q = \left(\frac{D-t-1+\varepsilon}{D-t}\right)X_t$. Let $Y_t = (1 - C\varepsilon/D)^t \frac{X_t}{D-t}$. Note

that $\mathbb{E}(Y_{t+1}|Y_0,\ldots,Y_t) = (1 - C\varepsilon/D)^{t+1}\frac{\mathbb{E}(X_{t+1}|X_0,\ldots,X_t)}{D-t-1} = (1-C\varepsilon/D)^{t+1}\frac{D-t-1+\varepsilon}{(D-t)(D-t-1)}X_t = (1-C\varepsilon/D)(1+\frac{\varepsilon}{D-t-1})Y_t \leqslant (1-C\varepsilon/D)(1+C\varepsilon/D)Y_t \leqslant Y_t.$ This shows that Y_t is a supermartingale.

Note that for $t \leqslant m$, we have

$$|Y_{t+1} - Y_t| = (1 - C\varepsilon/D)^t \left| \frac{(1-C\varepsilon/D)X_{t+1}}{D-t-1} - \frac{X_t}{D-t} \right|$$

$$= (1-C\varepsilon/D)^t \left| \frac{X_{t+1}-X_t}{(D-t-1)} - \frac{C\varepsilon X_{t+1}}{D(D-t-1)} + \frac{X_t}{(D-t-1)(D-t)} \right|$$

$$\leqslant (1-C\varepsilon/D)^t \left(\frac{1}{D-t-1} + \frac{C\varepsilon}{D-t-1} + \frac{D}{(D-t-1)(D-t)} \right)$$

$$\leqslant \frac{3C^2}{D}.$$

The third inequality uses the triangle inequality, the 1-Lipschitzness of X_t and $X_t \leqslant X_0 \leqslant D$. the fourth inequality comes from $D - t - 1 \geqslant D - m - 1 \geqslant \frac{D}{C}$ and $C\varepsilon \leqslant 0.1$. Hence, Y_t is $\frac{3C^2}{D}$-Lipschitz.

By Lemma 3.7, we have,

$$\mathbb{P}(Y_m \geqslant Y_0 + s) \leqslant e^{\frac{-s^2 D^2}{18C^4 m}}.$$

Substituting $Y_0 = X_0/D$, $s = \gamma X_0/D$ and using $X_0 \geqslant D/C$ and $m \leqslant D$ gives

$$\mathbb{P}\left(Y_m \geqslant (1+\gamma)\frac{X_0}{D} \right) \leqslant e^{\frac{-\gamma^2 X_0^2}{18C^4 m}} \leqslant e^{\frac{-\gamma^2 D}{18C^6}}. \tag{30}$$

Note that, as $m \leqslant D$ and $C\varepsilon \leqslant 0.01$, by (4) and (6), we have, $\frac{1+C\varepsilon}{(1-C\varepsilon/D)^m} \leqslant 1 + 3C\varepsilon$. Substituting $Y_m = (1-C\varepsilon/D)^m \frac{X_m}{D-m}$ and $\gamma = C\varepsilon$ into (30), and using this gives

$$\mathbb{P}\left(X_m \geqslant (1+3C\varepsilon)\left(1-\frac{m}{D}\right)X_0 \right) \leqslant \mathbb{P}\left(X_m \geqslant (1+C\varepsilon)\frac{(D-m)X_0}{(1-C\varepsilon/D)^m D} \right) \leqslant e^{\frac{-\varepsilon^2 D}{18C^4}}. \qquad \square$$

Analysis of the random process

Now we prove the first decomposition result of this paper. All our other decomposition results build on this. It produces a near-decomposition into nearly-perfect rainbow matchings in a δn-regular graph which is globally $(1-\sigma)\delta n$-bounded.

LEMMA 8.2. *Suppose that we have n, ℓ, δ, σ with $n^{-1} \overset{\text{POLY}}{\ll} \sigma \overset{\text{POLY}}{\ll} \delta \leqslant 1$ and $\ell \overset{\text{POLY}}{\ll} n$.*
Let G be a coloured balanced bipartite graph on $2n$ vertices which is δn-regular, locally ℓ-bounded and globally $(1-\sigma)\delta n$-bounded. Then, G has $(1-\sigma)\delta n$ edge-disjoint rainbow matchings $M_1,\ldots,M_{(1-\sigma)\delta n}$ with $e(M_i) = (1-\sigma)n$ for all i.

Proof. Choose k, γ, ν, p so that $n^{-1} \overset{\text{POLY}}{\ll} \gamma \overset{\text{POLY}}{\ll} p \overset{\text{POLY}}{\ll} \nu \overset{\text{POLY}}{\ll} k^{-1} \overset{\text{POLY}}{\ll} \sigma \overset{\text{POLY}}{\ll} \delta \leqslant 1$. Set $D = \lceil (1-k^{-1})\delta n \rceil$.

CLAIM 8.3. *There is a D-regular subgraph G_D of G and a $(6pnD, \nu D/6)$-dense graph E_D with G_D and E_D edge-disjoint.*

Proof. Apply Lemma 5.8 to G with $\mu = \nu\delta/8$ and $p' = k^{-1}\delta$ in order to find a $(\delta n - \lfloor k^{-1}\delta n \rfloor)$-regular subgraph G_D of G so that $E_D = \overline{G_D}$ is $(0.48k^{-1}\delta(\nu\delta n/8)^2, \nu\delta n/8)$-dense

(to apply Lemma 5.8, we use that $1 \geqslant k^{-1}\delta, \nu\delta/8 \overset{\text{POLY}}{\gg} n^{-1}$). Since $\delta n - \lfloor k^{-1}\delta n \rfloor = D$, the graph G_D is D-regular. Note that E_D is $(6pnD, \nu\delta n/8)$-dense (using $6pnD \leqslant 0.48k^{-1}\delta(\nu\delta n/8)^2$ which comes from $p \overset{\text{POLY}}{\ll} k^{-1}, \nu, \delta$). Now $\nu D/6 \geqslant \nu\delta n/8$ implies the claim. $\qquad\square$

We will define a random process producing spanning graphs $G_{D-1}, \ldots, G_{\nu D}, E_{D-1}, \ldots, E_{\nu D}$ and rainbow matchings $M_{D-1}, \ldots, M_{\nu D}$. They will always have the following properties:

(a) G_d is $(\gamma, d/n, n)$-regular;
(b) $M_d \subseteq E(G_d)$ with $e(M_d) \geqslant (1-p)n$;
(c) $\delta(G_d) \geqslant d$ and $e(G_d) \leqslant dn + kpn$;
(d) E_d is $(1, \frac{d-1}{2})$-dense and edge-disjoint from G_d;
(e) • if $k \nmid d$, then $G_{d-1} \subseteq G_d$;
 • if $k \mid d$, then $G_{d-1} \setminus G_d$ is a matching of size $\leqslant 4kpn$ in E_d. This matching has a dummy colour c_d.

Note (a) (c) and (d) hold for G_D and E_D: indeed G_D is D-regular (implying (a) and (c)), while E_D is $(6pnD, \nu D/6)$-dense (which combined with $(D-1)/2 \geqslant \nu D$ implies (d)).

The process producing $G_{D-1}, \ldots, G_{\nu D}, E_{D-1}, \ldots, E_{\nu D}$ and $M_{D-1}, \ldots, M_{\nu D}$ is the following.

• If G_d is globally d-bounded, then M_d is the (random) matching produced from Lemma 4.6 applied to G_d with $n = n$, $\gamma = \gamma$, $p' = p/20$, $\ell = \ell$, $\delta' = d/n$. Note that $1 \geqslant \delta' \overset{\text{POLY}}{\gg} p' \overset{\text{POLY}}{\gg} \gamma \overset{\text{POLY}}{\gg} n^{-1}$ and $n \overset{\text{POLY}}{\gg} \ell$ hold with these parameters (using $\delta \geqslant \delta' \geqslant \nu\delta/2$), allowing us to apply the lemma. Additionally:

 – if $k \nmid d$, then let $G_{d-1} = G_d \setminus M_d$ and $E_{d-1} = E_d$;
 – if $k \mid d$, then apply Lemma 6.9 to $G_d \setminus M_d$ and E_d with $d' = d-1$ and $m = e(G_d \setminus M_d) - (d-1)n$ (using that, by (d), E_d is $(1, d'/2)$-dense, by (c), $e(M_d) \geqslant (1-p)n$ and by $p \overset{\text{POLY}}{\ll} k^{-1}, \nu, \delta$ we have $2m = 2(e(G_d \setminus M_d) - (d-1)n) \leqslant 4kpn \leqslant \nu\delta n/2 \leqslant d'$). We get a subgraph H_d and a matching N_d in E_d of size $\leqslant 4kpn$. Let $G_{d-1} = H_d \cup N_d$ with the edges of N_d given the dummy colour c_d to get a $(d-1)$-regular graph. Let $E_{d-1} = E_d \setminus N_d$.

• If G_d is not globally d-bounded, then we stop the process.

We show that the properties we need in the process hold.

CLAIM 8.4. *As long as the process goes on, (a)–(e) hold.*

Proof. First we prove (b), (d) and (e).

(b) This is immediate since M_d was produced by applying Lemma 4.6 to G_d with $p' = p/20$.
(d) We have $E_d = E_D \setminus \bigcup_{i \in [d+1, D], k|i} N_i$ and $e(N_i) \leqslant 4kpn$. As E_D is $(6pnD, \nu D/6)$-dense, this implies that E_d is $(6pnD - 4kpn\lceil(D-d)/k\rceil, \nu D/6)$-dense (using Lemma 5.9). Since $\frac{d-1}{2} \geqslant \nu D/6$ and $6pnD - 4kpn\lceil(D-d)/k\rceil \geqslant 1$, this implies that E_d is $(1, \frac{d-1}{2})$-dense
(e) This is immediate from the construction of G_{d-1}: When $k \nmid d$, then $G_{d-1} = G_d \setminus M_d \subseteq G_d$ holds. When $k \mid d$, then $G_{d-1} = N_d \cup H_d \subseteq N_d \cup (G_d \setminus M_d)$ with N_d 'a matching in E_d of size $\leqslant 4kpn$' which has dummy colour c_d.

Next we prove (a) and (c). First recall that (a) and (c) hold for the starting graph G_D. If $k \mid d+1$, then, by construction, G_d is d-regular which implies both (a) and (c).

Suppose then that $k \nmid d+1$ and $d < D$. Fix $\hat{d} = \min(k\lceil(d+1)/k\rceil - 1, D)$. Note that $G_{\hat{d}}$ is $\hat{d}$-regular (as $\hat{d} = D$ or $k \mid \hat{d}+1$ in which case $G_{\hat{d}}$ is $\hat{d}$-regular from the application of Lemma 6.9). Also note that $0 < \hat{d} - d \leqslant k-1$ (the first inequality comes from $d < D$ and $k \nmid d+1$. The

second inequality comes from $k\lceil (d+1)/k\rceil - 1 \leqslant d + k - 1$). Because $k \nmid \hat{d}, \hat{d} - 1, \ldots, d + 1$, we have that

$$G_d = G_{\hat{d}} \setminus (M_{\hat{d}} \cup M_{\hat{d}-1} \cup \cdots \cup M_{d+1}). \tag{31}$$

Since the graphs M_i are matchings, this implies $\delta(G_d) \geqslant \delta(G_{\hat{d}}) - (\hat{d} - d) = d$. Next, (31) implies $\Delta(G_{d-1}) \leqslant \Delta(G_{\hat{d}}) = \hat{d} \leqslant d + k$, which combined with $\delta(G_d) \geqslant d$ implies that G_d is $(k/d, d/n, n)$-regular and hence $(\gamma, d/n, n)$-regular since $k/d \leqslant k/\nu D \leqslant 2k/\nu \delta n \leqslant \gamma$. Thus, (a) holds. Finally, (31) implies $e(G_d) \leqslant e(G_{\hat{d}}) - e(M_{\hat{d}}) - \cdots - e(M_{d+1}) \leqslant \hat{d}n - (\hat{d} - d)(1-p)n \leqslant dn + kpn$, completing the proof that (c) holds. $\square$

To show that the process does not end too early, it remains to show that G_d is globally d-bounded.

CLAIM 8.5.   *With probability* $\geqslant 1 - n^{-1}$, G_d *is globally* d-*bounded for* $d = D, \ldots, \nu D$.

Proof. Note that by (e) and $p \overset{\text{POLY}}{\ll} \nu, k^{-1}, \delta$, for any dummy colour c_i, we always have $|E_{G_d}(c_i)| \leqslant 4kpn \leqslant \nu D/2 \leqslant d$. For a non-dummy colour c, if $|E_{G_D}(c)| \leqslant \nu D/2$, then $|E_{G_d}(c)| \leqslant \nu D/2 \leqslant d$ for all d. It remains to prove the claim for non-dummy colours with $|E_{G_D}(c)| \geqslant \nu D/2$. Let c be such a colour.

Let Y_d^c be the number of colour c edges in G_d. From Lemma 4.6, we have that as long as the process goes on, we have $\mathbb{P}(e \in E(M_d) | e \in E(G_d)) \geqslant (1-p)\frac{1}{d}$. Since M_d is rainbow, for any colour c, the events '$e \in E(M_d)$' and '$f \in E(M_d)$' are disjoint for distinct $e, f \in E(c)$. This implies $\mathbb{P}(c \in C(M_d) | Y_d^c = y) \geqslant (1-p)\frac{y}{d}$. Recall that, G_{d-1} consists of a subgraph of $G_d \setminus M_d$ together with maybe some dummy colour edges. This implies that the event '$c \in C(M_d)$' is contained in the event '$Y_{d-1}^c \leqslant Y_d^c - 1$', which gives following.

$$Y_{d-1}^c = \begin{cases} \leqslant Y_d^c - 1 & \text{with probability} \geqslant (1-p)\frac{1}{d}Y_d^c \\[2mm] Y_d^c & \text{with probability} \leqslant 1 - (1-p)\frac{1}{d}Y_d^c. \end{cases}$$

Let $Z_D^c, \ldots, Z_{\nu D}^c$ be random variables with $Z_D^c = Y_D^c = |E_{G_D}(c)|$ and, for each $d = D, \ldots, \nu D + 1$, for any values of $Z_D^c, \ldots, Z_d^c$, we have

$$Z_{d-1}^c = \begin{cases} Z_d^c - 1 & \text{with probability } (1-p)\frac{1}{d}Z_d^c \\[2mm] Z_d^c & \text{with probability } 1 - (1-p)\frac{1}{d}Z_d^c. \end{cases}$$

Note that Y_d is stochastically dominated by Z_d, that is, specifically we have $\mathbb{P}(Y_d^c \geqslant x) \leqslant \mathbb{P}(Z_d^c \geqslant x)$ (to see this note that Y_d and Z_d can be coupled so that Y_d is always bounded above by Z_d).

Let $X_t^c = Z_{D-t}^c$ and note that $X_0^c, \ldots, X_m^c$ satisfy the assumptions of Lemma 8.1 with $D = D$, $\varepsilon = p$, $C = 2\nu^{-1}$, $m = D - d$. Therefore, we have,

$$\mathbb{P}\left(Z_d^c \geqslant \frac{d}{D}Z_D^c(1 + 6\nu^{-1}p)\right) = \mathbb{P}\left(X_{D-d}^c \geqslant \left(1 - \frac{D-d}{D}\right)X_0^c(1 + 6\nu^{-1}p)\right) \leqslant 4e^{-0.001\nu^4 p^2 D}$$

$$\leqslant n^{-4}.$$

The last inequality comes from $D \geqslant \delta n/2$ and $\nu^4 p^2 \delta \overset{\text{POLY}}{\gg} n^{-1}$. Note that $(1 + 6\nu^{-1}p)\frac{d}{D}Z_D^c \leqslant (1 + 6\nu^{-1}p)\frac{d}{D}(1-\sigma)\delta n \leqslant d$ (using the global $(1-\sigma)\delta n$-boundedness of G and $p\nu^{-1}, k^{-1} \overset{\text{POLY}}{\ll} \sigma$).

Thus, for every colour c with $|E_{G_D}(c)| \geqslant \nu D/2$, we have $\mathbb{P}(|E_{G_d}(c)| \geqslant d) = \mathbb{P}(Y_d \geqslant d) \leqslant \mathbb{P}(Z_d \geqslant d) \leqslant n^{-4}$ for all d. By the union bound over all c and d, we have that G_d is globally d-bounded for all $d \geqslant \nu D$ with probability $\geqslant 1 - n^{-1}$. $\qquad\square$

Since there are $\leqslant \lceil D/k \rceil \leqslant \sigma n/2$ dummy colours, we have that each M_i contains a submatching M_i' of size $\geqslant e(M_i) - \lceil D/k \rceil \geqslant (1-p)n - \sigma n/2 \geqslant (1-\sigma)n$ containing no dummy colours (using (b)). Finally note that we have $(1-\nu)D \geqslant (1-\sigma)\delta n$ matchings (using $\sigma \overset{\text{POLY}}{\gg} \nu, k^{-1}$). $\qquad\square$

The following lemma takes a decomposition into rainbow matchings (as in the previous lemma) and outputs another such decomposition where the matchings are nicely spread out around the vertex set.

