![CAMBRIDGE UNIVERSITY PRESS]

RESEARCH ARTICLE

The power of many colours

Noga Alon[1], Matija Bucić [ID][2], Micha Christoph[3] and Michael Krivelevich [ID][4]

[1]Department of Mathematics, Princeton University, Princeton, USA; E-mail: nalon@math.princeton.edu.
[2]School of Mathematics, Institute for Advanced Study and Department of Mathematics, Princeton University, Princeton, USA;
E-mail: matija.bucic@ias.edu (corresponding author).
[3]Department of Mathematics, ETH Zürich, Zürich, Switzerland; E-mail: micha.christoph@math.ethz.ch.
[4]School of Mathematical Sciences, Tel Aviv University, Tel Aviv, Israel; E-mail: krivelev@tauex.tau.ac.il.

Abstract
A classical problem, due to Gerencsér and Gyárfás from 1967, asks how large a monochromatic connected component can we guarantee in any r-edge colouring of K_n? We consider how big a connected component we can guarantee in any r-edge colouring of K_n if we allow ourselves to use up to s colours. This is actually an instance of a more general question of Bollobás from about 20 years ago which asks for a k-connected subgraph in the same setting. We complete the picture in terms of the approximate behaviour of the answer by determining it up to a logarithmic term, provided n is large enough. We obtain more precise results for certain regimes which solve a problem of Liu, Morris and Prince from 2007, as well as disprove a conjecture they pose in a strong form.

We also consider a generalisation in a similar direction of a question first considered by Erdős and Rényi in 1956, who considered given n and m, what is the smallest number of m-cliques which can cover all edges of K_n? This problem is essentially equivalent to the question of what is the minimum number of vertices that are certain to be incident to at least one edge of some colour in any r-edge colouring of K_n. We consider what happens if we allow ourselves to use up to s colours. We obtain a more complete understanding of the answer to this question for large n, in particular, determining it up to a constant factor for all $1 \le s \le r$, as well as obtaining much more precise results for various ranges including the correct asymptotics for essentially the whole range.

Contents

1. Introduction

What can one say about the monochromatic structures we are guaranteed to find in any r-edge colouring of a large graph G? This very general extremal question captures a large number of classical extensively studied topics including a large portion of modern Ramsey theory.

There are limitations to how much structure one can ensure by using only a single colour. For example, a very classical result of Gyárfás [27] and Füredi [23], essentially answering a question of Gerencsér and Gyárfás from 1967 [25], tells us that in an r-edge colouring of an n-vertex complete graph, one can always find a monochromatic tree, or equivalently, a connected component in a single colour, of order at least $\frac{n}{r-1}$, and one cannot hope to do better whenever an affine plane of order $r - 1$ exists (which is known to exist whenever $r - 1$ is a prime power). This question and its generalisations in many different directions have been extensively studied; see, for example, surveys [28, 21]. Let us, for example, highlight the question of how many monochromatic components one needs to cover every vertex of an r-edge coloured K_n, since due to a beautiful duality reduction of Füredi [23], this question is equivalent to the famous Ryser conjecture concerning cover and matching numbers in r-partite hypergraphs [30].

A very natural question that arises is how much more power do we get if we allow ourselves to use more than just a single colour? This type of question is very classical; for example, Erdős, Hajnal and Rado [16] already in 1965 raised a generalisation of the Ramsey question in this direction. Namely, how large a clique using edges of at most s different colours can one guarantee in an n-vertex r-edge coloured complete graph? This problem and its generalisations have been extensively studied ever since; see, for example, [2, 8, 9, 12, 29, 10].

In this paper, we will explore this 'power of many colours' paradigm for several classical extremal problems. The first one has as its starting point a folklore observation, usually attributed to Erdős and Rado (see, for example, [28]), which states that in any 2-edge colouring of K_n, the edges in one of the colours connect all n vertices. A natural generalisation asks how big a connected monochromatic subgraph can we guarantee if we are using more than two colours. This is precisely the subject of the question of Gerencsér and Gyárfás [25] from 1967 discussed above. While we will mostly focus on what happens with this question if we are allowed to use up to s colours, let us first mention yet another natural generalisation due to Bollobás and Gyárfás [7] from 2003, which also leads to the same question from a different angle. They consider how large a k-connected monochromatic subgraph can one guarantee in any r-edge colouring of K_n. See [40, 37, 39], and references therein, for details on what is known about this problem. In 2003, Bollobás (see [36]) asked the overarching 'power of many colours' generalisation – namely, what happens if one is allowed to use up to s colours when looking for a large highly connected subgraph. The benchmarks for this problem have been established by Liu, Morris and Prince [36]; for some improvements and extensions, see [35, 38, 34].

As already mentioned, we will mostly focus on the $k = 1$ instance of the question of Bollobás. The reasons for this are many-fold. We find it to be perhaps the most natural instance since it is a direct 'power of many colours' generalisation of the much more classical question of Gerencsér and Gyárfás. Also, when assuming the underlying complete graph is at least moderately large, it seems most of the difficulty of the general problem is already captured by this instance.

We will denote by $f(n, r, s)$ the answer to the above question, so it is the maximum number of vertices that we can guarantee to connect using edges in up to s colours in any r-edge colouring of K_n. For example, $f(n, 2, 1) = n$ is a restatement of the observation of Erdős and Rado, and determining $f(n, r, 1)$ precisely recovers the monochromatic component problem of Gerencsér and Gyárfás discussed above.

Gyárfás [27] and, independently, Füredi [23] showed $f(n, r, 1) \geq \frac{n}{r-1}$ – that this is tight given certain divisibility conditions on n and r and is always close to tight. See [24, 28] for more details on the plethora of results surrounding this classical problem. Turning to what is known about $f(n, r, s)$ for larger s, one can observe that the usual examples based on affine planes perform very poorly even for $s = 2$ since using any two colours one can connect almost the whole graph. Here, Liu, Morris and Prince [36] showed $f(n, r, 2) \geq \frac{4n}{r+1}$ – that this is tight given certain divisibility conditions on n and r and is always close to tight (in some well-defined quantitative sense). See also a paper of Liu and Person [38] obtaining more precise results for the $s = 2$ case of the more general k-connectivity version

of the question. For larger s, the best-known lower bound, also due to Liu, Morris and Prince [36], was $f(n, r, s) \geq (1 - e^{s^2/(3r)})n$, which reduces to $f(n, r, s) \geq \Omega(\frac{s^2}{r}) \cdot n$ when $s \leq \sqrt{r}$. In terms of upper bounds for very small values of s, the best bound they obtain is $f(n, r, s) \leq \frac{2s}{r+1} \cdot n$, given some divisibility conditions, which also implies a slightly weaker bound holds in general.

Our first result shows this upper bound is essentially tight, closing the gap between the quadratic in s lower bound and exponential in s upper bound for any fixed s, and r large enough, determining the answer up to an absolute constant factor.

Theorem 1.1. *Let $s \geq 1$, r large enough and $n \geq \Omega(r)$. Then $f(n, r, s) = \frac{\Theta(2^s)}{r} \cdot n$.*

We actually prove the same result holds for the more general question of Bollobás, requiring a k-connected subgraph, provided n is large enough compared to k and r; see Theorem 3.5 for the exact details. This solves a problem posed by Liu, Morris and Prince in 2007 asking whether the behaviour is exponential, polynomial or something in between in s when r is large compared to s. They asked this problem as an easier variant of a conjecture that a more precise version of Theorem 1.1 holds for any $s \leq \frac{1}{2} \log r$. We disprove this conjecture in a strong form, by improving the above-mentioned exponential upper bound in s to an essentially quadratic one. This bound also determines the answer up to a logarithmic factor for any $s \leq \sqrt{r}$.

Theorem 1.2. *For $1 \leq s \leq \sqrt{r}$ and n large enough, we have $f(n, r, s) = \widetilde{\Theta}_r\left(\frac{s^2}{r}\right) \cdot n$.*

Let us note that Theorem 1.1 holds for $s \leq (1 - o_r(1)) \log \log r$, and this is essentially tight since a more precise version of Theorem 1.2 (see Proposition 3.6) shows it cannot hold for $s \geq (1 + o_r(1)) \log \log r$.

The best known upper bound, also from [36], when $\sqrt{r} \ll s < \frac{r}{2}$, was $f(n, r, s) \leq \left\lceil \left(1 - \binom{r}{s}^{-1}\right)n \right\rceil$. They also show this bound is tight if r is odd and $s = \frac{r-1}{2}$, and it matches the above-stated lower bound from [36], up to an absolute constant in the exponent for $s \geq \Omega(r)$. We improve this upper bound for $\sqrt{r \log r} \leq s \ll r$ showing the above lower bound is tight up to an absolute constant in the exponent for the whole of this range.

Theorem 1.3. *For $\sqrt{r \log r} \leq s \ll r$ and n large enough, we have $f(n, r, s) = (1 - e^{\Theta(s^2/r)})n$.*

We also obtain an improved bound for $\sqrt{r} \ll s \leq \sqrt{r \log r}$, although for this regime, our new bound is not tight. For a summary theorem collecting the current state of affairs on the question of determining $f(n, r, s)$, including our new results, we point the reader to Theorem 4.1 in Section 4.

