Essentially tight bounds for rainbow cycles in proper edge-colourings

Noga Alon[*]   Matija Bucić[†]   Lisa Sauermann[‡]   Dmitrii Zakharov[§]   Or Zamir[¶]

Abstract

An edge-coloured graph is said to be *rainbow* if no colour appears more than once. Extremal problems involving rainbow objects have been a focus of much research over the last decade as they capture the essence of a number of interesting problems in a variety of areas. A particularly intensively studied question due to Keevash, Mubayi, Sudakov and Verstraëte from 2007 asks for the maximum possible average degree of a properly edge-coloured graph on n vertices without a rainbow cycle. Improving upon a series of earlier bounds, Tomon proved an upper bound of $(\log n)^{2+o(1)}$ for this question. Very recently, Janzer–Sudakov and Kim–Lee–Liu–Tran independently removed the $o(1)$ term in Tomon's bound, showing a bound of $O(\log^2 n)$. We prove an upper bound of $(\log n)^{1+o(1)}$ for this maximum possible average degree when there is no rainbow cycle. Our result is tight up to the $o(1)$ term, and so it essentially resolves this question.

In addition, we observe a connection between this problem and several questions in additive number theory, allowing us to extend existing results on these questions for abelian groups to the case of non-abelian groups.

1   Introduction

Roughly speaking, a general extremal question asks for the maximum or minimum possible value of a parameter of a large structure which suffices to guarantee a certain type of substructure. Such questions capture a vast number of interesting problems appearing naturally in a variety of areas in mathematics and beyond. The area of extremal graph theory is concerned with such questions when the structures of interest are graphs, and graphs can often also be used to encode the properties of other types of mathematical objects relevant in the context of extremal question. It is very common to consider colours on the edges of a graph, for example, this is the setting in the whole area of (graph) Ramsey theory. Colours on the edges can also be very important in order to keep track of additional data when using graphs to encode other mathematical objects and structures. For example, the papers [54] and [55] approached some questions in discrete geometry by considering edge-coloured graphs (given n points in the plane, they considered graphs whose vertices correspond to the given points and whose edges correspond to segments between the given points of a certain length and direction, with edges of the same colour corresponding to parallel segments). Another example is the classical Ryser–Brualdi–Stein Conjecture about Latin squares, which can be phrased as an extremal problem about edge-coloured graphs, see [40] for more details (and also more examples). In all of these examples, it is relevant to consider *proper*[1] edge-colourings.

Another interesting notion in the context of edge-colourings is being rainbow, with some surprising applications ranging from classical decomposition conjectures such as Ringel's conjecture [57] and the classical Hall-Paige conjecture in group theory (see e.g. [59]) to problems in bioinformatics [61]. An edge-coloured graph is said to be *rainbow* if no colour appears on more than one edge. There has been a lot of research on questions involving rainbow objects, with a paper of Keevash, Mubayi, Sudakov and Verstraëte [39] from 2007 being particularly influential. They raised

[1]An edge-colouring of a graph is called proper if at every vertex all incident edges have distinct colours.

rainbow analogues of a number of the perhaps most classical questions in extremal graph theory, and a particularly tantalising question they asked is the following: How many edges are needed in order to guarantee that in any proper edge-colouring of an n-vertex graph one can find a rainbow cycle? The non-rainbow version of this is the simple, yet often useful, fact that one needs n edges in an n-vertex graph in order to guarantee a cycle (this is optimal as evidenced by any n-vertex tree). In other words, if the average degree of a graph is at least 2, then the graph must contain a cycle. The rainbow version of this question, however, seems substantially more difficult and has resisted numerous attempts over the years. It is known, already due to [39], that the answer for the rainbow version is larger, and in particular that for an n-vertex graph an average degree of at least $\Omega(\log n)$ is needed to guarantee a rainbow cycle. This remains the best-known lower bound (up to the implicit constant factor). We will discuss the best-known construction in Section 9 and show how it arises in a perhaps more natural way. The upper bound has been the subject of a series of improvements, starting with an initial upper bound of Keevash, Mubayi, Sudakov and Verstraëte [39], who showed that any properly edge-coloured graph on n vertices without a rainbow cycle has average degree at most $O(n^{1/3})$ (i.e. it has at most $O(n^{4/3})$ edges). The first improvement of this bound was due to Das, Lee and Sudakov [18], showing an upper bound of the form $e^{(\log n)^{1/2+o(1)}}$ for the average degree. This was in turn improved by Janzer [35] to $O(\log^4 n)$ and subsequently to a bound of the form $(\log n)^{2+o(1)}$ by Tomon [73]. Very recently, the $o(1)$ term in the exponent was removed independently by Janzer and Sudakov [36] and by Kim, Lee, Liu and Tran [41], showing the current state-of-the-art bound of $O(\log^2 n)$. Our main result is an essentially tight answer to the question of Keevash, Mubayi, Sudakov and Verstraëte, determining the average degree needed in a properly edge-coloured graph on n vertices in order to guarantee a rainbow cycle up to lower order terms.

Theorem 1.1. *There exists a constant $C > 0$ such that every properly edge-coloured graph on $n \geq 3$ vertices with average degree at least $C \cdot \log n \cdot \log \log n$ contains a rainbow cycle.*

Our result is tight up to the $\log \log n$ factor. In particular, our result shows that every properly edge-coloured graph on n vertices without a rainbow cycle must have average degree at most $(\log n)^{1+o(1)}$, which is tight up to the $o(1)$-term. Our result can also be rephrased as saying that for a certain absolute constant $C > 0$, every properly edge-coloured graph on n vertices with at least $C \cdot n \cdot \log n \cdot \log \log n$ edges contains a rainbow cycle. The maximal possible number of edges in a properly edge-coloured graph on n vertices without a rainbow cycle may be viewed as the *rainbow Turán number* of the family of all cycles. Our result shows an upper bound of $O(n \cdot \log n \cdot \log \log n)$ for this rainbow Turán number, which is optimal up to the $\log \log n$ factor.

The previous work on this question can be split into two fairly different proof approaches. The previous state-of-the-art results [36, 41] both follow the "homomorphism counting" approach pioneered by Janzer in [35], which has found plenty of other applications, see e.g. [34, 37, 46]. The other approach of "passing to an expander", first applied to this problem by Das, Lee and Sudakov [18], has also found a number of applications and was significantly refined by Tomon in [73], see [13] and references therein for concrete examples. Our argument falls in the latter camp, and in the first step of our proof we pass to a subgraph which is a robust sublinear expander (see Section 3 for a definition) of essentially the same average degree as our initial graph. We note that our expander subgraph does not have quantitatively stronger expansion properties compared to that of [73], and our method of finding it is actually very similar to that of [73]. In fact, in general, finding an expander subgraph with substantially better quantitative expansion properties is impossible due to constructions obtained in [58]. The crucial difference in our argument compared to [73] is that in our setup the expansion properties are much more robust, in a sense that was first introduced in [29] and subsequently used and further developed in [13]. We introduce various new tools for working with such expanders, which may also be useful for other applications. The main part of our argument for finding a rainbow cycle in such an expander subgraph is a carefully designed random process that we analyse relying on the (robust) expansion properties and delicate multiple exposure arguments. In particular, a key point of our argument is a setup where we condition on the current state of the random process while keeping both the past and the future sufficiently random. At every step, we then show that (with high likelihood) we have significant expansion of a certain "rainbow-reachable" vertex set when going to the next step, or we had significant expansion when compared to the previous step. Interestingly, we cannot guarantee significant expansion of this "rainbow-reachable" vertex set for every step, but only going forwards *or* going backwards from any given step. Then, when comparing every other step, we have significant expansion and can complete our argument. See Section 2 for a more detailed proof sketch.

The bound obtained in Theorem 1.1 represents a hard limit for our approach in several ways. We will discuss this, as well as some circumstantial evidence that the $\log \log n$ factor in Theorem 1.1 might actually be necessary, in Section 10.

We also obtain a result bridging the gap between the answers in the rainbow and non-rainbow versions of the

problem. A cycle in a properly edge-coloured graph is said to be *r-almost rainbow* if no colour appears on more than r edges of the cycle. So, for example, an 1-almost rainbow cycle is the same as a rainbow cycle, and any cycle of length ℓ is trivially ℓ-almost rainbow. Somewhat surprisingly, the behaviour of the problem is already different from the rainbow case, if we allow colours to be used twice. From that point on, as we allow colours to be used more and more times, the answer gradually changes up to the point where we allow each colour to be used $\Theta(\log n)$ times. At this point already a linear number of edges suffices, like in the non-rainbow version of the problem. The precise behaviour is captured in the following result. Here, we denote by $\mathrm{Tr}_r(n, \mathcal{C})$ the r-almost rainbow Turán number for the family $\mathcal{C}$ of all cycles, namely the maximal possible number of edges in a properly edge-coloured n-vertex graph without an r-almost rainbow cycle. With this notation, Theorem 1.1 can be restated as $\mathrm{Tr}_1(n, \mathcal{C}) \leq O(n \log n \log \log n)$.

Proposition 1.2. *For any integers n and r with $2 \leq r < \frac{1}{10}\log n$, we have*

$$\mathrm{Tr}_r(n, \mathcal{C}) = \Theta\left(n \cdot \frac{(\log n)/r}{\log((\log n)/r)}\right).$$

This proposition determines the r-almost rainbow Turán number $\mathrm{Tr}_r(n, \mathcal{C})$ up to constant factors whenever $2 \leq r < \frac{1}{10}\log n$. Note that this automatically implies $\mathrm{Tr}_r(n, \mathcal{C}) = \Theta(n)$ for all $r \geq \frac{1}{10}\log n$. Indeed, for sufficiently large n we have $\mathrm{Tr}_r(n, \mathcal{C}) \leq \mathrm{Tr}_{\lfloor (\log n)/10\rfloor}(n, \mathcal{C}) \leq \Theta(n)$ whenever $r \geq \frac{1}{10}\log n$. In the opposite direction, for any n and r, it is clear that $\mathrm{Tr}_r(n, \mathcal{C}) \geq n - 1$ by considering an n-vertex tree. Hence, Proposition 1.2 actually determines the r-almost rainbow Turán number $\mathrm{Tr}_r(n, \mathcal{C})$ up to constant factors for any integers n and r with $r \geq 2$.

We note that several competing notions of almost rainbow cycles have been considered over the years. In particular, Janzer and Sudakov [36] consider what happens if only a small proportion of edges on a cycle are allowed to use non-unique colours. We note that Proposition 1.2 can not be improved in this direction, not even if we only insist on having at least one colour appearing only once, see the discussion at the end of Section 8.

1.1 Applications in additive number theory

Given an abelian group $(G, +)$, a subset $S \subseteq G$ is said to be *dissociated* if there is no solution to $\varepsilon_1 g_1 + \cdots + \varepsilon_m g_m = 0$ with distinct $g_1, \ldots, g_m \in G$ as well as $\varepsilon_1, \ldots, \varepsilon_m \in \{-1, 1\}$ and $m \geq 1$. One can view this as saying that there is no non-trivial "linear" relation between elements in S with coefficients ± 1. With this in mind, dissociated sets play to some extent a similar role in arbitrary abelian groups as linearly independent sets play in vector spaces. Taking the analogy even further, a maximal dissociated subset of a set $A \subseteq G$ spans the set A, using only coefficients 0 or ± 1, see [65]. In addition, all maximal dissociated subsets of any set $A \subseteq G$ have almost the same size, as shown in [49]. It is therefore not surprising that controlling the size of a maximal dissociated subset is relevant for various problems in additive number theory, see [7, 8, 9, 10, 11, 16, 45, 64, 66, 70]. Dissociated sets also play an important role in Harmonic analysis, see e.g. [44, 63]. The above-mentioned properties lead naturally to a notion of dimension of a finite subset $A \subseteq G$ of an abelian group G. More specifically, in [68] Schoen and Skhredov define the *additive dimension* $\dim A$ of A to be the size of the largest dissociated subset of A. The following theorem of Sanders [65] (see also [68, Theorem 1]) shows that a subset A with small doubling $A + A := \{a_1 + a_2 \mid a_1, a_2 \in A\}$ has small additive dimension and hence needs to be quite structured (namely, it can be "well approximated" by a "subspace" generated by its maximal dissociated subset).

Theorem 1.3. *Let G be an abelian group and let $A \subseteq G$ be a finite subset of size $|A| \geq 2$. If $|A + A| \leq K|A|$ for some positive integer K, then we have $\dim A \leq O(K \log |A|)$.*

The study of the structure of sets with small doubling traces back to a celebrated result of Freiman [22] from 1964 and has been the subject of extensive study ever since, see e.g. [25, 67] and references therein. There has been a lot of work extending results about sets with small doubling to non-abelian groups, see a foundational paper of Tao [72] fuelling the work in this direction, and see [12, 21, 24, 30, 31, 62] for some further examples. Given this, it is very natural to ask whether the result of Sanders stated in Theorem 1.3 can be extended to non-abelian groups. As a start, the above definition of additive dimension extends essentially verbatim to the non-abelian case, although we repeat it here in order to switch to the multiplicative notation (to highlight the fact we are working with general, not necessarily abelian groups).

Definition 1.4. *Let $(G, \cdot)$ be a group with identity element e. A subset $S \subseteq G$ is said to be* dissociated *if there is no solution to $g_1^{\varepsilon_1} \cdots g_m^{\varepsilon_m} = e$ with distinct $g_1, \ldots, g_m \in S$ as well as $\varepsilon_1, \ldots, \varepsilon_m \in \{-1, 1\}$ and $m \geq 1$. For a finite subset $A \subseteq G$, the* additive dimension $\dim A$ *is defined as the maximum cardinality of a dissociated subset of A.*

As an easy corollary of Theorem 1.1, we obtain a version of Sanders' result for *all* groups (not just abelian groups) with slightly weaker bounds.

Theorem 1.5. *Let $(G, \cdot)$ be a group and let $A \subseteq G$ be a finite subset of size $|A| > 2$. If $|A \cdot A| \leq K|A|$ for some positive integer K, then we have $\dim A \leq O(K \log |A| \log \log |A|)$.*

Here, the doubling $A \cdot A$ is defined by $A \cdot A := \{a_1 \cdot a_2 \mid a_1, a_2 \in A\}$. We do not know whether the $\log \log |A|$ term is necessary in the theorem above, but we do know the situation in the non-abelian case is at least slightly different compared to the abelian case. Namely, in the case of finite abelian groups, there is a stronger bound compared to Theorem 1.3 if we ask for the dimension of the whole group, namely $\dim G \leq \lceil \log_2 |G| \rceil$ (see [2]). On the other hand, this bound does not hold in the non-abelian case. Indeed, we have $\dim \mathcal{S}_3^k \geq \frac{3}{\log_2 6} \cdot \log_2 |\mathcal{S}_3^k|$ (see Observation 9.3), where $\mathcal{S}_3^k$ is the k-fold direct product of $\mathcal{S}_3$, the symmetric group on three elements. For finite non-abelian groups G, an upper bound for the additive dimension $\dim G$ can be obtained from Theorem 1.5. More precisely, Theorem 1.5 implies $\dim G \leq O(\log |G| \log \log |G|)$ for any finite group G of size $|G| \geq 3$, which is only slightly weaker than the bound in the abelian case.

We note that the additive dimension $\dim G$ of the whole (finite) group G agrees with a natural variant of the classical Davenport constant of a group, called the plus-minus weighted Davenport constant and denoted by $D_{\pm}(G)$. There has been a significant amount of work on plus-minus weighted Davenport constants, see [2, 53] or the survey [1], and see also [26] for several arithmetic interpretations. Our Theorem 1.5 implies $D_{\pm}(G) = \dim G \leq O(\log |G| \log \log |G|)$ for any finite group G of size $|G| \geq 3$, almost matching the upper bound $D_{\pm}(G) \leq \lceil \log_2 |G| \rceil$ in the abelian case.

While our manuscript was in its final stages we heard of some exciting progress by Conlon, Fox, Pham, and Yepremyan [17] towards a conjecture of Alon [3] concerning the existence of good Ramsey Cayley graphs for any group. A key step in their argument is a lemma counting the number of subsets A of a given group with the property that $|A \cdot A^{-1}| \leq K|A|$. An asymmetric version of Theorem 1.5 (see Theorem 9.2) gives a bound on the dimension of such sets which in turn implies a bound on the number of such subsets, which is only slightly weaker than the lemma in [17].

Remark. While our manuscript was under review, our main result Theorem 1.1 has been used as a key ingredient in a remarkable paper by Alrabiah and Guruswami [6] showing almost-tight bounds for the dimension of binary linear codes that are three-query locally correctable against a fixed fraction of errors. We refer to [6] for the relevant definitions and a detailed history of this well-known problem in coding theory. Alrabiah and Guruswami also used our result to obtain improved bounds for the case of r-query locally correctable codes for odd $r \geq 5$. They drew attention to the fact that the previously known bounds in the setting of Theorem 1.1 would not have led to any non-trivial bounds in this latter coding theory application at all.

Notation. For a graph G, we denote the set of edges by $E(G)$ and the set of vertices by $V(G)$. For a vertex subset $U \subseteq V(G)$, we denote the induced subgraph on U by $G[U]$. A proper edge-colouring of G with a colour palette $\mathcal{C}$ is an assignment $\gamma : E(G) \to \mathcal{C}$ such that for any two adjacent edges $e \neq e'$ we have $\gamma(e) \neq \gamma(e')$. For two adjacent vertices $v, v' \in V$, we slightly abuse notation to write $\gamma(v, v')$ for the colour of the edge vv'.

We denote the average degree of a graph G by $d(G)$. The neighbourhood of a vertex subset $U \subseteq V(G)$ is the set of vertices in $V(G) \setminus U$ that are adjacent to a vertex in U. For a subset $F \subseteq E(G)$ and a vertex $v \in V(G)$, we denote by $\deg_F(v)$ the number of edges in F incident to v (i.e. the degree of v in the subgraph described by the edges in F). Furthermore, $N_{G-F}(U)$ denotes the neighbourhood of a vertex subset $U \subseteq V(G)$ in the graph $G - F$ resulting from G when deleting the edges in F.

All our logarithms are to the base e unless otherwise specified. We make use of the standard asymptotic notation, namely, for functions $f : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ and $g : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$, we write $f = O(g)$ or $g = \Omega(f)$ if there exists an absolute constant C such that $f(x) \leq Cg(x)$ for all $x \in \mathbb{R}_{>0}$. We write $f = \Theta(g)$ if we have both $f = O(g)$ and $g = O(f)$. We also write $f = o(g)$ if $f(x)/g(x) \to 0$ as $x \to \infty$.

2 Proof Overview

As mentioned above, the first step of our proof of Theorem 1.1 is to pass to a well-connected subgraph of G with almost the same average degree as G. To make this precise, we work with a notion of "robust sublinear expansion", we discuss the rich history of this notion in Section 3. In such a robust sublinear expander graph with n vertices, for every non-empty vertex subset U of size at most $n^{1-\varepsilon}$ (for any $0 \leq \varepsilon \leq 1$) the neighbourhood of U has size at least $(\varepsilon/3) \cdot |U|$. Even more so, this remains true even when deleting up to $(\varepsilon/3) \cdot d(G) \cdot |U|$ edges, where $d(G)$

denotes the average degree of the graph (so the expansion property is in a sense quite robust). We prove that in any graph G, one can find a subgraph of similar average degree which is a robust sublinear expander in this sense (for more details, see Section 3). In order to prove Theorem 1.1, we can first choose such a subgraph of G with the aim of finding a rainbow cycle in this subgraph. This allows us to reduce Theorem 1.1 to the case that our graph is a robust sublinear expander.