LEMMA 8.6. *Let $t \geqslant pn$, $pt \geqslant 1$ and $p \leqslant \frac{1}{2}$. Let G be a properly coloured balanced bipartite graph on $2n$ vertices with $\delta(G) \geqslant t$ and $M_1, \ldots, M_t$ edge-disjoint rainbow matchings in G with $e(M_i) \geqslant (1-p^3)n$ for all i. Then, G has edge-disjoint rainbow matchings $M_1', \ldots, M_{(1-p)t}'$ with $e(M_i') \geqslant (1-10p)n$ for all i and also $\delta(M_1' \cup \cdots \cup M_t') \geqslant (1-101p)t$.*

Proof. For a bipartite graph H on $2n$ vertices with $\Delta(H) \leqslant t$, let $f_H(v) = \max((1-100p)t - d_H(v), 0)$ and $f(H) = \frac{1}{2} \sum_{v \in V(H)} f_H(v)$. Note that $f(H) \leqslant tn - e(H)$ (since $\Delta(H) \leqslant t$ implies that all vertices v have $f_H(v) \leqslant t - d_H(v)$). For $H = M_1 \cup \cdots \cup M_t$, note that $e(H) \geqslant (1-p^3)tn$ and hence $f(H) \leqslant tn - (1-p^3)tn = p^3 tn$.

Let $N_1, \ldots, N_t$ be a family of t edge-disjoint rainbow matchings in G such that $H = N_1 \cup \cdots \cup N_t$ has $e(H) - 4f(H)$ as large as possible. By the previous paragraph, we have $e(H) - 4f(H) \geqslant (1-5p^3)tn$.

CLAIM 8.7. $f(H) = 0$.

Proof. Suppose, for contradiction, that $f(H) > 0$. Let U be the set of vertices u in H with $d_H(u) \leqslant (1-10p)t$. Note that vertices v outside U have $(1-100p)t - d_H(v) \leqslant -90pt < -1$ (using $pt \geqslant 1$). We have $|U|(1-10p)t + (2n-|U|)t \geqslant 2e(H) \geqslant 2e(H) - 8f(H) \geqslant 2(1-5p^3)tn$, which implies that $|U| \leqslant p^2 n$.

Let u be a vertex with $f_H(u) > 0$. This implies that $d_H(u) < (1-100p)t$ and hence $d_G(u) - d_H(u) \geqslant 100pt > 4|U|$ (using $t \geqslant pn$). Choose some matching N_i with $u \notin V(N_i)$ (one must exist since $d_H(u) < (1-100p)t$). Let $N_i' \subset N_i$ be the set of edges in N_i that touch a vertex in U, so that $|N_i'| \leqslant |U|$. As $d_G(u) - d_H(u) \geqslant 4|N_i'|$ and G is properly coloured, we can pick some $y \in N_{G \setminus H}(u)$ so that y is in no edge in N_i' and uy has colour outside of $C(N_i')$.

Let $F_y \subset N_i'$ be the set of edges in N_i with the same colour as uy or which contain y, noticing that, by construction, $V(F_y) \cap U = \emptyset$. Let $N_i'' = (N_i \setminus F_y) \cup \{uy\}$ to get a rainbow matching. Let $H' = (H \setminus N_i) \cup N_i''$ to get another union of t rainbow matchings. Note that $f_{H'}(u) = f_H(u) - 1$ and $f_{H'}(w) = f_H(w)$ for $w \neq u$ (using $V(F_y) \cap U = \emptyset$), so that $f(H') \leqslant f(H) - 1/2$. We also have $e(H') \geqslant e(H) + 1 - e(F_y) \geqslant e(H) - 1$. Combining these gives $e(H') - 4f(H') \geqslant (e(H)-1) - 4(f(H) - \frac{1}{2}) > e(H) - 4f(H)$, which is a contradiction to the maximality of H. $\qquad\square$

Note that $f(H) = 0$ is equivalent to $\delta(H) \geqslant (1-100p)t$. Recall that $\sum_{i=1}^{t} e(N_i) = e(H) \geqslant e(H) - 4f(H) \geqslant (1-5p^3)tn$. These imply that H has at most pt matchings of size $\leqslant (1-10p)n$ (since otherwise we would have $e(H) \leqslant pt(1-10p)n + (t-pt)n = (1-10p^2)tn < (1-5p^3)tn$, giving a contradiction to $e(H) \geqslant (1-5p^3)tn$). The union of the remaining matchings has minimum degree $\geqslant (1-101p)t$ and so satisfies the lemma. $\qquad\square$

The following lemma strengthens our previous rainbow matching decomposition result (Lemma 8.2). It bootstraps that lemma in two different ways. First it removes the condition

that the host graph is regular, replacing this with the condition that it is (γ, δ, n)-regular. Secondly, the decomposition produced is nicely spread out as in Lemma 8.6.

LEMMA 8.8. *Let* $n^{-1} \overset{\text{POLY}}{\ll} \gamma \overset{\text{POLY}}{\ll} p, \delta \leqslant 1$. *Let* G *be a properly coloured balanced bipartite graph on* $2n$ *vertices, which is* (γ, δ, n)-*regular, and globally* $(1 - p)\delta n$-*bounded. Then,* G *has edge-disjoint rainbow matchings* $M_1, \ldots, M_{(1-p)\delta n}$ *with* $e(M_i) \geqslant (1 - p)n$ *for all* i *and also* $\delta(M_1 \cup \cdots \cup M_{(1-p)\delta n}) \geqslant (1 - 2p)\delta n$.

Proof. We will first prove the lemma when we additionally have '$p \overset{\text{POLY}}{\ll} \delta \leqslant 0.01$'. Assuming that, choose $n^{-1} \overset{\text{POLY}}{\ll} \gamma \overset{\text{POLY}}{\ll} \sigma \overset{\text{POLY}}{\ll} \xi \overset{\text{POLY}}{\ll} p \overset{\text{POLY}}{\ll} \delta \leqslant 0.01$.

By Lemma 6.11, there is a balanced bipartite, regular graph G' containing G which has $\leqslant 11\gamma n$ extra vertices in each part with $G'[V(G)] = G$. Colour the edges of G' so that the edges of G retain their colours, and any edge $e \notin E(G)$ gets a new colour c_e (which only occurs on e). Let $n' = |G'|/2 \leqslant (1 + 11\gamma)n$ be the size of the parts of G'. Since $G'[V(G)] = G$ and G was (γ, δ, n)-regular, the graph G' is $(\delta'n')$-regular for some $\delta'n' = (1 \pm \sigma)\delta n$ (using $n^{-1} \overset{\text{POLY}}{\ll} \gamma \overset{\text{POLY}}{\ll} \sigma \overset{\text{POLY}}{\ll} \delta \leqslant 1$). Note that G' is globally $(1 - \sigma)\delta'n'$-bounded (using $(1 - p)\delta n \leqslant (1 - \sigma)\delta'n'$ which comes from $\gamma \overset{\text{POLY}}{\ll} \sigma \overset{\text{POLY}}{\ll} p$).

By Lemma 8.2 applied to G' with $\ell = 1$, we have $(1 - \sigma)\delta'n' \geqslant (1 - 2\sigma)\delta n$ edge-disjoint rainbow matchings $M_1, \ldots, M_{(1-\sigma)\delta'n'}$ with $e(M_i) = (1 - \sigma)n'$. For each i, let $M_i' = M_i[V(G)]$ to get a rainbow matching in G with $e(M_i) \geqslant (1 - \sigma)n' - 22\gamma n \geqslant (1 - \xi^3)n$ (using $\gamma \overset{\text{POLY}}{\ll} \sigma \overset{\text{POLY}}{\ll} \xi$). Note that $\delta(G) \geqslant (1 - \gamma)\delta n \geqslant (1 - 2\sigma)\delta n$.

Apply Lemma 8.6 to G and $M_1', \ldots, M_{(1-2\sigma)\delta n}'$ with $t = (1 - 2\sigma)\delta n$, $n = n$, and $p = \xi$. This gives edge-disjoint rainbow matchings $M_1'', \ldots, M_{(1-\xi)(1-2\sigma)\delta n}''$ with $e(M_i'') = (1 - 10\xi)n$ for all i and also $\delta(M_1'' \cup \cdots \cup M_{(1-\xi)(1-2\sigma)\delta n}'') \geqslant (1 - 101\xi)(1 - 2\sigma)\delta n$. Note that $(1 - \xi)(1 - 2\sigma)\delta n \geqslant (1 - p)\delta n$ (using $\xi, \sigma \overset{\text{POLY}}{\ll} p$) and consider the matchings $M_1'', \ldots, M_{(1-p)\delta n}''$. Now the lemma holds for these matchings because $e(M_i'') = (1 - 10\xi)n \geqslant (1 - p)n$, and $\delta(M_1'' \cup \cdots \cup M_{(1-p)\delta n}'') \geqslant \delta(M_1'' \cup \cdots \cup M_{(1-\xi)(1-2\sigma)\delta n}'') - ((1 - \xi)(1 - 2\sigma)\delta n - (1 - p)\delta n) \geqslant (1 - 101\xi)(1 - 2\sigma)\delta n - (p - \xi - 2\sigma + 2\xi\sigma)\delta n \geqslant (1 - 2p)\delta n$ (using $\xi, \sigma \overset{\text{POLY}}{\ll} p$).

Now we will prove the general case when we just have $n^{-1} \overset{\text{POLY}}{\ll} \gamma \overset{\text{POLY}}{\ll} p, \delta \leqslant 1$. Choose $\hat{p}$ such that $n^{-1} \overset{\text{POLY}}{\ll} \gamma \overset{\text{POLY}}{\ll} \hat{p} \overset{\text{POLY}}{\ll} p, \delta \leqslant 1$. Apply Lemma 5.3 (b) to G with $p' = 0.01$ in order to partition the edges of G into 100 spanning subgraphs $G_1, \ldots, G_{100}$ with each G_i $(2\gamma, 0.01\delta, n)$-regular and globally $(1 + \gamma)(1 - p)0.01\delta n \leqslant (1 - \hat{p})0.01\delta n$-bounded. For $i = 1, \ldots, 100$, applying the '$\hat{p} \overset{\text{POLY}}{\ll} \delta \leqslant 0.01$' case of the lemma to G_i gives a family of rainbow matchings $M_1^i, \ldots, M_{(1-\hat{p})0.01\delta n}^i$ with $e(M_j^i) \geqslant (1 - \hat{p})n$ for all i and also $\delta(M_1^i \cup \cdots \cup M_{(1-\hat{p})\delta n}^i) \geqslant (1 - \hat{p})0.01\delta n$. Taking the union of these families for $i = 1, \ldots, 100$ gives the required edge-disjoint rainbow matchings (using $\hat{p} \leqslant p$). □

8.2. *Perfect matchings*

In this section, we find near-decompositions of graphs into perfect rainbow matchings. This is done by taking the near-decompositions into nearly-perfect rainbow matchings produced in the previous section and then using a completion lemma from Section 7 to turn them into perfect matchings. The most straightforward to prove version of this is the following.

LEMMA 8.9. *Let* $1 \overset{\text{POLY}}{\gg} \theta, p \overset{\text{POLY}}{\gg} \varepsilon \overset{\text{POLY}}{\gg} n^{-1}$ *and* $t \leqslant n$. *Suppose that we have the following edge-disjoint subgraphs in a properly coloured balanced bipartite graph on* $2n$ *vertices.*

- *Rainbow matchings* $M_1, \ldots, M_t$ *with* $e(M_i) \geqslant (1 - \varepsilon)n$ *and* $\delta(M_1 \cup \cdots \cup M_t) \geqslant t - 10\varepsilon n$.

- A $(2p(\theta n)^2, \theta n)$-dense graph E.
- Graphs D_X, D_Y with $\delta(D_X), \delta(D_Y) \geqslant 4\theta n$.

Additionally suppose that $M_1 \cup \cdots \cup M_t$, E, D_X, and D_Y are colour-disjoint. Then, there are edge-disjoint perfect rainbow matchings $M_1', \ldots, M_t'$ in $E \cup D_X \cup D_Y \cup M_1 \cup \cdots \cup M_t$.

Proof. Let X and Y be the parts of the bipartition. We construct the matchings $M_1', \ldots, M_t'$ one-by-one using Lemma 7.2. They will have the following properties.

(i) $xy \in E(D_X \cap M_i') \Rightarrow x \notin V(M_i) \cap X$.
(ii) $xy \in E(D_Y \cap M_i') \Rightarrow y \notin V(M_i) \cap Y$.
(iii) $e(E \cap M_i') \leqslant \varepsilon n$.

Suppose that we have constructed matchings $M_1', \ldots, M_s'$ satisfying the above properties. Let $E^s = E \setminus (M_1' \cup \cdots \cup M_s')$, $D_X^s = D_X \setminus (M_1' \cup \cdots \cup M_s')$, $D_Y^s = D_Y \setminus (M_1' \cup \cdots \cup M_s')$. Using (i), $\varepsilon \overset{\text{POLY}}{\ll} \theta$, and $\delta(M_1 \cup \cdots \cup M_t) \geqslant t - 10\varepsilon n$, we have $d_{D_X^s}(x) \geqslant d_{D_X}(x) - |\{i : x \notin V(M_i)\}| \geqslant \delta(D_X) - 10\varepsilon n \geqslant 3\theta n$ for any $x \in X$. Similarly, $d_{D_Y^s}(y) \geqslant 3\theta n$ for $y \in Y$. By (iii) and Lemma 5.9, the graph E_s is $(2p(\theta n)^2 - s\varepsilon n, \theta n)$-dense. Hence, using $s\varepsilon n \leqslant \varepsilon n^2 \leqslant p\theta^2 n^2$, E_s is $(p(\theta n)^2, \theta n)$-dense. By Lemma 7.2 applied to M_{s+1}, E^s, D_X^s and D_Y^s, there is a rainbow perfect matching M_{s+1}' satisfying (i)–(iii). □

Dense graphs are less convenient to work with than typical ones. The following lemma is a version of the previous one which replaces the colour-disjoint dense graphs by a single colour-disjoint typical one.

LEMMA 8.10. Let $1 \geqslant p \overset{\text{POLY}}{\gg} \varepsilon \overset{\text{POLY}}{\gg} \gamma \overset{\text{POLY}}{\gg} n^{-1}$. Suppose that we have the following edge-disjoint subgraphs in a properly coloured balanced bipartite graph on $2n$ vertices for some $t \leqslant n$.

- Rainbow matchings $M_1, \ldots, M_t$ with $e(M_i) \geqslant (1 - \varepsilon)n$ for each i, and $\delta(M_1 \cup \cdots \cup M_t) \geqslant t - 10\varepsilon n$.
- A (γ, p, n)-typical, balanced bipartite graph G which is colour-disjoint from $M_1, \ldots, M_t$.

Then, there are edge-disjoint perfect rainbow matchings $M_1', \ldots, M_t'$ in $M_1 \cup \cdots \cup M_t \cup G$.

Proof. Choose $1 \geqslant p \overset{\text{POLY}}{\gg} p_1 \overset{\text{POLY}}{\gg} \theta \overset{\text{POLY}}{\gg} \varepsilon \overset{\text{POLY}}{\gg} \gamma \overset{\text{POLY}}{\gg} n^{-1}$. Choose three disjoint sets of colours C_E, C_{D_X}, C_{D_Y} from G, with each colour put independently into C_E, C_{D_X} and C_{D_Y} with probability $2p_1 p^{-1}/0.99$, $5\theta p^{-1}$ and $5\theta p^{-1}$, respectively (this is possible since $p \overset{\text{POLY}}{\gg} p_1 \overset{\text{POLY}}{\gg} \theta$ implies $2p_1 p^{-1}/0.99 + 5\theta p^{-1} + 5\theta p^{-1} \leqslant 1$). Let E, D_X, D_Y be the subgraphs of G with colours from C_E, C_{D_X}, C_{D_Y}, respectively. By Lemma 5.3 (a), with high probability E is $(2\gamma, 2p_1/0.99, n)$-typical and D_X, D_Y are $(2\gamma, 5\theta, n)$-typical.

We have that $\delta(D_X), \delta(D_Y) \geqslant (1 - 2\gamma)5\theta n \geqslant 4\theta n$. By Lemma 5.7 applied with $\mu = \theta$, $p = p_1/0.99$ and $\gamma = 2\gamma$, E is $(2p_1(\theta n)^2, \theta n)$-dense. By Lemma 8.9 applied with $\theta = \theta$ and $p = p_1$, we obtain the required perfect matchings. □

Combining the above with Lemma 8.8, we get the following versatile lemma guaranteeing near-decompositions into perfect rainbow matchings.

LEMMA 8.11. Suppose that we have n, δ, p with $n^{-1} \overset{\text{POLY}}{\ll} \gamma \overset{\text{POLY}}{\ll} p, \delta \leqslant 1$. Let G be a properly coloured, (γ, δ, n)-regular, globally $(1 - p)\delta n$-bounded, balanced bipartite graph of order $2n$. Let H be a properly coloured, colour-disjoint, (γ, p, n)-typical, balanced bipartite graph on the same vertex set as G. Then, $G \cup H$ has edge-disjoint perfect rainbow matchings $M_1, \ldots, M_{(1-p)\delta n}$.

Proof. Choose $n^{-1} \overset{\text{POLY}}{\ll} \gamma \overset{\text{POLY}}{\ll} \varepsilon \overset{\text{POLY}}{\ll} p, \delta \leqslant 1$. By Lemma 8.8, there are edge-disjoint rainbow matchings $M_1, \ldots, M_{(1-\varepsilon)\delta n}$ with $e(M_i) = (1 - \varepsilon)n$ for all i with $\delta(M_1 \cup \cdots \cup M_{(1-\varepsilon)\delta n}) \geqslant (1 - 2\varepsilon)\delta n$. By Lemma 8.10 applied with $t = (1 - \varepsilon)\delta n$, there are edge-disjoint perfect rainbow matchings $M'_1, \ldots, M'_{(1-\varepsilon)n}$ in $H \cup G$. $\qquad\square$

As a corollary, we obtain that a typical properly coloured graph can be nearly-decomposed into perfect rainbow matchings as long as there is a gap between its global boundedness and its degrees.