The starting point for the second question we will explore is a problem dating back at least to the work of Erdős and Rényi from 1956 [19], and it asks, given n and m, to determine the smallest number of m-cliques which cover all edges of K_n. They observed an immediate counting lower bound of $\binom{n}{2}/\binom{m}{2}$ and that it is tight if and only if an appropriate combinatorial design exists, as well as proved a variety of asymptotic results. Let us also mention that this is perhaps the most basic and one of the oldest examples of the extensively studied 'clique covering' problem which asks for how many cliques one needs to cover all edges of an arbitrary graph (see, for example, a survey [43] for more details on the vast body of work in this direction). The question we will consider in this language asks what happens if instead of insisting that each clique has size m (or equivalently, size at most m), we restrict the size of the union of any s cliques to be at most m. We will, however, focus here on an equivalent 'inverse' formulation of the problem, which in the original $s = 1$ instance was first considered by Mills in 1979 [41] in order to understand what happens if n is not necessarily very large compared to m and the asymptotic results of Erdős and Rényi and later ones of Erdős and Hanani [17] do not yet kick in. The inverse of the classical problem asks, given a covering of the edges of K_n with r cliques, how large does a largest clique we used need to be? In this formulation, our question asks how large a union of some s cliques needs to be.

In fact, there is a third natural, equivalent reformulation which uncovers a relation to the largest monochromatic component question of Gerencsér and Gyárfás and the 'power of many colours' generalisation of it discussed above. The classical $s = 1$ instance is equivalent to asking how many vertices

are certain to be incident to at least one edge of some colour in any r-edge colouring of K_n. Indeed, given an r-edge colouring of K_n, we can obtain a clique covering by taking for each colour a clique consisting of all vertices incident to at least one edge of the said colour. For the other direction, given a clique covering, we assign colours to each clique and then colour an edge in the colour of a clique covering it (choosing arbitrarily if there is more than one option). In this setting, there is a natural 'power of many colours' generalisation which simply asks what happens if instead of insisting on a single colour we allow ourselves to use up to s colours.

Let us denote by $g(n, r, s)$ the answer to this question, so the minimum number of vertices that are certain to be incident to at least one edge in a colour from some set of s colours in any r-edge colouring of K_n. We note that determining $g(n, r, 1)$ corresponds to the classical problems discussed above and is reasonably well understood. There have been plenty of papers focusing on what happens for small values of r [20, 41, 42, 31, 47, 45]. There are also several results [46, 22] obtaining more general, precise results provided $r = p^2 + p + 1$, and a finite projective plane of order p exists, for example, if in addition, $r \mid n$, then $g(n, r, 1) = \frac{p+1}{p^2+p+1} \cdot n$. Füredi [22] proved in 1990 a very nice linear programming reduction to a question involving fractional cover numbers of intersecting hypergraphs which can be used to prove various approximate results on $g(n, r, 1)$. In terms of asymptotics, it is fairly easy to see that $g(n, r, 1) = \frac{1+o_r(1)}{\sqrt{r}} \cdot n$ for n large enough. The lower bound is the above-mentioned counting argument of Erdős and Rényi, while the basic idea of using projective planes to show the upper bound dates back to a paper of Erdős and Graham [15] from 1975, who worked on a closely related Ramsey theoretic question. See also a recent paper spelling this out in more detail [26]. For more information, we point the interested reader to Chapter 7 of a survey of Füredi [24].

Regarding $g(n, r, s)$, we determine the answer up to a constant factor for all $1 \le s \le r$, as well as obtain much more precise results for various ranges including the correct asymptotics for essentially the whole range.

In terms of how the two functions are related, we always have $f(n, r, s) \le g(n, r, s)$ since in a connected component using up to s colours, we know that every vertex of the component must be incident to an edge in one of these s colours. In some sense, this positions the question of determining $g(n, r, s)$ as the '0-connectivity' instance of the question of Bollobás, where in this context, we say a graph is 0-connected if it does not contain an isolated vertex. With this in mind, similarly, as upper bounds on $f(n, r, s)$ serve as often best-known bounds for any instance of the question of Bollobás, upper bounding $g(n, r, s)$ provides us with our improved upper bound on $f(n, r, s)$, which is behind Theorem 1.3.

The known classical results tell us that $f(n, r, 1)$ is roughly equal to $\frac{n}{r}$, while $g(n, r, 1)$ is roughly $\frac{n}{\sqrt{r}}$, showing that the functions have very different behaviour for $s = 1$. However, we show the functions have the same asymptotic behaviour for $s \ge \sqrt{r \log r}$. The main additional result for the g function is that we also understand its behaviour for all $s \ll \sqrt{r}$; namely, $g(n, r, s) = (1 + o_r(1)) \frac{s}{\sqrt{r}} \cdot n$ showing it grows linearly in s until the answer becomes an absolute constant times n, which at least for very small s is in stark contrast to the f function, which, by Theorem 1.1, initially grows exponentially fast in s. The following theorem collects most of our results for the g function and showcases its behaviour across the range.

Theorem 1.4. *Given integers $1 \le s \le r$ and n large enough compared to r, we have*

$$g(n, r, s) = \begin{cases} \frac{(1+o_r(1))s}{\sqrt{r}} \cdot n, & \text{for } s \ll \sqrt{r}; \\ \Theta(n), & \text{for } s = \Theta(\sqrt{r}); \\ \left(1 - e^{-\Theta(s^2/r)}\right) \cdot n, & \text{for } s \ge \sqrt{r \log r}; \\ \left(1 - (1 - o_r(1)) \cdot \binom{r-s-1}{s} / \binom{r}{s}\right) \cdot n, & \text{for } s \ge (1 - o_r(1)) \cdot \frac{r}{2}; \\ \left\lceil (1 - 1/\binom{r}{s})n \right\rceil, & \text{for } s = \frac{r-1}{2} \text{ and } r \text{ odd}; \\ n, & \text{for } s \ge \frac{r}{2}. \end{cases}$$

We note that in addition to the colouring formulation we mostly focused on, all our results can be translated to the alternative, clique-cover formulations of the problem mentioned above. There is one final Ramsey theoretic reformulation which we want to mention. To this end, let us denote by $\mathcal{T}_m$ the set of all trees on m vertices and by $\mathcal{F}_m$ the set of all spanning forests (without isolated vertices) on m vertices. Then determining the r-colour Ramsey number of these two families is equivalent to determining $f(n, r, 1)$ and $g(n, r, 1)$. Let us note that the question of obtaining bounds on Ramsey numbers of an arbitrary member of $\mathcal{F}_m$ as well as $\mathcal{T}_m$ has already been considered in 1975 by Erdős and Graham [15], and one can use the same argument to get an approximate understanding of the behaviour of $f(n, r, 1)$ and $g(n, r, 1)$. If one now considers the 'power of many colours' generalisation of these two Ramsey problems – namely, if instead of a monochromatic copy, we seek one which uses up to s colours, so precisely in the sense introduced by Erdős, Hajnal and Rado, and discussed at the beginning of the section – one recovers the problems of determining $f(n, r, s)$ and $g(n, r, s)$.

Most of our arguments use various probabilistic tools and ideas, often both for lower and upper bounds; in certain regimes, expansion arguments come in useful. We also prove a number of extremal set theoretic reductions which often lead to interesting questions in their own right, which we collect and discuss further in Section 4.

1.1. Organisation and notation

We begin with our results on the generalised clique covering function $g(n, r, s)$ in Section 2. We open the section by proving two general tools. The first one, Lemma 2.1, is a simple probabilistic bound that we will use to prove most of our lower bounds. The second one, Lemma 2.2, gives a reduction to an extremal set-theoretic problem, which is behind most of our upper bounds.

In Section 2.2, we prove the first two parts of Theorem 1.4, including a more precise lower bound, given certain divisibility conditions, in the form of Proposition 2.3. In the subsequent Section 2.3, we prove the remaining parts of Theorem 1.4 concerning larger values of s. The third and fourth parts follow from Proposition 2.6 and Corollary 2.8. The fifth part follows from Proposition 2.9 and Corollary 2.10 which is perhaps the most involved of the arguments related to the function g including an intriguing, somewhat nonstandard application of Janson's Inequality in tandem with the Lovász Local Lemma. The penultimate part is established in Proposition 2.11. The final part is the easy reduction to the Erdős-Rado observation, which was already discussed in the Introduction. In Section 2.4, we collect what our results tell us about the behaviour of $g(n, r, s)$ for small values of r.

In Section 3, we prove our results on the large connected component using few colours function $f(n, r, s)$. We begin with a part of the upper bound of Theorem 1.1, due to [36]. We include it for completeness and since it helps give motivation for the argument behind our new lower bound, which we prove next as Theorem 3.5. We note that we prove this result in the full generality of the question of Bollobás requiring k-connectivity of the subgraph we find. We continue with a slightly simpler argument for lower bounding $f(n, r, s)$, which improves the previously best constant in the exponent as Proposition 3.6, and which combined with our new upper bound for the g function proves Theorem 1.3. We conclude the section with our new upper bound for $f(n, r, s)$, for $s \leq \sqrt{r}$ which we prove as Proposition 3.7, and which combined with the known lower bound proves Theorem 1.3.

In Section 4, we make some final remarks and collect a number of interesting open problems.