Our approach to finding a rainbow cycle uses the following simple idea. Suppose we split the colour palette $\mathcal{C}$ of the proper edge-colouring of G into two disjoint subsets $\mathcal{C}_1$ and $\mathcal{C}_2$, say randomly (assigning each colour to $\mathcal{C}_1$ or $\mathcal{C}_2$ independently uniformly at random). If there exist two distinct vertices x and y, such that y can be reached from x with a rainbow walk with colours in $\mathcal{C}_1$ and also with a rainbow walk with colours in $\mathcal{C}_2$, then these two rainbow walks use disjoint sets of colours. Hence we can combine them to obtain a rainbow closed walk and hence a rainbow cycle. Here, by a rainbow walk, we mean a walk on which all edges are coloured using distinct colours.

Our approach to finding such vertices x and y is rather naive: We fix an arbitrary vertex x of the graph G, and look at the set $\mathrm{U}(\mathcal{C}_1)$ of all vertices y that can be reached by rainbow walks starting at x with colours in $\mathcal{C}_1$. We similarly look at the set $\mathrm{U}(\mathcal{C}_2)$ of all vertices y that can be reached by rainbow walks from x with colours in $\mathcal{C}_2$. If both of the sets $\mathrm{U}(\mathcal{C}_1)$ and $\mathrm{U}(\mathcal{C}_2)$ are large enough (more precisely, if they each have size larger than $\frac{n+1}{2}$), they must overlap in some vertex $y \neq x$. So our goal is to prove that with probability at least, say, $\frac{2}{3}$ we have $|\mathrm{U}(\mathcal{C}_1)| > \frac{n+1}{2}$. Then we analogously obtain $|\mathrm{U}(\mathcal{C}_2)| > \frac{n+1}{2}$ with probability at least $\frac{2}{3}$, and so with probability at least $\frac{1}{3}$ the sets $\mathrm{U}(\mathcal{C}_1)$ and $\mathrm{U}(\mathcal{C}_2)$ have a common vertex $y \neq x$ (meaning that we can find a rainbow cycle).

Of course, the crucial difficulty is to show that we indeed have $|\mathrm{U}(\mathcal{C}_1)| > \frac{n+1}{2}$ with sufficiently large probability. In order to show this, for various subsets $\mathcal{A} \subseteq \mathcal{C}_1$, we consider the set $\mathrm{U}(\mathcal{A})$ of vertices that can be reached by rainbow walks from x with colours in $\mathcal{A}$. More precisely, we consider a randomized chain of subsets $\mathcal{C}_1 = \mathcal{A}_0 \supseteq \mathcal{A}_1 \supseteq \cdots \supseteq \mathcal{A}_T$ of the colours. Then $\mathrm{U}(C_1) = \mathrm{U}(\mathcal{A}_0) \supseteq \mathrm{U}(\mathcal{A}_1) \supseteq \cdots \supseteq \mathrm{U}(\mathcal{A}_T)$, and we analyse how the size of $\mathrm{U}(\mathcal{A}_j)$ grows as we vary j from T to 0. We always have $|\mathrm{U}(\mathcal{A}_T)| \geq 1$ (as $x \in \mathrm{U}(\mathcal{A}_T)$), and our goal is proving that $|\mathrm{U}(\mathcal{A}_0)| > \frac{n+1}{2}$ holds with probability at least $\frac{2}{3}$.

The basic intuition is that each set $\mathrm{U}(\mathcal{A}_{j-1})$ is obtained from $\mathrm{U}(\mathcal{A}_j)$ by "expanding" via the edges with colours in $\mathcal{A}_{j-1} \setminus \mathcal{A}_j$. Note that $\mathrm{U}(\mathcal{A}_j) \subseteq \mathrm{U}(\mathcal{A}_{j-1})$. Furthermore, any vertex outside $\mathrm{U}(\mathcal{A}_j)$ is in $\mathrm{U}(\mathcal{A}_{j-1})$, if it is adjacent to a vertex in $\mathrm{U}(\mathcal{A}_j)$ via an edge with a colour in $\mathcal{A}_{j-1} \setminus \mathcal{A}_j$. If there are many such vertices outside $\mathrm{U}(\mathcal{A}_j)$, then $|\mathrm{U}(\mathcal{A}_{j-1})|$ is quite a bit larger than $|\mathrm{U}(\mathcal{A}_j)|$.

Assuming that our graph G is a "robust sublinear expander" in the sense discussed above, the neighbourhood of the set $\mathrm{U}(\mathcal{A}_j)$ has size at least $(\varepsilon/3) \cdot |\mathrm{U}(\mathcal{A}_j)|$, where $0 \leq \varepsilon \leq 1$ is such that $|\mathrm{U}(\mathcal{A}_j)| \leq n^{1-\varepsilon}$. Furthermore, there are a lot of edges between $\mathrm{U}(\mathcal{A}_j)$ and its neighbourhood. Let us expose the random colour set $\mathcal{A}_j$, which determines $\mathrm{U}(\mathcal{A}_j)$, and let us look at the edges between $\mathrm{U}(\mathcal{A}_j)$ and its neighbourhood. If a lot of these edges have colours in $\mathcal{C} \setminus \mathcal{A}_j$, then one can expect a reasonable fraction of these edges to have colours in $\mathcal{A}_{j-1} \setminus \mathcal{A}_j$ (the colour set $\mathcal{A}_{j-1}$ is still random, conditioned on our exposed outcome of $\mathcal{A}_j$). Using this, one can now expect a reasonable fraction of the at least $(\varepsilon/3) \cdot |\mathrm{U}(\mathcal{A}_j)|$ vertices in the neighbourhood of $\mathrm{U}(\mathcal{A}_j)$ to have an edge into $\mathrm{U}(\mathcal{A}_j)$ with colour in $\mathcal{A}_{j-1} \setminus \mathcal{A}_j$. These vertices are part of $\mathrm{U}(\mathcal{A}_{j-1})$ but not of $\mathrm{U}(\mathcal{A}_j)$, and so $|\mathrm{U}(\mathcal{A}_{j-1})|$ is indeed quite a bit larger than $|\mathrm{U}(\mathcal{A}_j)|$.

However, it can also happen that only a few of the edges between $\mathrm{U}(\mathcal{A}_j)$ and its neighbourhood have colours in $\mathcal{C} \setminus \mathcal{A}_j$. This means that a lot of the edges between $\mathrm{U}(\mathcal{A}_j)$ and its neighbourhood have colours in $\mathcal{A}_j$. If vv' is an edge with a colour $\gamma(v, v') \in \mathcal{A}_j$, and we have $v \in \mathrm{U}(\mathcal{A}_j)$ but $v' \notin \mathrm{U}(\mathcal{A}_j)$, then the colour $\gamma(v, v')$ of the edge vv' must appear on every rainbow walk from x to v with colours in $\mathcal{A}_j$ (indeed, otherwise we could append the edge vv' to a rainbow walk from x to v with colours in $\mathcal{A}_j \setminus \{\gamma(v, v')\}$, and we would obtain a rainbow walk from x to v' with colours in $\mathcal{A}_j$). Now, if $\gamma(v, v') \notin \mathcal{A}_{j+1}$, then v cannot be reached by a rainbow walk from x with colours in $\mathcal{A}_{j+1} \subseteq \mathcal{A}_j \setminus \{\gamma(v, v')\}$, and so $v \notin \mathrm{U}(\mathcal{A}_{j+1})$. If there are many edges between $\mathrm{U}(\mathcal{A}_j)$ and its neighbourhood with colours in $\mathcal{A}_j$, then we expect a reasonable fraction of these edges to have colours in $\mathcal{A}_j \setminus \mathcal{A}_{j+1}$. Using this, one can expect to have a reasonable number of vertices $v \in \mathrm{U}(\mathcal{A}_j)$ to have an edge into the neighbourhood of $\mathrm{U}(\mathcal{A}_j)$ with a colour in $\mathcal{A}_j \setminus \mathcal{A}_{j+1}$. All of these vertices are part of $\mathrm{U}(\mathcal{A}_j)$ but not of $\mathrm{U}(\mathcal{A}_{j+1})$. Hence $|\mathrm{U}(\mathcal{A}_{j+1})|$ is quite a bit smaller than $|\mathrm{U}(\mathcal{A}_j)|$.

In both of these cases (both the case where a lot of the edges between $\mathrm{U}(\mathcal{A}_j)$ and its neighbourhood have colours in $\mathcal{C} \setminus \mathcal{A}_j$ and the case where a lot of these edges have colours in $\mathcal{A}_j$), we can expect $|\mathrm{U}(\mathcal{A}_{j-1})|$ to be quite a bit larger than $|\mathrm{U}(\mathcal{A}_{j+1})|$ (note that we always have $|\mathrm{U}(\mathcal{A}_{j-1})| \geq |\mathrm{U}(\mathcal{A}_j)| \geq |\mathrm{U}(\mathcal{A}_{j+1})|$). More precisely, we heuristically expect an inequality like $|\mathrm{U}(\mathcal{A}_{j-1})| \geq (1 + \varepsilon/400) \cdot |\mathrm{U}(\mathcal{A}_{j+1})|$ if $|\mathrm{U}(\mathcal{A}_j)| \leq n^{1-\varepsilon}$. Iterating this, and using that we always have $|\mathrm{U}(\mathcal{A}_T)| \geq 1$, we can expect $|\mathrm{U}(\mathcal{A}_0)| = |\mathrm{U}(\mathcal{C}_1)|$ to be quite large, implying that $|\mathrm{U}(\mathcal{C}_1)| > \frac{n+1}{2}$ with sufficiently large probability.

However, there are various challenges in actually implementing this strategy. First, we have to be careful about the

5

notion of having "a lot of" edges between $\mathrm{U}(\mathcal{A}_j)$ and its neighbourhood with colours in $\mathcal{C} \setminus \mathcal{A}_j$, or "a lot of" such edges with colours in $\mathcal{A}_j$. Namely, the relevant condition here is not just about the total number of such edges, but about having many such edges which are not too clustered. This is because we want to conclude that we can expect many vertices in $\mathrm{U}(\mathcal{A}_{j-1}) \setminus \mathrm{U}(\mathcal{A}_j)$ or $\mathrm{U}(\mathcal{A}_j) \setminus \mathrm{U}(\mathcal{A}_{j+1})$, respectively (and so it is important that our edges under consideration are not clustered at a small number of vertices). This technical issue is resolved by Lemma 4.1 below, which shows that there are many "non-clustered" edges between $\mathrm{U}(\mathcal{A}_j)$ and its neighbourhood with colours in $\mathcal{C} \setminus \mathcal{A}_j$ or many such "non-clustered" edges with colours in $\mathcal{A}_j$.

A more severe issue is that we can only show $|\mathrm{U}(\mathcal{A}_{j-1})|$ to be quite a bit larger than $|\mathrm{U}(\mathcal{A}_{j+1})|$ in expectation. Indeed, there can be a lot of dependencies for different vertices for whether they are contained in the sets $\mathrm{U}(\mathcal{A}_{j-1})$ and $\mathrm{U}(\mathcal{A}_{j+1})$. Thus, we cannot establish concentration for $|\mathrm{U}(\mathcal{A}_{j-1})|$ or $|\mathrm{U}(\mathcal{A}_{j+1})|$ and so we only obtain our desired conclusions for the expectations of these random variables. A priori, this is not a problem (it essentially suffices to show that the expectation of $|\mathrm{U}(\mathcal{C}_1)|$ is large enough). However, the issue is that we need the assumption $|\mathrm{U}(\mathcal{A}_j)| \leq n^{1-\varepsilon}$ for our expansion properties, and it does not suffice to have this assumption in expectation (indeed, even if the expectation of $|\mathrm{U}(\mathcal{A}_j)|$ is at most $n^{1-\varepsilon}$, we could still have $|\mathrm{U}(\mathcal{A}_j)| > n^{1-\varepsilon}$ with large probability). Of course, if the expectation of $|\mathrm{U}(\mathcal{A}_j)|$ is small, then by Markov's inequality $|\mathrm{U}(\mathcal{A}_j)|$ is small with high probability. However, one loses so much in the quantitative bounds for "small" here that one cannot apply this argument at every step.

The solution to this issue is to only apply Markov's inequality for certain spaced-out indices. For those indices, we can then conclude from a bound for the expectation of $|\mathrm{U}(\mathcal{A}_j)|$ that $|\mathrm{U}(\mathcal{A}_j)|$ is small with sufficiently high probability (namely, $|\mathrm{U}(\mathcal{A}_j)| \leq n^{1-\varepsilon}$ for some suitably chosen ε depending on j). Whenever this is the case, we also have $|\mathrm{U}(\mathcal{A}_{j'})| \leq |\mathrm{U}(\mathcal{A}_j)| \leq n^{1-\varepsilon}$ for all $j' \geq j$ and so we get the desired expansion properties for all such sets $\mathrm{U}(\mathcal{A}_{j'})$. However, this requires performing the arguments sketched above while conditioning on $|\mathrm{U}(\mathcal{A}_j)| \leq n^{1-\varepsilon}$. This conditioning makes the argument more complicated and introduces various technical difficulties. Luckily, these difficulties can be overcome and we can carry out the above strategy even with this additional conditioning.

3 Robust sublinear expanders

In our proof of Theorem 1.1, we will first reduce to the case that G is a "robust sublinear expander" in the following sense.

Definition 3.1. *A graph G on $n \geq 1$ vertices is called a* robust sublinear expander *if for every $0 \leq \varepsilon \leq 1$ and every non-empty subset $U \subseteq V(G)$ of size $|U| \leq n^{1-\varepsilon}$ the following holds. For every subset $F \subseteq E(G)$ of $|F| \leq (\varepsilon/3) \cdot d(G) \cdot |U|$ edges, we have $|N_{G-F}(U)| \geq (\varepsilon/3) \cdot |U|$.*

Recall that $d(G)$ denotes the average degree of G and $N_{G-F}(U)$ is the neighbourhood of U in the graph resulting from G when deleting the edges in F. Note that every robust sublinear expander must have at least two vertices (since for $n = 1$ the condition in the definition above is not satisfied for $\varepsilon = 1$ and $U = V(G)$).

Various competing definitions of robust sublinear expander graphs have been considered over the past 30 years, with numerous applications [13, 15, 20, 23, 27, 28, 29, 38, 42, 47, 48, 50, 51, 52, 56, 69, 71, 73]. This general idea as well as one of the first definitions are due to Komlós and Szemerédi [43]. The above definition is a simplified version of the one in [29] in a similar manner as used in [13]. The following lemma is a strengthening of an observation dating back to the original paper of Komlós and Szemerédi. Our proof for this lemma is motivated by [73], where a similar lemma is obtained but for a weaker notion of sublinear expanders (while [73] considered a different type of expression to maximize, essentially the same maximization as in our proof below also occurs in the paper [74] optimizing the arguments from [73]).

The lemma states that in any graph G we can find a subgraph H with similar average degree as G, such that H is a robust sublinear expander. More precisely, with "similar average degree" we mean that the ratio $d(H)/\log|V(H)|$ is almost the same as the ratio $d(G)/\log|V(G)|$ (up to constant factors).

Lemma 3.2. *For every graph G with at least one edge, there exists a subgraph H which is a robust sublinear expander and has average degree*

$$d(H) \geq \frac{\log|V(H)| - (1/3)}{\log|V(G)| - (1/3)} \cdot d(G) \geq \frac{1}{3} \cdot \frac{\log|V(H)|}{\log|V(G)|} \cdot d(G).$$

When applying this lemma, we will only use the lower bound for $d(H)$ given by the term on the right-hand side, but we include the stronger middle term in the statement in case it becomes helpful for future applications.

Proof of Lemma 3.2. Among all subgraphs H of G (with at least one vertex), let us choose a subgraph H maximizing the expression

$$\frac{d(H)}{\log|V(H)| - (1/3)}.$$

Then we clearly have

$$\frac{d(H)}{\log|V(H)| - (1/3)} \geq \frac{d(G)}{\log|V(G)| - (1/3)}.$$

In particular, as $d(G) > 0$ and $\log|V(G)| \geq \log 2 > 1/3$, we must have $d(H) > 0$ and $|V(H)| \geq 2$. So we indeed obtain

$$d(H) \geq \frac{\log|V(H)| - (1/3)}{\log|V(G)| - (1/3)} \cdot d(G) \geq \frac{1}{3} \cdot \frac{\log|V(H)|}{\log|V(G)|} \cdot d(G),$$

where the second inequality follows from $\log|V(H)| - (1/3) \geq (1/3) \cdot \log|V(H)|$ since $\log|V(H)| \geq \log 2 \geq 1/2$.

It remains to show that H is a robust sublinear expander. Let $m = |V(H)| \geq 2$. Consider $0 \leq \varepsilon \leq 1$, and let $U \subseteq V(H)$ be a non-empty subset of size $|U| \leq m^{1-\varepsilon}$. Furthermore, let $F \subseteq E(H)$ be a subset of $|F| \leq (\varepsilon/3) \cdot d(H) \cdot |U|$ edges. We need to show that $|N_{H-F}(U)| \geq (\varepsilon/3) \cdot |U|$.

Suppose for contradiction that $|N_{H-F}(U)| < (\varepsilon/3) \cdot |U|$. Note that the number of edges in the graph H equals $\frac{1}{2} \cdot d(H) \cdot m$.

First, we claim that the subgraph of H induced by the vertex set $\overline{U} = V(H) \setminus U$ contains at most $\frac{1}{2} \cdot d(H) \cdot |\overline{U}|$ edges. Indeed, if $|U| = m$ or $|U| = m - 1$, this induced subgraph has at most one vertex and therefore it does not contain any edges. If $|U| \leq m - 2$, then $0 < \log|\overline{U}| - (1/3) \leq \log m - (1/3)$. By our choice of H, we have

$$\frac{d(H[\overline{U}])}{\log|\overline{U}| - (1/3)} \leq \frac{d(H)}{\log m - (1/3)},$$

and therefore $d(H[\overline{U}]) \leq d(H)$. Now, the number of edges of the subgraph $H[\overline{U}]$ induced by $\overline{U}$ is $\frac{1}{2} \cdot d(H[\overline{U}]) \cdot |\overline{U}| \leq \frac{1}{2} \cdot d(H) \cdot |\overline{U}|$, as claimed.

Since the graph H has at most $\frac{1}{2} \cdot d(H) \cdot |\overline{U}| = \frac{1}{2} \cdot d(H) \cdot (m - |U|)$ edges inside the set $\overline{U} = V(H) \setminus U$, there must be at least $\frac{1}{2} \cdot d(H) \cdot |U|$ edges in the graph H that are incident with at least one vertex in U. Hence, there must be at least

$$\frac{1}{2} \cdot d(H) \cdot |U| - |F| \geq \frac{1}{2} \cdot d(H) \cdot |U| - \frac{\varepsilon}{3} \cdot d(H) \cdot |U| = \left(\frac{1}{2} - \frac{\varepsilon}{3}\right) \cdot d(H) \cdot |U|$$

edges in the graph $H - F$ that are incident with at least one vertex in U. By definition of $N_{H-F}(U)$, all of these edges are contained in the set $X := U \cup N_{H-F}(U)$. Hence, the induced subgraph $H[X]$ on X has average degree

$$d(H[X]) = \frac{2 \cdot |E(H[X])|}{|X|} \geq \frac{2}{|X|} \cdot \left(\frac{1}{2} - \frac{\varepsilon}{3}\right) \cdot d(H) \cdot |U| = \left(1 - \frac{2\varepsilon}{3}\right) \cdot d(H) \cdot \frac{|U|}{|X|}.$$

By our assumption $|N_{H-F}(U)| < (\varepsilon/3) \cdot |U|$, we have $|X| = |U| + |N_{H-F}(U)| < (1 + (\varepsilon/3)) \cdot |U|$. So we can conclude

$$d(H[X]) > \left(1 - \frac{2\varepsilon}{3}\right) \cdot d(H) \cdot \frac{|U|}{(1 + (\varepsilon/3)) \cdot |U|} = \frac{1 - (2\varepsilon/3)}{1 + (\varepsilon/3)} \cdot d(H) \geq (1 - \varepsilon) \cdot d(H).$$

In particular, we obtain $d(H[X]) > 0$, so X contains at least one edge and in particular $|X| \geq 2$. Together with

$$\log|X| < \log\left(1 + \frac{\varepsilon}{3}\right) + \log|U| \leq \frac{\varepsilon}{3} + \log(m^{1-\varepsilon}) = \frac{\varepsilon}{3} + (1 - \varepsilon) \cdot \log m,$$

we now obtain

$$\frac{d(H[X])}{\log|X| - (1/3)} > \frac{(1-\varepsilon) \cdot d(H)}{(\varepsilon/3) + (1-\varepsilon) \cdot \log m - (1/3)} = \frac{(1-\varepsilon) \cdot d(H)}{(1-\varepsilon) \cdot \log m - (1-\varepsilon) \cdot (1/3)} = \frac{d(H)}{\log|V(H)| - (1/3)}.$$

This is a contradiction to H maximizing the expression on the right-hand side, which finishes the proof. $\qquad\square$

4 Preliminaries

4.1 A lemma for robust sublinear expanders

We will use the assumption that our graph is a robust sublinear expander by applying the following lemma (the property in the lemma is actually equivalent to the property of G being a robust sublinear expander up to changing the constants appearing in Definition 3.1).