COROLLARY 8.12. *Suppose that we have n, δ, p, γ with $n^{-1} \overset{\text{POLY}}{\ll} \gamma \overset{\text{POLY}}{\ll} p, \delta \leqslant 1$. Every properly coloured, (γ, δ, n)-typical, globally $(1 - p)\delta n$-bounded, balanced bipartite graph G of order $2n$ has $(1 - p)\delta n$ edge-disjoint perfect rainbow matchings.*

Proof. By Lemma 5.3 (a), G can be partitioned into a $(2\gamma, p\delta/2, n)$-typical graph H and a colour-disjoint $(2\gamma, \delta - p\delta/2, n)$-typical graph G'. Since $(1 - p)\delta n \leqslant (1 - p\delta/2)(1 - p/2)\delta n$, G' is globally $(1 - p\delta/2)(\delta - p\delta/2)n$-bounded. By Lemma 8.11 applied with $p' = p\delta/2$, and $\delta' = \delta - p\delta/2$, $G' \cup H$ has $(1 - p\delta/2)(\delta - p\delta/2)n \geqslant (1 - p)\delta n$ edge-disjoint perfect rainbow matchings. $\qquad\square$

Applying the above lemma when the host graph is $K_{n,n}$, we can show that any proper colouring of $K_{n,n}$ has a near-decomposition into perfect rainbow matchings under natural conditions on the sizes of the colour classes. This gives us Theorem 1.9, which we restate with our notation as Lemma 8.13.

LEMMA 8.13. *Let $1 \overset{\text{POLY}}{\gg} \varepsilon \overset{\text{POLY}}{\gg} n^{-1}$. Let $K_{n,n}$ be properly coloured with $\leqslant (1 - 20\varepsilon)n$ colours having $\geqslant (1 - 20\varepsilon)n$ edges. Then, $K_{n,n}$ has $(1 - \varepsilon)n$ edge-disjoint perfect rainbow matchings.*

Proof. Choose $1 \overset{\text{POLY}}{\gg} \varepsilon \overset{\text{POLY}}{\gg} p \overset{\text{POLY}}{\gg} \gamma \overset{\text{POLY}}{\gg} n^{-1}$. By Lemma 5.1, $K_{n,n}$ is $(\gamma, 1, n)$-typical. Apply Lemma 5.3 (a) with $p = p$, $\delta = 1$, and $\gamma = \gamma$ in order to partition $K_{n,n}$ into a $(2\gamma, p, n)$-typical graph J and a colour-disjoint graph G with $\delta(G) \geqslant (1 - 2\gamma)(1 - p)n \geqslant (1 - \varepsilon^2)n$. Apply Lemma 6.7 to G in order to find a subgraph G' which is globally $(1 - \varepsilon)n$-bounded and (γ, δ, n)-regular for some $\delta \geqslant 1 - \varepsilon + 9\varepsilon^2$. Since $\varepsilon \overset{\text{POLY}}{\gg} p$, we have that G' is globally $(1 - p)\delta n$-bounded. By Lemma 8.11 applied with $\gamma' = 2\gamma$ to G' and J, there are edge-disjoint perfect rainbow matchings $M_1, \ldots, M_{(1-p)\delta n}$ in $G' \cup J$. Since $\varepsilon \overset{\text{POLY}}{\gg} p$ and $\delta \geqslant 1 - \varepsilon + 9\varepsilon^2$, we have the required perfect rainbow matchings in $K_{n,n}$. $\qquad\square$

As a corollary, we show that having quadratically many colours guarantees perfect rainbow matchings.

LEMMA 8.14. *Let $1 \geqslant \varepsilon \overset{\text{POLY}}{\gg} n^{-1}$. Let $K_{n,n}$ be coloured with at least $2\varepsilon n^2$ colours. Then, $K_{n,n}$ has $\leqslant (1 - \varepsilon)n$ colours having $\geqslant (1 - \varepsilon)n$ edges.*

Proof. Suppose otherwise. Then, $K_{n,n}$ has $\leqslant n^2 - (1 - \varepsilon)n \cdot (1 - \varepsilon)n = 2\varepsilon n^2 - \varepsilon^2 n^2$ edges outside of the $(1 - \varepsilon)n$ largest colours. This means that $K_{n,n}$ has $\leqslant 2\varepsilon n^2 - \varepsilon^2 n^2 + (1 - \varepsilon)n$ colours. By $\varepsilon \overset{\text{POLY}}{\gg} n^{-1}$, this is smaller than $2\varepsilon n^2$, contradicting the lemma's assumption. $\qquad\square$

COROLLARY 8.15. *Let $1 \overset{\text{POLY}}{\gg} \varepsilon \overset{\text{POLY}}{\gg} n^{-1}$. Let $K_{n,n}$ be coloured with at least εn^2 colours. Then, $K_{n,n}$ has $(1 - \varepsilon)n$ edge-disjoint perfect rainbow matchings.*

Proof. Let $\varepsilon' = \varepsilon/40$. By Lemma 8.14, $K_{n,n}$ has $\leqslant (1 - 20\varepsilon')n$ colours having $\geqslant (1 - 20\varepsilon')n$ edges. By Lemma 8.13, $K_{n,n}$ has $(1 - \varepsilon')n \geqslant (1 - \varepsilon)n$ edge-disjoint perfect rainbow matchings. □

8.3. *2-factors*

Here, we use the perfect matching decomposition results from the previous section in order to show that suitable properly coloured complete graphs have near-decompositions into rainbow 2-factors. These 2-factor results are a stepping stone for finding Hamiltonian cycles. The basic idea of the proof is to join rainbow matchings together into 2-factors. Suppose that we have partitioned the vertices of a graph into sets $V_1, \ldots, V_k$ of equal size, and that we have rainbow matchings $M_1, \ldots, M_k$ with M_i going from V_i to $V_{i+1 \pmod{k}}$. Note that if the matchings $M_1, \ldots, M_k$ are all colour-disjoint, then $M_1 \cup \cdots \cup M_k$ is a rainbow 2-factor. The proof strategy in this section is to partition the edges of a general graph G into balanced bipartite subgraphs in which we can find perfect rainbow matchings using results from the previous section. Then, we can put these matchings together in the way just described and obtain 2-factors.

First we will need the following standard lemma which asserts that there exist complete graphs with rainbow Hamiltonian decompositions.

LEMMA 8.16. *For prime $n \geqslant 3$, there exist properly n-coloured K_n with decompositions into $(n-1)/2$ rainbow Hamiltonian cycles.*

Proof. Identify the vertices of K_n with $\mathbb{Z}/n\mathbb{Z}$. Colour ij by $i + j \pmod{n}$. For $i = 1, \ldots, (n-1)/2$, let $C_i = \{a(a+i) : a = 1, \ldots, n\}$. Note that C_i is rainbow since we have $a(a+i), a'(a'+i) \in C_i \Rightarrow c(a(a+i)) = 2a+i, c(a'(a'+i)) = 2a'+i$, and so distinct edges in C_i have distinct colours. Since n is prime and $i \leqslant (n-1)/2$, $a + ki = a \pmod{n} \Rightarrow n \mid k$ which implies that C_i is a cycle. The cycles $C_1, \ldots, C_{(n-1)/2}$ are disjoint because for $i \neq j$, $a(a+i) = b(b+j)$ implies that $b + j = a$ and $a + i = b$, which implies $i + j \equiv 0 \pmod{n}$. □

Using the above, and results about perfect matching decompositions, we can prove our first result about 2-factor decompositions. The following should be compared with Lemma 8.11. It shows that under analogous assumptions to that lemma, one can find a near-decomposition into rainbow 2-factors. It also has a divisibility condition on the size of the host graph. This divisibility condition will later be removed.

LEMMA 8.17. *Let $1 \geqslant \delta, p \overset{\text{POLY}}{\gg} k^{-1} \overset{\text{POLY}}{\gg} \gamma \overset{\text{POLY}}{\gg} n^{-1}$, with k prime and $k \mid n$. Let G be a properly coloured, globally $(1-p)\delta n/2$-bounded, (γ, δ, n)-regular graph, J a properly coloured (γ, p, n)-typical graph which is edge-disjoint and colour-disjoint from G. Then, $G \cup J$ has $(1 - 2p)\delta n/2$ edge-disjoint rainbow 2-factors with cycles of length $\geqslant k$.*

Proof. Partition $V(K_n)$ randomly into k sets $V_1, \ldots, V_k$ of size $m = n/k$. Partition $C(G \cup J)$ randomly into k sets $C_1, \ldots, C_k$, with each colour ending up in each set independently with probability $1/k$.

By Lemma 5.2(d), with probability $1 - o(k^2 n^{-1})$, the balanced bipartite graphs $G[V_i, V_j]$ are $(2\gamma, \delta, k^{-1}n)$-regular and globally $(1+\gamma)(1-p)\delta k^{-2}n$-bounded. Also by Lemma 5.2(d), with probability $1 - o(kn^{-1})$, the graphs $J[V_i, V_j]$ are $(2\gamma, p, k^{-1}n)$-typical. Let $G_{a,b,c} = G_{C_c}[V_a, V_b]$ and $J_{a,b,c} = J_{C_c}[V_a, V_b]$. By Lemma 5.3(a), with probability $1 - o(kn^{-1})$, the balanced bipartite graphs $G_{a,b,c}$ are $(4\gamma, k^{-1}\delta, k^{-1}n)$-regular, and the graphs $J_{a,b,c}$ are $(4\gamma, k^{-1}p, k^{-1}n)$-typical. By a union bound, with high probability these hold for all the graphs $G_{a,b,c}, J_{a,b,c}$ simultaneously (using $n^{-1} \overset{\text{POLY}}{\ll} k^{-1}$). Since $\gamma \overset{\text{POLY}}{\ll} p$, the graphs $G_{a,b,c}$ are globally $(1 - p/2)(k^{-1}\delta)(k^{-1}n)$-bounded.

For any $a, b, c \in \{1, \dots, k\}$ with $a \neq b$, apply Lemma 8.11 to $G_{a,b,c}$ and $J_{a,b,c}$ with $n' = k^{-1}n$, $\delta' = k^{-1}\delta$, $p' = k^{-1}p$, $\gamma' = 4\gamma$. This gives a family $\mathcal{M}_{a,b,c} = \{M^i_{a,b,c} : 1 \leqslant i \leqslant (1-p)k^{-2}\delta n\}$ of $(1-p)k^{-2}\delta n$ rainbow matchings with every matching $M^i_{a,b,c} \in \mathcal{M}_{a,b,c}$ having $V(M^i_{a,b,c}) = V_a \cup V_b$ and $C(M^i_{a,b,c}) \subseteq C_c$.

Consider a proper k-colouring of K_k with vertex set $\{1, \dots, k\}$ and a decomposition into rainbow Hamiltonian cycles $H_1, \dots, H_{(k-1)/2}$ (which exists by Lemma 8.16). For every $i \in [(k-1)/2]$, $j \in [(1-p)k^{-2}\delta n]$, and $t \in [k]$, let $F^j_{i,t} = \bigcup_{\substack{ab \in E(H_i) \\ c = c(ab)+t \pmod k}} M^j_{a,b,c}$.

CLAIM 8.18.   *For all i, j, t, $F^j_{i,t}$ is a rainbow 2-factor with cycles of length $\geqslant k$.*

Proof.   Without loss of generality, by reordering the vertex sets V_1 and colour sets C_j, we can suppose that K_k is ordered and coloured so that the vertex sequence of H_i is $1, 2, \dots, k$ and so that $c(a(a+1)) + t \mod k = a$. Then, $F^j_{i,t} = \bigcup_{a=1}^k M^j_{a,a+1,a}$.

Let $v \in V(F^j_{i,t})$ with $v \in V_a$. We claim that $N_{F^j_{i,t}}(v) = \{x, y\}$ for some $x \in V_{a+1}, y \in V_{a-1}$. To see this, note that since $M^j_{a,a+1,a}$, $M^j_{a-1,a,a-1}$ are perfect matchings between V_a and V_{a+1}, V_{a-1}, respectively, they each have one edge through v. Let these edges be vx and vy to get two vertices with $x, y \in N(v)$. To see that there are no edges other than vx, vy containing v, note that $v \in V_a$ and none of the matchings forming $F^j_{i,t}$ other than $M^j_{a,a+1,a}$ and $M^j_{a-1,a,a-1}$ touch V_a. We have shown that $F^j_{i,t}$ is 2-regular and so a 2-factor.

To see that $F^j_{i,t}$ has no cycles shorter than k, consider a cycle C with vertex sequence $v_1, v_2, \dots, v_s$. Without loss of generality, we can suppose that H_i is labelled so that $v_1 \in V_1$ and $v_2 \in V_2$. By the previous paragraph, this implies that $v_i \in V_{i \pmod k}$ must hold for all i, implying that $s \geqslant k$.

To see that $F^j_{i,t}$ is rainbow, note first that it is the union of rainbow sets of edges $M^j_{a,a+1,a}$ for $a = 1, \dots, k$. For any such matching, we have $C(M^j_{a,a+1,a}) \subseteq C_a$. Together with the colour-disjointness of C_a and $C_{a'}$, we get that $M^j_{a,a+1,a}$ and $M^j_{a',a'+1,a'}$ are colour-disjoint for $a \neq a'$. Thus, $F^j_{i,t}$ is rainbow.   □

Note that $F^j_{i,t}$ and $F^{j'}_{i',t'}$ are edge-disjoint for $(i, j, t) \neq (i', j', t')$ (since any matching $M^d_{a,b,c}$ is contained in exactly one of the 2-factors $F^j_{i,t}$, and matchings $M^d_{a,b,c}$, $M^{d'}_{a',b',c'}$ are edge-disjoint for $(a, b, c, d) \neq (a', b', c', d')$). The total number of 2-factors we have is $k \times (1-p)k^{-2}\delta n \times (k-1)/2 = (1-p)(1-k^{-1})\frac{\delta n}{2} \geqslant (1-2p)\frac{\delta n}{2}$.   □

In the remainder of this section, we prove that the above lemma is true even without the divisibility condition on n. The idea of the proof is to randomly partition the graph G into subgraphs which do satisfy the divisibility condition. Applying Lemma 8.17 to each of these subgraphs gives a decomposition of them into 2-factors. By carefully putting the 2-factors together, we get a decomposition of the whole graph into 2-factors. First, we need the following standard number-theoretic result.

LEMMA 8.19.   *Let $1 \overset{\text{POLY}}{\gg} \varepsilon \overset{\text{POLY}}{\gg} s^{-1} \overset{\text{POLY}}{\gg} k^{-1} \overset{\text{POLY}}{\gg} n^{-1}$. There exist prime numbers $k_1, k_2 \in [k, (1+\varepsilon)k]$, integers $s' = (1 \pm \varepsilon)s$ and $n_1, \dots, n_{s'} = (1 \pm \varepsilon)n/s$ so that $n_1 + \cdots + n_{s'} = n$ and for each $i = 1, \dots, s'$ either $k_1 \mid n_i$ or $k_2 \mid n_i$.*

Proof.   Since $1 \overset{\text{POLY}}{\gg} \varepsilon \overset{\text{POLY}}{\gg} k^{-1}$, we can choose two distinct primes $k_1, k_2 \in [k, (1+0.2\varepsilon)k]$. (When ε is constant, this is a consequence of the Prime Number Theorem. More generally, we need the result of Hoheisel that there is some fixed number $\alpha > 0$ such that for sufficiently large n, there is a prime in the interval $[n, (1+n^{-\alpha})n]$. (See [11]). Since $\varepsilon \overset{\text{POLY}}{\gg} n^{-1}$ implies $0.2\varepsilon \geqslant n^{-\alpha}$,

we get that there is a prime in $[x, (1 + 0.2\varepsilon)x]$ for sufficiently large x). As $\varepsilon \overset{\text{POLY}}{\gg} k_1^{-1}, k_2^{-1}, s^{-1} \overset{\text{POLY}}{\gg} n^{-1}$, there are integers $z_1, z_2 \geqslant 100n/\varepsilon sk$ with $k_1 z_1 + k_2 z_2 = n$. (See [**47**, Corollary 2, p. 25, 26]. Applying this corollary with $a = k_1, b = k_2$, $n' = n - (k_1 + k_2)\lceil 100n/\varepsilon sk \rceil$ gives non-negative integers x, y with $k_1 x + k_2 y = n - (k_1 + k_2)\lceil 100n/\varepsilon sk \rceil$. Letting $z_1 = x + \lceil 100n/\varepsilon sk \rceil$, $z_2 = y + \lceil 100n/\varepsilon sk \rceil$ gives the numbers we want.)

For some appropriate $s'' \geqslant 50\varepsilon^{-1}$, pick integers $m_1, \ldots, m_{s''}$ so that $m_i = (1 \pm 0.5\varepsilon)n/sk_1$ and $\sum_{i=1}^{s''} m_i = z_1$. This is possible as $z_1 \geqslant 100n/\varepsilon sk$. Similarly, for some appropriate s', pick integers $m_{s''+1}, \ldots, m_{s'}$ so that $m_i = (1 \pm 0.5\varepsilon)n/sk_2$ and $\sum_{i=s''+1}^{s'} m_i = z_2$.

For each $1 \leqslant i \leqslant s''$, let $n_i = k_1 m_i$, and for each $s'' < i \leqslant s'$, let $n_i = k_2 m_i$. Then, for each $1 \leqslant i \leqslant s'$, $n_i = (1 \pm 0.5\varepsilon)n/s$. Thus, as $n_1 + \cdots + n_{s'} = n$, we have $s' = (1 \pm \varepsilon)s$. The numbers k_1, k_2, s' and $n_1, \ldots, n_{s'}$, then satisfy the conditions of the lemma. $\qquad \square$

The following lemma shows how any large set can be evenly covered by subsets whose sizes satisfy the divisibility condition of Lemma 8.17.