Notation. Given a graph or a hypergraph H, we will denote by $V(H)$ and $E(H)$ its vertex and edge set, respectively, and by $d(H)$ its average degree. We use the standard asymptotic notation. Some conventions are that we write $g = O(f)$ if there exists an absolute constant $C > 0$ such that $g(x) \leq C \cdot f(x)$ for all x. We also write $g = \Omega(f)$ if $f = O(g)$, and $\Theta(f)$ if both $g = O(f)$ and $f = O(g)$. We index the asymptotic notation by a variable we are letting tend to infinity when we wish to emphasise or clarify which one drives the asymptotics. For example, we write $o_r(f)$ to be a function which tends to zero when divided by f as the parameter r tends to infinity. We also write $f \ll g$ to mean $f = o(g)$ when both f and g depend only on a single parameter which is then assumed to tend to infinity. We write $f = \tilde{\Theta}_r(g)$ if $f = \Omega(g)$ and $f \leq O(g) \cdot \log r$. All our logarithms are in base two. Given two events A and B in the same probability space, we write $A \sim B$ if they are dependent.

2. Power of many colours for covering vertices

In this section, we prove our results on the generalisation of the clique-covering problem, (i.e., determining the behaviour of the function $g(n, r, s)$). We note though that most of the upper bounds we will prove in this section are also the best upper bounds we have for the large connected component using few colours function $f(n, r, s)$ using the easy bound $f(n, r, s) \leq g(n, r, s)$ explained in the introduction. We begin with the general tools proven in the following subsection. We then proceed to deduce various parts of our main result, Theorem 1.4, using them.

2.1. General tools

We begin with the following simple lemma which is behind most of our lower bounds on the generalised clique covering function $g(n, r, s)$.

Lemma 2.1. *Let r, s and d be positive integers such that $s \leq d < r - s$. Then*

$$g(n, r, s) \geq \min\left\{1 + \frac{s(n - 1)}{d}, \left(1 - \binom{r - d - 1}{s} \bigg/ \binom{r}{s}\right)n\right\}.$$

Proof. Consider an r-edge colouring of K_n. Our goal is to show that there exist s colours such that there are at least $1 + \frac{s(n-1)}{d}$ vertices incident to an edge coloured in one of these s colours or at least $n - n \cdot \binom{r-d-1}{s} / \binom{r}{s}$.

Suppose first that there is a vertex v for which there are at most d different colours used on the edges incident to v. In this case, the s colours with most edges incident to v account for at least $\frac{s(n-1)}{d}$ edges incident to v. Counting, in addition, v this implies that there are at least $1 + \frac{s(n-1)}{d}$ vertices incident to an edge coloured using one of these s colours, as desired.

Hence, we may assume that for every vertex v, there are at least $d + 1$ different colours used on the edges incident to v. We will choose our s colours uniformly at random. The probability that we did not pick any of the at least $d + 1$ colours incident to v is at most $\binom{r-d-1}{s} / \binom{r}{s}$. So the expected number of vertices not incident to an edge using one of our random s colours is at most $n \cdot \binom{r-d-1}{s} / \binom{r}{s}$. In particular, there exists a choice of s colours for which there are at most this many vertices not incident to an edge using one of them. We conclude that there are at least $n - n \cdot \binom{r-d-1}{s} / \binom{r}{s}$ vertices incident to an edge using one of these colours.   □

We note that strictly speaking, the assumptions $s \leq d$ and $d < r - s$ are not necessary in the above lemma. This follows since the terms in the minimum are decreasing and increasing in d, respectively, and evaluate to n in the case $d = s$ and $d = r - s$, respectively. We will only ever use the lemma in the specified range since the same reasoning explains one cannot obtain a better bound by choosing d outside this range.

We next turn to a lemma which will help us prove our upper bounds on $g(n, r, s)$.

Lemma 2.2. *Let $1 \leq s \leq r$ be integers. Suppose there exists a hypergraph H with vertex set $[r]$ such that*

1. *for any $e, f \in E(H)$, we have $e \cap f \neq \emptyset$ (i.e., H is intersecting), and*
2. *any set of $r - s$ vertices of H contains at least t distinct edges.*

Then

$$g(n, r, s) \leq n - t \cdot \left\lfloor \frac{n}{|E(H)|} \right\rfloor.$$

Proof. Our goal is to colour the edges of K_n into r colours in such a way that for any s colours, there are quite a few vertices of K_n not incident to an edge coloured using one of the s colours. To do this, we will exploit the existence of the auxiliary hypergraph H.

Let $m := |E(H)|$, and let us partition the vertices of K_n into m parts A_e, one for each edge $e \in E(H)$ with each A_e of size either $\lfloor n/m \rfloor$ or $\lceil n/m \rceil$. We note that the r vertices of H represent the r colours we are going to use, while an edge $e \in E(H)$ should be viewed as the set of colours we are 'permitted' to use on the edges of our K_n incident to a vertex belonging to A_e. With this in mind, we now colour all the edges of K_n as follows. Given any two distinct vertices belonging to A_e and $A_{e'}$, respectively (with possibly $e = e'$), we colour the edge between them using an arbitrary colour in $e \cap e'$. We know such a colour always exists since $e \cap e' \neq \emptyset$, by the assumption that H is intersecting – namely, property 1.

Now to show this colouring satisfies the desired property suppose S is an arbitrary set of s colours. By property 2, we know the set of $r - s$ colours in $V(H) \setminus S$ contains at least t edges, the set of which we denote by T. Observe now that any vertex v belonging to A_e with $e \in T$ is not incident to an edge coloured in a colour in S. Indeed, by the definition of our colouring, we are only ever colouring an edge incident to a vertex $v \in A_e$ by a colour which belongs to e, and since $e \in T$, we know $e \subseteq V(H) \setminus S$, so, in particular, no colour in S is allowed to be used. This implies the desired claim since there are at least $t \cdot \lfloor n/m \rfloor$ vertices of K_n contained in one of the sets A_e with $e \in T$.   □

Both above lemmas might seem somewhat wasteful, but as we will soon see, both can be tight, and they suffice to give a pretty good understanding of the behaviour of $g(n, r, s)$ in general.

Let us also note that if one starts with an r-edge colouring of K_n, one can define a hypergraph H with the vertex set being the set of r colours and an edge $e(v)$ of H for each vertex v of K_n, where $e(v)$ consists of all the colours appearing on edges of K_n incident to v. We note that certain edges might appear multiple times, so H is, strictly speaking, a multi-hypergraph. By construction, H must be intersecting, and the number of vertices of K_n not incident to an edge of colour in some set S of s colours is precisely equal to n minus the number of edges of H induced by $V(H) \setminus S$. This shows that the question of determining $g(n, r, s)$ is equivalent to finding a multi hypergraph H as in Lemma 2.2 with the maximum possible value of t.

Let us also observe that we are not specifying the uniformity of H; in fact, we are free to choose it according to our heart's desire. That said, if any edge of H has size at most s then its complement is a set of at least $r - s$ vertices not containing any edges (since H must be intersecting), so in that case the best t we can choose is 0 and we only obtain the trivial bound. In particular, if we wish to obtain nontrivial bounds we should always choose every edge of H to have a size of at least $s + 1$. In most of our applications of the lemma, we will choose H to be $(s + 1)$-uniform.

2.2. Bounds when s is small

In this section, we will show our results on $g(n, r, s)$ for $s \leq \sqrt{r}$. We begin with an upper bound under some divisibility conditions.

Proposition 2.3. *Let s be a positive integer, and let p be a power of a prime number. Setting $r = p^2 + p + 1$, we have for any n divisible by r*

$$g(n, r, s) \leq (sp + 1) \cdot \frac{n}{r}.$$

Proof. Let $r = p^2 + p + 1$. If $s \geq p + 1$, then the inequality is implied by the trivial upper bound of $g(n, r, s) \leq n$, so we may assume $s \leq p$. By the assumption that p is a power of a prime, we know there exists a projective plane of order p. In particular, there exists an r-vertex, r-edge, $(p + 1)$-uniform hypergraph H in which every vertex belongs to exactly $p + 1$ edges and every pair of vertices belongs to a unique edge.

Observe first that if we fix a set S of s vertices of H, it intersects at most $sp + 1$ edges of H. To see this, fix some $v \in S$. Note that v is contained in precisely $p + 1$ edges, but each $w \in S \setminus \{v\}$ belongs to at most p edges not already accounted for (since one of the edges containing w contains v as well). Therefore, there are at most $p + 1 + (s - 1)p = sp + 1$ edges intersecting S in total.

This in particular implies that any set of $r - s$ vertices contains at least $r - sp - 1$ edges of H, so Lemma 2.2 implies

$$g(n, r, s) \le n - (r - sp - 1) \cdot \frac{n}{r} = (sp + 1) \cdot \frac{n}{r},$$

as desired.   □

Proposition 2.3 is tight for $s = 1$ and $s = 2$ by applying Lemma 2.1 with $d = p$. We will also soon see that it is asymptotically tight for any $s \ll \sqrt{r}$. But let us first extract from it an approximate upper bound which applies without any divisibility assumptions and is similarly asymptotically optimal for $s \ll \sqrt{r}$ and somewhat large n.

Corollary 2.4. *Suppose $n \gg r^{3/2}/s$. Then*

$$g(n, r, s) \le (1 + o_r(1)) \cdot \frac{s}{\sqrt{r}} \cdot n,$$

where the $o_r(1)$ term is a function of r which tends to 0 as $r \to \infty$.