Lemma 4.1. *Let G be a robust sublinear expander (in the sense of Definition 3.1) with n vertices, let $0 \leq \varepsilon \leq 1$ and let $U \subseteq V(G)$ be a vertex subset of size $|U| \leq n^{1-\varepsilon}$. Consider a (not necessarily proper) colouring of the edges of G between U and $V(G) \setminus U$ with the colours red and blue. Then at least one of the following two statements holds:*

- *we can find a subset $F_{red} \subseteq E(G)$ of red edges in G with $|F_{red}| \geq (\varepsilon/7) \cdot d(G) \cdot |U|$ and $\deg_{F_{red}}(v) \leq \lceil d(G) \rceil$ for all $v \in U$, or*

- *we can find a subset $F_{blue} \subseteq E(G)$ of blue edges in G with $|F_{blue}| \geq (\varepsilon/7) \cdot d(G) \cdot |U|$ and $\deg_{F_{blue}}(v') \leq \lceil d(G) \rceil$ for all $v' \in V(G) \setminus U$.*

Proof. Let $E_{\text{red}} \subseteq E(G)$ be the set of red edges in G, and $E_{\text{blue}} \subseteq E(G)$ be the set of blue edges in G. Recall that all edges in E_{red} and all edges in E_{blue} are between U and $V(G) \setminus U$, so they are of the form (v, v') with $v \in U$ and $v' \in V(G) \setminus U$.

Let $F'_{\text{red}} \subseteq E_{\text{red}}$ be the set of red edges $(v, v') \in E_{\text{red}}$ with $\deg_{E_{\text{red}}}(v) \leq d(G)$. Similarly, let $F'_{\text{blue}} \subseteq E_{\text{blue}}$ be the set of blue edges $(v, v') \in E_{\text{blue}}$ with $\deg_{E_{\text{blue}}}(v') \leq d(G)$. If $|F'_{\text{red}}| \geq (\varepsilon/7) \cdot d(G) \cdot |U|$, then F'_{red} satisfies the first alternative in the lemma statement. Similarly, if $|F'_{\text{blue}}| \geq (\varepsilon/7) \cdot d(G) \cdot |U|$, then F'_{blue} satisfies the second alternative in the lemma statement. Hence, we may assume that $|F'_{\text{red}}| < (\varepsilon/7) \cdot d(G) \cdot |U|$ and $|F'_{\text{blue}}| < (\varepsilon/7) \cdot d(G) \cdot |U|$, and therefore $|F'_{\text{red}} \cup F'_{\text{blue}}| < (2\varepsilon/7) \cdot d(G) \cdot |U|$.

Let $U_{\text{red}} \subseteq U$ be the set of vertices $v \in U$ with $\deg_{E_{\text{red}}}(v) \geq d(G)$, and let $V_{\text{blue}} \subseteq V(G) \setminus U$ be the set of vertices $v' \in V(G) \setminus U$ with $\deg_{E_{\text{blue}}}(v') \geq d(G)$. Note that every edge $(v, v') \in E(G)$ with $v \in U \setminus U_{\text{red}}$ and $v' \in V(G) \setminus (U \cup V_{\text{blue}})$ is contained in F'_{red} or F'_{blue} (depending on whether the edge is red or blue).

If $|U_{\text{red}}| \geq (\varepsilon/7) \cdot |U|$, then we can choose a set F_{red} as in the first alternative in the lemma statement by choosing $\lceil d(G) \rceil$ edges $(v, v') \in E_{\text{red}}$ for each $v \in U_{\text{red}}$. Similarly, if $|V_{\text{blue}}| \geq (\varepsilon/7) \cdot |U|$, then we can choose a set F_{blue} as in the second alternative by choosing $\lceil d(G) \rceil$ edges $(v, v') \in E_{\text{red}}$ for each $v' \in V_{\text{blue}}$. Hence, we may assume that $|U_{\text{red}}| < (\varepsilon/7) \cdot |U|$ and $|V_{\text{blue}}| < (\varepsilon/7) \cdot |U|$, and therefore in particular $|U_{\text{red}} \cup V_{\text{blue}}| < (2\varepsilon/7) \cdot |U|$.

We will now show that not all of these assumptions can hold at the same time by deriving a contradiction to G being a robust sublinear expander. To obtain this contradiction, let $U' = U \setminus U_{\text{red}}$, then $|U'| \geq (1 - \varepsilon/7) \cdot |U| \geq (6/7) \cdot |U|$. Recall that all edges between $U' = U \setminus U_{\text{red}}$ and $V(G) \setminus (U \cup V_{\text{blue}})$ are contained in F'_{red} or F'_{blue}. Hence, setting $F = F'_{\text{red}} \cup F'_{\text{blue}}$, we have $N_{G-F}(U') \subseteq U_{\text{red}} \cup V_{\text{blue}}$ and hence $|N_{G-F}(U')| \leq |U_{\text{red}} \cup V_{\text{blue}}| < (2\varepsilon/7) \cdot |U| \leq (\varepsilon/3) \cdot |U'|$. On the other hand, $|U'| \leq |U| \leq n^{1-\varepsilon}$ and $|F| = |F'_{\text{red}} \cup F'_{\text{blue}}| < (2\varepsilon/7) \cdot d(G) \cdot |U| \leq (\varepsilon/3) \cdot d(G) \cdot |U'|$, so this is a contradiction to G being a robust sublinear expander. $\square$

4.2 Probabilistic lemmas

In this subsection we collect some simple probabilistic lemmas which will be used in the proof of Theorem 1.1. The first lemma gives bounds for the conditional probabilities for various colours to be contained in various sets in the randomized nested colour set sequence $\mathcal{C} \supseteq \mathcal{A}_0 \supseteq \mathcal{A}_1 \supseteq \cdots \supseteq \mathcal{A}_T$ mentioned in the proof overview in Section 2 (when conditioning on the outcome of some $\mathcal{A}_j$).

Lemma 4.2. *For a given set $\mathcal{C}$, some integer T and $1 - (1/T) \leq p < 1$, consider a random sequence of nested subsets $\mathcal{C} \supseteq \mathcal{A}_0 \supseteq \mathcal{A}_1 \supseteq \cdots \supseteq \mathcal{A}_T$ obtained as follows: Let $\mathcal{A}_0 \subseteq \mathcal{C}$ be a random subset obtained by including each element of $\mathcal{C}$ independently with probability $1/2$. For any outcome of $\mathcal{A}_i$ for some $i = 0, \ldots, T - 1$, we define $\mathcal{A}_{i+1}$ to be a random subset of $\mathcal{A}_i$ obtained by including each element of $\mathcal{A}_i$ into $\mathcal{A}_{i+1}$ with probability p independently. Then both of the following two statements hold.*

(a) *For integers $0 \leq i \leq j \leq T$, let us consider any outcome of $\mathcal{A}_j$. Then, conditional on this outcome of $\mathcal{A}_j$, for each element $x \in \mathcal{C}$, with probability at least $(j - i) \cdot (1 - p)/6$ we have $x \in \mathcal{A}_i$.*

(b) *For integers $0 \leq i < j \leq T$ with $j - i + 1 \leq T/2$, let us consider any outcomes of A_i and $\mathcal{A}_j$. Then, conditional on these outcomes of $\mathcal{A}_i$ and $\mathcal{A}_j$, for each element $x \in \mathcal{A}_i$, with probability at least $(T/(j - i)) \cdot (1 - p)/2$ we have $x \in \mathcal{A}_{j-1}$.*

Proof. Note that the restrictions $\mathcal{A}_0 \cap \{x\} \supseteq \mathcal{A}_1 \cap \{x\} \supseteq \cdots \supseteq \mathcal{A}_T \cap \{x\}$ of our random sequence are mutually independent between different choices for the element $x \in \mathcal{C}$ (because the random sequence was chosen in a way that decisions of whether to include different elements of $\mathcal{C}$ are completely independent of each other).

For (a), note that for every $x \in \mathcal{A}_j$, when conditioning on $\mathcal{A}_j$, we have $x \in \mathcal{A}_i$ with probability $1 \geq T \cdot (1 - p) \geq (j - i) \cdot (1 - p)/6$. So let us assume that $x \notin \mathcal{A}_j$. When looking at the event $x \in \mathcal{A}_i$, it is equivalent to condition on any given outcome of $\mathcal{A}_j$ which does not contain x or to just condition on the event $x \notin \mathcal{A}_j$. Thus, for (a) it suffices to show that $\mathbb{P}[x \in \mathcal{A}_i \mid x \notin \mathcal{A}_j] \geq (j - i) \cdot (1 - p)/6$ for each $x \in \mathcal{C}$.

So let $x \in \mathcal{C}$, then

$$\mathbb{P}[x \in \mathcal{A}_i \mid x \notin \mathcal{A}_j] = \frac{\mathbb{P}[x \in \mathcal{A}_i \setminus \mathcal{A}_j]}{\mathbb{P}[x \notin \mathcal{A}_j]} \geq \mathbb{P}[x \in \mathcal{A}_i \setminus \mathcal{A}_j] = \frac{1}{2} p^i (1 - p^{j-i}) \geq \frac{p^T}{2} \cdot \frac{1 - p^{j-i}}{p^{j-i}} = \frac{p^T}{2}((1/p)^{j-i} - 1).$$

Applying Bernoulli's inequality, this implies

$$\mathbb{P}[x \in \mathcal{A}_i \mid x \notin \mathcal{A}_j] \geq \frac{p^T}{2} \cdot (j - i)((1/p) - 1) = \frac{p^{T-1}}{2} \cdot (j - i)(1 - p) \geq (j - i)(1 - p)/6,$$

where in the last step we used that $p^{T-1} \geq (1 - \frac{1}{T})^{T-1} \geq 1/e \geq 1/3$. This proves (a).

For (b), we are conditioning on the outcomes of $\mathcal{A}_i$ and $\mathcal{A}_j$, and for some $x \in \mathcal{A}_i$ we are looking at the event $x \in \mathcal{A}_{j-1}$. If $x \in A_j$, then under this conditioning, we have $x \in \mathcal{A}_{j-1}$ with probability $1 \geq T \cdot (1 - p) \geq (T/(j - i)) \cdot (1 - p)/2$. So let us assume that $x \in \mathcal{A}_i \setminus \mathcal{A}_j$. It now suffices to prove that $\mathbb{P}[x \in \mathcal{A}_{j-1} \mid x \in \mathcal{A}_i \setminus \mathcal{A}_j] \geq (T/(j - i)) \cdot (1 - p)/2$ (assuming that $j - i + 1 \leq T/2$). Note that

$$\mathbb{P}[x \in \mathcal{A}_{j-1} \mid x \in \mathcal{A}_i \setminus \mathcal{A}_j] = \frac{\mathbb{P}[x \in \mathcal{A}_{j-1} \setminus \mathcal{A}_j]}{\mathbb{P}[x \in \mathcal{A}_i \setminus \mathcal{A}_j]} = \frac{p^{j-1}(1 - p)}{p^i(1 - p^{j-i})} = (1 - p) \cdot \frac{p^{j-i-1}}{1 - p^{j-i}} \geq (1 - p) \cdot \frac{p^{j-i}}{1 - p^{j-i}}.$$

Recalling that $p^{T-1} \geq (1 - \frac{1}{T})^{T-1} \geq 1/e$, we have $p^{j-i} \geq \exp(-(j - i)/(T - 1)) \geq 1 - (j - i)/(T - 1)$ and hence $1 - p^{j-i} \leq (j - i)/(T - 1)$. Hence, the desired probability $\mathbb{P}[x \in \mathcal{A}_{j-1} \mid x \in \mathcal{A}_i \setminus \mathcal{A}_j]$ is at least

$$(1 - p) \cdot \frac{p^{j-i}}{1 - p^{j-i}} = (1 - p)\left(\frac{1}{1 - p^{j-i}} - 1\right) \geq (1 - p)\left(\frac{T - 1}{j - i} - 1\right) = (1 - p) \cdot \frac{T - 1 - j + i}{j - i} \geq \frac{1 - p}{2} \cdot \frac{T}{j - i},$$

where in the last step we used that $T - 1 - j + i \geq T/2$ as $j - i + 1 \leq T/2$. $\qquad\square$

We will also need the following very easy probabilistic lemma.

Lemma 4.3. *Let m be a positive integer, let $0 \leq p \leq 1/m$ and let $\mathcal{E}_1, \ldots, \mathcal{E}_m$ be independent events such that for $i = 1, \ldots, m$ the event $\mathcal{E}_i$ holds with probability $p_i \geq p$. Then with probability at least $pm/2$, at least one of the events $\mathcal{E}_1, \ldots, \mathcal{E}_m$ holds.*

Proof. Since decreasing p_i decreases the relevant probability in the conclusion of the lemma, we can assume that $p_1 = \cdots = p_m = p$. Now,

$$\mathbb{P}[\mathcal{E}_1 \vee \cdots \vee \mathcal{E}_m] \geq \sum_{i=1}^m \mathbb{P}[\mathcal{E}_i] - \sum_{i<j} \mathbb{P}[\mathcal{E}_i \wedge \mathcal{E}_j] = pm - p^2\binom{m}{2} \geq pm - \frac{p^2 m^2}{2} = pm\left(1 - \frac{pm}{2}\right) \geq \frac{pm}{2},$$

as desired. $\qquad\square$

We will use the following version of the Chernoff bound for sums of independent indicator random variables, see [33, Theorem 2.8] or [5, Appendix A].

Lemma 4.4. *Let $X_1, \ldots, X_n$ be independent $\{0, 1\}$-valued random variables. Let $X = X_1 + \cdots + X_n$, and $\mu = \mathbb{E}[X]$. Then for every $\delta \geq 0$, we have*

$$\mathbb{P}[X \leq (1 - \delta)\mu] \leq e^{-\delta^2 \mu/2} \quad \text{and} \quad \mathbb{P}[X \geq (1 + \delta)\mu] \leq e^{-\delta^2 \mu/(2+\delta)}.$$

In particular, the following corollary will be convenient for us.

Corollary 4.5. *Let X and μ be as in Lemma 4.4, and let $\tau \geq 2\mu$. Then we have*

$$\mathbb{P}[X \geq \tau] \leq e^{-\tau/6}.$$

9

Proof. If $\mu = 0$, then we have $X = 0$ with probability 1, and so the statement is trivially true. So let us assume that $\mu > 0$. Let $\delta = (\tau/\mu) - 1 \geq 1$, and note that then $2 + \delta = 1 + (\tau/\mu) = (\tau + \mu)/\mu$. Now, by the second inequality in Lemma 4.4 we have

$$\mathbb{P}[X \geq \tau] = \mathbb{P}[X \geq (1 + \delta)\mu] \leq \exp\left(-\frac{\delta^2}{2 + \delta} \cdot \mu\right) = \exp\left(-\frac{(\tau - \mu)^2/\mu^2}{(\tau + \mu)/\mu} \cdot \mu\right) = \exp\left(-\frac{(\tau - \mu)^2}{\tau + \mu}\right)$$

$$\leq \exp\left(-\frac{(\tau - (\tau/2))^2}{\tau + (\tau/2)}\right) = \exp\left(-\frac{\tau^2/4}{(3/2) \cdot \tau}\right) = \exp(-\tau/6),$$

where for the second inequality we used that $\mu \leq \tau/2$. $\qquad\square$

4.3 A probabilistic lemma for bipartite graphs

Our last preparatory lemma is a probabilistic lemma concerning certain events assigned to the edges of some bipartite graph.

Lemma 4.6. *Let $0 \leq \varepsilon \leq 1$ and let $1 \leq \ell \leq d$ be integers, and consider a bipartite graph G with vertex set U on one side and vertex set W on the other side. Let us assume that G has at least $\varepsilon d|U|$ edges, but that every vertex $w \in W$ satisfies $\deg(w) \leq d$. For every edge $e \in E(G)$, consider an event $\mathcal{E}_e$ which holds with probability at least ℓ/d. Suppose that for every vertex $v \in U \cup W$ the events $\mathcal{E}_e$ are mutually independent for all edges e incident with v. Then with probability at least $1 - 15 \exp(-\varepsilon\ell/16)$, there exists a subset $E' \subseteq E(G)$ of size $|E'| \geq (\varepsilon/12) \cdot \ell \cdot |U|$ such that $\mathcal{E}_e$ holds for each $e \in E'$ and such that for every vertex $w \in W$ we have $\deg_{E'}(w) \leq 2\ell$.*

Proof. We may assume that we have $\mathbb{P}[\mathcal{E}_e] = \ell/d$ for all edges $e \in E(G)$. Indeed, suppose we already proved the lemma in this special case. Then, in the general case that $\mathbb{P}[\mathcal{E}_e] \geq \ell/d$ for every edge $e \in E(G)$, let us consider an event $\mathcal{E}'_e$ with $\mathbb{P}[\mathcal{E}'_e] = (\ell/d)/\mathbb{P}[\mathcal{E}_e]$ for each $e \in E(G)$ such that these events $\mathcal{E}'_e$ are mutually independent and also independent from the events $\mathcal{E}_e$. Note that then for all $e \in E(G)$ we have $\mathbb{P}[\mathcal{E}_e \wedge \mathcal{E}'_e] = \mathbb{P}[\mathcal{E}_e] \cdot \mathbb{P}[\mathcal{E}'_e] = \ell/d$. Furthermore, for every vertex $v \in U \cup W$, the events $\mathcal{E}_e \wedge \mathcal{E}'_e$ are mutually independent for all edges e incident with v. Applying the special case of the lemma to the events $\mathcal{E}_e \wedge \mathcal{E}'_e$ (with $\mathbb{P}[\mathcal{E}_e \wedge \mathcal{E}'_e] = \ell/d$ for all $e \in E(G)$) we find that with probability at least $1 - 15 \exp(-\varepsilon\ell/16)$, there exists a subset $E' \subseteq E(G)$ of size $|E'| \geq (\varepsilon/12) \cdot \ell \cdot |U|$ such that $\mathcal{E}_e \wedge \mathcal{E}'_e$ holds for each $e \in E'$ and such that for every vertex $w \in W$ we have $\deg_{E'}(w) \leq 2\ell$. Any such set E' now has the desired conditions, since having $\mathcal{E}_e \wedge \mathcal{E}'_e$ for each $e \in E'$ in particular implies having $\mathcal{E}_e$ for each $e \in E'$. Thus, it indeed suffices to prove the lemma in the special case that $\mathbb{P}[\mathcal{E}_e] = \ell/d$ for all $e \in E(G)$.