LEMMA 8.20. *Let* $1 \overset{\text{POLY}}{\gg} \varepsilon \overset{\text{POLY}}{\gg} \hat{s}^{-1} \overset{\text{POLY}}{\gg} k^{-1} \overset{\text{POLY}}{\gg} \gamma \overset{\text{POLY}}{\gg} n^{-1}$. *There exists a family* $\mathcal{H}$ *of partitions of* $[n]$ *and a number* $s = (1 \pm \varepsilon)\hat{s}$ *with the following properties.*

(i) $\mathcal{H} = \mathcal{M}_1 \cup \cdots \cup \mathcal{M}_{s^2 \log^2 n}$ *with each* $\mathcal{M}_i = \{M_i^1, \ldots, M_i^s\}$ *for disjoint sets* $M_i^1, \ldots, M_i^s$ *satisfying* $|M_i^j| = (1 \pm \varepsilon)n/s$ *and* $\bigcup_{j=1}^{s} M_i^j = [n]$.

(ii) *For each* i, j *there is some prime number* $k_i^j \in [k, (1 + \varepsilon)k]$ *with* $k_i^j \mid |M_i^j|$.

(iii) *For each distinct pair* $x, y \in [n]$, *there are* $(1 \pm 5\varepsilon)s \log^2 n$ *sets* M_i^j *containing both* x *and* y.

Proof. Apply Lemma 8.19 in order to find a number $s = (1 \pm \varepsilon)\hat{s}$, primes $k_1, k_2 = (1 \pm \varepsilon)k$ and numbers $n_1, \ldots, n_s = (1 \pm \varepsilon)n/s$ so that $n_1 + \cdots + n_s = n$ and, for each $i = 1, \ldots, s$, either $k_1 \mid n_i$ or $k_2 \mid n_i$. Let $M^1, \ldots, M^s$ be disjoint subsets of $[n]$ with $|M^i| = n_i$.

Choose $s^2 \log^2 n$ permutations $\sigma_1, \ldots, \sigma_{s^2 \log^2 n}$ of $[n]$ uniformly at random. Let $\mathcal{M}_i = \{\sigma_i(M^1), \ldots, \sigma_i(M^s)\}$ and $\mathcal{H} = \mathcal{M}_1 \cup \cdots \cup \mathcal{M}_{s^2 \log^2 n}$. Note that as a consequence of the properties from Lemma 8.19, all the conditions of the lemma hold for $\mathcal{H}$ aside from (iii). We will show that this condition holds with high probability.

Let $x, y \in [n]$ be distinct vertices. We have $\mathbb{P}(x, y \in \mathcal{M}_i) = \sum_{j=1}^{s} \binom{n_j}{2}/\binom{n}{2} = (1 \pm 4\varepsilon)s^{-1}$. Let X be the number of families $\mathcal{M}_i$ which contain x, y. We have that X is bounded above and below by random variables with distributions Binomial$(s^2 \log^2 n, (1 + 4\varepsilon)s^{-1})$ and Binomial$(s^2 \log^2 n, (1 - 4\varepsilon)s^{-1})$, respectively. From Chernoff's Bound, we have $\mathbb{P}(|X - s\log^2 n| > 5\varepsilon s \log^2 n) \leqslant 4e^{-\frac{\varepsilon^2 s \log^2 n}{100}} = o(n^{-2})$ (using $\varepsilon \overset{\text{POLY}}{\gg} \hat{s}^{-1}$). By the union bound taken over all pairs x, y, we have that with high probability all pairs $x, y \in [n]$ have $(1 \pm 5\varepsilon)s \log^2 n$ families $\mathcal{M}_i$ containing both x and y. $\qquad \square$

By combining the lemmas of this section, we can prove Lemma 8.17 without the divisibility assumption.

LEMMA 8.21. *Let* $1 \geqslant \delta, p, \log^{-1} n \overset{\text{POLY}}{\gg} k^{-1} \overset{\text{POLY}}{\gg} \gamma \overset{\text{POLY}}{\gg} n^{-1}$. *Let* G *be a properly coloured, globally* $(1 - p)\delta n/2$-*bounded,* (γ, δ, n)-*regular graph and let* J *be a properly coloured* (γ, p, n)-*typical graph which is edge-disjoint and colour-disjoint from* G *but has the same vertex set. Then,* $G \cup J$ *has* $(1 - p)\delta n/2$ *edge-disjoint rainbow 2-factors with cycles of length* $\geqslant k$.

Proof. Choose $1 \geqslant \delta, p, \log^{-1} n \overset{\text{POLY}}{\gg} \varepsilon \overset{\text{POLY}}{\gg} \hat{s}^{-1} \overset{\text{POLY}}{\gg} k^{-1} \overset{\text{POLY}}{\gg} \gamma \overset{\text{POLY}}{\gg} n^{-1}$.

Apply Lemma 8.20 to find a family $\mathcal{H}$ of partitions of $V(G)$ and a number $s = (1 \pm \varepsilon)\hat{s}$ so that the properties in that lemma hold (with the associated notation). Let σ be a random permutation of $V(G)$. Lemma 5.2(c) implies that for each i, j with probability $1 - o(n^{-1})$, we have that $G[\sigma(M_i^j)]$ is $(2\gamma, \delta, |M_i^j|)$-regular and globally $(1 + \gamma)(1 - p)\frac{\delta|M_i^j|^2}{2n}$-bounded, while $J[\sigma(M_i^j)]$ is $(2\gamma, p, |M_i^j|)$-typical. Noticing that there are $s^3 \log n \leqslant n$ choices for (i, j), fix σ then, so that this holds for each i, j. Note that, for all distinct vertices x and y, there are $(1 \pm 5\varepsilon)s \log^2 n$ sets M_i^j containing $\sigma^{-1}(x)$ and $\sigma^{-1}(y)$, and hence there are $(1 \pm 5\varepsilon)s \log^2 n$ sets $\sigma(M_i^j)$ containing x and y. Relabelling the sets $\sigma(M_i^j)$ by M_i^j, we can now assume that the properties from Lemma 8.20 hold for the family $\mathcal{H}$ (with the associated notation), $G[M_i^j]$ is $(2\gamma, \delta, |M_i^j|)$-regular and globally $(1 + \gamma)^2(1 - p)\frac{\delta|M_i^j|^2}{2n}$-bounded, while $J[M_i^j]$ is $(2\gamma, p, |M_i^j|)$-typical.

Partition the colours independently at random into sets $C_1, \ldots, C_s$. Let G^{C_j} and J^{C_j} denote the subgraphs of G and J, respectively, consisting of colour C_j edges. By Lemma 5.2(a), for all i, j, t, with probability $1 - o(sn^{-1})$ the graph $G^{C_t}[M_i^j]$ is $(4\gamma, \delta/s, |M_i^j|)$-regular and globally $(1 + \gamma)^2(1 - p)\frac{\delta|M_i^j|^2}{2n}$-bounded, while $J^{C_t}[M_i^j]$ is $(4\gamma, p/s, |M_i^j|)$-typical. Thus, as there are at most $s^4 \log n \leqslant ns^{-1}$ choices of (i, j, t), we can fix sets $C_1, \ldots, C_s$ so that this holds for all i, j, t.

For any edge e, let $d_{\mathcal{H}}(e)$ be the number of partitions in $\mathcal{H}$ which contain some set containing e. From the properties from Lemma 8.20, we have $d_{\mathcal{H}}(e) = (1 \pm 5\varepsilon)s \log^2 n$. For each edge e, choose an arbitrary injection $f_e : [d_{\mathcal{H}}(e)] \to [s^2 \log^2 n]$ for which e is containing in some set in $\mathcal{M}_{f_e(m)}$ for all $m \in [d_{\mathcal{H}}(e)]$. For each edge e, choose a number m_e out of $1, \ldots, (1 + 5\varepsilon)s \log^2 n$ at random. For $i = 1, \ldots, s^2 \log^2 n$, let G_i and J_i be subgraphs of G and J, respectively, consisting of edges e with $f_e(m_e) = i$ (here it is possible that $m_e > d_{\mathcal{H}}(e)$, in which case $f_e(m_e)$ is undefined. When this happens, the edge e is placed in neither of the graphs G_i, J_i). Note that edges of G are placed into G_i and J_i independently with probability $1/(1 + 5\varepsilon)s \log^2 n$.

For $i = 1, \ldots, s^2 \log^2 n$ and $j, t = 1, \ldots, s$, define $G_{i,j,t}$ and $J_{i,j,t}$ to be the subgraphs of G and J with vertex set M_i^j consisting of edges xy which are simultaneously contained in $E(G_i)$, contained in M_i^j, and whose colours are in $C_{j+t \pmod{s}}$. Note that $G_{i,j,t}$ and $J_{i,j,t}$ are formed from $G^{C_t}[M_i^j]$ and $J^{C_t}[M_i^j]$ by choosing every edge with probability $1/(1 + 5\varepsilon)s \log^2 n$. By Lemma 5.2(b) with probability $1 - o(n^{-1}s)$, the graph $G_{i,j,t}$ is $(8\gamma, \frac{\delta}{(1+5\varepsilon)s^2 \log^2 n}, |M_i^j|)$-regular and globally $(1 + \gamma)^3(1 - p)\frac{\delta|M_i^j|^2}{(1+5\varepsilon)2ns \log^2 n}$-bounded, while $J_{i,j,t}$ is $(8\gamma, \frac{p}{(1+5\varepsilon)s^2 \log^2 n}, |M_i^j|)$-typical. Using $|M_i^j| = (1 \pm \varepsilon)\frac{n}{s}$ and $p \overset{\text{POLY}}{\gg} \varepsilon, \gamma$, we have that $G_{i,j,t}$ is globally $(1 - p/4)\frac{\delta|M_i^j|}{(1+5\varepsilon)2s^2 \log^2 n}$-bounded. Noting again that there are at most $s^4 \log n \leqslant ns^{-1}$ choices of (i, j, t), we can fix the choices of m_e so that this holds for each i, j, t.

Apply Lemma 8.17 with $\delta' = \frac{\delta}{(1+5\varepsilon)s^2 \log^2 n}$, $p' = \frac{p}{(1+5\varepsilon)s^2 \log^2 n}$, $n' = |M_i^j|$, $\gamma' = 12\gamma$ and $k' = k_i^j$ in order to find a family $\mathcal{F}_{i,j,t}$ of $(1 - p/2)\frac{\delta|M_i^j|}{(1+5\varepsilon)2s^2 \log^2 n}$ edge-disjoint rainbow 2-factors with cycles of length $\geqslant k$ in $G_{i,j,t} \cup J_{i,j,t}$ (for this application, we are using $1 \geqslant \delta, p, \log^{-1} n \overset{\text{POLY}}{\gg} s^{-1} \overset{\text{POLY}}{\gg} k^{-1} \overset{\text{POLY}}{\gg} \gamma \overset{\text{POLY}}{\gg} n^{-1}$ to conclude that $1 \geqslant \delta', p' \overset{\text{POLY}}{\gg} k'^{-1} \overset{\text{POLY}}{\gg} \gamma' \overset{\text{POLY}}{\gg} n'^{-1}$. The divisibility condition in Lemma 8.17 comes from the property of k_i^j in Lemma 8.20). Since $|M_i^j| = (1 \pm \varepsilon)\frac{n}{s}$ and $p \overset{\text{POLY}}{\gg} \varepsilon$, we can choose a subfamily $\mathcal{F}'_{i,j,t}$ of size $(1 - p)\frac{\delta n}{2s^3 \log^2 n}$. Let $\mathcal{F}_{i,t} = \bigcup_{j=1}^s \mathcal{F}'_{i,j,t}$ to get a family of $(1 - p)\frac{\delta n}{2s^2 \log^2 n}$ edge-disjoint rainbow 2-factors in $G \cup J$ with cycles of length $\geqslant k$. To see that these are rainbow 2-factors note that for $j \neq j'$ the graphs $G_{i,j,t} \cup J_{i,j,t}$ and $G_{i,j',t} \cup J_{i,j',t}$ are vertex-disjoint (their vertex sets are M_i^j and $M_i^{j'}$, respectively) and colour-disjoint (their colours are contained in $C_{j+t \pmod{s}}$ and $C_{j'+t \pmod{s}}$,

respectively). Since the 2-factors in $\{\mathcal{F}_{i,t} : 1 \leqslant i \leqslant s^2 \log^2 n, 1 \leqslant t \leqslant s\}$ are all edge-disjoint, we have a total of $(1-p)\delta n/2$ edge-disjoint rainbow 2-factors as required. $\square$

By combining the above with a regularization lemma, we can find 2-factor decompositions in nearly-complete graphs which have few large colours.

LEMMA 8.22. *Let* $1 \geqslant \varepsilon, \log^{-1} n \overset{\text{\tiny POLY}}{\gg} k^{-1} \overset{\text{\tiny POLY}}{\gg} \gamma \overset{\text{\tiny POLY}}{\gg} n^{-1}$. *Let* H *be a properly coloured,* $(\gamma, 1 - \varepsilon^2, n)$-*typical graph with* $\leqslant (1 - 60\varepsilon)n$ *colours having* $\geqslant (1 - 60\varepsilon)n/2$ *edges. Then,* H *has* $(1 - 3\varepsilon)n/2$ *edge-disjoint rainbow spanning 2-factors with cycles of length* $\geqslant k$.

Proof. Choose $1 \geqslant \varepsilon, \log^{-1} n \overset{\text{\tiny POLY}}{\gg} p \overset{\text{\tiny POLY}}{\gg} k^{-1} \overset{\text{\tiny POLY}}{\gg} \gamma \overset{\text{\tiny POLY}}{\gg} n^{-1}$. Apply Lemma 5.2(a) with $p = p/(1 - \varepsilon^2)$, $\delta = (1 - \varepsilon^2)$ and $\gamma = \gamma$ in order to partition H into a $(2\gamma, p, n)$-typical graph J and a colour-disjoint graph G with $\delta(G) \geqslant \delta(H) - \Delta(J) \geqslant (1 - 3\varepsilon^2)n$.

Apply Lemma 6.6 to G with $\varepsilon' = 3\varepsilon$, $\gamma = \gamma$ in order to find a subgraph G' which is globally $(1 - 3\varepsilon)n/2$-bounded and (γ, δ, n)-regular for some $\delta \geqslant 1 - 3\varepsilon + 81\varepsilon^2$. Note that G' is globally $(1-p)\delta n/2$-bounded (since $p \overset{\text{\tiny POLY}}{\ll} \varepsilon$).

By Lemma 8.21 applied to G' and J with $\gamma' = 2\gamma$, $\delta = \delta$, $p = p$, $k = k$, there are $(1 - p)\delta n/2 \geqslant (1 - 3\varepsilon)n/2$ edge-disjoint rainbow 2-factors with cycles of length $\geqslant k$. $\square$

8.4. *Hamiltonian cycles*

Here we take the 2-factor decompositions from the previous section and modify them into Hamiltonian decompositions. The proofs and results in this section are very similar to the ones in Section 8.2 where we took nearly-perfect matchings and modified them into perfect matchings. The following is the first result we prove about turning a family of 2-factors into a family of Hamiltonian cycles. It parallels Lemma 8.9 for turning nearly-perfect matchings into perfect matchings.

As in Lemmas 7.4 and 7.5, the following lemma has a mix of directed and undirected graphs. As before in the cycles we build, we do not care about the directions of their edges.

LEMMA 8.23. *Let* $1 \overset{\text{\tiny POLY}}{\gg} \delta \overset{\text{\tiny POLY}}{\gg} p \overset{\text{\tiny POLY}}{\gg} \theta \overset{\text{\tiny POLY}}{\gg} k^{-1} \overset{\text{\tiny POLY}}{\gg} n^{-1}$ *and* $t \leqslant n$. *Suppose that we have the following edge-disjoint, properly coloured graphs on a set of* n *vertices.*

- $F_1, \ldots, F_t$ *rainbow 2-factors with cycles of length* $\geqslant k$.
- $(3p(\theta n)^2, \theta n)$-*dense graphs* E_1, E_2, E_3.
- *Digraphs* D_X, D_Y *with* $\delta^+(D_X), \delta^+(D_Y) \geqslant 3\delta n$.

Additionally suppose that $F_1 \cup \cdots \cup F_t$, E_1, E_2, E_3, D_X *and* D_Y *are colour-disjoint. Then, there are edge-disjoint rainbow Hamiltonian cycles* $C_1, \ldots, C_t$ *in* $E_1 \cup E_2 \cup E_3 \cup D_X \cup D_Y \cup F_1 \cup \cdots \cup F_t$.

Proof. For $i = 1, \ldots, t$, let m_i be the number of cycles in F_i, and note that $m_i \leqslant k^{-1}n \leqslant p\theta n$.

CLAIM 8.24. *There are matchings* $M_1, \ldots, M_t$ *with* $\Delta(M_1 \cup \cdots \cup M_t) \leqslant 4n/k$ *such that* $M_i \subseteq F_i$ *is a matching of size* m_i *containing exactly one edge from each cycle of* F_i.