Proof. Note that the desired inequality holds by the trivial upper bound of $g(n, r, s) \le n$ if $s \ge \sqrt{r}$, so we may assume $s < \sqrt{r}$. Let r' be the largest integer smaller or equal to r of the form $p^2 + p + 1$ where p is a prime. By well-known results on the difference between consecutive primes (see, for example, [4]), we know that $p = (1 - o_r(1))\sqrt{r}$. Let finally $n' = n - (n \bmod r')$ so that $r' \mid n'$ and $n' = (1 - o_r(1))n$. We now have

$$g(n, r, s) \le g(n, r', s) \le g(n', r', s) + n - n' \le (sp + 1) \cdot \frac{n'}{r'} + r' \le (1 + o_r(1)) \cdot \frac{s}{\sqrt{r}} \cdot n,$$

where in the penultimate inequality we used Proposition 2.3 and in the final the assumption that $n \gg r^{3/2}/s$.   □

We note that some lower bound on n in terms of s and r is clearly necessary, and one can easily extend the range for which the above result applies by weakening it after replacing the $1 + o_r(1)$ term with a $O_r(1)$ term.

We now turn to the lower bounds. We begin by showing that Corollary 2.4 is asymptotically tight.

Proposition 2.5. *Given positive integers n, r, s such that $s \le \sqrt{r}$, we have*

$$g(n, r, s) \ge \frac{1}{2} \cdot \frac{s}{\sqrt{r}} \cdot n;$$

furthermore, if $s \ll \sqrt{r}$, then

$$g(n, r, s) \ge (1 - o_r(1)) \cdot \frac{s}{\sqrt{r}} \cdot n.$$

Proof. Let us apply Lemma 2.1 with $d = \lfloor \sqrt{r} \rfloor \ge s$. This shows that either $g(n, r, s) \ge 1 + \frac{s(n-1)}{\lfloor \sqrt{r} \rfloor} \ge \frac{sn}{\lfloor \sqrt{r} \rfloor}$ in which case we are done or

$$\frac{g(n, r, s)}{n} \ge 1 - \frac{\binom{r-d-1}{s}}{\binom{r}{s}} \ge 1 - \left(1 - \frac{d+1}{r}\right)^s \ge \frac{s(d+1)}{r} - \binom{s}{2} \cdot \frac{(d+1)^2}{r^2}$$

$$= \frac{s(d+1)}{r} \cdot \left(1 - \frac{(s-1)(d+1)}{2r}\right) \ge \frac{s}{\sqrt{r}} \cdot \left(1 - \frac{sd}{2r}\right),$$

which in turn is at least $(1 - o_r(1)) \cdot \frac{s}{\sqrt{r}}$ for $s \ll \sqrt{r}$, and at least $\frac{1}{2} \cdot \frac{s}{\sqrt{r}}$ for $s \le \sqrt{r}$, as desired.   □

2.3. Bounds when s is large

In this section, we show our bounds on $g(n, r, s)$ for $s \geq \sqrt{r}$. We begin with the lower bounds.

Proposition 2.6. *Let $s \leq r$ be positive integers. Then, for any n we have*

$$g(n, r, s) \geq n \cdot \left(1 - \binom{r-s-1}{s} \Big/ \binom{r}{s}\right) \geq n \cdot \left(1 - e^{-\frac{s(s+1)}{r}}\right).$$

Proof. Let us apply Lemma 2.1 with $d = s$. Since the first term is equal to n and hence always at least as large as the second, we always have

$$g(n, r, s) \geq n - n \cdot \binom{r-s-1}{s} \Big/ \binom{r}{s} \geq n - n \cdot \left(1 - \frac{s+1}{r}\right)^s \geq n \cdot \left(1 - e^{-\frac{s(s+1)}{r}}\right),$$

as desired. □

The following result will be behind an almost matching upper bound we obtain for $s \geq \sqrt{r \log r}$.

Proposition 2.7. *Let s, r be positive integers such that $\sqrt{r \log r} \leq s \leq \frac{r}{32}$. Then there exists an r-vertex intersecting hypergraph of uniformity $8s$ and with cover number larger than s with at most $e^{O(s^2/r)}$ edges.*

Proof. Let $u := 8s$. Let H be a random hypergraph obtained by picking $m = \left\lfloor e^{u^2/(2r)} \right\rfloor$ random edges $e_1, \ldots, e_m \subseteq [r]$, each sampled uniformly at random among all subsets of $[r]$ of size u, independently from all other edges.

First, observe that for any $1 \leq i, j \leq m$, the probability that e_i and e_j do not intersect is at most

$$\binom{r-u}{u} \Big/ \binom{r}{u} = \frac{(r-u) \cdot \ldots \cdot (r-2u+1)}{r \cdot \ldots \cdot (r-u+1)} < \left(1 - \frac{u}{r}\right)^u < e^{-u^2/r}.$$

So by a union bound, the probability that there exists a non-intersecting pair of edges is less than

$$\binom{m}{2} \cdot e^{-u^2/r} \leq \frac{m^2}{2} \cdot e^{-u^2/r} \leq \frac{1}{2}.$$

However, given a fixed set of $r - s$ vertices, the probability that e_i is contained in it equals

$$\binom{r-s}{u} \Big/ \binom{r}{u} \geq \left(1 - \frac{s}{r-u+1}\right)^u \geq \left(1 - \frac{4s}{3r}\right)^u \geq e^{-3su/r+1},$$

where in the penultimate inequality we used $u = 8s \leq \frac{r}{4}$ and in the final inequality we used that $1 - x \geq e^{-2x}$ for any real $0 \leq x \leq \frac{1}{2}$ and that $\frac{su}{3r} \geq 1$.

Hence, the probability that none of our random m edges is contained in our set of size $r - s$ is at most

$$\left(1 - e^{-3su/r+1}\right)^m \leq \exp\left(-m \cdot e^{-3su/r+1}\right) \leq \exp\left(-e^{su/r}\right).$$

Since there are $\binom{r}{r-s} = \binom{r}{s} \leq \frac{r^s}{s!} \leq \frac{r^s}{2}$ subsets of size $r - s$, another union bound implies that the probability that there exists a set of $r - s$ vertices not containing any edge is at most

$$\frac{r^s}{2} \cdot \exp\left(-e^{su/r}\right) = \frac{1}{2} \cdot \exp\left(s \log r - e^{su/r}\right) \leq \frac{1}{2},$$

where the final inequality follows since $su/r = 8s^2/r \geq 8 \log r \geq \log s + \log \log r$.

A final union bound implies that there exists an outcome in which the hypergraph we obtain is intersecting and any subset of $r - s$ vertices contains an edge; in other words, there is no cover of size at most s, as desired.   □

We note that Proposition 2.7 is tight up to the constant factor in the exponent as can be seen by an easy double-counting argument. Indeed, there needs to be at least $\binom{r}{r-s}$ pairs of a set of $r - s$ vertices containing an edge of our hypergraph, and each edge can belong to at most $\binom{r-u}{r-s-u}$ sets. Therefore, our hypergraph needs to have at least $\binom{r}{r-s}/\binom{r-u}{r-s-u} = \binom{r}{s}/\binom{r-u}{s} \geq \left(1 + \frac{u}{r-u}\right)^s \geq e^{\Omega(s^2/r)}$ edges.

Corollary 2.8. *Let s, r be positive integers such that $\sqrt{r \log r} \leq s \leq \frac{r}{32}$. Then there exists $C > 0$ such that for any $n \geq e^{Cs^2/r}$, we have*

$$g(n, r, s) \leq \left(1 - e^{-O(s^2/r)}\right) \cdot n.$$

Proof. Proposition 2.7 provides us with an r-vertex hypergraph with at most $e^{O(s^2/r)}$ edges which is intersecting, and any set of $r - s$ vertices contains at least one edge. This allows us to apply Lemma 2.2 with $t = 1$ to obtain

$$g(n, r, s) \leq n - \left\lfloor \frac{n}{e^{O(s^2/r)}} \right\rfloor \leq \left(1 - e^{-O(s^2/r)}\right) \cdot n,$$

as desired.   □

We note that by being a bit more careful in Proposition 2.7, one can find a hypergraph of uniformity $u = (2 + o_r(1))s$ and ensure via Chernoff's inequality that every set of $r - s$ vertices contains not just one but an $e^{-(1+o_r(1))su/r}$ proportion of the edges of H. This allows one to improve the constant in the exponent of the lower order term in the above corollary to $2 + o_r(1)$ which is tight up to this factor of about two, thanks to Proposition 2.6. Let us also note that Corollary 2.8 is tight in terms of the bound on n, up to the constant C, since if $n < e^{\frac{s(s+1)}{r}}$, Proposition 2.6 implies that $g(n, r, s) = n$.

Remark. We believe Corollary 2.8 should hold also when $\sqrt{r} \leq s \leq \sqrt{r \log r}$. In this case, one can prove a weaker bound than the one in Corollary 2.8, which is still better than $g(n, r, s) \leq (1 - r^{-O(1)})n$ which one gets from monotonicity and Corollary 2.8 with $s = \sqrt{r \log r}$. Namely, by choosing $m = r^8$ random edges of size $u = 4\sqrt{r \log r}$, we can guarantee the hypergraph is intersecting with probability more than a half, and by using Chernoff, we can also guarantee with probability at least a half that any set of $r - s$ vertices contains $e^{-O(su/r)} = e^{-O\left(s\sqrt{\frac{\log r}{r}}\right)}$ proportion of the edges. This shows via Lemma 2.2 that in this range, $g(n, r, s) \leq \left(1 - e^{-O\left(s\sqrt{\frac{\log r}{r}}\right)}\right)n$.