So let us now assume that $\mathbb{P}[\mathcal{E}_e] = \ell/d$ for all $e \in E(G)$. Let $U' \subseteq U$ be the set of vertices $u \in U$ with $\deg(u) \geq \varepsilon d/2$. Then

$$\sum_{u \in U \setminus U'} \deg(u) < |U \setminus U'| \cdot \frac{\varepsilon d}{2} \leq \frac{\varepsilon d|U|}{2} \leq \frac{|E(G)|}{2},$$

and consequently

$$\sum_{u \in U'} \deg(u) \geq \sum_{u \in U} \deg(u) - \frac{|E(G)|}{2} = |E(G)| - \frac{|E(G)|}{2} = \frac{|E(G)|}{2}.$$

Let us call a vertex $u \in U'$ *bad* if there are fewer than $\ell \deg(u)/(2d)$ edges $e \in E(G)$ incident with u for which the event $\mathcal{E}_e$ holds (note that this depends on the outcomes of these random events). Recall that the events $\mathcal{E}_e$ are independent for all edges e incident with u. The expected number of edges $e \in E(G)$ incident with u with $\mathcal{E}_e$ holding is precisely $(\ell/d) \cdot \deg(u) = \ell \deg(u)/d$. Hence, by the Chernoff bound (see Lemma 4.4), for each vertex $u \in U'$ the probability that u is bad is at most $\exp(-\ell \deg(u)/(8d)) \leq \exp(-\varepsilon\ell/16)$ (recall that $\deg(u) \geq \varepsilon d/2$ since $u \in U'$).

Thus, we obtain

$$\mathbb{E}\left[\sum_{u \in U' \text{ bad}} \deg(u)\right] \leq \exp(-\varepsilon\ell/16) \cdot \sum_{u \in U'} \deg(u).$$

So by Markov's inequality, we have

$$\mathbb{P}\left[\sum_{u \in U' \text{ bad}} \deg(u) \geq \frac{1}{3} \cdot \sum_{u \in U'} \deg(u)\right] \leq 3 \exp(-\varepsilon\ell/16),$$

10

and hence
$$\mathbb{P}\left[\sum_{u \in U' \text{ not bad}} \deg(u) \geq \frac{2}{3} \cdot \sum_{u \in U'} \deg(u)\right] \geq 1 - 3\exp(-\varepsilon\ell/16).$$

Recalling $\sum_{u \in U'} \deg(u) \geq |E(G)|/2$, this implies
$$\mathbb{P}\left[\sum_{u \in U' \text{ not bad}} \deg(u) \geq \frac{|E(G)|}{3}\right] \geq 1 - 3\exp(-\varepsilon\ell/16). \tag{4.1}$$

Let us call an edge $e \in E(G)$ of the form (u, w) with $u \in U$ and $w \in W$ *crowded* if the event $\mathcal{E}_e$ holds and there are at least 2ℓ other edges $e' \in E(G) \setminus \{e\}$ incident with w for which $\mathcal{E}_{e'}$ holds. We claim that every edge $e \in E(G)$ (of the form (u, w) with $u \in U$ and $w \in W$) is crowded with probability at most $\exp(-\ell/3) \cdot (\ell/d)$. Indeed, the expected number of edges $e' \in E(G) \setminus \{e\}$ incident with w for which $\mathcal{E}_{e'}$ holds is $(\deg(w) - 1) \cdot (\ell/d) < \ell$ (recalling that $\deg(w) \leq d$). Furthermore, the events $\mathcal{E}_{e'}$ are independent for all $e' \in E(G) \setminus \{e\}$ incident with w and are also independent from $\mathcal{E}_e$. So by the Chernoff bound (see Corollary 4.5) the probability that there are at least 2ℓ edges $e' \in E(G) \setminus \{e\}$ incident with w for which $\mathcal{E}_{e'}$ holds is at most $\exp(-2\ell/6) = \exp(-\ell/3)$. This is independent of the event $\mathcal{E}_e$, which has probability ℓ/d. Hence, the probability that both of these conditions hold, i.e. that e is crowded, is at most $\exp(-\ell/3) \cdot (\ell/d)$.

Since every edge $e \in E(G)$ is crowded with probability at most $\exp(-\ell/3) \cdot (\ell/d)$, the expected number of crowded edges $e \in E(G)$ is at most $\exp(-\ell/3) \cdot (\ell/d) \cdot |E(G)|$. Hence, by Markov's inequality, the probability that there are at least $(\ell/d) \cdot (|E(G)|/12)$ crowded edges $e \in E(G)$ is at most $12\exp(-\ell/3) \leq 12\exp(-\varepsilon\ell/16)$.

Together with (4.1) this implies that with probability at least $1 - 15\exp(-\varepsilon\ell/16)$ there are at most $(\ell/d) \cdot (|E(G)|/12)$ crowded edges $e \in E(G)$ and we have
$$\sum_{u \in U' \text{ not bad}} \deg(u) \geq \frac{|E(G)|}{3}.$$

It now suffices to show that whenever this happens there is a subset $E' \subseteq E$ satisfying the conditions in the statement of the lemma. Let $E' \subseteq E(G)$ be the set of edges $e \in E(G)$ such that $\mathcal{E}_e$ holds but e is not crowded.

To check that $|E'| \geq (\varepsilon/12) \cdot \ell \cdot |U|$, we first claim that there are at least $(\ell/d) \cdot (|E(G)|/6)$ edges $e \in E(G)$ for which $\mathcal{E}_e$ holds. Indeed, for every vertex $u \in U'$ which is not bad there are at least $\ell\deg(u)/(2d)$ edges e incident with u for which the event $\mathcal{E}_e$ holds. So the total number of edges $e \in E(G)$ for which $\mathcal{E}_e$ holds is at least
$$\sum_{u \in U' \text{ not bad}} \frac{\ell\deg(u)}{2d} = \frac{\ell}{2d} \sum_{u \in U' \text{ not bad}} \deg(u) \geq \frac{\ell}{2d} \cdot \frac{|E(G)|}{3} = \frac{\ell}{d} \cdot \frac{|E(G)|}{6}.$$

Recalling our assumption that there are at most $(\ell/d) \cdot (|E(G)|/12)$ crowded edges $e \in E(G)$, we can conclude that there are at least $(\ell/d) \cdot (|E(G)|/12)$ edges $e \in E(G)$ such that $\mathcal{E}_e$ holds but e is not crowded. Hence, $|E'| \geq (\ell/d) \cdot (|E(G)|/12) \geq (\varepsilon/12) \cdot \ell \cdot |U|$.

By definition of E' we already know that for all edges $e \in E'$ the event $\mathcal{E}_e$ holds. It remains to show that $\deg_{E'}(w) \leq 2\ell$ for all $w \in W$. Suppose that there is a vertex $w \in W$ with $\deg_{E'}(w) \geq 2\ell + 1$, and let $e \in E'$ be an edge incident with w. Then the event $\mathcal{E}_e$ holds and there are at least 2ℓ other edges in $e' \in E(G) \setminus \{e\}$ incident with w for which the event $\mathcal{E}_{e'}$ holds. Thus, e is crowded, but this is a contradiction to $e \in E'$. Thus, we indeed have $\deg_{E'}(w) \leq 2\ell$ for all $w \in W$. $\qquad\square$

5 Finding a rainbow cycle

In this section, we prove Theorem 1.1, but the key part of the proof will actually be to prove various lemmas (whose proofs are postponed to the next two sections). In order to prove the theorem, it suffices to prove the following proposition.

Proposition 5.1. *For sufficiently large n, assume that G is a graph on n vertices with average degree $d(G) \geq 10^7 \cdot \log n \cdot \log\log n$ which is a robust sublinear expander (in the sense of Definition 3.1). Then any proper edge-colouring $\gamma : E(G) \to \mathcal{C}$ of the graph G has a rainbow cycle.*

Indeed, this proposition implies Theorem 1.1 via the following easy reduction.

Proof of Theorem 1.1 assuming Proposition 5.1. We may assume that n is sufficiently large (note that this can be ensured by choosing C sufficiently large). So let us assume that n is large enough such that the statement in Proposition 5.1 holds for graphs on at least $\log \log n$ vertices. We show that then the statement in Theorem 1.1 holds with $C = 10^8$, so let G be a properly edge-coloured graph on n vertices with average degree $d(G) \geq 10^8 \cdot \log n \cdot \log \log n$. Applying Lemma 3.2, we obtain a subgraph H of G which is a robust sublinear expander and has average degree

$$d(H) \geq \frac{1}{3} \cdot \frac{\log |V(H)|}{\log n} \cdot d(G) \geq \frac{1}{3} \cdot \frac{\log |V(H)|}{\log n} \cdot 10^8 \cdot \log n \cdot \log \log n > 10^7 \cdot \log |V(H)| \cdot \log \log n.$$

In particular, $d(H) \geq 10^7 \cdot \log |V(H)| \cdot \log \log |V(H)|$. So by Proposition 5.1 (noting that $|V(H)| \geq d(H) \geq \log \log n$), the subgraph H must contain a rainbow cycle, and then the original graph G also contains a rainbow cycle. $\qquad\square$

The rest of this section is devoted to proving Proposition 5.1, but we will postpone the proof of a key lemma to the next section.

Proof of Proposition 5.1. Let G and $\gamma : E(G) \to \mathcal{C}$ be as in the proposition. Let us fix an arbitrary vertex $x \in V(G)$ for the rest of this proof. A *rainbow walk* in G is a walk on which all edges have distinct colours.

Definition 5.2. *For a subset $\mathcal{A} \subseteq \mathcal{C}$ of the colour palette, let $\mathrm{U}(\mathcal{A}) \subseteq V(G)$ be the set of vertices that can be reached from x via a rainbow walk with colours in $\mathcal{A}$.*

Note that for $\mathcal{A} \subseteq \mathcal{A}' \subseteq \mathcal{C}$, we have $\mathrm{U}(\mathcal{A}) \subseteq \mathrm{U}(\mathcal{A}')$. Also, note that for all $\mathcal{A} \subseteq \mathcal{C}$ we have $x \in \mathrm{U}(\mathcal{A}) \subseteq V(G)$ (since x can always be reached by the trivial walk with no edges), and hence $|\mathrm{U}(\mathcal{A})| \geq 1$.

Let

$$K = \lceil \log \log n \rceil, \qquad L = 10^5 \cdot \lceil \log n \rceil, \quad \text{and} \quad T = KL,$$

and note that $T \leq 4 \cdot 10^5 \cdot \log n \cdot \log \log n \leq d(G)/12$ for n sufficiently large. Also, note that L is divisible by 4.

Let $\mathcal{A}_0 \subseteq \mathcal{C}$ be a random subset of the colour palette, obtained by including each colour in $\mathcal{C}$ with probability $1/2$. It now suffices to prove that

$$\mathbb{P}\big[|\mathrm{U}(\mathcal{A}_0)| \geq n/\sqrt{e}\,\big] \geq 2/3. \tag{5.1}$$

Indeed, applying (5.1) to the complement $\mathcal{C} \setminus \mathcal{A}_0$ (which has the same distribution as $\mathcal{A}_0$), we also have

$$\mathbb{P}\big[|\mathrm{U}(\mathcal{C} \setminus \mathcal{A}_0)| \geq n/\sqrt{e}\,\big] \geq 2/3.$$

Thus, with probability at least $1/3$, we have $|\mathrm{U}(\mathcal{A}_0)| + |\mathrm{U}(\mathcal{C} \setminus \mathcal{A}_0)| \geq (2/\sqrt{e}) \cdot n > n + 1$ (assuming that n is sufficiently large), so there must be a vertex $v \in V(G) \setminus \{x\}$ with $v \in \mathrm{U}(\mathcal{A}_0) \cap \mathrm{U}(\mathcal{C} \setminus \mathcal{A}_0)$. This means that v can be reached from x by a rainbow walk with colours in $\mathcal{A}_0$ and also by a rainbow walk with colours in $\mathcal{C} \setminus \mathcal{A}_0$. Concatenating these walks gives a (non-trivial) rainbow closed walk starting and ending at x. Thus, there must be a rainbow cycle. So it indeed suffices to prove (5.1).

Let us extend $\mathcal{A}_0$ to a nested sequence of random colour sets $\mathcal{A}_0 \supseteq \mathcal{A}_1 \supseteq \cdots \supseteq \mathcal{A}_T$, where for $i = 0, \ldots, T-1$, for a given outcome of $\mathcal{A}_i$, we define $\mathcal{A}_{i+1}$ to be a random subset of $\mathcal{A}_i$ obtained by including each colour in $\mathcal{A}_i$ into $\mathcal{A}_{i+1}$ with probability $1 - (1/T)$ independently.

Furthermore, let $s : \{0, \ldots, K\} \to \mathbb{R}$ be the function given by

$$s(k) = n \cdot \exp\left(-\frac{1}{2} \cdot 10^k\right)$$

for all $k \in \{0, \ldots, K\}$. Note that we have $s(0) = n/\sqrt{e}$ and $s(K) \leq n \cdot \exp(-(1/2) \cdot (\log n)^2) < 1$.

In particular, we always have $|\mathrm{U}(\mathcal{A}_{KL})| \geq 1 \geq s(K)$. The following lemma states (roughly speaking) that for given $k = 1, \ldots, K$, if we have $|\mathrm{U}(\mathcal{A}_{kL})| \geq s(k)$, then we likely also have $|\mathrm{U}(\mathcal{A}_{(k-1)L})| \geq s(k-1)$. Iterating this will then give us a lower bound for $|\mathrm{U}(\mathcal{A}_0)|$ with high probability.

Lemma 5.3. *For every $k \in \{1, \ldots, K\}$, we have*

$$\mathbb{P}\big[|\mathrm{U}(\mathcal{A}_{(k-1)L})| < s(k-1) \text{ and } |\mathrm{U}(\mathcal{A}_{kL})| \geq s(k)\big] \leq \frac{1}{4^k}.$$

We postpone the proof of this lemma to the next section. Intuitively speaking, the relevance of the indices kL for $k = 0, \ldots, K$ appearing in Lemma 5.3 is that these are the "spaced-out" indices for which we apply Markov's inequality (recall the discussion at the end of the proof overview in Section 2). Basically, for each such index, we will apply Markov's inequality to show that if the expectation $\mathbb{E}[|U(\mathcal{A}_{kL})|]$ is small, then $|U(\mathcal{A}_{kL})|$ is small with high probability .

Assuming Lemma 5.3, we have (recalling that $|U(\mathcal{A}_{KL})| \geq s(K)$ always holds)

$$\mathbb{P}[|U(\mathcal{A}_0)| < s(0)] = \mathbb{P}\big[|U(\mathcal{A}_0)| < s(0) \text{ and } |U(\mathcal{A}_{KL})| \geq s(K)\big]$$

$$\leq \sum_{k=1}^{K} \mathbb{P}\big[|U(\mathcal{A}_{(k-1)L})| < s(k-1) \text{ and } |U(\mathcal{A}_{kL})| \geq s(k)\big] \leq \sum_{k=1}^{K} \frac{1}{4^k} \leq \frac{1}{3}.$$

Hence, $\mathbb{P}[|U(\mathcal{A}_0)| \geq n/\sqrt{e}] = \mathbb{P}[|U(\mathcal{A}_0)| \geq s(0)] \geq 2/3$, which shows (5.1) and thereby finishes the proof of Proposition 5.1. $\qquad\square$

6 Proof of Lemma 5.3

In this section, we prove Lemma 5.3, apart from postponing the proof of another lemma to the next section. As in the statement of Lemma 5.3, let $k \in \{1, \ldots, K\}$. Let $\mathcal{E}$ be the event that $|U(\mathcal{A}_{(k-1)L})| < s(k-1)$, and note that this event $\mathcal{E}$ is determined by the outcome of $\mathcal{A}_{(k-1)L}$. We need to show that

$$\mathbb{P}\big[\mathcal{E} \text{ and } |U(\mathcal{A}_{kL})| \geq s(k)\big] \leq 1/4^k.$$

This inequality is automatically satisfied if $\mathbb{P}[\mathcal{E}] \leq 1/4^k$. So let us assume from now on that $\mathbb{P}[\mathcal{E}] \geq 1/4^k$.

Note that whenever $\mathcal{E}$ holds, we have $1 \leq |U(\mathcal{A}_{(k-1)L})| < s(k-1)$, so $\mathbb{P}[\mathcal{E}] \geq 1/4^k > 0$ in particular implies $s(k-1) > 1$.

Now, assume that the conditional expectation $\mathbb{E}\big[|U(\mathcal{A}_{kL})| \,\big|\, \mathcal{E}\big]$ of $|U(\mathcal{A}_{kL})|$ conditioned on the event $\mathcal{E}$ satisfies $\mathbb{E}\big[|U(\mathcal{A}_{kL})| \,\big|\, \mathcal{E}\big] < s(k)/4^k$. Then by Markov's inequality (in the probability space obtained by conditioning on $\mathcal{E}$), we have

$$\mathbb{P}\big[|U(\mathcal{A}_{kL})| \geq s(k) \,\big|\, \mathcal{E}\big] \leq \frac{\mathbb{E}\big[|U(\mathcal{A}_{kL})| \,\big|\, \mathcal{E}\big]}{s(k)} < \frac{s(k)/4^k}{s(k)} = \frac{1}{4^k}.$$

Thus,

$$\mathbb{P}\big[\mathcal{E} \text{ and } |U(\mathcal{A}_{kL})| \geq s(k)\big] \leq \mathbb{P}\big[|U(\mathcal{A}_{kL})| \geq s(k) \,\big|\, \mathcal{E}\big] \leq \frac{1}{4^k},$$

as desired in the lemma statement.

So, in addition to $\mathbb{P}[\mathcal{E}] \geq 1/4^k$, we may from now on assume that $\mathbb{E}\big[|U(\mathcal{A}_{kL})| \,\big|\, \mathcal{E}\big] \geq s(k)/4^k$. We will show that these assumptions lead to a contradiction.

Let

$$\varepsilon = \frac{10^{k-1}}{2 \log n},$$

and observe that then $s(k-1) = n^{1-\varepsilon}$. We have $0 < \varepsilon < 1$ (since $s(k-1) > 1$).

The key step to prove Lemma 5.3 is showing the following lemma, whose proof we postpone to the next section. Basically, this lemma states that when conditioning on the event $\mathcal{E}$, for each $j = 1, \ldots, L/2-1$ we expect $|U(\mathcal{A}_{kL-j-1})|$ to be quite a bit larger than $|U(\mathcal{A}_{kL-j+1})|$, namely larger by a factor of $1 + \varepsilon/400$ apart from some error term (also recall the discussion in the proof overview in Section 2).