Proof. Choose each matching M_i uniformly at random from all matchings containing exactly one edge from each cycle of F_i. If $t \leqslant 4n/k$, then we trivially have $\Delta(M_1 \cup \cdots \cup M_t) \leqslant 4n/k$. Otherwise, note that for any vertex v, its degree in $M_1 \cup \cdots \cup M_t$ is stochastically dominated by Binomial$(t, 2k^{-1})$. By Chernoff's Bound with $\varepsilon = 1/2$, the union bound and $t \leqslant n$, we have

$\mathbb{P}(\Delta(M_1 \cup \cdots \cup M_t) > 4t/k) \leqslant 4ne^{-2t/12k} < 1)$ (using $n^{-1} \overset{\text{POLY}}{\ll} k^{-1}$ and $4n/k \leqslant t$). A choice of matchings satisfying the claim thus exists. $\qquad \square$

Let $M_i = \{x_i^1 y_i^1, \ldots, x_i^{m_i} y_i^{m_i}\}$. We construct the Hamiltonian cycles $C_1, \ldots, C_t$ one-by-one using Lemma 7.5. They will have the following properties.

(i) $xy \in D_X \cap C_i \Rightarrow x = x_i^j$ for some $j \in \{1, \ldots, m_i\}$.
(ii) $yx \in D_Y \cap C_i \Rightarrow y = y_i^j$ for some $j \in \{1, \ldots, m_i\}$.
(iii) $e(E_i \cap C_i) \leqslant k^{-1}n$ for $i = 1, 2, 3$.

Suppose that we have constructed Hamiltonian cycles $C_1, \ldots, C_s$ satisfying the above properties. Let $E_i^s = E_i \setminus (C_1 \cup \cdots \cup C_s)$ for $i = 1, 2, 3$, $D_X^s = D_X \setminus (C_1 \cup \cdots \cup C_s)$, $D_Y^s = D_Y \setminus (C_1 \cup \cdots \cup C_s)$. Using (i), $\delta^+(D_X) \geqslant 3\delta n$, and $\Delta(M_1 \cup \cdots \cup M_t) \leqslant 4n/k \leqslant \delta n$, we have $d_{D_X^s}^+(x) \geqslant d_{D_X}^+(x) - d_{D_X \cap (C_1 \cup \cdots \cup C_s)}^+(x) \geqslant \delta^+(D_X) - 4n/k \geqslant \delta n + nk^{-1} \geqslant \delta n + m_{s+1}$ for any $x \in X$. Similarly, $d_{D_Y^s}^+(y) \geqslant \delta n + m_{s+1}$ for $y \in Y$. Using Lemma 5.9 and (iii), E_1^s, E_2^s, E_3^s are $(3p(\theta n)^2 - s(k^{-1}n), \theta n)$-dense. Since $k^{-1}ns, m_{s+1}\theta n \leqslant p(\theta n)^2$, they are also $(p(\theta n)^2 + m_{s+1}\theta n, \theta n)$-dense. By Lemma 7.5 applied to F_{s+1}, $E_1^s, E_2^s, E_3^s, D_X^s, D_Y^s$ with $\theta = \theta$, $p = p$, $m = m_{s+1}$, $\delta = \delta$, $(x_j, y_j) = (x_i^j, y_i^j)$, there is a rainbow Hamiltonian cycle C_{s+1} satisfying (i)–(iii). $\qquad \square$

We will need the following easy lemma.

LEMMA 8.25. *Let $\delta \overset{\text{POLY}}{\gg} n^{-1}$. Every graph G with $\delta(G) \geqslant \delta n$ has an orientation D such that $\delta^+(D) \geqslant \delta n/3$.*

Proof. Orient the graph at random. Note that for any vertex $d^+(v) \sim \text{Binomial}(0.5, d(v))$. By Chernoff's Bound, we have $\mathbb{P}(d^+(v) < d(v)/3) \leqslant 2e^{-d(v)/54} \leqslant 2e^{-\delta n/54} = o(n^{-1})$ (using $\delta \overset{\text{POLY}}{\gg} n^{-1}$). Taking a union bound over all vertices shows that some suitable orientation exists. $\qquad \square$

The following version of Lemma 8.23 will be easier to apply.

LEMMA 8.26. *Let $1 \overset{\text{POLY}}{\gg} p \overset{\text{POLY}}{\gg} \gamma, k^{-1} \overset{\text{POLY}}{\gg} n^{-1}$. For $t \leqslant n$, let $F_1, \ldots, F_t$ be edge-disjoint rainbow 2-factors with cycles of length $\geqslant k$. Let G be an edge-disjoint, colour-disjoint (γ, p, n)-typical graph. Then, there are edge-disjoint rainbow Hamiltonian cycles $C_1, \ldots, C_t$ in $F_1 \cup \cdots \cup F_t \cup G$.*

Proof. Choose $1 \overset{\text{POLY}}{\gg} p \overset{\text{POLY}}{\gg} \delta \overset{\text{POLY}}{\gg} p_1 \overset{\text{POLY}}{\gg} \theta \overset{\text{POLY}}{\gg} \gamma, k^{-1} \overset{\text{POLY}}{\gg} n^{-1}$. Choose five disjoint sets of colours $C_{E_1}, C_{E_2}, C_{E_3}, C_{D_X}, C_{D_Y}$ from G, with each colour put independently into $C_{E_1}, C_{E_2}, C_{E_3}, C_{D_X}, C_{D_Y}$ with probabilities $4p_1 p^{-1}, 4p_1 p^{-1}, 4p_1 p^{-1}, 10\delta p^{-1}, 10\delta p^{-1}$, respectively (this is possible since $p \overset{\text{POLY}}{\gg} \delta \overset{\text{POLY}}{\gg} p_1$ implies $4p_1 p^{-1} + 4p_1 p^{-1} + 4p_1 p^{-1} + 10\delta p^{-1} + 10\delta p^{-1} \leqslant 1$). Let E_1, E_2, E_3, D_X, D_Y be the subgraphs of G with colours from $C_{E_1}, C_{E_2}, C_{E_3}, C_{D_X}, C_{D_Y}$, respectively. By Lemma 5.2 (a), with positive probability E_1, E_2, E_3 are $(2\gamma, 4p_1, n)$-typical and D_X, D_Y are $(2\gamma, 10\delta, n)$-typical. By Lemma 8.25 and $(2\gamma, 10\delta, n)$-typicality, D_X and D_Y can be oriented so that $\delta^+(D_X), \delta^+(D_Y) \geqslant 3\delta n$.

By Lemma 5.7 applied with $\mu = \theta$, $\gamma' = \gamma/2$, E_1, E_2, E_3 are $(3p_1(\theta n)^2, \theta n)$-dense. By Lemma 8.23 applied with $\theta = \theta$, $p = p_1$, we obtain the required Hamiltonian cycles. $\qquad \square$

The following lemma should be compared with Lemmas 8.11 and 8.21. It produces a near-decomposition into Hamiltonian cycles under a similar assumption to those lemmas.

LEMMA 8.27. *Let* $1 \geqslant \delta, p, \log^{-1} n \overset{\text{POLY}}{\gg} \gamma \overset{\text{POLY}}{\gg} n^{-1}$. *Let* G *be a properly coloured* (γ, δ, n)-*typical graph which is globally* $(1-p)\delta n/2$-*bounded. Then,* G *has* $(1-p)\delta n/2$ *edge-disjoint rainbow Hamiltonian cycles.*

Proof. Choose $1 \overset{\text{POLY}}{\gg} \delta, p, \log^{-1} n \overset{\text{POLY}}{\gg} p_1 \overset{\text{POLY}}{\gg} k^{-1} \overset{\text{POLY}}{\gg} \gamma \overset{\text{POLY}}{\gg} n^{-1}$. Apply Lemma 5.2(a) with $p' = p_1$, $\delta = \delta$ and $\gamma = \gamma$ in order to partition H into three colour-disjoint graphs G', J_1 and J_2 so that J_1 and J_2 are $(2\gamma, p_1\delta, n)$-typical and G' is $(2\gamma, (1-2p_1)\delta, n)$-typical. Setting $\delta_1 = (1-2p_1)\delta$ and using $p \overset{\text{POLY}}{\gg} p_1$, we have that G' is $(2\gamma, \delta_1, n)$-typical and globally $(1-p_1)\delta_1 n/2$-bounded.

Apply Lemma 8.21 to G' and J_1 with $\gamma' = 2\gamma, \delta' = \delta_1$, $p' = p_1$, $k = k$ in order to find $(1-p_1)\delta_1 n/2$ edge-disjoint rainbow spanning 2-factors $F_1, \ldots, F_{(1-p_1)\delta_1 n/2}$ in G' whose cycles have length $\geqslant k$.

Apply Lemma 8.26 to $F_1, \ldots, F_{(1-p_1)n/2}$ and J_2 with $p = \delta p_1$, $\gamma' = 2\gamma$, $k = k$, $t = (1-p_1)\delta_1 n/2$ in order to find $(1-p_1)\delta_1 n/2$ edge-disjoint rainbow Hamiltonian cycles in G. Since $p \overset{\text{POLY}}{\gg} p_1$ implies $(1-p_1)\delta_1 n/2 \geqslant (1-p)\delta n/2$, we have enough cycles for the lemma. $\square$

The following lemma should be compared with Lemma 8.22. Under similar assumptions, it produces a near-decomposition into Hamiltonian cycles rather than 2-factors.

LEMMA 8.28. *Let* $1 \overset{\text{POLY}}{\gg} \varepsilon, \log^{-1} n \overset{\text{POLY}}{\gg} \gamma \overset{\text{POLY}}{\gg} n^{-1}$. *Let* G *be a properly coloured,* $(\gamma, 1 - \varepsilon^2, n)$-*typical graph with* $\leqslant (1 - 180\varepsilon)n$ *colours having* $\geqslant (1 - 180\varepsilon)n/2$ *edges. Then,* G *has* $(1 - 6\varepsilon)n/2$ *edge-disjoint rainbow Hamiltonian cycles.*

Proof. Choose $1 \overset{\text{POLY}}{\gg} \varepsilon, \log^{-1} n \overset{\text{POLY}}{\gg} p \overset{\text{POLY}}{\gg} k^{-1} \overset{\text{POLY}}{\gg} \gamma \overset{\text{POLY}}{\gg} n^{-1}$. Apply Lemma 5.2 (a) with $p' = p/(1-\varepsilon^2)$, $\delta = (1-\varepsilon^2)$, and $\gamma = \gamma$ in order to partition H into a $(2\gamma, p, n)$-typical graph J and a colour-disjoint $(2\gamma, 1 - p - \varepsilon^2, n)$-typical graph G'.

Let $\varepsilon' = \sqrt{p + \varepsilon^2}$, and note that $\varepsilon \overset{\text{POLY}}{\gg} p$ implies $\varepsilon \leqslant \varepsilon' \leqslant 2\varepsilon$. Hence, G' has $\leqslant (1 - 60\varepsilon')n$ colours having $\geqslant (1 - 60\varepsilon')n/2$ edges. Apply Lemma 8.22 to G' with $\gamma' = 2\gamma, \varepsilon' = \varepsilon'$, $k = k$ in order to find $(1 - 3\varepsilon')n/2$ edge-disjoint rainbow spanning 2-factors $F_1, \ldots, F_{(1-3\varepsilon')n/2}$ in G' whose cycles have length $\geqslant k$.

Apply Lemma 8.26 to $F_1, \ldots, F_{(1-3\varepsilon')n/2}$ and J with $p = p$, $\gamma' = 2\gamma$, $k = k$, $t = (1 - 3\varepsilon')n/2$ in order to find $(1 - 3\varepsilon')n/2$ edge-disjoint rainbow Hamiltonian cycles in G. Since $(1 - 3\varepsilon')n/2 \geqslant (1 - 6\varepsilon)n/2$, we have enough cycles. $\square$

We can show that when a properly coloured K_n has few large colours, then it has a near-decomposition into Hamiltonian cycles. This is 'half' of our proof of the asymptotic version of the Brualdi–Hollingsworth and Kaneko–Kano–Suzuki Conjectures. The other half will be in the case when there are many large colours, which is performed in Section 9. The following lemma is a restatement, in our notation, of Theorem 1.10.

LEMMA 8.29. *Let* $1 \overset{\text{POLY}}{\gg} \varepsilon \overset{\text{POLY}}{\gg} n^{-1}$. *Let* K_n *be properly coloured with* $\leqslant (1 - \varepsilon)n$ *colours having* $\geqslant (1 - \varepsilon)n/2$ *edges. Then* K_n *has* $(1 - \varepsilon)n/2$ *edge-disjoint rainbow Hamiltonian cycles.*

Proof. Choose $1 \overset{\text{POLY}}{\gg} \varepsilon, \log^{-1} n \overset{\text{POLY}}{\gg} \varepsilon_1 \overset{\text{POLY}}{\gg} \gamma \overset{\text{POLY}}{\gg} n^{-1}$. Let G be an arbitrary $(\gamma, 1 - \varepsilon_1^2, n)$-typical subgraph of K_n (it exists, for example, by Lemmas 5.1 and 5.2 (a) or (b)). Note that since $\varepsilon \overset{\text{POLY}}{\gg} \varepsilon_1$, G has $\leqslant (1 - 180\varepsilon_1)n$ colours having $\geqslant (1 - 180\varepsilon_1)n/2$ edges. By Lemma 8.28, G has $(1 - 6\varepsilon_1)n/2 \geqslant (1 - \varepsilon)n/2$ edge-disjoint rainbow Hamiltonian cycles. $\square$

We now restate, in our notation, Corollary 1.11, and deduce it from Lemma 8.29.

COROLLARY 8.30. *Let $1 \overset{\text{\tiny POLY}}{\gg} \varepsilon \overset{\text{\tiny POLY}}{\gg} n^{-1}$. Given a properly coloured K_n, let U be a random subset of $(1 - \varepsilon)n$ vertices. Then, with high probability, the subgraph induced by U has $(1 - 2\varepsilon)n/2$ edge-disjoint rainbow Hamiltonian cycles.*

Proof. Take η with $\varepsilon \overset{\text{\tiny POLY}}{\gg} \eta \overset{\text{\tiny POLY}}{\gg} 1/n$. By Lemma 5.2(c), $G[U]$ is with high probability globally $(1 + \eta)(1 - \varepsilon)^2 n/2$-bounded, and hence $(1 - \varepsilon/2)|U|/2$-bounded. Thus, by Lemma 8.29, $G[U]$ has $(1 - \varepsilon/2)|U|/2 \geqslant (1 - 2\varepsilon)n/2$ edge-disjoint rainbow Hamiltonian cycles. □

9. Rainbow trees

In this section, we show that the Brualdi–Hollingsworth and Kaneko–Kano–Suzuki Conjectures hold asymptotically. Part of this result was already proved in Lemma 8.29, which shows that the asymptotic versions of the Brualdi–Hollingsworth and Kaneko–Kano–Suzuki Conjectures hold in colourings of K_n which have few large colours. In this section, we focus on colourings of K_n which have many large colours. Such colourings should be thought of as being close to 1-factorizations.

The basic idea of the proof is to note that for *any* properly coloured K_n, we know how to find a large set of vertices S so that the induced subgraph $K_n[S]$ has a near-decomposition into Hamiltonian paths. Indeed, a random set S will have this property (by combining Lemmas 5.2(c) and 8.27). To find a near-decomposition into spanning trees, we modify the paths in $K_n[S]$ by extending them one vertex at a time to cover all of $V(K_n)$.

9.1. Small rainbow trees

Here we prove a result about near-decompositions of globally bounded graphs into rainbow forests which are sufficiently small. We remark that this lemma is only needed to deal with properly coloured complete graphs which are not 1-factorizations — if one only wants to prove an asymptotic version of the Brualdi–Hollingsworth Conjecture, then this section can be omitted.

The result we prove in this section is essentially the following: for $m \overset{\text{\tiny POLY}}{\gg} k$, every properly coloured globally m-bounded graph with $\geqslant (1 + o(1))mk$ edges has a near-decomposition into m rainbow k-edge forests $F_1, \ldots, F_m$. This is relatively straightforward (see Lemma 9.1), however, we need to find such a near-decomposition that interacts well with a large *vertex cover*. Here, a *vertex cover* S is a set of vertices which contains at least one vertex in each edge. We develop Lemma 9.1 through Lemma 9.3 to arrive at the result we need, Lemma 9.5.

LEMMA 9.1. *Let $1 \overset{\text{\tiny POLY}}{\gg} \beta \overset{\text{\tiny POLY}}{\gg} k/n$, $m \geqslant \beta n$ and $0 \leqslant \ell < m$. Let G be a properly coloured, globally m-bounded, n-vertex graph, with $e(G) \geqslant (1 + \beta)(km + \ell)$. Then, G has m edge-disjoint rainbow forests $F_1, \ldots, F_m$, so that each F_i has $k + \mathbf{1}_{\{i \leqslant \ell\}}$ edges.*

Proof. Note that if $k = 0$, then selecting ℓ edges gives the required forests. Assume then that $k \geqslant 1$. By deleting edges if necessary, assume that $e(G) = (1 + \beta)(km + \ell)$.

Choose integers $d_c, c \in C(G)$, such that $\lfloor \frac{|E_G(c)|}{1 + \beta} \rfloor \leqslant d_c \leqslant \lceil \frac{|E_G(c)|}{1 + \beta} \rceil$ and $\sum_{c \in C(G)} d_c = km + \ell$, where we have used that

$$\sum_{c \in C(G)} \left\lfloor \frac{|E_G(c)|}{1 + \beta} \right\rfloor \leqslant \sum_{c \in C(G)} \frac{|E_G(c)|}{1 + \beta} = \frac{e(G)}{1 + \beta} = km + \ell \leqslant \sum_{c \in C(G)} \left\lceil \frac{|E_G(c)|}{1 + \beta} \right\rceil.$$

Let $C_1, \ldots, C_m$ be sets in $C(G)$, so that each C_i has size $k + \mathbf{1}_{\{i \leqslant \ell\}}$, and each colour c appears in d_c sets C_i. Note that this is possible as $d_c \leqslant m$ for each $c \in C(G)$ and $\sum_{i=1}^{m}(k + \mathbf{1}_{\{i \leqslant \ell\}}) = km + \ell$. Let $F_1', \ldots, F_m'$ be edge-disjoint rainbow forests in G with $C(F_i') \subset C_i$ for each i and so

that $|\cup_{i=1}^m E(F_i')|$ is maximized. Suppose that $|\cup_{i=1}^m E(F_i')| < km + \ell$, for otherwise $F_1', \ldots, F_m'$ satisfy the lemma.