The following proposition allows us to obtain an even more precise result when s is close to $r/2$.

Proposition 2.9. *Let s, r be positive integers such that $r/2 - o_r((r/\log r)^{1/3}) \leq s < r/2 - 1$. Then there exists an r-vertex intersecting hypergraph H of uniformity $s + 1$ such that every set of $r - s$ vertices of H contains at least $(1 - o_r(1))\frac{\binom{r-s-1}{s}}{\binom{r}{s}} \cdot |E(H)|$ edges.*

Proof. Let $\varepsilon > 0$, and let us also for convenience set

$$x := s + 1 \qquad \text{and} \qquad d := r/2 - x = r/2 - s - 1 = o((r/\log r)^{1/3}).$$

Our general strategy is similar to the one we used in Proposition 2.7. However, in order to obtain a result at our desired level of precision, we cannot afford to simply sample every edge with an appropriate probability. What we do is for each $X \in \binom{[r]}{x}$, sample a Bernoulli random variable z_X with probability

$p = \frac{1}{\binom{r-x}{x}} \leq \frac{1}{2}$, where the inequality holds since $r - x > x$. We now choose X to be an edge if $z_X = 1$, but $z_{X'} = 0$ for all $X' \in \binom{[r]\setminus X}{x}$. Let A_X be the indicator random variable for whether we selected X as an edge, so

$$A_X = z_X \cdot \prod_{X' \in \binom{[r]\setminus X}{x}} (1 - z_{X'}).$$

Note first that H is intersecting. Indeed, if $A_X = 1$ for some $X \in \binom{[r]}{x}$, then $z_X = 1$ and $z_{X'} = 0$ for all $X' \in \binom{[r]\setminus X}{x}$, which in turn guarantees $A_{X'} = 0$, so H contains no edge disjoint from X.

Let B be the number of edges of H, so $B = \sum_{X \in \binom{[r]}{x}} A_X$. Note that since for any $X \in \binom{[r]}{x}$, $\mathbb{P}(z_X = 1) = p$, and since the variables z_X are independent, we have

$$\mathbb{P}(A_X = 1) = p(1 - p)^{\binom{r-x}{x}} \geq p \cdot e^{-2p\binom{r-x}{x}} = \frac{p}{e^2},$$

where in the inequality, we used that $1 - p \geq e^{-2p}$ which holds since $p \leq \frac{1}{2}$. This combined with the linearity of expectation implies

$$\mathbb{E}[B] = \binom{r}{x} \cdot \mathbb{P}(A_X = 1) \geq \binom{r}{x} \cdot \frac{p}{e^2}.$$

Our probability space is given by $\binom{r}{x}$ independent random variables, and changing one of them changes the value of B by at most $\binom{r-x}{x}$. In other words, B is $\binom{r-x}{x}$-Lipschitz, implying by Azuma's inequality (see, for example, [3, Chapter 7]) that

$$\mathbb{P}(B > (1 + \varepsilon)\mathbb{E}[B]) \leq \exp\left(-\frac{\varepsilon^2 \mathbb{E}[B]^2}{2 \sum_{X \in \binom{[r]}{x}} \binom{r-x}{x}^2}\right) \leq \exp\left(-\binom{r}{x} \cdot \frac{\varepsilon^2 p^4}{2e^4}\right) \leq \exp\left(-\binom{r}{x} \cdot \Omega(r^{-8d})\right), \quad (1)$$

where we used that $\frac{1}{p} = \binom{r-x}{x} = \binom{r-x}{r-2x} = \binom{r-x}{2d} \leq r^{2d}$.

We next turn to the number of edges contained by vertex subsets of size $r - s$. For each $T \in \binom{[r]}{r-s}$, let B_T count the number of edges of the subgraph of H induced by T, so $B_T = \sum_{X \in \binom{T}{x}} A_X$. Hence, for any T,

$$\mathbb{E}[B_T] = \binom{|T|}{x} \cdot \mathbb{P}(A_X = 1) \geq \binom{r-s}{x} \cdot \frac{p}{e^2} = \frac{\binom{r-s}{x}}{e^2 \cdot \binom{r-s-1}{x}} = \frac{1}{e^2} \cdot \frac{r-s}{r-s-x} > \frac{1}{2e^2} \cdot \frac{r}{2d+1}. \quad (2)$$

Now, let us fix some $T \in \binom{[r]}{r-s}$. Note that the random variable B_T is increasing in the product probability space given by random variables z_X for $X \in \binom{T}{x}$ and $1 - z_X$ for $X \in \binom{[r]}{x} \setminus \binom{T}{x}$ since $r - s < 2x$, and therefore, for no $X \in T$ does A_X depend on $z_{X'}$ for some $X' \in \binom{T}{x} \setminus \{X\}$. This will allow us to apply Janson's inequality (see [3, Chapter 8]). Before applying it, we first need to understand the dependency structure.

Given some $X, X' \in \binom{T}{x}$, A_X and $A_{X'}$ are functions of disjoint sets of independent variables if and only if $r - |X \cup X'| < x$ since this is equivalent to there being no $X'' \in \binom{[r]}{x}$ disjoint from both X and X' (and $r - s < 2x$ guarantees X and X' are not disjoint). It follows that A_X is not independent of $A_{X'}$ only if

$$r - x \geq |X \cup X'| = |X| + |X'| - |X \cap X'| = 2x - |X \cap X'|.$$

This is equivalent to $|X \cap X'| \geq 3x - r = x - 2d$. Therefore, for each $X \in \binom{T}{x}$, A_X is independent of all but at most

$$\sum_{i=x-2d}^{x} \binom{x}{i} \binom{r-s-x}{x-i} = \sum_{i=x-2d}^{x} \binom{x}{x-i} \binom{2d+1}{x-i}$$

$$= \sum_{i=0}^{2d} \binom{x}{i} \binom{2d+1}{i}$$

$$\leq (2d+1) \cdot \sum_{i=0}^{2d} \binom{x}{x-i} \binom{2d}{i}$$

$$= (2d+1) \cdot \binom{x+2d}{x} = (2d+1) \cdot \binom{r-x}{x} = \frac{2d+1}{p}$$

events $A_{X'}$. The above allows us to calculate the following term, with the goal of applying the lower tail version of Janson's inequality [32]:

$$\Delta(B_T) = \sum_{\substack{X, X' \in \binom{T}{x} \\ \text{s.t. } A_X \sim A_{X'}}} \mathbb{E}[A_X A_{X'}] \leq \mathbb{E}[B_T] + \binom{r-s}{x} \cdot \frac{2d+1}{p} \cdot \mathbb{P}[A_X] \cdot p = (2d+2) \cdot \mathbb{E}[B_T],$$

where in the inequality, we used that $\mathbb{P}(A_X \cap A_{X'}) \leq \mathbb{P}[A_X] \cdot p$, which holds since for $A_{X'}$ to happen, we need to have $z_{X'} = 1$, which happens with probability p, independently of A_X since $r - s - x < x$. We get

$$\frac{\mathbb{E}[B_T]^2}{\Delta(B_T)} \geq \frac{\mathbb{E}[B_T]}{2d+2} \geq \frac{r}{2e^2(2d+1)(2d+2)},$$

where we used (2) in the final inequality. Finally, the lower tail version of Janson's inequality [32] tells us

$$\mathbb{P}[B_T \leq (1-\varepsilon)\mathbb{E}[B_T]] \leq \exp\left(-\frac{\varepsilon^2 \mathbb{E}[B_T]^2}{2\Delta(B_T)}\right) \leq \exp\left(-\Omega\left(\frac{r}{d^2}\right)\right). \tag{3}$$

The last step is to apply the quantitative version of the Lovász Local Lemma (see [3, Chapter 5]) with 'bad events' being $B_T < (1-\varepsilon)\mathbb{E}[B_T]$ for $T \in \binom{[r]}{r-s}$ to show that all these events can be avoided with probability which is greater than the probability that $B > (1+\varepsilon)\mathbb{E}[B]$ estimated in (1). Towards this, we first need to estimate, given $T \in \binom{[r]}{r-s}$, for how many $T' \in \binom{[r]}{r-s}$ do B_T and $B_{T'}$ both depend on the same z_X for some $X \in \binom{[r]}{x}$, since B_T is independent of all other $B_{T'}$. Towards this, note that $B_T = \sum_{X \in \binom{T}{x}} A_X$, and that each A_X only depends on z_X and $z_{X'}$ for X' disjoint from X. This means that B_T depends only on z_X when $X \subseteq T$ or $|T \setminus X| \geq x$ (as otherwise X is not disjoint from any subset of T of size x). To summarise, B_T (and similarly $B_{T'}$) depends on z_X if and only if one of the following holds:

○ $X \subseteq T$ or
○ $|T \setminus X| \geq x$, or equivalently $|T \cap X| \leq r - s - x = 2d + 1$.