Lemma 6.1. *Assume that $\mathbb{P}[\mathcal{E}] \geq 1/4^k$. Then for each $j = 1, \ldots, L/2-1$, we have*

$$\mathbb{E}\big[|U(\mathcal{A}_{kL-j-1})| \,\big|\, \mathcal{E}\big] \geq \left(1 + \frac{\varepsilon}{400}\right) \cdot \mathbb{E}\big[|U(\mathcal{A}_{kL-j+1})| \,\big|\, \mathcal{E}\big] - 15 \cdot 4^k \cdot \frac{\varepsilon}{400} \cdot n \cdot \exp\left(-\frac{\varepsilon L}{128}\right).$$

This lemma leads to a contradiction as follows. First, note that

$$n \cdot \exp\left(-\frac{\varepsilon L}{128}\right) \leq n \cdot \exp\left(-\frac{10^{k-1}}{2 \log n} \cdot \frac{10^5 \log n}{128}\right) \leq n \cdot \exp\left(-\frac{1}{2} \cdot 3 \cdot 10^k\right) \leq n \cdot \exp\left(-\frac{1}{2} \cdot 10^k\right) \cdot e^{-10k} \leq \frac{s(k)}{60 \cdot 16^k}.$$

13

Recalling our assumption $\mathbb{E}\big[|\mathrm{U}(\mathcal{A}_{kL})|\,\big|\,\mathcal{E}\big] \geq s(k)/4^k$, this implies

$$15 \cdot 4^k \cdot \frac{\varepsilon}{400} \cdot n \cdot \exp\left(-\frac{\varepsilon L}{128}\right) \leq \frac{\varepsilon}{1600} \cdot \frac{s(k)}{4^k} \leq \frac{\varepsilon}{1600} \cdot \mathbb{E}\big[|\mathrm{U}(\mathcal{A}_{kL})|\,\big|\,\mathcal{E}\big] \leq \frac{\varepsilon}{1600} \cdot \mathbb{E}\big[|\mathrm{U}(\mathcal{A}_{kL-j+1})|\,\big|\,\mathcal{E}\big]$$

for each $j = 1, \ldots, L/2 - 1$. Thus, the inequality in Lemma 6.1 implies

$$\mathbb{E}\big[|\mathrm{U}(\mathcal{A}_{kL-j-1})|\,\big|\,\mathcal{E}\big] \geq \left(1 + \frac{3\varepsilon}{1600}\right) \cdot \mathbb{E}\big[|\mathrm{U}(\mathcal{A}_{kL-j+1})|\,\big|\,\mathcal{E}\big] \geq \left(1 + \frac{\varepsilon}{600}\right) \cdot \mathbb{E}\big[|\mathrm{U}(\mathcal{A}_{kL-j+1})|\,\big|\,\mathcal{E}\big]$$

for each $j = 1, \ldots, L/2 - 1$. By applying this iteratively to $j = 1, 3, 5, \ldots, L/2 - 1$, we can deduce that

$$\mathbb{E}\big[|\mathrm{U}(\mathcal{A}_{kL-L/2})|\,\big|\,\mathcal{E}\big] \geq \left(1 + \frac{\varepsilon}{600}\right)^{L/4} \cdot \mathbb{E}\big[|\mathrm{U}(\mathcal{A}_{kL})|\,\big|\,\mathcal{E}\big] \geq \left(1 + \frac{\varepsilon}{600}\right)^{L/4} \cdot \frac{s(k)}{4^k}.$$

Note that for all real numbers $0 \leq t \leq 1$ we have $1 + t \geq \exp(t/2)$. Hence, if n (and hence K) is sufficiently large, we have

$$\left(1 + \frac{\varepsilon}{600}\right)^{L/4} \geq \exp\left(\frac{\varepsilon}{1200} \cdot \frac{L}{4}\right) \geq \exp\left(\frac{10^{k-1}}{2\log n} \cdot \frac{10^5 \log n}{4 \cdot 1200}\right) \geq \exp(10^k) \geq \exp\left(\frac{1}{2} \cdot 10^k\right) \cdot 2 \cdot 4^k.$$

So we obtain

$$\mathbb{E}\big[|\mathrm{U}(\mathcal{A}_{kL-L/2})|\,\big|\,\mathcal{E}\big] \geq \left(1 + \frac{\varepsilon}{600}\right)^{L/4} \cdot \frac{s(k)}{4^k} \geq \exp\left(\frac{1}{2} \cdot 10^k\right) \cdot 2 \cdot 4^k \cdot \frac{s(k)}{4^k} = 2 \cdot \exp\left(\frac{1}{2} \cdot 10^k\right) \cdot s(k) = 2n.$$

But since always $|\mathrm{U}(\mathcal{A}_{kL-L/2})| \leq |V(G)| \leq n$, we must clearly have $\mathbb{E}\big[|\mathrm{U}(\mathcal{A}_{kL-L/2})|\,\big|\,\mathcal{E}\big] \leq n$, so we indeed obtain a contradiction. This contradiction finishes the proof of Lemma 5.3, apart from proving Lemma 6.1.

7   Proof of Lemma 6.1

In this section, we prove Lemma 6.1 which is the last missing piece for the proof of Proposition 5.1 (and hence for the proof of our main theorem).

Let $j \in \{1, \ldots, L/2 - 1\}$. Recall that we chose $k \in \{1, \ldots, K\}$ at the start of the previous section, and also recall the definitions of the event $\mathcal{E}$ and the number $0 < \varepsilon < 1$ in the previous section (and recall that $s(k-1) = n^{1-\varepsilon}$). As in the statement of Lemma 6.1, let us assume that $\mathbb{P}[\mathcal{E}] \geq 1/4^k$.

Let us consider the event $\mathcal{F}$, determined by $\mathcal{A}_{(k-1)L}$ and $\mathcal{A}_{kL-j}$, defined as follows. Let us say that the event $\mathcal{F}$ holds if both of the following two conditions are satisfied:

(i) There exists a set $F \subseteq E(G)$ with $|F| \geq (\varepsilon/7) \cdot d(G) \cdot |\mathrm{U}(\mathcal{A}_{kL-j})|$ and $\deg_F(v') \leq \lceil d(G) \rceil$ for all $v' \in V(G) \setminus \mathrm{U}(\mathcal{A}_{kL-j})$, such that each edge in F is of the form (v, v') with $v \in \mathrm{U}(\mathcal{A}_{kL-j} \setminus \{\gamma(v, v')\}) \subseteq \mathrm{U}(\mathcal{A}_{kL-j})$ and $v' \notin \mathrm{U}(\mathcal{A}_{kL-j})$.

(ii) There does not exist a set $F' \subseteq E(G)$ with $|F'| \geq (\varepsilon/100) \cdot L \cdot |\mathrm{U}(\mathcal{A}_{kL-j})|$ and $\deg_{F'}(v') \leq 2L$ for all $v' \in V(G) \setminus \mathrm{U}(\mathcal{A}_{kL-j})$, such that each edge in F' is of the form (v, v') with $v \in \mathrm{U}(\mathcal{A}_{kL-j} \setminus \{\gamma(v, v')\}) \subseteq \mathrm{U}(\mathcal{A}_{kL-j})$ and $v' \notin \mathrm{U}(\mathcal{A}_{kL-j})$ and satisfies $\gamma(v, v') \in \mathcal{A}_{(k-1)L}$.

Let $\overline{\mathcal{F}}$ denote the complementary event to $\mathcal{F}$, i.e. the event that $\mathcal{F}$ does not hold. We will deduce the inequality in Lemma 6.1 from the following three claims. Intuitively speaking, the first claim says that the event $\mathcal{F}$ is very unlikely, the second claim says that if $\mathcal{F}$ does not hold (which is normally the case) then "in expectation" we have $|\mathrm{U}(\mathcal{A}_{kL-j-1})| \geq (1 + \varepsilon/400) \cdot |\mathrm{U}(\mathcal{A}_{kL-j+1})|$ and the third claim specifies an error term for this inequality in the unlikely case that $\mathcal{F}$ holds.

Claim 7.1. *We have*

$$\mathbb{P}[\mathcal{F}] \leq 15 \exp\left(-\frac{\varepsilon L}{128}\right).$$

Claim 7.2. *We have $\mathbb{P}[\mathcal{E} \text{ and } \overline{\mathcal{F}}] > 0$ and*

$$\mathbb{E}\left[|\mathrm{U}(\mathcal{A}_{kL-j-1})| - \left(1 + \frac{\varepsilon}{400}\right) \cdot |\mathrm{U}(\mathcal{A}_{kL-j+1})| \,\Big|\, \mathcal{E}, \overline{\mathcal{F}}\right] \geq 0.$$

14

Claim 7.3. *If $\mathbb{P}[\mathcal{E} \text{ and } \mathcal{F}] > 0$, then*

$$\mathbb{E}\left[|\mathrm{U}(\mathcal{A}_{kL-j-1})| - \left(1 + \frac{\varepsilon}{400}\right) \cdot |\mathrm{U}(\mathcal{A}_{kL-j+1})| \,\Big|\, \mathcal{E}, \mathcal{F}\right] \geq -\frac{\varepsilon}{400} \cdot n.$$

Let us now show that these three claims indeed imply the inequality in Lemma 6.1. Recalling our assumption that $\mathbb{P}[\mathcal{E}] \geq 1/4^k$, in the case that $\mathbb{P}[\mathcal{E} \text{ and } \mathcal{F}] > 0$ we obtain

$$\mathbb{E}\left[|\mathrm{U}(\mathcal{A}_{kL-j-1})| - \left(1 + \frac{\varepsilon}{400}\right) \cdot |\mathrm{U}(\mathcal{A}_{kL-j+1})| \,\Big|\, \mathcal{E}\right]$$

$$= \mathbb{E}\left[|\mathrm{U}(\mathcal{A}_{kL-j-1})| - \left(1 + \frac{\varepsilon}{400}\right) \cdot |\mathrm{U}(\mathcal{A}_{kL-j+1})| \,\Big|\, \mathcal{E}, \mathcal{F}\right] \cdot \mathbb{P}[\mathcal{F} \mid \mathcal{E}]$$

$$+ \mathbb{E}\left[|\mathrm{U}(\mathcal{A}_{kL-j-1})| - \left(1 + \frac{\varepsilon}{400}\right) \cdot |\mathrm{U}(\mathcal{A}_{kL-j+1})| \,\Big|\, \mathcal{E}, \overline{\mathcal{F}}\right] \cdot \mathbb{P}[\overline{\mathcal{F}} \mid \mathcal{E}]$$

$$\geq -\frac{\varepsilon}{400} \cdot n \cdot \mathbb{P}[\mathcal{F} \mid \mathcal{E}] + 0 \cdot \mathbb{P}[\overline{\mathcal{F}} \mid \mathcal{E}]$$

$$\geq -\frac{\varepsilon}{400} \cdot n \cdot \frac{\mathbb{P}[\mathcal{F}]}{\mathbb{P}[\mathcal{E}]} \geq -15 \cdot 4^k \cdot \frac{\varepsilon}{400} \cdot n \cdot \exp\left(-\frac{\varepsilon L}{128}\right).$$

In the case $\mathbb{P}[\mathcal{E} \text{ and } \mathcal{F}] = 0$, we simply have

$$\mathbb{E}\left[|\mathrm{U}(\mathcal{A}_{kL-j-1})| - \left(1 + \frac{\varepsilon}{400}\right) \cdot |\mathrm{U}(\mathcal{A}_{kL-j+1})| \,\Big|\, \mathcal{E}\right] = \mathbb{E}\left[|\mathrm{U}(\mathcal{A}_{kL-j-1})| - \left(1 + \frac{\varepsilon}{400}\right) \cdot |\mathrm{U}(\mathcal{A}_{kL-j+1})| \,\Big|\, \mathcal{E}, \overline{\mathcal{F}}\right] \geq 0.$$

So in either case we obtain

$$\mathbb{E}\left[|\mathrm{U}(\mathcal{A}_{kL-j-1})| - \left(1 + \frac{\varepsilon}{400}\right) \cdot |\mathrm{U}(\mathcal{A}_{kL-j+1})| \,\Big|\, \mathcal{E}\right] \geq -15 \cdot 4^k \cdot \frac{\varepsilon}{400} \cdot n \cdot \exp\left(-\frac{\varepsilon L}{128}\right),$$

which by linearity of expectation and rearranging yields the inequality in the statement of Lemma 6.1.

It remains to prove Claims 7.1 to 7.3. We start by proving Claim 7.3, which is the easiest.

Proof of Claim 7.3. Note that we always have $\mathrm{U}(\mathcal{A}_{kL-j-1}) \supseteq \mathrm{U}(\mathcal{A}_{kL-j+1})$ and therefore

$$|\mathrm{U}(\mathcal{A}_{kL-j-1})| - \left(1 + \frac{\varepsilon}{400}\right) \cdot |\mathrm{U}(\mathcal{A}_{kL-j+1})| \geq -\frac{\varepsilon}{400} \cdot |\mathrm{U}(\mathcal{A}_{kL-j+1})| \geq -\frac{\varepsilon}{400} \cdot n.$$

This clearly implies the inequality for the conditional expectation in the claim (the assumption $\mathbb{P}[\mathcal{E} \text{ and } \mathcal{F}] > 0$ ensures that this conditional expectation is well-defined). $\qquad\square$

Next, let us prove Claim 7.1.

Proof of Claim 7.1. It suffices to prove the conditional probability bound $\mathbb{P}[\mathcal{F} \mid \mathcal{A}_{kL-j}] \leq 15 \exp(-\varepsilon L/128)$ for every possible outcome of $\mathcal{A}_{kL-j}$. So let us fix any outcome of $\mathcal{A}_{kL-j}$.

Note that condition (i) in the definition of the event $\mathcal{F}$ only depends on $\mathcal{A}_{kL-j}$ (and not on $\mathcal{A}_{(k-1)L}$). If our fixed outcome $\mathcal{A}_{kL-j}$ does not satisfy (i), then $\mathbb{P}[\mathcal{F} \mid \mathcal{A}_{kL-j}] = 0$. Hence, we may assume that the fixed outcome of $\mathcal{A}_{kL-j}$ satisfies (i). Let us fix a subset $F \subseteq E(G)$ as in condition (i). Let $U = \mathrm{U}(\mathcal{A}_{kL-j})$, then $|F| \geq (\varepsilon/7) \cdot d(G) \cdot |U| \geq (\varepsilon/8) \cdot \lceil d(G) \rceil \cdot |U|$ and every edge in F is of the form (v, v') with $v \in U$ and $v' \in V(G) \setminus U$ and $v \in \mathrm{U}(\mathcal{A}_{kL-j} \setminus \{\gamma(v, v')\})$. Furthermore, we have $\deg_F(v') \leq \lceil d(G) \rceil$ for all $v' \in V(G) \setminus U$.

We can now apply Lemma 4.6 to the bipartite graph with edge set F between the vertex sets U and $V(G) \setminus U$, where for each edge $(v, v') \in F$ we consider the event that $\gamma(v, v') \in \mathcal{A}_{(k-1)L}$. These events are independent for edges of different colours, and so in particular they are independent for edges at the same vertex. Furthermore, for every edge $(v, v') \in F$, conditioned on our fixed outcome of $\mathcal{A}_{kL-j}$, by Lemma 4.2(a) (applied with $p = 1 - 1/T$) we have $\gamma(v, v') \in \mathcal{A}_{(k-1)L}$ with probability at least $(L-j)/(6T) \geq L/(12T) \geq L/\lceil d(G) \rceil$ (recall that $j \leq L/2$ and $d(G) \geq 12T$).

Lemma 4.6 now shows that with probability at least $1 - 15 \exp(-(\varepsilon/8)L/16) = 1 - 15 \exp(-\varepsilon L/128)$ there exists a subset $F' \subseteq F \subseteq E(G)$ of size $|F'| \geq \varepsilon/(8 \cdot 12) \cdot L \cdot |U| \geq (\varepsilon/100) \cdot L \cdot |\mathrm{U}(\mathcal{A}_{kL-j})|$ such that $\gamma(v, v') \in \mathcal{A}_{(k-1)L}$ for all $(v, v') \in F'$ and $\deg_{F'}(v') \leq 2L$ for all $v' \in V(G) \setminus U$. Since $F' \subseteq F$, every edge in F' is of the form (v, v') with $v \in \mathrm{U}(\mathcal{A}_{kL-j} \setminus \{\gamma(v, v')\}) \subseteq \mathrm{U}(\mathcal{A}_{kL-j})$ and $v' \in V(G) \setminus \mathrm{U}(\mathcal{A}_{kL-j})$. Hence, F' is as in condition (ii) in the definition of the event $\mathcal{F}$ and so $\mathcal{F}$ does not hold if such a set F' exists. This shows that $\mathcal{F}$ holds with probability at most $15 \exp(-\varepsilon L/128)$. $\qquad\square$

It remains to prove Claim 7.2.

Proof of Claim 7.2. For the first part of the claim recall our assumption $\mathbb{P}[\mathcal{E}] \geq 1/4^k$ and recall from Claim 7.1 that

$$\mathbb{P}[\mathcal{F}] \leq 15 \exp\left(-\frac{\varepsilon L}{128}\right) \leq 15 \exp\left(-\frac{10^{k-1}}{2 \cdot \log n} \cdot \frac{10^5 \log n}{128}\right) \leq 15 e^{-10k} < 1/4^k.$$

Hence, $\mathbb{P}[\mathcal{E}] > \mathbb{P}[\mathcal{F}]$ and therefore $\mathbb{P}[\mathcal{E} \text{ and } \overline{\mathcal{F}}] > 0$.

For the second part of the claim, we need to prove that the expectation of $|\mathrm{U}(\mathcal{A}_{kL-j-1})| - (1 + \varepsilon/400) \cdot |\mathrm{U}(\mathcal{A}_{kL-j+1})|$ conditional on $\mathcal{E}$ and $\overline{\mathcal{F}}$ is non-negative. It suffices to prove that the expectation of this quantity is non-negative upon further conditioning on any given outcomes of $\mathcal{A}_{(k-1)L}$ and $\mathcal{A}_{kL-j}$ satisfying $\mathcal{E}$ and $\overline{\mathcal{F}}$. In other words, it suffices to prove that

$$\mathbb{E}\left[|\mathrm{U}(\mathcal{A}_{kL-j-1})| - \left(1 + \frac{\varepsilon}{400}\right) \cdot |\mathrm{U}(\mathcal{A}_{kL-j+1})| \,\Big|\, \mathcal{A}_{(k-1)L}, \mathcal{A}_{kL-j}\right] \geq 0$$

for every possible outcomes of $\mathcal{A}_{(k-1)L}$ and $\mathcal{A}_{kL-j}$ satisfying $\mathcal{E}$ and $\overline{\mathcal{F}}$. So let us fix $\mathcal{A}_{(k-1)L} \supseteq \mathcal{A}_{kL-j}$ satisfying $\mathcal{E}$ and $\overline{\mathcal{F}}$. Let $U = \mathrm{U}(\mathcal{A}_{kL-j})$.

Since the event $\mathcal{E}$ holds, we have $|U| = |\mathrm{U}(\mathcal{A}_{kL-j})| \leq |\mathrm{U}(\mathcal{A}_{(k-1)L})| < s(k-1) = n^{1-\varepsilon}$. We will now apply Lemma 4.1 with the following auxiliary colouring of the edges of G between U and $V(G) \setminus U$ with the colours red and blue (assume that red and blue are not part of the colour palette $\mathcal{C}$ of our original colouring $\gamma : E(G) \to \mathcal{C}$). For an edge $(v, v') \in E(G)$ with $v \in U = \mathrm{U}(\mathcal{A}_{kL-j})$ and $v' \in V(G) \setminus U$, let us colour (v, v') red in the auxiliary colouring if $v \notin \mathrm{U}(\mathcal{A}_{kL-j} \setminus \{\gamma(v, v')\})$ and let us colour (v, v') blue in the auxiliary colouring if $v \in \mathrm{U}(\mathcal{A}_{kL-j} \setminus \{\gamma(v, v')\})$. Now, at least one of the two alternatives in Lemma 4.1 must hold, and we split into cases accordingly.

Case 1: The first alternative in Lemma 4.1 holds. In this case, there exists a subset $F_{\mathrm{red}} \subseteq E(G)$ of red edges in G with $|F_{\mathrm{red}}| \geq (\varepsilon/7) \cdot d(G) \cdot |U|$ and $\deg_{F_{\mathrm{red}}}(v) \leq \lceil d(G) \rceil$ for all $v \in U$.

After conditioning on $\mathcal{A}_{(k-1)L}$ and $\mathcal{A}_{kL-j}$, for every $(v, v') \in F_{\mathrm{red}}$ the probability of having $\gamma(v, v') \notin \mathcal{A}_{kL-j+1}$ is at least $1/T \geq 1/\lceil d(G) \rceil$ (since each colour in $\mathcal{A}_{kL-j}$ is missing in $\mathcal{A}_{kL-j+1}$ with probability $1/T$, and each colour that is not in $\mathcal{A}_{kL-j}$ is missing in $\mathcal{A}_{kL-j+1}$ with probability 1). Furthermore, these events $\gamma(v, v') \notin \mathcal{A}_{kL-j+1}$ are independent for edges (v, v') of different colours (even after conditioning on $\mathcal{A}_{(k-1)L}$ and $\mathcal{A}_{kL-j}$).