CLAIM 9.2. *For each colour c, $|E_G(c) \setminus (\cup_{i=1}^m E(F_i'))| \leqslant \frac{\beta m}{1+\beta} + 2k$.*

Proof. Fixing a colour c, let M be the edges with colour c not in $\cup_{i=1}^m E(F_i')$, and suppose that $|M| \geqslant |E_G(c)| - d_c + k + 1$. As $\sum_{i=1}^m |E(F_i') \cap E_G(c)| < d_c$, there is some j for which $c \in C_j$ but F_j' contains no colour c edge, so that, furthermore, $|V(F_j')| \leqslant 2k$. But then, as $|M| \geqslant k+1$, there is some colour c edge in M which is not contained in $V(F_j')$, contradicting the maximality of $|\cup_{i=1}^m E(F_i')|$.

Thus, we must have $|M| < |E_G(c)| - d_c + k + 1 \leqslant |E_G(c)| - \lfloor \frac{|E_c(G)|}{1+\beta} \rfloor + 2k \leqslant m - \lfloor \frac{m}{1+\beta} \rfloor + 2k = \lceil \frac{\beta m}{1+\beta} \rceil + 2k$. $\qquad\square$

Next, let $F_1, \ldots, F_m$ be a set of edge-disjoint rainbow forests in G with $F_i' \subset F_i$ and $|E(F_i)| \leqslant k + \mathbf{1}_{\{i \leqslant \ell\}}$ for each i, so that $|\cup_{i=1}^m E(F_i)|$ is maximized. Suppose that there is some $1 \leqslant j \leqslant m$ for which $|E(F_j)| < k + \mathbf{1}_{\{j \leqslant \ell\}}$. Any edge outside of $\cup_{i=1}^m E(F_j)$ must be contained in $V(F_j)$ or share a colour with F_j. Thus, by Claim 9.2, we have

$$e(G) \leqslant |\cup_{i=1}^m E(F_i)| + \binom{2k}{2} + k \left(\frac{\beta m}{1+\beta} + 2k \right)$$

$$\leqslant km + \ell + 4k^2 + k \frac{\beta m}{1+\beta}$$

$$\leqslant km + \ell + 4k^2 + k\beta m(1 - \beta/2)$$

$$\leqslant (1+\beta)(km + \ell) + k(4k - \beta^2 m/2)$$

$$\leqslant (1+\beta)(km + \ell) + k(4k - \beta^3 n/2)$$

$$< (1+\beta)(km + \ell),$$

where we have used that $m \geqslant \beta n$ and $1 \overset{\text{POLY}}{\gg} \beta \overset{\text{POLY}}{\gg} k/n$. This contradicts $e(G) = (1+\beta)(km + \ell)$, and thus there is no such j with $|E(F_j)| < k + \mathbf{1}_{\{i \leqslant \ell\}}$. $\qquad\square$

Given a large vertex cover S in a graph G, we wish to find edge-disjoint k-edge rainbow forests so that large degree vertices outside S are in every forest while small degree vertices outside S have degree at most 1 in every forest. Lemma 9.1 can almost cover the edges within S with forests. We now expand this to prove a lemma almost covering these edges as well as the edges next to vertices with small degree in $A := V(G) \setminus S$.

LEMMA 9.3. *Let $1 \overset{\text{POLY}}{\gg} \beta \overset{\text{POLY}}{\gg} \varepsilon \overset{\text{POLY}}{\gg} k/n$ and $m \geqslant \beta n$. Let G be a properly coloured, globally m-bounded, n-vertex graph, with $e(G) \geqslant (1+\beta)km$ and let A be a set of vertices with $|A| \leqslant \varepsilon n$ which contains no edges in G. Furthermore, suppose $d(v) \leqslant m + 2k$ for each $v \in A$.*

Then, G has edge-disjoint k-edge rainbow forests $F_1, \ldots, F_m$, where, additionally, for any $v \in A$ and $1 \leqslant i \leqslant m$, $d_{F_i}(v) \leqslant 1$.

Proof. Note that the lemma is trivial if $k = 0$, and follows immediately from Lemma 9.1 when $A = \emptyset$. Suppose then that $k, |A| \geqslant 1$. Let G_1 be the subgraph of edges of G with no vertices in A. Let k' and ℓ be integers with $0 \leqslant \ell < m$ maximizing $k'm + \ell$ subject to $e(G_1) \geqslant (1+\beta^2)(k'm + \ell)$ and $k'm + \ell \leqslant km$. By Lemma 9.1, G_1 contains edge-disjoint rainbow forests $F_1', \ldots, F_m'$ so that F_i' has $k' + \mathbf{1}_{\{i \leqslant \ell\}}$ edges. If $k'm + \ell = km$, then these forests satisfy the lemma, so suppose that $k'm + \ell < km$, and therefore $e(G_1) \leqslant (1+\beta^2)(k'm + \ell) + 2$.

Let $G_2 = G - G_1$ and pick an integer λ so that $\beta km/4 \geqslant \lambda|A| \geqslant \beta km/8$ (which is possible as $|A| \leqslant \varepsilon n$, $m \geqslant \beta n$ and $\beta \overset{\text{POLY}}{\gg} \varepsilon$). Note that

$$e(G_2) \geqslant (1+\beta)km - (1+\beta^2)(k'm + \ell) - 2 \geqslant (1+\beta^2)(km - k'm + \ell) + \beta km/2 - 2$$

$$\geqslant (1+\beta^2)(km - k'm + \ell) + \beta km/4 \geqslant (1+\beta^2)(km - k'm - \ell) + \lambda|A|.$$

By deleting edges if necessary, assume that $e(G_2) = (1+\beta^2)(km - k'm - \ell) + \lambda|A|$. Choose integers d_v, $v \in A$, such that $\lfloor \frac{d_{G_2}(v) - \lambda}{1+\beta^2} \rfloor \leqslant d_v \leqslant \lceil \frac{d_{G_2}(v) - \lambda}{1+\beta^2} \rceil$ and $\sum_{v \in A}(d_v - \lambda) = km - k'm - \ell$. Let $A_1, \ldots, A_m$ be sets in A, so that each A_i has size $k - k' - \mathbf{1}_{\{i \leqslant \ell\}}$ and each vertex v appears in d_v sets A_i, noting that this is possible as for each v we have, as $\beta \overset{\text{POLY}}{\gg} k/n$ and $m \overset{\text{POLY}}{\gg} \beta n$,

$$d_v \leqslant 1 + \frac{d(v) - \lambda}{1+\beta^2} \leqslant 1 + d(v)(1 - \beta^2/2) \leqslant 1 + (m + 2k)(1 - \beta^2/2) \leqslant m,$$

and $\sum_{i=1}^{m}(k - k' - \mathbf{1}_{\{i \leqslant \ell\}}) = km - k'm - \ell$.

Let $F_1'', \ldots, F_m''$ be a set of edge-disjoint rainbow forests in G with, for each i, $F_i' \subset F_i'' \subset F_i' \cup G_2$ and $d_{F_i'}(v) \leqslant 1$ for each $v \in A_i$ and $d_{F_i'}(v) = 0$ for each $v \in A \setminus A_i$. Furthermore, suppose $|\cup_{i=1}^{m} E(F_i'')|$ is maximized subject to these conditions.

CLAIM 9.4.   For each $v \in A$, $|E_G(v) \setminus (\cup_{i=1}^{m} E(F_i''))| \leqslant \frac{\beta^2 m + \lambda}{1+\beta^2} + 2k$.

Proof.   Fixing a vertex $v \in A$, let E be the edges through v not in $\cup_{i=1}^{m} E(F_i'')$, and suppose $|E| \geqslant d_{G_2}(v) - d_v + k$. As $\sum_{i=1}^{m}|V(F_i'') \cap \{v\}| = d_{G_2}(v) - |E| < d_v$, there is some j for which $v \in A_j$ but F_j'' contains no edge adjacent to v. But then, as $|E| \geqslant k$ and G is properly coloured, there is some edge in E with colour outside of $C(F_j'')$, contradicting the maximality of $|\cup_{i=1}^{m} E(F_i'')|$.

Thus, we must have $|E| \leqslant d_{G_2}(v) - d_c + k \leqslant d_{G_2}(v) - \lfloor \frac{d_{G_2}(v) - \lambda}{1+\beta^2} \rfloor + k \leqslant m + k - \lfloor \frac{m - \lambda}{1+\beta^2} \rfloor + k \leqslant \lceil \frac{\beta^2 m - \lambda}{1+\beta^2} \rceil + 2k$. $\qquad \square$

Let $F_1, \ldots, F_m$ be a set of edge-disjoint rainbow forests in G with $F_i'' \subset F_i \subset F_i'' \cup G_2 = F_i' \cup G_2$ and $|E(F_i)| \leqslant k$ for each $1 \leqslant i \leqslant m$, and $d_{F_i}(v) \leqslant 1$ for each $1 \leqslant i \leqslant m$ and $v \in A$, so that $|\cup_{i=1}^{m} E(F_i)|$ is maximized. Suppose that there is some $1 \leqslant j \leqslant m$ for which $|E(F_j)| < k$. Any edge in G_2 outside of $\cup_{i=1}^{m} E(F_i)$ must contain a vertex in $V(F_j) \cap A$ or share a colour with F_j. Recall that there are no edges in G within A, and note that every edge in $E(F_j) \setminus E(F_j')$ is in G_2. Thus, as $F_j' \subset E(F_j)$ is a forest with $k' + \mathbf{1}_{\{i \leqslant \ell\}}$ edges and no vertices in A, we have,

$$|V(F_j) \cap A| = |E(F_j) \setminus E(F_j')| < k - k' - \mathbf{1}_{\{i \leqslant \ell\}} \leqslant k - k',$$

and, hence, $|V(F_j) \cap A| \leqslant \min\{|A|, k - k'\}$. Thus, by Claim 9.4, and noticing that G_2 is globally $|A|$-, and hence εn-, bounded, we have

$$e(G_2) \leqslant |\cup_{i=1}^{m} E(F_i) \setminus E(F_i')| + \min\{|A|, k - k'\} \cdot \left(\frac{\beta^2 m + \lambda}{1+\beta^2} + 2k\right) + k\varepsilon n$$

$$\leqslant km - (k'm + \ell) + (k - k')\frac{\beta^2 m}{1+\beta^2} + |A|\frac{\lambda}{1+\beta^2} + 2k^2 + k\varepsilon n$$

$$\leqslant km - k'm - \ell + \beta^2 m(k - k') + \lambda(1 - \beta^2/2)|A| + 2k^2 + k\varepsilon n$$

$$\leqslant (1+\beta^2)(km - k'm - \ell) + \lambda|A| + (2k^2 + k\varepsilon n - \lambda\beta^2|A|/2)$$

$$\leqslant (1+\beta^2)(km - k'm - \ell) + \lambda|A| + (2k^2 + k\varepsilon n - \beta^4 kn/16)$$

$$< (1+\beta^2)(km - k'm - \ell) + \lambda|A|,$$

where we have used that $1 \overset{\text{POLY}}{\gg} \beta \overset{\text{POLY}}{\gg} \varepsilon, k/n$ and $\lambda|A| \geqslant \beta km/8 \geqslant \beta^2 kn/8$. This contradicts $e(G_2) = (1 + \beta^2)(km - k'm - \ell) + \lambda|A|$, so $F_1, \ldots, F_m$ must satisfy the lemma. $\qquad \square$

We can now prove the main result of this section.

LEMMA 9.5 (Near-decomposition into small rainbow trees). Let $1 \overset{\text{POLY}}{\gg} \beta \overset{\text{POLY}}{\gg} \varepsilon \overset{\text{POLY}}{\gg} k/n$ and $m \geqslant \beta n$. Let G be a properly coloured, globally m-bounded, n-vertex graph, with $e(G) \geqslant (1 + \beta)km$ and S a vertex cover of G with $|S| \geqslant (1 - \varepsilon)n \geqslant 2m$. Then, G has edge-disjoint k-edge rainbow forests $F_1, \ldots, F_m$, where, additionally, for any $v \notin S$ either $v \in V(F_i)$ for all i, or $d_{F_i}(v) \leqslant 1$ for all i.

Proof. Let $k' = 0$ and let G_1 be the empty graph with vertex set $V(G)$. Iteratively, for each $i \geqslant 1$, if $k' < k$ and there is a vertex $v_i \notin S$ with degree in G at least $m + 2(k - k')$, do the following. Let $d_i = \lceil \frac{d_G(v_i) - 2(k-k')}{m} \rceil$, let $d'_i = \min\{d_i, k - k'\}$ and pick a set E_i of $d'_i m + 2(k - k')$ edges next to v_i in G, remove them from G and add them to G_1, and increase k' by d'_i. Let G_2 be the remains of G at the end of this process.

Pick ℓ so that this produces a series of vertices $v_1, \ldots, v_\ell$, integers $d'_1, \ldots, d'_\ell$ and disjoint edge sets $E_1, \ldots, E_\ell$. Note that $\ell \leqslant n - |S| \leqslant \varepsilon n$ and, for each $i < \ell$, we have $|E_i| = d'_i m + 2(k - \sum_{j=1}^{i-1} d'_i)$. Note further that, if $k' < k$, then $d_{G_2}(v) \leqslant m + 2(k - k')$ for each $v \notin S$. Furthermore, if $k' < k$, then

$$e(G_2) \geqslant e(G) - k'm - 2k \cdot \ell \geqslant (k - k')m + \beta km - 2k\varepsilon n \geqslant (1 + \beta/2)(k - k')m,$$

where we have used that $m \geqslant \beta n$ and $\beta \overset{\text{POLY}}{\gg} \varepsilon$. Therefore, by Lemma 9.3, if $k' < k$, then G_2 contains m edge-disjoint rainbow $(k - k')$-edge forests $F'_1, \ldots, F'_m$, in which $d_{F'_i}(v) \leqslant 1$ for each $v \notin S$. Note that we can also assume this when $k' = k$.

For each $i = \ell, \ell - 1, \ldots, 1$ in turn, greedily, for each $j \in [m]$, add d'_i edges from E_i to F'_j, so that the resulting subgraphs, $F_1, \ldots, F_m$ say, are still edge-disjoint rainbow forests. This is possible as, firstly, when we come to add edges from E_i to each forest F'_j, v_i is in the $d'm_i + 2(k - \sum_{j<i} d'_j)$ edges in E_i, which are disjoint from $E(\cup_{i=1}^m F'_i) \subset E(G_1)$ and E_j, $j \geqslant i$. Each forest to which we add d'_i edges from E_i will have $(k - k') + \sum_{j>i} d'_j = k - \sum_{j \leqslant i} d'_j$ edges. We add in total, from E_i, d'_i edges to each of m forests, so at each addition there are at least $d'_i + 2(k - \sum_{j<i} d'_j)$ unused edges left in E_i, and therefore we can add d'_i edges from E_i without any vertices or colours in the forest we are adding edges to. Note that, in this process, if an edge is added to a forest F'_j adjacent to $v \notin S$, then an edge is added adjacent to v in all other forests F_i as well.

Thus, noting that the resulting forests each have k edges, $F_1, \ldots, F_m$ are edge-disjoint rainbow k-edge forests, with $\{v \notin S : d'_v \geqslant 1\} \subset F_i$ for each $1 \leqslant i \leqslant m$ and $d_{F_i}(v) \leqslant 1$ for each $1 \leqslant i \leqslant m$ and $v \in S$ with $d'_v = 0$. $\qquad \square$

9.2. Completion

In this section, we show how to modify nearly-spanning rainbow trees into spanning ones. The starting point of this section is Lemma 8.27. That lemma implies that in any properly coloured K_n, there is a set S of size $(1 - o(1))n$ such that $K_n[S]$ has a near-decomposition into rainbow Hamiltonian paths $P_1, \ldots, P_{(1-o(1))n/2}$. Indeed a random set S will have this property since it satisfies the assumptions of Lemma 8.27 by Lemma 5.2. Our goal in this section is to take such a family of rainbow paths, and modify them into a near-decomposition of K_n into rainbow spanning trees.

The paths $P_1, \ldots, P_{(1-o(1))n/2}$ are modified into spanning trees gradually, that is, we switch edges on them one at a time to get bigger and bigger rainbow trees. During this modification procedure, we always have a family of rainbow trees $T_1, \ldots, T_{(1-o(1))n/2}$ which satisfy several

properties that guarantee that it is possible to keep extending them. We will now informally go through these properties and explain why each is natural. The first property is the following:

(a) $S \subseteq V(T_i)$ for all i.

This property simply comes from the fact that the trees T_i are formed by enlarging the paths P_i, and the paths P_i had $V(P_i) = S$. Property (a) is useful to have because we will have more control over vertices outside S due to the fact that they were untouched by the starting paths $P_1, \ldots, P_{(1-o(1))n/2}$.

(b) For every $v \notin S$, the tree T_i has at most one edge through v.

Since we aim to produce trees which are spanning in K_n, every vertex $v \notin S$ will eventually need to be added to every tree. Condition (b) will ensure that every vertex $v \notin S$ always has enough free edges to be added to every tree. Without it, it is possible that all the edges in K_n through v lie in some small subfamily of trees $T_1, \ldots, T_m$, preventing the addition of v to the other trees.