It follows that there exists a z_X such that both B_T and $B_{T'}$ depend on it only if one of the following holds:

1. $|T \cap T'| \geq x$ (there exists $X \in \binom{[r]}{x}$ such that $X \subseteq T \cap T'$);
2. $|T \cap T'| \leq 4d + 2$ (there exists X with $|X \cap T| \leq 2d + 1$ and $X \subseteq T'$ or vice versa);
3. $|T \cap T'| \geq r/2 - 3d$ (there exists X with $|X \cap T|, |X \cap T'| \leq 2d + 1$),

where 2 follows since $|X \cap T| \leq 2d + 1$ and $X \subseteq T'$ imply $x - (2d+1) \leq |X \cap (T' \setminus T)| \leq r - s - |T \cap T'|$. Similarly, 3 follows since $|X \cap T|, |X \cap T'| \leq 2d + 1$ imply that $x \leq r - |T \cup T'| + |X \cap T| + |X \cap T'|$, which is equivalent to $r/2 - 3d \leq |T \cap T'|$.

If we fix T of size $r - s$, then there are

$$\binom{r-s}{i} \binom{s}{r-s-i} = \binom{r-s}{i} \binom{s}{2s-r+i} = \binom{r-s}{i} \binom{s}{i-2d-2}$$

choices for T' of size $r - s$ such that $|T \cap T'| = i$.

Since $x = r/2 - d$, 1 is implied by 3, so B_T is independent of all but at most

$$\sum_{i=2d+2}^{4d+2} \binom{r-s}{i}\binom{s}{i-2d-2} + \sum_{i=r/2-3d}^{r-s} \binom{r-s}{r-s-i}\binom{s}{r-s-i} \le$$

$$(2d+1)\binom{r-s}{4d+2}\binom{s}{2d} + \sum_{i=0}^{4d+2} \binom{r-s}{i}\binom{s}{i} \le$$

$$2(4d+2)\binom{r-s}{4d+2}\binom{s}{4d+2} \le r^{8d+4}$$

B'_T. Note next that since $d^3 = o(r/\log r)$, we get from (3)

$$\mathbb{P}[B_T \le (1-\varepsilon)\mathbb{E}[B_T]] \le \exp\left(-\Omega\left(\frac{r}{d^2}\right)\right) = r^{-\omega(d)} \le r^{-8d-4}(1 - r^{-8d-4})^{r^{8d+4}},$$

allowing us to apply the Lovász Local Lemma to conclude

$$\mathbb{P}\left[\bigcap_{T \in \binom{[r]}{r-s}} B_T > (1-\varepsilon)\mathbb{E}[B_T]\right] \ge (1 - r^{-8d-4})^{\binom{r}{r-s}} \ge \exp\left(-2\binom{r}{r-s}/r^{8d+4}\right) > \exp\left(-\binom{r}{x}\cdot\Omega(r^{-8d})\right)$$

$$\ge \mathbb{P}(B > (1+\varepsilon)\mathbb{E}[B]).$$

This implies that there exists an outcome in which for each $T \in \binom{[r]}{r-s}$, we have $B_T > (1-\varepsilon)\mathbb{E}[B_T]$, and $B \le (1+\varepsilon)\mathbb{E}[B]$. In other words, there exists an $(s+1)$-uniform intersecting hypergraph with r vertices such that any $r - s$ vertices contain at least $(1-\varepsilon)\mathbb{E}[B_T]$ edges, while the whole hypergraph has at most $(1+\varepsilon)\mathbb{E}[B]$ edges. The conclusion follows from

$$\frac{(1-\varepsilon)\mathbb{E}[B_T]}{(1+\varepsilon)\mathbb{E}[B]} \ge (1-\varepsilon)^2 \frac{\binom{r-s}{x}}{\binom{r}{x}} = (1-\varepsilon)^2 \frac{\binom{r-s-1}{s}}{\binom{r}{s}}$$

and letting $\varepsilon \to 0$. □

Corollary 2.10. *Let s, r be positive integers such that $r/2 - o_r((r/\log r)^{1/3}) \le s < r/2 - 1$. Then for large enough n,*

$$g(n, r, s) \le n\left(1 - (1 - o_r(1)) \cdot \binom{r-s-1}{s}/\binom{r}{s}\right).$$

Proof. Proposition 2.9 provides us with an r-vertex intersecting hypergraph H such that any set of $r - s$ vertices contains at least $(1 - o_r(1))\dfrac{\binom{r-s-1}{s}}{\binom{r}{s}} \cdot |E(H)|$ edges. Lemma 2.2 applied with $t = (1 - o_r(1))\dfrac{\binom{r-s-1}{s}}{\binom{r}{s}}$ now implies

$$g(n, r, s) \le n - (1 - o_r(1))\frac{\binom{r-s-1}{s}}{\binom{r}{s}} \cdot |E(H)| \cdot \left\lfloor \frac{n}{|E(H)|} \right\rfloor \le n - (1 - o_r(1))n \cdot \frac{\binom{r-s-1}{s}}{\binom{r}{s}},$$

where we used the assumption that n is sufficiently large compared to r (and hence also $|E(H)|$). □

Finally, for the very end of the range, we obtain the following simple, precise result.

Proposition 2.11. *For any odd $r \geq 3$, we have*

$$g(n, r, \lfloor r/2 \rfloor) = \left\lceil \left(1 - \frac{1}{\binom{r}{\lfloor r/2 \rfloor}}\right) n \right\rceil.$$

Proof. Let $s = \lfloor r/2 \rfloor$. For the upper bound, we use Lemma 2.2 with the auxiliary hypergraph being the complete $(s + 1)$-uniform hypergraph on $2s + 1$ vertices.

For the lower bound, we use Lemma 2.1 with $d = s$, which shows

$$g(n, 2s + 1, s) \geq \left(1 - \frac{1}{\binom{2s+1}{s}}\right) n,$$

and since $g(n, 2s + 1, s)$ is always an integer, we may put a ceiling function here. □

2.4. Precise results for small values of r

Given the interest various reformulations of the question of determining $g(n, r, 1)$ attracted for the small values of r, we collect here what our various results imply for several small values of r. We note that for $r \leq 4$, the behaviour is completely determined by the known results for $g(n, r, 1)$ and the observation that $g(n, r, s) = n$ whenever $s \geq \frac{r}{2}$. Hence, in the following theorem, we solve the next three cases – namely, for $5 \leq r \leq 7$ given certain divisibility conditions on n. For $s = 1$, it is known and not too hard to remove the divisibility conditions, although including it in the following statement would make the statement somewhat unwieldy. We suspect it would not be too hard to figure out the precise answers regardless of the divisibility conditions on n in the $s = 2$ cases below, although we do not do this here. Let us also note that the below results always serve as lower bounds regardless of the divisibility conditions.

Theorem 2.12. *We have*

$g(n, 5, 1) = \frac{5}{9}n$	if $9 \mid n$,	$g(n, 6, 1) = \frac{1}{2}n$	if $4 \mid n$,	$g(n, 7, 1) = \frac{3}{7}n$	if $7 \mid n$,
$g(n, 5, 2) = \lceil \frac{9}{10}n \rceil$,		$g(n, 6, 2) = \frac{4}{5}n$	if $10 \mid n$,	$g(n, 7, 2) = \frac{5}{7}n$	if $7 \mid n$,
$g(n, 5, s) = n$	$\forall s \geq 3$,	$g(n, 6, s) = n$	$\forall s \geq 3$,	$g(n, 7, 3) = \lceil \frac{34}{35}n \rceil$	
				$g(n, 7, s) = n$	$\forall s > 3$.

Proof. Curiously, $g(n, 5, 1)$ is the only result on the above list for which the lower bound does not follow from Lemma 2.1 directly, so we point the reader to, for example, [45]. For the upper bound, one may take the (multi) hypergraph with vertex set [5] and edge (multi) set $\{145, 145, 145, 234, 234, 235, 235, 12, 13\}$. This is an intersecting hypergraph in which any four vertices contain at least 4 edges, implying via Lemma 2.2 that $g(n, 5, 1) \leq n - 4 \cdot \lfloor \frac{n}{9} \rfloor = \frac{5}{9}n$ if $9 \mid n$.

The lower bounds for $g(n, 6, 1), g(n, 6, 2), g(n, 7, 1)$ and $g(n, 7, 2)$ all follow from Lemma 2.1 by choosing $d = 2$. The upper bound on $g(n, 6, 1)$ follows by taking the Fano plane and removing a vertex to obtain a six-vertex hypergraph with four edges in which every vertex has degree two so any five vertices contain at least two edges. The upper bound on $g(n, 6, 2)$ follows by applying Lemma 2.2 to the hypergraph with vertex set [6] and edge set $\{123, 124, 346, 345, 256, 135, 245, 236, 146, 156\}$, which has the property that any four vertices contain at least two edges. The upper bounds on $g(n, 7, 1)$ and $g(n, 7, 2)$ follow from Proposition 2.3. The results for $g(n, 5, 2)$ and $g(n, 7, 3)$ follow from Proposition 2.11. □

3. Power of many colours for connected components

In this section, we turn to our results on the large connected component using only a few colours question – namely, determining $f(n, r, s)$.

3.1. Behaviour for very small s

We start with the following simple upper bound from [36]. We include its short proof here for completeness and since it motivates the argument behind the subsequent lower bound.

Proposition 3.1 [36]. *Suppose $r = 2^d - 1$ and $1 \le s \le d$. Then we have $f(n, r, s) \le 2^s \cdot \left\lceil \frac{n}{r+1} \right\rceil$.*

Proof. Let $V = \mathbb{Z}_2^d$, so $|V| = r + 1$. Let us consider the complete graph with vertex set V and let us colour the edge between distinct $\mathbf{x}, \mathbf{y} \in \mathbb{Z}_2^d$ in colour identified with $\mathbf{x} - \mathbf{y}$. This gives a colouring of an $(r + 1)$-vertex complete graph using r colours.