Thus, for each $v \in U$ we can apply Lemma 4.3 to the events $\gamma(v, v') \notin \mathcal{A}_{kL-j+1}$ for the $\deg_{F_{\mathrm{red}}}(v) \leq \lceil d(G) \rceil$ different edges $(v, v') \in F_{\mathrm{red}}$. By the lemma, with probability at least $\deg_{F_{\mathrm{red}}}(v)/(2\lceil d(G) \rceil)$, we have $\gamma(v, v') \notin \mathcal{A}_{kL-j+1}$ for at least one edge $(v, v') \in F_{\mathrm{red}}$. Thus, for every $v \in U$, we showed that

$$\mathbb{P}\big[\gamma(v, v') \notin \mathcal{A}_{kL-j+1} \text{ for some edge } (v, v') \in F_{\mathrm{red}} \text{ at } v \,\big|\, \mathcal{A}_{(k-1)L}, \mathcal{A}_{kL-j}\big] \geq \frac{\deg_{F_{\mathrm{red}}}(v)}{2\lceil d(G) \rceil}. \tag{7.1}$$

We claim that if $(v, v') \in F_{\mathrm{red}}$ (with $v \in U$ and $v' \in V(G) \setminus U$) satisfies $\gamma(v, v') \notin \mathcal{A}_{kL-j+1}$, then $v \notin \mathrm{U}(\mathcal{A}_{kL-j+1})$. Indeed, under the assumption $\gamma(v, v') \notin \mathcal{A}_{kL-j+1}$ we have $\mathcal{A}_{kL-j+1} \subseteq \mathcal{A}_{kL-j} \setminus \{\gamma(v, v')\}$ and hence $\mathrm{U}(\mathcal{A}_{kL-j+1}) \subseteq \mathrm{U}(\mathcal{A}_{kL-j} \setminus \{\gamma(v, v')\})$. Since (v, v') is red in the auxiliary colouring, we know that $v \notin \mathrm{U}(\mathcal{A}_{kL-j} \setminus \{\gamma(v, v')\})$ and can conclude that $v \notin \mathrm{U}(\mathcal{A}_{kL-j+1})$.

Thus, (7.1) implies

$$\mathbb{P}\big[v \notin \mathrm{U}(\mathcal{A}_{kL-j+1}) \,\big|\, \mathcal{A}_{(k-1)L}, \mathcal{A}_{kL-j}\big] \geq \frac{\deg_{F_{\mathrm{red}}}(v)}{2\lceil d(G) \rceil}.$$

for every $v \in U$. Summing this up for all $v \in U$ yields

$$\mathbb{E}\big[|U \setminus \mathrm{U}(\mathcal{A}_{kL-j+1})| \,\big|\, \mathcal{A}_{(k-1)L}, \mathcal{A}_{kL-j}\big] = \sum_{v \in U} \mathbb{P}\big[v \notin \mathrm{U}(\mathcal{A}_{kL-j+1}) \,\big|\, \mathcal{A}_{(k-1)L}, \mathcal{A}_{kL-j}\big]$$

$$\geq \sum_{v \in U} \frac{\deg_{F_{\mathrm{red}}}(v)}{2\lceil d(G) \rceil} = \frac{|F_{\mathrm{red}}|}{2\lceil d(G) \rceil} \geq \frac{(\varepsilon/7) \cdot d(G) \cdot |U|}{2\lceil d(G) \rceil} \geq \frac{\varepsilon}{15} \cdot |U|$$

(where in the last step we used that $\lceil d(G) \rceil \leq 15/14 \cdot d(G)$ since n is sufficiently large). Therefore we obtain (using that always $\mathrm{U}(\mathcal{A}_{kL-j+1}) \subseteq \mathrm{U}(\mathcal{A}_{kL-j}) = U$)

$$\mathbb{E}\big[|\mathrm{U}(\mathcal{A}_{kL-j+1})| \,\big|\, \mathcal{A}_{(k-1)L}, \mathcal{A}_{kL-j}\big] = |U| - \mathbb{E}\big[|U \setminus \mathrm{U}(\mathcal{A}_{kL-j+1})| \,\big|\, \mathcal{A}_{(k-1)L}, \mathcal{A}_{kL-j}\big] \leq \left(1 - \frac{\varepsilon}{15}\right) \cdot |U|.$$

On the other hand, we always have $U(\mathcal{A}_{kL-j-1}) \supseteq U(\mathcal{A}_{kL-j}) = U$ and hence $\mathbb{E}\big[|U(\mathcal{A}_{kL-j-1})| \mid \mathcal{A}_{(k-1)L}, \mathcal{A}_{kL-j}\big] \geq |U|$. Thus,

$$\mathbb{E}\Big[|U(\mathcal{A}_{kL-j-1})| - \Big(1 + \frac{\varepsilon}{400}\Big) \cdot |U(\mathcal{A}_{kL-j+1})| \;\Big|\; \mathcal{A}_{(k-1)L}, \mathcal{A}_{kL-j}\Big] \geq |U| - \Big(1 + \frac{\varepsilon}{400}\Big) \cdot \Big(1 - \frac{\varepsilon}{15}\Big) \cdot |U| \geq 0,$$

as desired.

Case 2: The second alternative in Lemma 4.1 holds. In this case, there exists a subset $F_{\text{blue}} \subseteq E(G)$ of blue edges in G with $|F_{\text{blue}}| \geq (\varepsilon/7) \cdot d(G) \cdot |U|$ and $\deg_{F_{\text{blue}}}(v') \leq \lceil d(G) \rceil$ for all $v' \in V(G) \setminus U$. Note that this set F_{blue} satisfies condition (i) in the definition of the event $\mathcal{F}$. Indeed, for the last part of condition (i) recall that every edge in F_{blue} is of the form (v, v') with $v \in U = U(\mathcal{A}_{kL-j})$ and $v' \notin U(\mathcal{A}_{kL-j})$, and also $v \in U(\mathcal{A}_{kL-j} \setminus \{\gamma(v, v')\})$ since (v, v') is blue.

Now, since we chose outcomes of $\mathcal{A}_{(k-1)L}$ and $\mathcal{A}_{kL-j}$ that do not satisfy $\mathcal{F}$, condition (ii) in the definition of $\mathcal{F}$ must fail. This means that there exists a set $F' \subseteq E(G)$ with $|F'| \geq (\varepsilon/100) \cdot L \cdot |U|$ and $\deg_{F'}(v') \leq 2L$ for all $v' \in V(G) \setminus U$, such that each edge in F' is of the form (v, v') with $v \in U(\mathcal{A}_{kL-j} \setminus \{\gamma(v, v')\}) \subseteq U$ and $v' \notin U$ and satisfies $\gamma(v, v') \in \mathcal{A}_{(k-1)L}$.

We claim for every $(v, v') \in F'$ the probability (conditioning on $\mathcal{A}_{(k-1)L}$ and $\mathcal{A}_{kL-j}$) of having $\gamma(v, v') \in \mathcal{A}_{kL-j-1}$ is at least $1/(2L)$. Indeed, since $(v, v') \in F'$, we have $\gamma(v, v') \in \mathcal{A}_{(k-1)L}$. So by Lemma 4.2(b) (applied with $p = 1 - 1/T$), the probability of having $\gamma(v, v') \in \mathcal{A}_{kL-j-1}$ is at least $T/(L - j) \cdot 1/(2T) \geq 1/(2L)$. Furthermore, these events $\gamma(v, v') \in \mathcal{A}_{kL-j-1}$ are independent for edges (v, v') of different colours (even after conditioning on $\mathcal{A}_{(k-1)L}$ and $\mathcal{A}_{kL-j}$).

Now, for each $v' \in V(G) \setminus U$ we can apply Lemma 4.3 to the events $\gamma(v, v') \in \mathcal{A}_{kL-j-1}$ for the $\deg_{F'}(v') \leq 2L$ different edges $(v, v') \in F'$. By Lemma 4.3, with probability at least $\deg_{F'}(v')/(4L)$, we have $\gamma(v, v') \in \mathcal{A}_{kL-j-1}$ for at least one edge $(v, v') \in F'$. Thus, for every $v' \in V(G) \setminus U$, we showed that

$$\mathbb{P}\big[\gamma(v, v') \in \mathcal{A}_{kL-j-1} \text{ for some edge } (v, v') \in F' \text{ at } v' \mid \mathcal{A}_{(k-1)L}, \mathcal{A}_{kL-j}\big] \geq \frac{\deg_{F'}(v)}{4L}. \tag{7.2}$$

We claim that if $(v, v') \in F'$ (with $v \in U$ and $v' \in V(G) \setminus U$) satisfies $\gamma(v, v') \in \mathcal{A}_{kL-j-1}$, then $v' \in U(\mathcal{A}_{kL-j-1})$. Indeed, because $(v, v') \in F'$, by the choice of F' we have $v \in U(\mathcal{A}_{kL-j} \setminus \{\gamma(v, v')\})$. This means that there exists a rainbow walk from x to v with colours in $\mathcal{A}_{kL-j} \setminus \{\gamma(v, v')\} \subseteq \mathcal{A}_{kL-j-1} \setminus \{\gamma(v, v')\}$. Adding the edge (v, v') gives a rainbow walk from x to v' with colours in $\mathcal{A}_{kL-j-1}$ (here, we used that $\gamma(v, v') \in \mathcal{A}_{kL-j-1}$). Thus, indeed $v' \in U(\mathcal{A}_{kL-j-1})$.

Thus, (7.2) implies

$$\mathbb{P}\big[v' \in U(\mathcal{A}_{kL-j-1}) \mid \mathcal{A}_{(k-1)L}, \mathcal{A}_{kL-j}\big] \geq \frac{\deg_{F'}(v')}{4L}.$$

for every $v' \in V(G) \setminus U$. Summing this up for all $v' \in V(G) \setminus U$ yields

$$\begin{aligned}
\mathbb{E}\big[|U(\mathcal{A}_{kL-j-1}) \setminus U| \mid \mathcal{A}_{(k-1)L}, \mathcal{A}_{kL-j}\big] &= \sum_{v' \in V(G) \setminus U} \mathbb{P}\big[v' \in U(\mathcal{A}_{kL-j-1}) \mid \mathcal{A}_{(k-1)L}, \mathcal{A}_{kL-j}\big] \\
&\geq \sum_{v' \in V(G) \setminus U} \frac{\deg_{F'}(v')}{4L} = \frac{|F'|}{4L} \geq \frac{(\varepsilon/100) \cdot L \cdot |U|}{4L} \geq \frac{\varepsilon}{400} \cdot |U|.
\end{aligned}$$

Therefore we obtain (using that always $U(\mathcal{A}_{kL-j-1}) \supseteq U(\mathcal{A}_{kL-j}) = U$)

$$\mathbb{E}\big[|U(\mathcal{A}_{kL-j-1})| \mid \mathcal{A}_{(k-1)L}, \mathcal{A}_{kL-j}\big] = |U| + \mathbb{E}\big[|U(\mathcal{A}_{kL-j-1}) \setminus U| \mid \mathcal{A}_{(k-1)L}, \mathcal{A}_{kL-j}\big] \geq \Big(1 + \frac{\varepsilon}{400}\Big) \cdot |U|.$$

On the other hand, we always have $U(\mathcal{A}_{kL-j+1}) \subseteq U(\mathcal{A}_{kL-j}) = U$ and hence $\mathbb{E}\big[|U(\mathcal{A}_{kL-j+1})| \mid \mathcal{A}_{(k-1)L}, \mathcal{A}_{kL-j}\big] \leq |U|$. Thus,

$$\mathbb{E}\Big[|U(\mathcal{A}_{kL-j-1})| - \Big(1 + \frac{\varepsilon}{400}\Big) \cdot |U(\mathcal{A}_{kL-j+1})| \;\Big|\; \mathcal{A}_{(k-1)L}, \mathcal{A}_{kL-j}\Big] \geq \Big(1 + \frac{\varepsilon}{400}\Big) \cdot |U| - \Big(1 + \frac{\varepsilon}{400}\Big) \cdot |U| \geq 0,$$

as desired.

This finishes the proof of Claim 7.2 in both cases. $\qquad\square$

This completes the proof of Lemma 6.1 and hence of Proposition 5.1.

8 Almost rainbow cycles

A cycle in a properly edge-coloured graph is said to be *r-almost rainbow* if there are at most r edges of each colour on the cycle. So, for example, a 1-almost rainbow cycle is the same as a rainbow cycle, and any cycle of length ℓ is trivially ℓ-almost rainbow. This offers a natural intermediate notion between a rainbow cycle and an unrestricted cycle, shedding light on the transition from the simple fact that n edges in an n-vertex graph guarantee a cycle to the fact that at least $\Omega(n \log n)$ edges are needed to guarantee a rainbow cycle. Somewhat surprisingly, the behaviour is already slightly different if we allow colours to be used twice rather than once, as shown in Proposition 1.2 for $r = 2$. As we allow colours to be used more and more times, we observe a gradual progression up to allowing $\Theta(\log n)$ edges of any colour, at which point already a linear number of edges suffices. This transition is described by Proposition 1.2, which we prove below.

Recall that $\mathrm{Tr}_r(n, \mathcal{C})$ denotes the maximal possible number of edges in a properly edge-coloured n-vertex graph without an r-almost rainbow cycle.

Proof of Proposition 1.2. As in the statement of the proposition, let $2 \le r \le \frac{1}{10} \log n$. Let us start with the lower bound for $\mathrm{Tr}_r(n, \mathcal{C})$. By a classical result of Erdős and Sachs [19, p. 252], for any integers $d \ge 2$ and $r \ge 2$ and $n \ge 4d^{(d+1)r}$, there exists an n-vertex graph with girth at least $(d+1)r + 1$, maximum degree d and at least $\Omega(nd)$ edges (in fact, they proved that for even $n \ge 4d^{(d+1)r}$ there is a d-regular graph on n vertices with girth at least $(d+1)r + 1$). By Vizings's theorem, this graph has a proper edge-colouring using at most $d+1$ colours. In this colouring, every cycle contains some colour at least $r+1$ times (since the cycle has length more than $(d+1)r$ and there are at most $d+1$ colours in total). Choosing $d := \left\lfloor \frac{\log n}{2r} / \log \frac{\log n}{r} \right\rfloor \ge 2$ (recalling that $r \le \frac{1}{10} \log n$), we obtain (noting that then we indeed have $4d^{(d+1)r} \le 4d^{(3/2)dr} \le 4n^{3/4} \le n$)

$$\mathrm{Tr}_r(n, \mathcal{C}) \ge \Omega(nd) \ge \Omega\left(n \cdot \frac{(\log n)/r}{\log((\log n)/r)} \right).$$

Turning to the upper bound, let $d := 16 \left\lceil \frac{\log n}{r} / \log \frac{\log n}{r} \right\rceil$, so that $\frac{1}{8} rd \log \frac{d}{2} > \log n$ and hence $(d/2)^{dr/8} > n$. It suffices to show that $\mathrm{Tr}_r(n, \mathcal{C}) \le nd$, since then we have

$$\mathrm{Tr}_r(n, \mathcal{C}) \le nd = 16n \left\lceil \frac{\log n}{r} / \log \frac{\log n}{r} \right\rceil \le O\left(n \cdot \frac{(\log n)/r}{\log((\log n)/r)} \right),$$

recalling that $(\log n)/r \ge 10$. To show $\mathrm{Tr}_r(n, \mathcal{C}) \le nd$, we need to show that in any properly edge-coloured graph with average degree more than $2d$ there exists a cycle using each colour at most r times. To this end, let us repeatedly remove any vertex of degree at most d until we are left with a (non-empty) graph G of minimum degree at least $d+1$. Now fix some vertex $v_0 \in V(G)$. We call a sequence of colours $(c_1, \ldots, c_\ell)$ *admissible* if no colour appears more than $r/2$ many times in the sequence, $c_i \ne c_{i+1}$ for all $i = 1, \ldots, \ell - 1$, and there exists a walk $v_0 v_1 \ldots v_\ell$ in the graph G (starting at v_0) such that the edge $v_{i-1} v_i$ has colour c_i for all $i = 1, \ldots, \ell$. We note that such a walk is unique for any admissible sequence and we refer to it as the corresponding walk and to its endpoint v_ℓ as the corresponding endpoint of the admissible sequence. It now suffices to show that there are more than n different admissible sequences. Indeed, then there must be two different admissible sequences with the same corresponding endpoint. The union of all edges of the corresponding walks of these two sequences contains at most r edges of each colour, so it includes an r-almost rainbow cycle unless this union of the two walks forms a tree. However, if this union was a tree, then both of the walks would need to be the unique path in this tree from v_0 to the common endpoint since we forbid backtracking by insisting that $c_i \ne c_{i-1}$. Since the two walks cannot be identical, this is a contradiction, so the union of the two walks cannot form a tree and we have an r-almost rainbow cycle as desired.

To see that we have more than n different admissible sequences, let us define $L = \frac{d}{2} \cdot \lfloor \frac{r}{2} \rfloor$, and show by induction that there are at least $(d/2)^\ell$ admissible sequences for any given length $\ell = 1, \ldots, L$. For $\ell = 1$, any of the colours appearing on the at least $d+1$ edges incident to v_0 gives rise to a length-1 admissible sequence, establishing the base case of the induction. Now suppose that we have at least $(d/2)^{\ell-1}$ admissible sequences of length $\ell - 1$. For each such sequence $(c_1, \ldots, c_{\ell-1})$, the corresponding endpoint $v_{\ell-1}$ is incident to at least d edges with distinct colours other than $c_{\ell-1}$. At most $d/2$ of these at least d colours can appear $\lfloor r/2 \rfloor$ times among $(c_1, \ldots, c_{\ell-1})$. Any other colour can be used to extend our sequence into an admissible sequence of length ℓ, implying that there are indeed at least $(d - d/2) \cdot (d/2)^{\ell-1} = (d/2)^\ell$ admissible sequences of length ℓ. Finally, observe that by $r \ge 2$ we have $L = \frac{d}{2} \cdot \lfloor \frac{r}{2} \rfloor \ge dr/8$, so in total we have at least $(d/2)^L \ge (d/2)^{dr/8} > n$ different admissible sequences. $\qquad\square$

Note that Proposition 1.2 shows that for any $r \geq 2$, the r-almost rainbow Turán number $\mathrm{Tr}_r(n, \mathcal{C})$ is by at least a $\Theta(\log \log n)$ factor smaller than the rainbow Turán number $\mathrm{Tr}_1(n, \mathcal{C})$. Indeed, we have $\mathrm{Tr}_1(n, \mathcal{C}) \geq \Omega(n \log n)$ (which is equivalent to the $\Omega(\log n)$ lower bound for the best possible average degree in the setting of Theorem 1.1). This can be seen by considering a hypercube graph with vertex set $\{0, 1\}^m$, where two vertices are connected by an edge if and only if they differ in exactly one coordinate. Taking the colour palette $\{1, \ldots, m\}$, let us colour any such edge with the colour given by the index of the coordinate in which the two vertices differ (that is, the colour corresponds to the coordinate direction of the edge when visualizing the hypercube geometrically). This gives a proper edge-colouring of the hypercube graph which does not have any rainbow cycles, as every cycle must contain every colour an even number of times.

This example also shows that Proposition 1.2 cannot be extended to the notion of almost rainbow cycles considered by Janzer and Sudakov [36], where most colours appear only once, but a small proportion of edges on a cycle are allowed to use non-unique colours. The hypercube example above shows that there are properly edge-coloured graphs on n vertices with $\Omega(n \log n)$ edges that do not have any such cycles, even if we only insist on having at least one colour appearing exactly once on the cycle.

9 Applications and connections to additive number theory

In this section, we continue the discussion of applications and connections to additive number theory started in Section 1.1, and in particular show how Theorem 1.5 follows from Theorem 1.1.