(c) For a tree T_i, there are $n - |T_i|$ colours c outside T_i with $|E_{K_n}(c)| \geqslant (1-o(1))n/2$.

For a vertex $v \notin V(T_i)$, it will not always be possible to add an edge from v to T_i in order to produce a rainbow tree. While properties (a) and (b) ensure that there are free edges from v to T_i, it is conceivable that the colours of all these edges are already present on T_i, so v cannot simply be added while maintaining a rainbow tree. We get around this by finding some colour c outside of T_i and two edges $e \in E(c)$, $f \in e(T_i)$ so that $T_i - f + e$ is a rainbow tree, that is, we switch an edge on T_i for an edge of some previously unused colour. This operation frees up the colour $c(f)$, which we might be able to use to attach v. Property (c) ensures that there are *many* colours $c(f)$ which can be freed using this operation.

(d) There is a graph H disjoint from $T_1, \ldots, T_{(1-o(1))n/2}$ in which any set of k colours covers at least $(1-o(1))n$ vertices (where k is a large constant).

Property (d) plays a similar role to property (c), that is, it allows us to free up more colours, with the hope that eventually we free a colour which is present at some vertex $v \notin V(T_i)$ (and then add that vertex to the tree T_i). The reason we need both properties (c) and (d) is a bit technical. In general, property (d) is more powerful, except that to invoke it we need k colours outside the tree T_i. This will not happen towards the end of our process when there might be only one colour outside the tree. On the other hand, (c) can always be invoked to free up a small number of colours. The strategy is to combine the applications of (c) and (d), that is, first we apply (c) to free up k colours, and then we use (d) to free up enough colours to add v.

The following lemma is what we use to exchange edges on a tree with edges outside it.

LEMMA 9.6. *Let T be a tree and G a graph with no isolated vertices with $V(G) \subseteq V(T)$. Then, for every $v \in V(G)$, there are edges $xv \in E(T)$ and $yv \in E(G)$ with $T - xv + yv$ a tree. In particular, there are $\geqslant |G|/2$ edges $e \in T$ for which there is an edge $f \in E(G)$ with $T - e + f$ a tree.*

Proof. Let yv be an arbitrary edge of G containing v. If $yv \in E(T)$, then $T - yv + yv$ is the required tree, so assume that $yv \notin E(T)$. Since T is a tree and $\{y, v\} \subseteq V(T)$, $T + yv$ has a cycle C containing the edge yv. Let $xv \neq yv$ be the other edge of C containing v. Now $T + yv - xv$ is the required tree.

Thus, to every $v \in V(G)$, we can assign a pair of edges $e_v \in T$, $f_v \in G$ containing v with $T - e_v + f_v$ a tree. Since $v \in e_v$, for an edge $e \in E(T)$, there can be at most two vertices $v \in V(G)$ with $e = e_v$. This gives $|\{e_v : v \in V(G)\}| \geqslant |G|/2$ as required. $\square$

The following is the basic extension lemma which drives our proof. Under conditions to be compared to (b)–(d), it shows how to extend a tree by one vertex. The idea of the proof of the lemma is to show that by performing two switches as in Lemma 9.6, we can free up nearly half of the colours on T. At least one of these colours will have an edge going to T, which can be added to extend the tree.

LEMMA 9.7.   *In a properly coloured n-vertex graph G, suppose that we have:*

- *T a rainbow tree with $|T| = n - 1$;*
- *$v \notin V(T)$ with $d(v) \geqslant \frac{1}{2}n + b$;*
- *$c \notin C(T)$ with $e(c) \geqslant b$;*
- *H a graph on $V(T)$ in which any set of b colours of $C(T)$ covers $\geqslant n - 2b$ vertices.*

Then, there is a rainbow tree T' in $T \cup H \cup E(c) \cup E(v)$ with $V(T') = V(T) \cup \{v\}$, $e(T' \setminus T) \leqslant 3$ and $d_{T'}(v) = 1$.

Proof.   If there is a colour c edge next to v, then clearly we can add such an edge to T to get the required tree. Assume, then, that every colour in $C(v)$ is on T, and thus, in particular, $V(c) \subset V(G) \setminus \{v\} = V(T)$.

Let J be the set of edges $j \in E(T)$ for which there is a colour c edge e_j so that $T_j := T - j + e_j$ is a (rainbow) tree. By Lemma 9.6 and $e(c) \geqslant b$, we have $e(J) \geqslant b$ (for the application of this lemma, we take G to be the set of colour c edges).

For each $j \in J$, let H_j be the graph of colour $c(j)$ edges in H with no isolated vertices. By Lemma 9.6, we have

$$V(H_j) \subset V(\{e \in E(T_j) : \exists e' \in E(H_j) \text{ s.t. } T_j - e + e' \text{ is a tree}\}). \tag{32}$$

Note that the trees $T_j - e + e'$ above are always rainbow (since T_j is a rainbow tree on $V(T)$ missing colour $c(j)$ and $V(H_j) \subseteq V(T)$). Let

$$J' = \{e \in E(c) \cup E(T) : \exists \text{ a rainbow tree } T'_e \text{ in } T \cup H \cup E(c) \text{ with } V(T') = V(T),$$

$$e(T' \setminus T) \leqslant 2,\ c(e) \notin C(T'_e)\}.$$

Then, for each $j \in J$, by (32), we have $V(H_j) \subset V(J')$. Therefore, $V(\bigcup_{j \in J} H_j) \subset V(J')$.

As $|J| \geqslant b$, we have $|V(\bigcup_{j \in J} H_j)| \geqslant n - 2b$. Thus, $|V(J')| \geqslant n - 2b$, so that $|J'| \geqslant \frac{1}{2}n - b$. As $C(v) \subset C(T)$, $C(J') \subset \{c\} \cup C(T)$, $|T| = n - 1$ and $d(v) \geqslant \frac{1}{2}n + b$, there is some edge e adjacent to v and $f \in J'$ with $c(e) = c(f)$. Then, using the tree T'_f from the definition of J', the tree $T'_f + e$ satisfies the conditions in the lemma. $\square$

Iterating the above lemma, we can turn nearly-spanning trees into spanning trees. The conditions we need are to be compared with (a)–(d).

LEMMA 9.8 (Completing rainbow trees).   *Let $1 \overset{\text{POLY}}{\gg} \beta, k^{-1} \overset{\text{POLY}}{\gg} \varepsilon \overset{\text{POLY}}{\gg} n^{-1}$. In a properly coloured K_n, suppose that we have the following:*

- (i) *$S \subseteq V(K_n)$ with $|S| \geqslant n - \varepsilon n$;*
- (ii) *$T_1, \ldots, T_{n(1-8\beta)/2}$ rainbow trees with $V(T_i) \supseteq S$;*
- (iii) *for each T_i, there is a set C_L^i of $n - |T_i|$ colours outside of $C(T_i)$, where each colour in C_L^i has $\geqslant n(1-\beta)/2$ edges;*
- (iv) *for each $v \notin S$, $d_{T_i}(v) \leqslant 1$ for all i;*
- (v) *H a subgraph on S disjoint from $T_1, \ldots, T_{n(1-8\beta)/2}$ in which any set of k colours covers at least $n(1-\beta)$ vertices.*

Then, there are $n(1-8\beta)/2$ spanning rainbow trees in K_n.

Proof. Set $r = n - |S| \leqslant \varepsilon n$.

CLAIM 9.9. *Let H' be a subgraph of H with $e(H') \geqslant e(H) - 4rn$. Any set of $\beta n/2$ colours in H' covers at least $n(1 - 2\beta)$ vertices.*

Proof. Consider a set Y of $\beta n/2$ colours in H'. Since $\beta, k^{-1} \overset{\text{POLY}}{\gg} \varepsilon$ and $r \leqslant \varepsilon n$, we have $8\beta^{-1} r \cdot k \leqslant \beta n/2$ and thus Y can be partitioned into disjoint subsets $Y_1, \ldots, Y_{8\beta^{-1}r}$ of order $\geqslant k$. Since $e(H \setminus H') \leqslant 4rn$, one of these subsets Y_i has $\leqslant 4rn/(8\beta^{-1}r) = \beta n/2$ edges in $E(H) \setminus E(H')$. Since $|Y_i| \geqslant k$, by the assumptions of the lemma, Y_i covers at least $n(1 - \beta)$ vertices in H. At most βn of these might be uncovered in H' (any uncovered vertex like this must have a colour Y_i edge of $H \setminus H'$ passing through it. There are $\leqslant \beta n/2$ such edges). This shows that Y_i covers at least $n(1 - 2\beta)$ vertices in H'. $\qquad\square$

Let $T'_1, \ldots, T'_{n(1-8\beta)/2}$ be a set of edge-disjoint rainbow trees in K_n satisfying (ii)–(iv) and also

(vi) $e(T'_i \setminus T_i) \leqslant 3(|T'_i| - |T_i|)$.

Additionally, choose this family of trees so that $\sum_{i=1}^{n(1-8\beta)/2} e(T'_i)$ is as large as possible. We claim that all the rainbow trees T'_i are spanning. Suppose for the sake of contradiction that there is a vertex $v \notin V(T'_j)$ for some j. By (iii), and as $|T'_j| < n$, there is a colour $c \in C_L^j$ outside $C(T'_j)$ with $\geqslant n(1 - \beta)/2$ edges. Since T'_j satisfies (ii), we have $v \notin S$. Let G^- be the subgraph of K_n on $V(T_j) \cup \{v\}$ with the edges of T'_i deleted for all i, the edges not touching S deleted, and edges with colour in $C_L^j \setminus \{c\}$ deleted. Let $G = G^- \cup T'_j$.

Since the trees T'_i satisfy (iv), the number of trees is $n(1 - 8\beta)/2$, and $|S| \geqslant n - \varepsilon n$, we have $d_G(v) \geqslant \frac{1}{2}|G| + \beta n$. Since the trees T'_i are rainbow, $|S| \geqslant n - \varepsilon n$, and $|E_{K_n}(c)| \geqslant n(1 - \beta)/2$, we have $|E_G(c)| \geqslant \beta n$. Let $H' = H \cap G$ to get a graph with $e(H') \geqslant e(H) - \sum_{c' \in C_L^j} |E(c')| - \sum_{i=1}^{n(1-8\beta)/2} e(T'_i \setminus T_j) \geqslant e(H) - rn/2 - 3rn$ (using $|C_L^j| \leqslant r$, (vi), $|T_j| \geqslant |S|$, and $|T'_j| \leqslant n$). By Claim 9.9, any set of $\beta n/2$ colours in H' covers at least $n(1 - 2\beta)$ vertices.

Apply Lemma 9.7 to G, with the tree T'_j, vertex v, colour c, graph H', $n' = |T'_j| + 1$ and $b = \beta n$. This gives a rainbow spanning tree T''_j in G containing at most three edges outside T'_j and having $d_{T''_j}(v) = 1$. Note that the family of trees $\{T'_i : i \neq j\} \cup \{T''_j\}$ satisfies (ii)–(iv) and (vi). Indeed $S \subseteq V(T'_j) \subseteq V(T''_j)$ implies that (ii) holds. For (iii), we have that $C_L^i \setminus \{c\}$ is a set of $n - |T'_j| - 1 = n - |T''_j|$ colours outside $C(G) \cup C(T'_j) \supseteq C(T''_j)$ with $\geqslant n(1 - \beta)/2$ edges. For (iv), we have $d_{T''_i}(v) \leqslant 1$ by the property from Lemma 9.7 and $d_{T''_j}(u) \leqslant d_{T'_i}(u) \leqslant 1$ for $u \notin S \setminus \{v\}$ since there are no edges in $G \setminus T'_j$ through such u. Finally, (vi) comes from the properties from Lemma 9.7 since $e(T''_j \setminus T'_j) \leqslant 3$. Thus, we have a larger family of trees satisfying (ii)–(iv) and (vi), contradicting the maximality of the original family. $\qquad\square$

9.3. Near-decompositions into spanning rainbow trees

Now we combine everything from this section to prove the asymptotic version of the Brualdi–Hollingsworth and Kaneko–Kano–Suzuki Conjectures. We will need the following standard lemma.

LEMMA 9.10. *Every graph G with $e(G) \geqslant (1 - (\varepsilon/2)^2)n^2/2$ has an induced subgraph H with $\delta(H) \geqslant (1 - \varepsilon)n$.*

Proof. Let S be the set of vertices v in G with $d(v) \leqslant (1 - \varepsilon/2)n$. We have $2e(G) \leqslant (n - |S|)n + |S|(1 - \varepsilon/2)n$ which combined with $e(G) \geqslant (1 - (\varepsilon/2)^2)n^2/2$ gives $|S| \leqslant \varepsilon n/2$. Let $H = G \setminus S$ to get a graph with $\delta(H) \geqslant (1 - \varepsilon/2)n - |S| \geqslant (1 - \varepsilon)n$. $\qquad\square$

We will also need the following lemma about switching edges between a tree and a forest.

LEMMA 9.11. *Let T be a tree and F a forest all of whose edges touch $V(T)$. Then, there is a tree T' which contains F and is contained in $T \cup F$.*

Proof. Note that $T \cup F$ is connected since T is a tree and all edges of F touch T. Let T' be a connected subgraph of $T \cup F$ which contains F and has $e(T')$ as small as possible. If T' is acyclic, then we are done. Otherwise, T' contains a cycle C. Since F is a forest, C must contain at least one edge of T. Deleting this edge gives a smaller connected graph contradicting the minimality of $e(T')$. $\qquad\square$

By combining everything in this section with our earlier Hamiltonian decompositions, we can show that the Brualdi–Hollingsworth and Kaneko–Kano–Suzuki Conjectures hold asymptotically when the colouring on K_n is close to a 1-factorization.

LEMMA 9.12. *Let $1 \overset{\text{POLY}}{\gg} \varepsilon, \log^{-1} n \overset{\text{POLY}}{\gg} \gamma \overset{\text{POLY}}{\gg} n^{-1}$. Let K_n be properly coloured with $\geqslant (1 - \gamma)n$ colours having $\geqslant (1 - \gamma)n/2$ edges. Then, K_n has $(1 - 8\varepsilon)n/2$ edge-disjoint spanning rainbow trees.*

Proof. Choose $1 \overset{\text{POLY}}{\gg} \varepsilon, \log^{-1} n \overset{\text{POLY}}{\gg} \eta \overset{\text{POLY}}{\gg} \beta, \hat{k}^{-1} \overset{\text{POLY}}{\gg} \nu \overset{\text{POLY}}{\gg} \gamma_1 \overset{\text{POLY}}{\gg} \gamma \overset{\text{POLY}}{\gg} n^{-1}$.

Set aside small colours: Let C be the set of colours with $\geqslant (1 - \gamma)n/2$ edges in K_n. By the assumption of the lemma and $\gamma_1 \overset{\text{POLY}}{\gg} \gamma$, we have $e_{K_n}(C) \geqslant (1 - \gamma)^2 n^2/2 \geqslant (1 - (\gamma_1/2)^2)n^2/2$.

By Lemma 9.10 applied to $K_n[C]$ with $\varepsilon = \gamma_1$, there is a subgraph G of K_n with $\delta(G) \geqslant (1 - \gamma_1)n$, having only colours of C. Set $n_1 = |G| \geqslant (1 - \gamma_1)n$ and note that G is $(3\gamma_1, 1, n_1)$-typical.

Choose set S: Fix $n_2 = \lceil(1 - \nu)n\rceil$. Apply Lemma 5.2 (c) with $p = n_2/n_1$, $n' = n_1$, $\mu = 1/2$ and $\gamma' = 3\gamma_1$ in order to find a set of vertices $S \subseteq V(G)$ of order n_2 with $G[S]$ globally $(1 + 3\gamma_1)(n_2/2n_1)n_2$-bounded, and $G[S]$ $(6\gamma_1, 1, n_2)$-typical. Note that $G[S]$ is globally $(1 - 0.9\nu)n_2/2$-bounded (using $n_1 \geqslant (1 - \gamma_1)n$, $n_2 = \lceil(1 - \nu)n\rceil$ and $1 \overset{\text{POLY}}{\gg} \nu \overset{\text{POLY}}{\gg} \gamma_1$). Note that in $G[S]$ any colour of C covers at least $\geqslant (1 - \gamma)n - (n - n_2) \geqslant (1 - 2\nu)n_2$ vertices.

Set aside a pseudo-random graph H: Partition $G[S]$ into subgraphs G_1 and H with every edge placed in H independently with probability η. By Lemma 5.2 (b), G_1 is $(12\gamma_1, 1 - \eta, n_2)$-typical and globally $(1 + 6\gamma_1)(1 - \eta)(1 - 0.9\nu)n_2/2$-bounded (for the application take $p = 1 - \eta$, $\mu = (1 - 0.9\nu)/2$, $n' = n_2$, $\delta = 1$, $\gamma' = 6\gamma_1$). Since $\nu \overset{\text{POLY}}{\gg} \gamma_1$, G_1 is globally $(1 - 0.5\nu)(1 - \eta)n_2/2$-bounded. By Lemma 5.4 applied with $p = \eta$, $\varepsilon' = \varepsilon$, $k' = \hat{k}$, $\nu' = 2\nu$, H has the property that any set of $\hat{k}$ colours of C cover $\geqslant (1 - \varepsilon)n$ vertices.

Find near-decomposition of $K_n[S]$ into rainbow paths: Apply Lemma 8.27 to G_1 with $n' = n_2$, $\gamma' = 12\gamma_1$, $p = 0.5\nu$, $\delta = 1 - \eta$ in order to find $(1 - 0.5\nu)(1 - \eta)n_2$ edge-disjoint rainbow Hamiltonian paths in G_1. Using $(1 - 0.5\nu)(1 - \eta)n_2 \geqslant (1 - \varepsilon)n/2$, choose a subcollection $P_1, \ldots, P_{\lfloor(1-\varepsilon)n/2\rfloor}$ of these paths. Since G_1 is a subgraph of G, these paths only use edges with colour in C.