The key property of this colouring is that in any s colours, the graph consisting only of edges using these s colours is a vertex disjoint union of connected components, each consisting of at most 2^s vertices. Indeed, any set of s colours corresponds to a subset $S \subseteq \mathbb{Z}_2^d$ of size s, and the components correspond to cosets of $\mathrm{Span}(S)$ which partition the vertices and have size at most 2^s as claimed.

Now let us consider a complete graph K_n and partition its vertices into r parts of sizes as equal as possible, so each of size $\left\lfloor \frac{n}{r+1} \right\rfloor$ or $\left\lceil \frac{n}{r+1} \right\rceil$. We then identify the parts with V and colour the edges between parts according to the colouring described above. We colour the edges inside the parts in an arbitrary colour used between it and some other part. In this colouring for any s colours, at most 2^s parts become connected showing that the maximum size of a connected component in the graph consisting only of edges using these s colours is at most $2^s \cdot \left\lceil \frac{n}{r+1} \right\rceil$, as claimed. □

We next show that the upper bound in the above proposition is tight up to a constant factor for very small values of s. We will prove it in the full generality of the question of Bollobás, so instead of finding just a large connected component using few colours, we find a highly connected subgraph using edges of only a few colours. To help us find such highly connected subgraphs, we are going to use the following classical theorem of Mader; see [13, Theorem 1.4.3].

Theorem 3.2 (Mader 1972). *For every integer $k \ge 1$, any graph G with $d(G) \ge 4k$ has a $(k + 1)$-connected subgraph H with $d(H) > d(G) - 2k$.*

The main idea behind our lower bound result for small s is to show that instead of having just a single large highly connected subgraph using up to s colours, we will find many such subgraphs, using the same set of s colours. While this is in a sense stronger than what we need, the main benefit is for the next step, in which we want to again find many even larger subgraphs using some fixed $s + 1$ colours. We begin with a lemma which quantifies how many such large highly connected subgraphs one can find in a single colour. We note that a result in a similar direction was derived in [6] and turned out to be quite useful over the years; see, for example, [1]. The key distinction in the following lemma is that it works with average degree assumption rather than a minimum degree one and that it tracks the tradeoff between the actual size of the components and the number of them we can find.

Lemma 3.3. *Let $r \ge 2, k \ge 1, n > 16r^2(k - 1) + 1$ and G be an n-vertex graph with at least $\frac{1}{r} \binom{n}{2}$ edges. Then there exists $j \ge 2$ such that we can find at least $\frac{r}{4^j \log r}$ vertex disjoint k-connected subgraphs, each with at least $2^{j-3} \cdot \frac{n}{r} \ge 2$ vertices.*

Proof. We will construct our subgraphs greedily. Let V_1 be the vertex set of a largest k-connected subgraph of G. In general, given $V_1, \ldots, V_{i-1}$, we let V_i be the vertex set of the largest k-connected subgraph of $G \setminus (V_1 \cup \ldots \cup V_{i-1})$, so long as such a subgraph exists of size more than $\frac{n}{2r} + 1$. Let $V_1 \cup \ldots \cup V_t$ be the disjoint sets that the above process produced, and let $V' = V(G) \setminus (V_1 \cup \ldots \cup V_t)$. Let us also write $s_i = |V_i|$.

Since by adding to a k-connected graph a vertex adjacent to at least k of its vertices keeps the graph k-connected, we know that the total number of edges in $G \setminus (V_1 \cup \ldots \cup V_{i-1})$ with at least one endpoint in V_i is at most $\binom{s_i}{2} + n(k - 1)$. Turning to V', if the average degree in $G[V']$ is at least $\frac{n}{2r} + 2k - 2$, then we claim it contains a k-connected subgraph with at least $\frac{n}{2r} + 1$ vertices, contradicting the definition of V'. If $k = 1$, this holds by taking a vertex of maximum degree and all of its at least $\frac{n}{2r}$ neighbours. For $k \ge 2$, by Mader's Theorem (Theorem 3.2), it contains a subgraph with connectivity at least k, since

$\frac{n}{8r} \geq k$, and average degree larger than $\frac{n}{2r}$. Hence, we may conclude that the average degree of $G[V']$ is less than $\frac{n}{2r} + 2k - 2 \leq \frac{n-1}{r}$ (which guarantees $V' \neq V(G)$ and $|V'| \leq n - 1$) and in particular $G[V']$ has at most $\frac{n(n-1)}{4r} + (k-1)n$ edges.

Using this estimate on the number of edges, we get that there are at most

$$\frac{1}{r}\binom{n}{2} \leq \frac{n(n-1)}{4r} + (k-1)n + \sum_{i=1}^{t}\left(\binom{s_i}{2} + n(k-1)\right) \leq \frac{1}{2r}\binom{n}{2} + (t+1)(k-1)n + \sum_{i=1}^{t}\binom{s_i}{2}$$

edges in G. Since $s_i > \frac{n}{2r}$, we have $t < 2r$. So $(t+1)(k-1)n \leq 2r(k-1)n \leq \frac{1}{4r}\binom{n}{2}$, using that $n > 16r^2(k-1)$. This implies

$$\sum_{i=1}^{t}\binom{s_i}{2} \geq \frac{1}{4r} \cdot \binom{n}{2},$$

which in turn implies $s_1^2 + \ldots + s_t^2 \geq \frac{n^2}{4r}$.

Since for all $i \in [t]$ we have $n \geq s_i \geq \max\{\frac{n}{2r}, 2\}$, we know that there exists $2 \leq j \leq \lceil \log r \rceil + 2$ such that the set $I \subseteq [t]$ consisting of all i such that $s_i \in [2^{j-1} \cdot \frac{n}{4r}, 2^j \cdot \frac{n}{4r}]$ satisfies

$$\frac{n^2}{4r(\log r + 2)} \leq \sum_{i \in I} s_i^2 \leq |I| \cdot 2^{2j} \cdot \frac{n^2}{16r^2}.$$

This implies $|I| \geq \frac{r}{4^j \log r}$. Finally, note that since $s_i \geq 2$, we may also assume that $2^{j-1} \cdot \frac{n}{4r} \geq 2$.   □

Let us note that if $k = 1$, the above lemma does not require n to be large, besides the trivial bound $n > 0$. This will be important in our applications since we are going to apply the above lemma multiple times and may lose control of the number of vertices we are still working with after the first application; luckily, it will suffice for us to apply the lemma with $k = 1$ from the second instance on.

Before turning to the main result of this section, we need a final prerequisite, an immediate observation determining the bipartite Ramsey number for matchings.

Observation 3.4. *Given positive integers r, k and $n > r(k-1)$ in any r-colouring of $K_{n,n}$, we can find a monochromatic matching of size k.*

Proof. Pick an arbitrary perfect matching and observe it contains more than $r(k-1)$ edges. By the pigeonhole principle, this means some k of these edges have the same colour, providing us with the desired matching.   □

It is easy to see that the above lemma is tight. We simply split one side into r parts of size at most $k - 1$ each and colour all edges incident to part i in the colour i.

We are now ready to prove our lower bound theorem for small values of s. We will denote by $f(n, r, s, k)$ the largest number of vertices in a k-connected subgraph using up to s colours that we can guarantee in any r-edge colouring of K_n.

Theorem 3.5. *Suppose $1 \leq s \leq (1 - o_r(1)) \log \log r$ and $n > 16r^2(k-1) + 1$. Then we have*

$$f(n, r, s, k) \geq 2^{s-2} \cdot \frac{n}{r}.$$

Proof. We will prove that the desired inequality holds for $s \leq \log \log r - \log(4 + 3 \log \log r)$.

Let $n_0 = n$ and let us apply Lemma 3.3 to the graph consisting of the edges in the majority colour c_1 to find $j_1 \geq 2$ such that there exists at least $n_1 = \left\lceil \frac{r}{4^{j_1} \log r} \right\rceil$ k-connected subgraphs, each with at least $s_1 \geq 2^{j_1 - 3} \cdot \frac{n}{r} \geq 2$ vertices. We now contract each of these graphs into a single vertex to obtain a (multi) coloured K_{n_1}, where each vertex corresponds to a set of at least s_1 vertices in our original graph,

spanning a k-connected subgraph in colour c_1. Next, we keep an edge of only one colour c between any two vertices $u, v \in V(K_{n_1})$, where c is selected in the following way. Let $U, V \subseteq V(K_n)$ be the sets of vertices in our original graph corresponding to u and v, respectively. We choose c to be such that the original graph contains a matching of size k in c between U and V. Note that there exists such a colour by Observation 3.4 since $s_1 \geq \frac{n}{2r} > r(k-1)$. We then repeat, applying Lemma 3.3, but from now on always with $k = 1$, to the subgraph consisting of the majority colour in the current K_{n_i}, for as long as we have at least two vertices left. So, after step i for some $j_i \geq 0$, we have at least $n_i = \left\lceil \frac{r}{4^{j_i} \log r} \right\rceil$ vertices each corresponding to at least $s_1 \cdot s_2 \cdots s_i$ vertices of our original K_n, which induce a k-connected graph using the edges of colours $c_1, \ldots, c_i$ and where $s_i \geq 2^{j_i-3} \cdot \frac{n_{i-1}}{r} \geq 2$. Note that placing a matching of size k between two k-connected graphs produces a new k-connected graph. Therefore, our choice for the colouring of K_{n_1} guarantees that a connected component in any colour c in K_{n_i}, when $i \geq 2$, corresponds to a k-connected subgraph of our original graph using colours $c_1, \ldots, c_{i-1}$ and c.