9.1 Upper bounds on the additive dimension

Recall that given an arbitrary group $(G, \cdot)$ with identity element e, we say a subset $S \subseteq G$ is dissociated if there is no solution to $g_1^{\varepsilon_1} \cdots g_m^{\varepsilon_m} = e$ with distinct $g_1, \ldots, g_m \in S$ as well as $\varepsilon_1, \ldots, \varepsilon_m \in \{-1, 1\}$ and $m \geq 1$. The additive dimension $\dim A$ of a finite set $A \subseteq G$ is defined as the maximum cardinality of a dissociated subset of A. We will refer to a product of the form $g_1^{\varepsilon_1} \cdots g_m^{\varepsilon_m} = e$ with distinct $g_1, \ldots, g_m \in G$ and $\varepsilon_1, \ldots, \varepsilon_m \in \{-1, 1\}$ and $m \geq 1$, as a *non-trivial product*.

The following observation shows how a dissociated subset of a group can be used to construct a properly edge-coloured graph without rainbow cycles. With the same idea, Theorem 1.1 can be used to obtain upper bounds for the size of dissociated subsets, and in particular to prove Theorem 1.5.

Observation 9.1. *Let G be a group of size $|G| = n$, and let $S \subseteq G$ be a dissociated subset of G. Then there exists a properly edge-coloured $|S|$-regular graph on $2n$ vertices without a rainbow cycle. Furthermore, if S consists only of elements of even order in the group G, then there also exists a properly edge-coloured $|S|$-regular graph on n vertices without a rainbow cycle.*

Proof. For the first part of the observation, let us consider a bipartite graph on $2n$ vertices, where each part consists of n vertices corresponding to the elements of G. For every $s \in S$ and every $g \in G$, let us draw an edge from vertex g on the left side to vertex gs on the right side and colour this edge with a colour corresponding to s. This way, we obtain an $|S|$-regular graph with a proper edge-colouring (with $|S|$ colours). We claim that this graph does not contain a rainbow cycle. Indeed, suppose there is a rainbow cycle, and let $g_1, \ldots, g_m \in S$ be associated with the colours on the edges of the cycle in the order they appear on the cycle (starting at some vertex $g \in G$ on the left side). Note that m is even (as the graph is bipartite) and the elements $g_1, \ldots, g_m \in S$ are distinct (as the cycle is rainbow). Now, the vertices on the cycle, starting at g, are labelled by $g, gg_1, gg_1g_2^{-1}, gg_1g_2^{-1}g_3, \ldots$, and we must have $g = gg_1g_2^{-1}g_3g_4^{-1} \cdots g_{m-1}g_m^{-1}$. Thus, we obtain $g_1g_2^{-1}g_3g_4^{-1} \cdots g_{m-1}g_m^{-1} = e$, contradicting the assumption that the set S is dissociated.

For the second part of the observation, assume that all elements of S have even order. Let $S^{-1} := \{g^{-1} \mid g \in S\}$, and consider the Cayley graph of G with the generator set $S \cup S^{-1}$ with $|G| = n$ vertices. For any edge hh' in this graph, there is a unique element $g \in S$ such that $h' = hg$ or $h = h'g$ (otherwise, if $h' = hg_1$ and $h = h'g_2$ for two distinct $g_1, g_2 \in S$, we would obtain $g_1g_2 = e$, contradicting the assumption that S is dissociated). Let us now colour any such edge hh' with a colour corresponding to the element $g \in S$ such that $h' = hg$ or $h = h'g$. For every $g \in S$ of order 2, the edges with the colour corresponding to G form a perfect matching (i.e. every vertex is incident to exactly one edge of this colour). For every $g \in S$ of order larger than 2, the edges with the colour corresponding to g form a collection of vertex-disjoint cycles of even length (their length is precisely the order of g) covering all vertices in the graph. For each such cycle, let us delete every other edge so that in the resulting

graph every vertex is incident to exactly one edge of each colour. Hence the resulting graph is $|S|$-regular with a proper edge-colouring. It remains to show that there is no rainbow cycle in this graph. Indeed, assume that the vertices $h_1, h_2, \ldots, h_m$ form a rainbow cycle in this order. Then for $i = 1, \ldots, m$, we can find an element $g_i \in S$ such that $h_{i+1} = h_i g_i$ or $h_{i+1} = h_i g_i^{-1}$ (where $h_{m+1} := h_1$). The elements $g_1, \ldots, g_m \in S$ must be distinct since they correspond to the colours appearing on the rainbow cycle with vertices $h_1, h_2, \ldots, h_m$. Thus, we obtain a solution to $h_1 g_1^{\varepsilon_1} \cdots g_m^{\varepsilon_m} = h_1$, i.e. to $g_1^{\varepsilon_1} \cdots g_m^{\varepsilon_m} = e$, with distinct $g_1, \ldots, g_m \in S$ as well as $\varepsilon_1, \ldots, \varepsilon_m \in \{-1, 1\}$ and $m \geq 3$. This contradicts the assumption that S is dissociated. $\qquad\square$

Combining Observation 9.1 with our main result in Theorem 1.1 implies that any dissociated subset $S \subseteq G$ of any group G of size $n \geq 2$ satisfies $|S| \leq C \cdot \log(2n) \cdot \log\log(2n)$ for the absolute constant $C > 0$ in Theorem 1.1. In other words, this shows an upper bound $\dim G \leq O(\log |G| \cdot \log\log |G|)$ for the additive dimension of any finite group G of size $|G| \geq 3$ (see also the discussion in Section 1.1).

The same idea as for Observation 9.1 can also be used to deduce Theorem 1.5 from Theorem 1.1. It will be convenient for us to have the following asymmetric variant of Theorem 1.5, see also [68, Theorem 1] for a similar asymmetric variant of Theorem 1.3 in the abelian setting.

Theorem 9.2. *Let $(G, \cdot)$ be a group and consider finite subsets $A, B \subseteq G$. If $2 < |A \cdot B| \leq K|A|$ for some positive integer K, then we have*
$$\dim B \leq O(K \log(K|A|) \log\log(K|A|)).$$

Proof. Note that $2 < |A \cdot B| \leq K|A|$ implies that A and B are non-empty and $\log((K+1)|A|) \cdot \log\log((K+1)|A|) \leq O(\log(K|A|) \log\log(K|A|))$. Now suppose towards a contradiction that there exists a dissociated subset $S \subseteq B$ of size $|S| \geq C \cdot K \cdot \log((K+1)|A|) \cdot \log\log((K+1)|A|)$, where C is the constant given by Theorem 1.1.

Consider a bipartite graph H with parts A and $A \cdot B$ and with edge set consisting of all pairs (a, ab) with $a \in A$ and $b \in S$. We consider a proper edge-colouring where each edge (a, ab) is given a colour corresponding to b. Note that the resulting properly-edge coloured graph H is a subgraph of the bipartite graph considered in the proof of the first part of Observation 9.1, and in particular H does not contain a rainbow cycle.

The graph H has $|A| + |A \cdot B| \leq (K+1)|A|$ vertices and $|A||S|$ edges, so its average degree is
$$\frac{2|A||S|}{|A| + |A \cdot B|} \geq \frac{2|S|}{K+1} \geq C \cdot \log((K+1)|A|) \cdot \log\log((K+1)|A|) \geq C \cdot \log |V(H)| \cdot \log\log |V(H)|.$$

But by Theorem 1.1 this implies there exists a rainbow cycle in H, which is a contradiction. Thus, every dissociated dissociated subset $S \subseteq B$ has size $|S| < C \cdot K \cdot \log((K+1)|A|) \cdot \log\log((K+1)|A|) \leq O(K \log(K|A|) \log\log(K|A|))$. $\quad\square$

Theorem 1.5 follows from Theorem 9.2 by setting $B = A$ and observing that we may always assume $K \leq |A|$ (since $|A \cdot A| \leq |A|^2$) and hence $\log(K|A|) \leq 2 \log |A|$.

9.2 Lower bounds on the additive dimension

In this subsection, we briefly turn our attention to lower bounds. Given a group G we denote by G^k the k-fold direct product of G. Furthermore, as usual, $\mathcal{S}_k$ denotes the symmetric group consisting of the permutations of a set of size k.

Observation 9.3. *Suppose G is a group of size m with $\dim G = d$. Then $\dim G^k \geq \frac{d}{\log m} \cdot \log |G^k|$ for any integer $k \geq 1$. In particular, $\dim \mathcal{S}_3^k \geq \frac{3}{\log 6} \cdot \log |\mathcal{S}_3^k|$ for any integer $k \geq 1$. Furthermore, for any integer $k \geq 1$, there is a dissociated set $Z \subseteq \mathcal{S}_3^k$ of size $|Z| = \frac{3}{\log 6} \cdot \log |\mathcal{S}_3^k|$ such that every element of Z has order 2 in the group $\mathcal{S}_3^k$.*

Proof. Let $Y \subseteq G$ be a dissociated subset of size $|Y| = \dim G = d$. Let $Z \subseteq G^k$ consist of the dk elements of the form $(e, \ldots, e, g, e, \ldots, e) \in G^k$ with $g \in Y$, where e is the identity element of G. We claim that the set $Z \subseteq G^k$ is dissociated. Indeed, any non-trivial product of elements of Z or their inverses would give, in at least one coordinate, a non-trivial product of elements of Y or their inverses. Since $Y \subseteq G$ is dissociated, such a non-trivial product cannot exist, and hence $Z \subseteq G^k$ is dissociated. Thus, we have $\dim G^k \geq |Z| = dk = d \cdot \frac{\log |G^k|}{\log |G|} = \frac{d}{\log m} \cdot \log |G^k|$. Also note that the order of any element $(e, \ldots, e, g, e, \ldots, e) \in Z$ in the group G^k is the same as the order of the element $g \in Y$ in the group G.

In the case of $G = \mathcal{S}_3$, we can take the set Y to consist of the three transpositions in $\mathcal{S}_3$. Since this set Y is dissociated, we have $|\dim \mathcal{S}_3| \geq 3$ and therefore $\dim \mathcal{S}_3^k \geq \frac{3}{\log 6} \cdot \log |\mathcal{S}_3^k|$. Since the transpositions in Y each have

order 2 in $\mathcal{S}_3$, the set $Z \subseteq \mathcal{S}_3^k$ constructed above consists of elements of order 2 in $\mathcal{S}_3^k$. As discussed above, this set is dissociated and has size $|Z| = \frac{3}{\log 6} \cdot \log |\mathcal{S}_3^k|$. $\hfill\square$

Given a lower bound for $\dim G$ for some group G, via Observation 9.1 one obtains a lower bound for the problem of determining the minimum possible average degree of a properly edge-coloured graph on n vertices that guarantees the existence of a rainbow cycle. In particular, by the last part of Observation 9.3 combined with Observation 9.1, when n is a power of 6, there is a properly edge-coloured n-vertex graph with average degree at least $\frac{3}{\log 6} \cdot \log n \approx 1.674 \cdot \log n$, which does not contain a rainbow cycle. This is the best-known lower bound for this problem (this bound was obtained in [39] with a different description of the same edge-coloured graph, see also the discussion in the introduction).

A simpler example for a properly edge-coloured n-vertex graph with average degree $\Omega(\log n)$ without a rainbow cycle is given by taking hypercube graph and colouring the edges according to the coordinate directions (see the end of Section 8 for a more detailed description of this construction). This construction can also be obtained by combining Observations 9.3 and 9.1, starting with the group $G = \mathcal{S}_2 = \mathbb{Z}_2$, whose additive dimension is 1 (in $\mathcal{S}_2$, the single transposition forms a dissociated set).

Both these examples have in common that they are formed by taking powers of a symmetric group, namely $\mathcal{S}_2$ or $\mathcal{S}_3$, and that the generator sets for the Cayley graphs consist of transpositions. This raises the natural question of studying the additive dimension $\dim \mathcal{S}_k$ of symmetric groups, which might help in understanding whether the $\log \log n$ term in our main result in Theorem 1.1 is actually needed. In particular, recalling that both of the constructions above are based on transpositions, one may ask about the additive dimension of the set of transpositions of $\mathcal{S}_k$. We discuss this question in the next subsection.

9.3 Additive dimension of the set of transpositions

Here, we discuss the question of determining the additive dimension of the set T_k consisting of all transpositions in the symmetric group $\mathcal{S}_k$. We initially considered this question in order to try to understand whether the $\log \log n$ term is necessary in our result in Theorem 1.1, but it is also a natural question in itself. In particular, since transpositions are self-inverse, the question can equivalently be phrased as follows: What is the minimum number m such that among any set of m transpositions in $\mathcal{S}_k$ we can choose distinct transpositions $\tau_1, \ldots, \tau_\ell$ for some $\ell \geq 1$ such that $\tau_1 \circ \cdots \circ \tau_\ell = \mathrm{id}$? In other words, how many transpositions do we need in $\mathcal{S}_k$ to always be able to multiply some of them in some order, without repetitions, to obtain the identity?

This minimum number m is precisely the additive dimension $\dim T_k$ of the set T_k of transpositions in $\mathcal{S}_k$. Since there are only $\binom{k}{2}$ transpositions in total, we trivially have $\dim T_k \leq |T_k| = \binom{k}{2}$. There is a simple upper bound of $\dim T_k \leq O(k^{7/4})$, based on the observation that for a dissociated set $S \subseteq T_k \subseteq \mathcal{S}_k$, the k-vertex graph with edges corresponding to the transpositions in S must be $K_{4,4}$-free. Indeed, suppose this graph would contain a copy of $K_{4,4}$, say with vertices $\{1, 2, 3, 4\}$ on one side and vertices $\{5, 6, 7, 8\}$ on the other side. Noting that

$$(15)(26)(16)(25) = (12)(56) \qquad\qquad (35)(46)(36)(45) = (34)(56)$$
$$(17)(28)(18)(27) = (12)(78) \qquad\qquad (37)(48)(38)(47) = (34)(78),$$

and hence

$$(15)(26)(16)(25)(35)(46)(36)(45)(17)(28)(18)(27)(37)(48)(38)(47) = (12)(56)(34)(56)(12)(78)(34)(78) = \mathrm{id},$$

this would mean that a product of distinct elements of S yields the identity, which is a contradiction to S being dissociated. With the same idea, one can more generally show that for a dissociated set $S \subseteq T_k \subseteq \mathcal{S}_k$, the graph corresponding to S cannot contain a 2-blow-up of a cycle as a subgraph. Using a result of Janzer [35, Theorem 1.14] on the Turán number of this family of blow-ups, one then obtains $\dim T_k \leq O(k^{3/2}(\log k)^{7/2})$. Our Theorem 9.2 implies the following much stronger bound.

Proposition 9.4. *There exists a constant $C > 0$ such that for any integer $k \geq 3$, among any set of at least $C \cdot k \cdot (\log k)^2$ transpositions in $\mathcal{S}_k$, we can choose a non-empty sequence of distinct transpositions whose product is the identity. In other words, we have $\dim T_k \leq C \cdot k \cdot (\log k)^2$.*

Proof. We apply Theorem 9.2 to $A = \mathcal{S}_k$ and $B = T_k$ and $K = 1$, noting that that $|A \cdot B| = |A|$, and obtain $\dim T_k \leq O(\log |A| \cdot \log \log |A|) = O(\log k! \cdot \log \log k!) \leq O(k \log^2 k)$. $\hfill\square$

The following lower bound for $\dim T_k$ shows the above result is not too far from being tight.

Proposition 9.5. *For any integer $k \geq 3$, there exists a set of $\left(\frac{1}{2} + o(1)\right) \cdot \frac{k \log k}{\log \log k}$ transpositions in $\mathcal{S}_k$ containing no non-empty sequence with product equal to the identity. In other words, we have $\dim T_k \geq \left(\frac{1}{2} + o(1)\right) \cdot \frac{k \log k}{\log \log k}$.*

Proof. Given that $\dim T_k \geq \dim T_{k-1}$, it suffices to prove the result for even k. As shown by Erdős and Sachs [19, p. 252], for $d \geq 2$ and $g \geq 4$ and any even $k \geq 4d^{g-1}$ there is a d-regular graph on k vertices with girth at least g. Hence there is a d-regular graph G on the vertex set $\{1, 2, \ldots, k\}$ with girth $g > d$ and $d \geq (1 + o(1))\frac{\log k}{\log \log k}$. Let Z be the set of transpositions $(i\, j)$ where ij is an edge of G, then $|Z| = \frac{1}{2}kd \geq \left(\frac{1}{2} + o(1)\right) \cdot \frac{k \log k}{\log \log k}$. We claim that there is no sequence of distinct transpositions in Z whose product is the identity. Indeed, assuming there is such a sequence, let M be the set of all vertices $i \in \{1, \ldots, k\}$ appearing in at least one transposition in the sequence. For each such vertex i, when the first transposition containing i is applied, i is swapped with a neighbour in the graph. Since the same transposition (edge) cannot appear again, and the graph G has girth g, every $i \in M$ has to move at least g steps in order to return to its original position. Altogether this corresponds to at least $|M|g$ single moves, and each transposition provides two such moves. The total number of transpositions in the sequence is at most the number of edges in the induced subgraph $G[M]$, which is at most $|M|d/2$ (since G is d-regular). It follows that $|M|g \leq |M|d/2 \cdot 2$ and therefore $g \leq d$, contradicting $g > d$. $\square$

There is a natural correspondence between a set S of transpositions in $\mathcal{S}_k$ and a graph with vertex set $[k] = \{1, \ldots, k\}$ where the edges are formed by the pairs of vertices making the transpositions. In the literature, this graph is called the Schreier graph $\mathrm{Sch}(\mathcal{S}_k \curvearrowright [k], S)$ corresponding to the natural action of $\mathcal{S}_k$ on the index set $[k]$ with generator set S. The proof above used that in order for S to contain a (non-empty) sequence whose product is the identity, the corresponding graph needs to contain a subgraph with a cycle double cover, i.e. a set of cycles with the property that any edge belongs to precisely two of these cycles (arising from the trajectories of individual elements of the ground set while applying the sequence of transpositions). Such "decompositions" have attracted a lot of attention in connection to the famous cycle double-cover conjecture of Szekeres and Seymour, for more details see e.g. the survey [32]. However, in our setting, there are further restrictions. In particular, considering the trajectories of the individual elements of the ground set with their natural direction, we obtain a collection of directed cycles. So if we form a directed graph from a sequence of transposition multiplying to the identity, where for every transposition we draw two directed edges in opposite directions between the corresponding pair of vertices, then this directed graph must have a decomposition into directed cycles. This condition has also been considered in a directed strengthening of the cycle double cover conjecture due to Jaeger [32] from 1985. This relationship and various constructions considered in connection to the cycle double cover conjecture and its directed strengthening might be useful in trying to improve the lower bound in Proposition 9.5. There is also a further restriction imposed in our setting arising from the ordering of our sequence of transpositions whose product is the identity, but this additional condition seems somewhat difficult to use.

We note that the Schreier graph $\mathrm{Sch}(\mathcal{S}_k \curvearrowright [k], S)$, which featured in the proof of the lower bound for $\dim T_k$ in Proposition 9.5, is not the same graph as the Cayley graph $\mathrm{Cay}(\mathcal{S}_k, S)$ featuring in a more direct deduction of the upper bound for $\dim T_k$ in Proposition 9.4 from Theorem 1.1 via Observation 9.1. Indeed, the former graph has only k vertices, while the latter has $k!$ vertices. It is worth noting, however, that the relationship between these two graphs plays an important role in understanding the interchange random process, see e.g. [60] for more details. In particular, these two graphs have the same spectral gap, which is the subject of the famous Aldous conjecture, proved by Caputo, Liggett, and Richthammer [14]. Given the close connection between the spectral gap and the expansion properties of a graph, as well as the fact that expansion plays a fundamental role in our proof, it is natural to wonder whether one can exploit the relationship between the graphs $\mathrm{Sch}(\mathcal{S}_k \curvearrowright [k], S)$ and $\mathrm{Cay}(\mathcal{S}_k, S)$ to close the gap between our results in Propositions 9.4 and 9.5.