Add small colours into trees: Let C_L be the set of colours with $\geqslant (1 - \varepsilon)n/2$ edges in K_n. Choose $k = \max(n - 1 - |C_L|, 0)$. By assumption, we have $k \leqslant \gamma n$. Let G_2 be the subgraph of K_n consisting of edges with colour outside C_L which touch S. We claim that $e(G_2) \geqslant (1 + \eta)k\lfloor(1 - \varepsilon)n/2\rfloor$. When $k = 0$, this is obvious. Otherwise since $\delta(K_n) = n - 1$ and K_n is properly coloured, we have $e(G_2) \geqslant \frac{1}{2}\sum_{v \in S} d_{C(K_n)\setminus C_L}(v) \geqslant |S|(\delta(K_n) - |C_L|)/2 = k\lceil(1 - \nu)n\rceil/2 \geqslant (1 + \eta)k(1 - \varepsilon)n/2$. By definition of C_L, the graph G_2 is globally $\lfloor(1 - \varepsilon)n/2\rfloor$-bounded. Apply Lemma 9.5 to G_2 with $m = \lfloor(1 - \varepsilon)n/2\rfloor$, $\varepsilon' = 1.01\nu$, $\beta' = \eta$, $S = S$. This gives us edge-disjoint rainbow forests $F_1, \ldots, F_{\lfloor(1-\varepsilon)n/2\rfloor}$ of size k in G_2.

Apply Lemma 9.11 for $i = 1, \ldots, \lfloor(1-\varepsilon)n/2\rfloor$ to P_i and F_i in order to find a rainbow tree T_i containing F_i and contained in $P_i \cup F_i$ (T_i is rainbow since P_i and F_i are colour-disjoint which happens because $C(P_i) \subseteq C \subseteq C_L$ and $C(F_i) \cap C_L = \emptyset$). In particular, each T_i contains k edges outside C_L (the edges of F_i). Since $k \geqslant n - 1 - |C_L|$, this implies that each T_i avoids $k + |C_L| - e(T_i) \geqslant n - 1 - e(T_i)$ colours of C_L, each of which has $\geqslant (1-\varepsilon)n/2$ edges in K_n. Additionally, from Lemma 9.5, we have that for every vertex $v \notin S$, either $v \in T_i$ for all i or $d_{T_i}(v) \leqslant 1$ for all i. Let $S' = S \cup \{v \notin S : v \in T_i \text{ for each } i\}$ and note that $|S'| \geqslant |S| = \lceil(1-\nu)n\rceil$. Now for each i and $v \notin S'$, we have $d_{T_i}(v) \leqslant 1$ and also $S' \subseteq V(T_i)$.

Make trees spanning: Observe that H is disjoint from G_1 and G_2. (The former holds by construction of G_1. The latter by $C(H) \subseteq C \subseteq C_L$ and $C(G_2) \cap C_L = \emptyset$), and hence H is disjoint from the trees $T_1, \ldots, T_{\lfloor(1-\varepsilon)n/2\rfloor}$. Apply Lemma 9.8 with $S = S'$, trees $T_1, \ldots, T_{(1-8\varepsilon)n/2}$, $H = H$, $\beta' = \varepsilon$, $k' = \hat{k}$ and $\varepsilon' = \nu$ in order to find $(1-8\varepsilon)n/2$ edge-disjoint spanning rainbow trees in K_n, where we have used that $\varepsilon, 1/\hat{k} \overset{\text{\tiny POLY}}{\gg} \nu$. $\qquad\square$

Combining the above with our earlier result about Hamiltonian decompositions, we prove that the Brualdi–Hollingsworth and Kaneko–Kano–Suzuki Conjectures hold asymptotically.

THEOREM 9.13. *Let* $1 \overset{\text{\tiny POLY}}{\gg} \varepsilon \overset{\text{\tiny POLY}}{\gg} n^{-1}$. *Every properly coloured* K_n *has* $(1-\varepsilon)n/2$ *edge-disjoint spanning rainbow trees.*

Proof. Fix $1 \overset{\text{\tiny POLY}}{\gg} \varepsilon, \log^{-1} n \overset{\text{\tiny POLY}}{\gg} \gamma \overset{\text{\tiny POLY}}{\gg} n^{-1}$. If K_n has $\geqslant (1-\gamma)n$ colours having $\geqslant (1-\gamma)n/2$ edges, then the theorem follows from Lemma 9.12. Otherwise, K_n has $\leqslant (1-\gamma)n$ colours having $\geqslant (1-\gamma)n/2$ edges, and the theorem follows from Lemma 8.29. $\qquad\square$

10. *Concluding remarks*

There are various other areas in which our results have implications. We mention some of them here.

- Constantine made the following generalization of the Brualdi–Hollingsworth Conjecture.

CONJECTURE 10.1 (Constantine [**21**]). Every properly $(2n-1)$-coloured K_{2n} can be decomposed into edge-disjoint rainbow spanning trees which are all isomorphic to each other.

The best known result about this is due to the second and third author [**44**] who showed that it is possible to find $10^{-12}n$ edge-disjoint rainbow copies of some particular tree. While we did not do this, our results still have implications for Constantine's Conjecture. In particular, Corollary 1.11 is relevant — it shows that under the assumptions of Constantine's Conjecture, we can nearly-decompose the graph into nearly-spanning rainbow paths.

Additionally, we expect that the methods in this paper can be generalized to prove the true asymptotic version of Constantine's Conjecture, that is, to find $(1 - o(1))n$ edge-disjoint isomorphic spanning rainbow trees under the assumption of the theorem. We think this is plausible as the trees we find in the proof of Theorem 1.14 are all quite similar to each other — they are all built from a length $(1 - o(1))n$ path by making $o(n)$ modifications. It seems likely that, with some additional ideas, the modifications can be controlled in order to give a copy of the same tree.

- Note a parallel between Theorem 1.14 and Lemma 9.5 — both of these results give a near-decomposition of a graph into forests of the same size. We wonder if there is a common generalization of these results.

CONJECTURE 10.2.   For $\varepsilon > 0$, there exists an $m \in \mathbb{N}$ so that the following holds for all k. Every properly coloured, globally m-bounded graph G with km edges has $\geqslant (1 - \varepsilon)m$ edge-disjoint rainbow forests of order k.

Currently, there are two extremes of this conjecture which are known to be true. Theorem 1.14 shows that it holds when $|G| = 2m$ and $k = |G| - 1$. Lemma 9.5 shows that it holds when $k = o(|G|)$. It would be interesting to prove or disprove it in general.

• Recall that the randomized rainbow matching M in Lemma 4.6 behaves like a uniformly random perfect matching in a sense that any edge of G ends up in M with (approximately) at least the expected probability $\mathrm{d}(G)^{-1}$. One can ask whether more can be proven, that is, whether M shares more features with a uniformly random perfect matching. This is indeed the case — Lemma 4.6 can easily be strengthened to say more about the matching M. For example, with some work the following can be added to that lemma.

$$\mathbb{P}(e \in E(M)) = (1 \pm p)\frac{1}{\delta n} \quad \text{ for each } e \in E(G).$$

$$\mathbb{P}(e, f \in E(M)) = (1 \pm p)\frac{1}{\delta^2 n^2} \quad \text{ for each } e \neq f \in E(G).$$

$$\mathbb{P}(v \notin V(H)) = (1 \pm p)p \quad \text{ for each } v \in V(G).$$

The randomness of the matching produced in Lemma 4.6 may have applications in future work.

• Note that some of our results (particularly Lemma 8.2) are about graphs which may not be properly coloured, but are only locally n^ϵ-bounded. It is natural to ask whether our other theorems can be proved with 'proper colouring' replaced by 'local boundedness', or perhaps even with the proper colouring assumption removed entirely. Some results in this direction were recently obtained by Kim, Kühn, Kupavskii and Osthus in [40] (see note below).

It would be extremely interesting to prove new results about spanning rainbow structures in graphs with *no local boundedness assumptions at all*. For example in [43], the second and third authors asked whether every globally $(1 - o(1))n$ bounded $K_{n,n}$ has a perfect rainbow matching. If true, this would be a natural weakening of the recently disproved Stein's Equi-n-Square Conjecture (see [43]).

Note added in proof

The results of Theorem 1.2 and its corollaries were presented at the 'Workshop on Probabilistic and Extremal Combinatorics' in Harvard on 07 February 2018 (see [42]). After the presentation, we learned from Keevash and Yepremyan that they also found a proof of the Akbari–Alipour Conjecture (Conjecture 1.3) for large n (see [39]).

Also after hearing our Theorem 1.2 at the workshop, Kim, Kühn, Kupavskii, and Osthus published the preprint [40] on 22 May 2018. In this paper, they proved (amongst others) that every coloured $K_{n,n}$ which is globally $(1 - o(1))n$-bounded and locally $o(n/\log^2 n)$-bounded has $(1 - o(1))n$ edge-disjoint rainbow perfect matchings. This is on one hand stronger than Theorem 1.2 since it also works for locally bounded colorings, but it is also weaker since it requires all (rather than just few) colors to have size less than $(1 - o(1))n$. In particular, it does not imply the Akbari–Alipour Conjecture or our results on multiplication tables of groups. Independently from our work, Kim, Kühn, Kupavskii and Osthus also proved results similar to our Theorem 1.10 about decompositions into rainbow Hamiltonian cycles (that are both stronger and weaker as we explain above). The main focus of their work is quite different from ours and they deduce their result from a general theorem about rainbow F-factors for arbitrary graphs F.

Note added to proof

Recently a conjecture of Constantine (Conjecture 10.1) was proved for large n by Glock, Kühn, Montgomery and Osthus [**28**].

References

1. S. Akbari and A. Alipour, 'Transversals and multicolored matchings', *J. Combin. Des.* 12 (2004) 325–332.

2. S. Akbari and A. Alipour, 'Multicolored trees in complete graphs', *J. Graph Theory* 54 (2007) 221–232.

3. S. Akbari, O. Etesami, H. Mahini and M. Mahmoody, 'On rainbow cycles in edge colored complete graphs', *Australas. J. Combin.* 37 (2007) 33.

4. N. Alon, 'Additive Latin transversals', *Israel J. Math.* 117 (2000) 125–130.

5. N. Alon, M. Krivelevich and B. Sudakov, 'List coloring of random and pseudo-random graphs', *Combinatorica* 19 (1999) 453–472.

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

7. N. Alon and J. Spencer, *The probabilistic method*, 4th edn (John Wiley & Sons, Hoboken, NJ, 2004).

8. N. Alon, J. Spencer and P. Tetali, 'Covering with Latin transversals', *Discrete Appl. Math.* 57 (1995) 1–10.

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

10. B. Arsovski, 'A proof of Snevily's conjecture', *Israel J. Math.* 182 (2011) 505–508.

11. R. C. Baker and G. Harman, 'The difference between consecutive primes', *Proc. Lond. Math. Soc.* (3) 3 (1996) 261–280.

12. J. Balogh, H. Liu and R. Montgomery, 'Rainbow spanning trees in properly coloured complete graphs', *Discrete Appl. Math.* 247 (2018) 97–101.

13. J. Barát and Z. L. Nagy, 'Transversals in generalized Latin squares', *Ars Math. Contemp.* 16 (2018) 39–47.

14. R. Bose, S. Shrikhande and E. Parker, 'Further results on the construction of mutually orthogonal Latin squares and the falsity of Euler's conjecture', *Canad. J. Math.* 12 (1960) 189–203.

15. A. Brouwer, A. de Vries and R. Wieringa, 'A lower bound for the length of partial transversals in a Latin square', *Nieuw Arch. Wiskd.* (5) 26 (1978) 330.

16. R. A. Brualdi and S. Hollingsworth, 'Multicolored trees in complete graphs', *J. Combin. Theory Ser. B* 68 (1996) 310–313.

17. R. A. Brualdi and H. J. Ryser, *Combinatorial matrix theory* (Cambridge University Press, Cambridge, 1991).

18. J. M. Carraher, S. G. Hartke and P. Horn, 'Edge-disjoint rainbow spanning trees in complete graphs', *European J. Combin.* 57 (2016) 71–84.

19. H. Chen and X. Li, 'Long rainbow path in properly edge-colored complete graphs', Preprint, 2015, arXiv:1503.04516.

20. D. Christofides, D. Kühn and D. Osthus, 'Edge-disjoint Hamilton cycles in graphs', *J. Combin. Theory Ser. B* 102 (2012) 1035–1060.

21. G. M. Constantine, 'Multicolored parallelisms of isomorphic spanning trees', *Discrete Math. Theor. Comput. Sci.* 5 (2002) 121–126.

22. S. Dasgupta, G. Károlyi, O. Serra and B. Szegedy, 'Transversals of additive Latin squares', *Israel J. Math.* 126 (2001) 17–28.

23. P. Erdős and J. Spencer, 'Lopsided Lovász local lemma and Latin transversals', *Discrete Appl. Math.* 30 (1991) 151–154.

24. L. Euler, 'Recherches sur une nouvelle espéce de quarrés magiques', *Verh. Zeeuwsch. Gennot. Weten. Vliss.* 9 (1782) 85–239.

25. J. Fournier, 'Sharpness in young's inequality for convolution', *Pacific J. Math.* 72 (1977) 383–397.

26. H.-L. Fu, Y.-H. Lo, K. Perry and C. Rodger, 'On the number of rainbow spanning trees in edge-colored complete graphs', *Discrete Math.* 341 (2018) 2343–2352.

27. H. Gebauer and F. Mousset, 'On rainbow cycles and paths', Preprint, 2012, arXiv:1207.0840.

28. S. Glock, D. Kühn, R. Montgomery, D. Osthus, 'Decompositions into isomorphic rainbow spanning trees', Preprint, 2019, arXiv:1903.04262.

29. B. Green and A. Wigderson, 'Lecture notes for the 22nd McGill invitational workshop on computational complexity', 2010, https://www.cs.mcgill.ca/denis/additive-lectures-v2.pdf.

30. C. Greenhill, M. Isaev, M. Kwan and B. D. McKay, 'The average number of spanning trees in sparse graphs with given degrees', *European J. Combin.* 63 (2017) 6–25.

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

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

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

34. M. Hall and L. Paige, 'Complete mappings of finite groups', *Pacific J. Math.* 5 (1955) 541–549.

35. P. Hatami and P. W. Shor, 'A lower bound for the length of a partial transversal in a Latin square', *J. Combin. Theory Ser. A* 115 (2008) 1103–1113.

36. P. Horn, 'Rainbow spanning trees in complete graphs colored by one-factorizations', *J. Graph Theory* 87 (2018) 333–346.

37. A. Kaneko, M. Kano and K. Suzuki, 'Three edge disjoint multicolored spanning trees in complete graphs', Preprint, 2003.

38. A. Keedwell and J. Dénes, *Latin squares and their applications* (Elsevier Science, Amsterdam, 2015).

39. P. Keevash and L. Yepremyan, 'On the number of symbols that forces a transversal', Preprint, 2018, arXiv:1805.10911.

40. J. Kim, D. Kühn, A. Kupavskii and D. Osthus, 'Rainbow structures in locally bounded colourings of graphs', 2018, arXiv:1805.08424.

41. O. Ore, *Theory of graphs*, vol. 38 (American Mathematical Society, Providence, RI, 1962).

42. A. Pokrovskiy, 'Latin squares and rainbow subgraphs', 2018, https://people.math.ethz.ch/~palexey/Slides/RainbowTreesPresentation.pdf.

43. A. Pokrovskiy and B. Sudakov, 'A counterexample to Stein's equi-n-square conjecture', *Proc. Amer. Math. Soc.* 2018, to appear.

44. A. Pokrovskiy and B. Sudakov, 'Linearly many rainbow trees in properly edge-coloured complete graphs', *J. Combin. Theory Ser. B* 132 (2018) 134–156.

45. H. Ryser, 'Neuere probleme der kombinatorik', *Vorträge über Kombinatorik* (Mathematisches Forschungsinstitut, Oberwolfach, 1967) 69–91.

46. H. J. Ryser, *Combinatorial mathematics*, vol. 14 (Mathematical Association of America, Washington, DC, 1963). (distributed by Wiley, New York)

47. W. Sierpinski, *Elementary theory of numbers: second English edition*, vol. 31 (ed. A. Schinzel; Elsevier, Amsterdam, 1988).

48. H. S. Snevily, 'The Cayley addition table of Z_n', *Amer. Math. Monthly* 106 (1999) 584–585.

49. S. K. Stein, 'Transversals of Latin squares and their generalizations', *Pacific J. Math.* 59 (1975) 567–575.

50. S. Wilcox, 'Reduction of the Hall–Paige conjecture to sporadic simple groups', *J. Algebra* 321 (2009) 1407–1428.

51. D. Woolbright, 'An $n \times n$ Latin square has a transversal with at least $n - \sqrt{n}$ distinct symbols', *J. Combin. Theory Ser. A* 24 (1978) 235–237.

Richard Montgomery
School of Mathematics
University of Birmingham
Edgbaston
Birmingham B15 2TT
United Kingdom

r.h.montgomery@bham.ac.uk

Alexey Pokrovskiy
Department of Mathematics
Economics and Statistics
Birkbeck, University of London
London WC1E 7HX
United Kingdom

dr.alexey.pokrovskiy@gmail.com

Benjamin Sudakov
Department of Mathematics
ETH
8092 Zurich
Switzerland

benjamin.sudakov@math.ethz.ch