10   Concluding remarks

In this paper we give an essentially tight bound for the maximum possible average degree of a properly edge-coloured graph on n vertices without a rainbow cycle, a problem raised by Keevash, Mubayi, Sudakov and Verstraëte [39]. In particular, we show that for some absolute constant $C > 0$ any n-vertex properly edge-coloured graph with average degree at least $C \log n \log \log n$ must contain a rainbow cycle, which is tight up to the $\log \log n$ term. The immediate open problem is to decide whether the $\log \log n$ term is necessary.

Question 10.1. *Is the $\log \log n$ factor in the bound for the average degree in Theorem 1.1 necessary in order to guarantee a rainbow cycle?*

In our proof of Theorem 1.1, we first pass to a robust sublinear expander subgraph. Our approach for finding a rainbow cycle is then to randomly split the colour palette into two parts and show that from some fixed starting vertex x we are likely to reach most of the other vertices by rainbow walks using only edges of colours in the first part of the palette, and the same holds for the colours in the second part of the palette (see the discussion in Section 2, and note that in the proof this statement is encapsulated in (5.1)). This approach does require the $\log\log n$ term, which may be viewed as circumstantial evidence that the $\log\log n$ term might indeed be needed in Theorem 1.1. In fact, for any large enough n, there exists an n-vertex properly edge-coloured robust sublinear expander with average degree $\log n(\log\log n)^{1-o(1)}$ with the property that when sampling the colours independently, each with probability $\frac{1}{2}$, and only taking the edges with the sampled colours, with high probability the resulting subgraph has no large connected components. Examples of such robust sublinear expander graphs can be formed by taking Hamming graphs with appropriately chosen parameters, the details of the construction can be found in [55] (these graphs are discussed in [55] as bottlenecks to a similar, but much simpler expansion argument, and the paper [55] only verifies a weaker expansion property for these graphs, but essentially the same argument proves these graphs are also robust sublinear expanders according to our definition). This shows that one cannot hope to push our argument to a bound on the order of $\log n$ without substantial modification (in particular, (5.1) fails when omitting the $\log\log n$ term in the bound for the average degree in Theorem 1.1). As some illustration of the issue our argument encounters, note that the vertex expansion properties of our robust sublinear expanders only guarantee that we reach a constant proportion of vertices, starting from an arbitrary initial vertex, after taking $\Theta(\log n\log\log n)$ steps. Therefore our number of steps needs to be on the order of $\log n\log\log n$ (see the choice of T at the start of the proof of Proposition 5.1), but we need this number to be smaller than the average degree.

On the other hand, we note that these caveats are mostly issues with the probabilistic approach in our proof, and do not rule out the possibility an average degree on the order of $\log n$ suffices to guarantee a rainbow cycle. In fact, even the approach of splitting the colour palette into two parts and showing that from some fixed starting vertex x, when restricting to either part of the colour palette, we can reach most of the other vertices by rainbow walks, might work for an average degree on the order of $\log n$. The above discussion shows that one cannot just assign the colours to the two parts of the colour palette randomly, but potentially a more careful choice of the parts of the colour palettes could work. As a first step in this direction, the following question asks whether it is possible to split the colour palette into two parts such that both of the resulting subgraphs corresponding to the two parts of the colour palette are connected.

Question 10.2. *Does there exist a constant $C > 0$ such that the edges of any n-vertex properly edge-coloured robust sublinear expander G with average degree at least $C\log n$ can be decomposed into two spanning connected subgraphs in such a way that every colour appears on only one of them?*

As mentioned above, the perhaps most natural approach of forming two subgraphs by assigning every colour independently at random to the first or to the second subgraph with probability $\frac{1}{2}$ each, fails (since there are examples for G, where the resulting subgraphs are disconnected with high probability). We note that properties of robust sublinear expanders under this type of "subsampling" have been quite useful in a number of applications (see e.g. [13] and the discussion on this topic there). We note that a positive answer to the above question would not quite suffice to remove the $\log\log n$ term in Theorem 1.1. What would be enough is to ensure there is a pair of vertices that are connected by rainbow paths in both subgraphs. This in turn would be implied if we could show that rather than just being connected, the subgraphs are (perhaps somewhat weaker) robust sublinear expanders. A positive answer to Question 10.2 would nevertheless give a tight answer to a certain "X-Ray reconstruction" problem in discrete geometry, when combined with a reduction from [55]. It would also provide an alternative solution to a question of Matoušek [54] (recently solved by the authors of [4]), which in turn would give a somewhat weaker version of a result concerning the Erdős Unit Distance problem for "typical" norms obtained recently in [4].

In Section 9, we discussed a number of interesting connections of the problem of determining the maximum possible average degree of a properly edge-coloured graph on n vertices without a rainbow cycle, including a number of open problems. Let us highlight the problem of determining the maximum possible additive dimension, or equivalently the plus-minus weighted Davenport constant, of a group with n elements.

Question 10.3. *Given a positive integer n, what is the smallest number $d = d(n)$ so that in any group G of size $|G| = n$, given any subset $S \subseteq G$ of size $|S| \geq d$, we can always find a solution to $g_1^{\varepsilon_1} \cdot \ldots \cdot g_m^{\varepsilon_m} = e$ with distinct $g_1, \ldots, g_m \in S$ as well as $\varepsilon_1, \ldots, \varepsilon_m \in \{-1, 1\}$ and $m \geq 1$.*

Theorem 1.5 shows an upper bound of $d(n) \leq O(\log n\log\log n)$ and it is again unclear whether the $\log\log n$ term is necessary (it is easy to show a lower bound of the form $d(n) \geq \Omega(\log n)$). We find the instance of this question when $G = \mathcal{S}_k$ to be of particular interest since any k-element group can be realised as a subgroup of $\mathcal{S}_k$, since the

symmetric group $\mathcal{S}_k$ is far from being abelian, and since the best known lower bounds for $d(n)$ come from exploring dissociated subsets of $\mathcal{S}_k$. More precisely, the best-known lower bounds for $d(n)$ are obtained from dissociated subsets consisting of transpositions, which motivates the following question of determining the additive dimension of the set of all transpositions in the symmetric group $\mathcal{S}_k$. We find this problem very natural and appealing, and it has some interesting connections discussed in Section 9.3.

Question 10.4. *Given $k \geq 2$, determine the minimum number $t = t(k)$ such that any set of t transpositions in $\mathcal{S}_k$ contains a non-empty subset whose elements can be multiplied in some order to give the identity.*

Propositions 9.4 and 9.5 show that $k(\log k)^{1-o(1)} \leq t(k) \leq O(k(\log k)^2)$. We note that this question might be helpful in deciding whether the $\log \log n$ term is necessary in our main result in Theorem 1.1. Note that the exponent 2 in the upper bound $t(k) \leq O(k(\log k)^2)$ appears because of applying our main result to a graph on $k!$ vertices, and so the $\log \log(k!)$ term leads to a factor of roughly $\log k$ (while the "main term" $\log(k!)$ is roughly $k \log k$). For similar reasons, the previous best bounds in the setting of Theorem 1.1 in [36, 41] only lead to upper bounds for Question 10.4 that are worse than the trivial bound $t(k) \leq \binom{k}{2}$.

References

[1] S. D. Adhikari, *Plus-minus weighted zero-sum constants: a survey*, Analytic number theory, modular forms and q-hypergeometric series, Springer Proc. Math. Stat., vol. 221, Springer, Cham, 2017, pp. 1–9.

[2] S. D. Adhikari, D. J. Grynkiewicz, and Z. W. Sun, *On weighted zero-sum sequences*, Adv. in Appl. Math. **48** (2012), no. 3, 506–527.

[3] N. Alon, *Graph powers*, Contemporary combinatorics, Bolyai Soc. Math. Stud., vol. 10, János Bolyai Math. Soc., Budapest, 2002, pp. 11–28.

[4] N. Alon, M. Bucić, and L. Sauermann, *Unit and distinct distances in typical norms*, Geom. Funct. Anal. **35** (2025), 1–42.

[5] N. Alon and J. H. Spencer, *The Probabilistic Method, 4th ed.*, John Wiley & Sons, 2016.

[6] O. Alrabiah and V. Guruswami, *Near-Tight Bounds for 3-Query Locally Correctable Binary Linear Codes via Rainbow Cycles*, preprint arXiv:2404.05864 (2024).

[7] M. Bateman and N. H. Katz, *Structure in additively nonsmoothing sets*, preprint arXiv:1104.2862 (2011).

[8] M. Bateman and N. H. Katz, *New bounds on cap sets*, J. Amer. Math. Soc. **25** (2012), no. 2, 585–613.

[9] T. F. Bloom and O. Sisask, *Breaking the logarithmic barrier in Roth's theorem on arithmetic progressions*, preprint arXiv:2007.03528 (2020).

[10] J. Bourgain, *On triples in arithmetic progression*, Geom. Funct. Anal. **9** (1999), no. 5, 968–984.

[11] J. Bourgain, *Roth's theorem on progressions revisited*, J. Anal. Math. **104** (2008), 155–192.

[12] E. Breuillard, B. Green, and T. Tao, *Approximate subgroups of linear groups*, Geom. Funct. Anal. **21** (2011), no. 4, 774–819.

[13] M. Bucić and R. Montgomery, *Towards the Erdős-Gallai Cycle Decomposition Conjecture*, Adv. Math. **437** (2024), Paper No. 109434, 40 pp.

[14] P. Caputo, T. M. Liggett, and T. Richthammer, *Proof of Aldous' spectral gap conjecture*, J. Amer. Math. Soc. **23** (2010), no. 3, 831–851.

[15] D. Chakraborti, J. Kim, J. Kim, M. Kim, and H. Liu, *Well-mixing vertices and almost expanders*, Proc. Amer. Math. Soc. **150** (2022), no. 12, 5097–5110.

[16] M.-C. Chang, *A polynomial bound in Freiman's theorem*, Duke Math. J. **113** (2002), no. 3, 399–419.

[17] D. Conlon, J. Fox, L. Yepremyan, and H. T. Pham, *On the clique number of random Cayley graphs and related topics*, preprint arXiv:2412.21194 (2024).

[18] S. Das, C. Lee, and B. Sudakov, *Rainbow Turán problem for even cycles*, European J. Combin. **34** (2013), no. 5, 905–915.

[19] P. Erdős and H. Sachs, *Reguläre Graphen gegebener Taillenweite mit minimaler Knotenzahl*, Wiss. Z. Martin-Luther-Univ. Halle-Wittenberg Math.-Natur. Reihe **12** (1963), 251–257.

[20] I. G. Fernández, J. Kim, Y. Kim, and H. Liu, *Nested cycles with no geometric crossings*, Proc. Amer. Math. Soc. Ser. B **9** (2022), 22–32.

[21] D. Fisher, N. H. Katz, and I. Peng, *Approximate multiplicative groups in nilpotent Lie groups*, Proc. Amer. Math. Soc. **138** (2010), no. 5, 1575–1580.

[22] G. A. Freĭman, *On the addition of finite sets*, Dokl. Akad. Nauk SSSR (1964), 1038–1041.

[23] I. Gil Fernández and H. Liu, *How to build a pillar: a proof of Thomassen's conjecture*, J. Combin. Theory Ser. B **162** (2023), 13–33.

[24] N. Gill and H. A. Helfgott, *Growth in solvable subgroups of* $\mathrm{GL}_r(\mathbb{Z}/p\mathbb{Z})$, Math. Ann. **360** (2014), no. 1-2, 157–208.

[25] B. Green and I. Z. Ruzsa, *Freiman's theorem in an arbitrary abelian group*, J. Lond. Math. Soc. **75** (2007), no. 1, 163–175.

[26] F. Halter-Koch, *Arithmetical interpretation of weighted Davenport constants*, Arch. Math. (Basel) **103** (2014), no. 2, 125–131.

[27] J. Haslegrave, J. Hu, J. Kim, H. Liu, B. Luan, and G. Wang, *Crux and long cycles in graphs*, SIAM J. Discrete Math. **36** (2022), no. 4, 2942–2958.

[28] J. Haslegrave, J. Hyde, J. Kim, and H. Liu, *Ramsey numbers of cycles versus general graphs*, Forum Math. Sigma **11** (2023), Paper No. e10, 18 pp.

[29] J. Haslegrave, J. Kim, and H. Liu, *Extremal density for sparse minors and subdivisions*, Int. Math. Res. Not. (2022), no. 20, 15505–15548.

[30] H. A. Helfgott, *Growth and generation in* $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$, Ann. of Math. **167** (2008), no. 2, 601–623.

[31] H. A. Helfgott, *Growth in* $\mathrm{SL}_3(\mathbb{Z}/p\mathbb{Z})$, J. Eur. Math. Soc. **13** (2011), no. 3, 761–851.

[32] F. Jaeger, *A survey of the cycle double cover conjecture*, Cycles in graphs (Burnaby, B.C., 1982), North-Holland Math. Stud., vol. 115, North-Holland, Amsterdam, 1985, pp. 1–12.

[33] S. Janson, T. Łuczak, and A. Rucinski, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000.

[34] O. Janzer, *Disproof of a Conjecture of Erdős and Simonovits on the Turán Number of Graphs with Minimum Degree 3*, Int. Math. Res. Not. (2023), no. 10, 8478–8494.

[35] O. Janzer, *Rainbow Turán number of even cycles, repeated patterns and blow-ups of cycles*, Israel J. Math. (2022), 1–28.

[36] O. Janzer and B. Sudakov, *On the Turán number of the hypercube*, Forum Math. Sigma **12** (2024), Paper No. e38, 19 pp.

[37] O. Janzer, B. Sudakov, and I. Tomon, *Regular subgraphs of linear hypergraphs*, Int. Math. Res. Not. (2024), no. 17, 12366–12381.

[38] T. Jiang, A. Methuku, and L. Yepremyan, *Rainbow Turán number of clique subdivisions*, European J. Combin. **110** (2023), Paper No. 103675, 8 pp.

[39] P. Keevash, D. Mubayi, B. Sudakov, and J. Verstraëte, *Rainbow Turán problems*, Combin. Probab. Comput. **16** (2007), no. 1, 109–126.

[40] P. Keevash, A. Pokrovskiy, B. Sudakov, and L. Yepremyan, *New bounds for Ryser's conjecture and related problems*, Trans. Amer. Math. Soc. Ser. B **9** (2022), 288–321.

[41] J. Kim, J. Lee, H. Liu, and T. Tran, *Rainbow cycles in properly edge-colored graphs*, Combinatorica **44** (2024), no. 4, 909–919.

[42] J. Kim, H. Liu, M. Sharifzadeh, and K. Staden, *Proof of Komlós's conjecture on Hamiltonian subsets*, Proc. Lond. Math. Soc. **115** (2017), no. 5, 974–1013.

[43] J. Komlós and E. Szemerédi, *Topological cliques in graphs II*, Combinatorics, Probability and Computing **5** (1996), no. 1, 79–90.

[44] S. V. Konyagin and I. D. Shkredov, *A quantitative version of the Beurling-Helson theorem*, Funct. Anal. Appl. **49** (2015), no. 2, 110–121, Translation of Funktsional. Anal. i Prilozhen. **49** (2015), no. 2, 39–53.

[45] S. V. Konyagin and I. D. Shkredov, *On subgraphs of random Cayley sum graphs*, European J. Combin. **70** (2018), 61–74.

[46] A. Kupavskii, A. Polyanskii, I. Tomon, and D. Zakharov, *The Extremal Number of Surfaces*, Int. Math. Res. Not. (2022), no. 17, 13246–13271.

[47] S. Letzter, *Separating paths systems of almost linear size*, Trans. Amer. Math. Soc. **377** (2024), no. 8, 5583–5615.

[48] S. Letzter and A. Girão, *Immersion of complete digraphs in Eulerian digraphs*, Israel J. Math. **260** (2024), no. 1, 401–425.

[49] V. F. Lev and R. Yuster, *On the size of dissociated bases*, Electron. J. Combin. **18** (2011), no. 1, Paper 117, 5 pp.

[50] H. Liu and R. Montgomery, *A proof of Mader's conjecture on large clique subdivisions in C_4-free graphs*, Journal of the London Mathematical Society **95** (2017), no. 1, 203–222.

[51] H. Liu and R. Montgomery, *A solution to Erdős and Hajnal's odd cycle problem*, J. Amer. Math. Soc. **36** (2023), no. 4, 1191–1234.

[52] H. Liu, G. Wang, and D. Yang, *Clique immersion in graphs without a fixed bipartite graph*, J. Combin. Theory Ser. B **157** (2022), 346–365.

[53] L. E. Marchan, O. Ordaz, and W. A. Schmid, *Remarks on the plus-minus weighted Davenport constant*, Int. J. Number Theory **10** (2014), no. 5, 1219–1239.

[54] J. Matoušek, *The number of unit distances is almost linear for most norms*, Adv. Math. **226** (2011), no. 3, 2618–2628.

[55] J. Matoušek, A. Přívětivý, and P. Škovroň, *How many points can be reconstructed from k projections?*, SIAM J. Discrete Math. **22** (2008), no. 4, 1605–1623.

[56] R. Montgomery, *Logarithmically small minors and topological minors*, J. Lond. Math. Soc. **91** (2015), no. 1, 71–88.

[57] R. Montgomery, A. Pokrovskiy, and B. Sudakov, *Embedding rainbow trees with applications to graph labelling and decomposition*, J. Eur. Math. Soc. **22** (2020), no. 10, 3101–3132.

[58] G. Moshkovitz and A. Shapira, *Decomposing a graph into expanding subgraphs*, Random Structures Algorithms **52** (2018), no. 1, 158–178.

[59] A. Müyesser and A. Pokrovskiy, *A random Hall–Paige conjecture*, preprint arXiv:2204.09666 (2022).

[60] O. Parzanchevski and D. Puder, *Aldous's spectral gap conjecture for normal sets*, Trans. Amer. Math. Soc. **373** (2020), no. 10, 7067–7086.

[61] A. Popa, *Better lower and upper bounds for the minimum rainbow subgraph problem*, Theoret. Comput. Sci. **543** (2014), 1–8.

[62] L. Pyber and E. Szabó, *Growth in finite simple groups of Lie type*, J. Amer. Math. Soc. **29** (2016), no. 1, 95–146.

[63] W. Rudin, *Fourier analysis on groups*, Wiley Classics Library, John Wiley & Sons, Inc., New York, 1990, Reprint of the 1962 original, A Wiley-Interscience Publication.

[64] T. Sanders, *Structure in sets with logarithmic doubling*, Canad. Math. Bull. **56** (2013), no. 2, 412–423.

[65] T. Sanders, *On a theorem of Shkredov*, Online J. Anal. Comb. (2010), no. 5, 4 pp.

[66] T. Sanders, *On Roth's theorem on progressions*, Ann. of Math. **174** (2011), no. 1, 619–636.

[67] T. Sanders, *The structure theory of set addition revisited*, Bull. Amer. Math. Soc. **50** (2013), no. 1, 93–127.

[68] T. Schoen and I. D. Shkredov, *Additive dimension and a theorem of Sanders*, J. Aust. Math. Soc. **100** (2016), no. 1, 124–144.

[69] A. Shapira and B. Sudakov, *Small complete minors above the extremal edge density*, Combinatorica **35** (2015), no. 1, 75–94.

[70] I. D. Shkredov and S. Yekhanin, *Sets with large additive energy and symmetric sets*, J. Combin. Theory Ser. A **118** (2011), no. 3, 1086–1093.

[71] B. Sudakov and I. Tomon, *The extremal number of tight cycles*, Int. Math. Res. Not. (2022), no. 13, 9663–9684.

[72] T. Tao, *Product set estimates for non-commutative groups*, Combinatorica **28** (2008), no. 5, 547–594.

[73] I. Tomon, *Robust (rainbow) subdivisions and simplicial cycles*, Adv. Comb. (2024), Paper No. 1, 37 pp.

[74] Y. Wang, *Rainbow clique subdivisions*, European J. Combin. **116** (2024), Paper No. 103868, 9 pp.