Let us suppose that the process runs for t steps so that $n_t = 1$ and $n_i \geq 2$ for all $i < t$. Since $n_t = 1$, we have $\frac{r}{\log r} \leq 4^{j_t}$, or after taking a logarithm that $j_t \geq \frac{1}{2} \log r - \frac{1}{2} \log \log r$. Using that $n_{i-1} \geq 2$ for all $i \leq t$, we get

$$2 \leq 2^{j_i-3} \cdot \frac{n_{i-1}}{r} \leq 2^{j_i-3} \cdot \frac{\left\lceil \frac{r}{4^{j_{i-1}} \log r} \right\rceil}{r} \leq 2^{j_i-3} \cdot \frac{2}{4^{j_{i-1}} \log r}.$$

After taking a logarithm of both sides, this implies $j_i \geq 2j_{i-1} + 3 + \log \log r$.

If $t \geq s$, then we know that after s steps, we connected at least $s_1 \cdot s_2 \cdots s_s \geq 2^{s-2} \cdot \frac{n}{r}$ vertices, since $s_1 \geq \frac{1}{2} \cdot \frac{n}{r}$ and $s_i \geq 2$ for all i, and are done. So we may assume that the total number of steps t satisfies $t < s$, and hence, it suffices to show that $s_1 \cdot s_2 \cdots s_t \geq 2^{s-2} \cdot \frac{n}{r}$.

Finally, we prove by induction on i that for $i \leq t$, we have $s_1 \cdot s_2 \cdots s_i \geq 2^{j_i/2^{i-1} - (1-1/2^{i-1})(3+\log \log r)} \cdot \frac{n}{8r}$. Indeed, the base case $i = 1$ holds since $s_1 \geq 2^{j_1-3} \cdot \frac{n}{r}$, and for $i > 1$, we get

$$s_1 \cdot s_2 \cdots s_i \geq 2^{j_{i-1}/2^{i-2} - (1-1/2^{i-2})(3+\log \log r)} \cdot \frac{n}{8r} \cdot s_i$$

$$\geq 2^{j_{i-1}/2^{i-2} - (1-1/2^{i-2})(3+\log \log r)} \cdot \frac{n}{8r} \cdot \frac{2^{j_i-3}}{4^{j_{i-1}} \log r}$$

$$= 2^{j_i - (2-1/2^{i-2})(j_{i-1}+3+\log \log r)} \cdot \frac{n}{8r}$$

$$\geq 2^{j_i/2^{i-1} - (1-1/2^{i-1})(3+\log \log r)} \cdot \frac{n}{8r},$$

where in the second inequality, we used $s_i \geq 2^{j_i-3} \cdot \frac{n_{i-1}}{r} = 2^{j_i-3} \cdot \frac{\left\lceil \frac{r}{4^{j_{i-1}} \log r} \right\rceil}{r} \geq \frac{2^{j_i-3}}{4^{j_{i-1}} \log r}$, and in the final inequality, we used $(1-1/2^{i-1})j_i \geq (1-1/2^{i-1})(2j_{i-1}+3+\log \log r)$. This completes the induction step.

Applying this inequality for $i = t$ gives $s_1 \cdot s_2 \cdots s_t \geq 2^{j_t/2^{t-1} - 3 - \log \log r} \cdot \frac{n}{8r} > 2^{s-2} \cdot \frac{n}{r}$, since

$$\frac{j_t}{2^{t-1}} - 3 - \log \log r \geq \frac{\log r - \log \log r}{2^t} - 3 - \log \log r \geq \frac{\log r}{2^t} - 3 - 2 \log \log r > \log \log r + 1 \geq s + 1,$$

where in the penultimate inequality, we used $t < s \leq \log \log r - \log(4 + 3 \log \log r)$. $\square$

3.2. Behaviour for larger s

The following proposition is a slight improvement compared to a similar result from [36] showing that $f(n, r, s) \geq \left(1 - e^{-\frac{s^2}{3r}}\right) \cdot n$. We include it here mostly for completeness and since it is a very short and nice argument and, as we will see, is close to being tight.

Proposition 3.6. *For any* $1 \le s \le \frac{r}{2}$, *we have*

$$f(n, r, s) \ge n - n \cdot \left(1 - \frac{s}{r-s+1}\right) \cdots \left(1 - \frac{2}{r-1}\right) \cdot \left(1 - \frac{1}{r}\right) \ge \left(1 - e^{-\frac{s(s+1)}{2r}}\right) \cdot n.$$

Proof. We will prove the stronger bound by induction on s. The basis, for $s = 1$, the desired lower bound of $\frac{n}{r}$ follows by looking at a vertex and the colour which appears the most often on the edges incident to it. Suppose now that the result holds for $s - 1$ colours, and let S be the largest component one can obtain using $s - 1$ colours.

For any vertex v outside of S, if v sends edges of at most $s - 1$ different colours towards S, then we can connect $S \cup \{v\}$ using the edges of the star between v and S and their at most $s - 1$ colors, contradicting the maximality of S. Hence, every vertex outside of S sends edges of at least s different colours to S. Now by picking a uniformly random new colour, we will connect any such vertex to S with probability at least $\frac{s}{r-s+1}$. This means that the expected number of vertices we connect to S is $\frac{s}{r-s+1} \cdot (n - |S|)$. Therefore, we can choose a colour in such a way that the new set of s colours connects at least

$$|S| + \frac{s}{r-s+1} \cdot (n - |S|) = n - (n - |S|) \cdot \left(1 - \frac{s}{r-s+1}\right)$$

vertices, which after invoking the induction assumption on $|S|$ gives the desired bound. □

Remark. Let us also note that under certain regularity assumptions on the underlying colouring. the above argument performs significantly better. For example, if the underlying colouring is approximately uniformly balanced, meaning that no vertex has a degree in any colour larger than $C \cdot \frac{n}{r}$, for some $C \ge 1$, then we obtain a much stronger lower bound. Indeed, if $|S| = D \cdot \frac{n}{r} \le \frac{n}{2}$, then by our assumption we know that every vertex needs to send edges of at least $\frac{|S|}{Cn/r} = \frac{D}{C}$ different colours to S, so in expectation. we reach $\frac{D/C}{r-s+1} \cdot (n - |S|) \ge \frac{1}{2C} \cdot |S|$ new vertices. In particular, the size of the largest component in some s colours is at least $\left(1 + \frac{1}{2C}\right)^{s-1} \cdot \frac{n}{r}$. So, for example, already with $2C \log r$ colours, we can connect $\frac{n}{2}$ vertices, in stark contrast to the general case where we need at least about $\sqrt{r}$ colours to achieve this.

We conclude by showing that the lower bound given in Proposition 3.6 is actually close to tight for the entire range of s. It is tight up to a constant in the exponent of the lower order term when $s \ge \sqrt{r \log r}$ by our upper bounds for the g function – namely, by Corollary 2.8. For $s \le \sqrt{r}$, the bound in Proposition 3.6 reduces to $f(n, r, s) \ge \Omega\left(\frac{s^2}{r}\right) \cdot n$. The following proposition shows this is tight up to a logarithmic factor, disproving in quite a strong form a conjecture from [36].

Proposition 3.7. *For any* $2 \le s \le r$, *provided* $n \ge \frac{r}{s}$, *we have*

$$f(n, r, s) \le \frac{8s(s+1) \log r}{r} \cdot n.$$

Proof. We may assume that $8(s+1) \log r \le \frac{r}{s}$, or the inequality is trivial. This implies $\frac{r}{\log r} \ge 8(s+1)s \ge 48$, which in turn implies $\log r \ge 6$. Let us define $m := \left\lfloor \frac{r}{6s} \right\rfloor \ge \frac{r}{7s}$. To start with, we want to show that there exists an r-edge colouring of K_m with no s colours connecting more than $(s + 1) \log r$ vertices. To do so, let us consider a random colouring in which every edge of our K_m chooses one of the r colours uniformly at random, independently between different edges.

If there are $k = \lceil (s + 1) \log r \rceil$ vertices connected using up to s colours, then there exists a k-vertex tree using up to s colours. There are $\binom{m}{k} \cdot k^{k-2}$ possible choices for the tree, using Cayley's formula, and $\binom{r}{s}$ choices for the set of s colours used on the tree. Given these choices, the probability a given tree is coloured using the given set of s colours is $\left(\frac{s}{r}\right)^{k-1}$. Now a union bound implies that the probability that

there exists a k-vertex tree using up to s colours is at most

$$\binom{m}{k} \cdot k^{k-2} \cdot \binom{r}{s} \cdot \left(\frac{s}{r}\right)^{k-1} \leq r^{s+1} \cdot \left(\frac{ems}{r}\right)^k \leq 2^k \cdot \left(\frac{e}{6}\right)^k < 1,$$

where in the first inequality, we used the standard estimates $\binom{x}{y} \leq \min\left\{\left(\frac{ex}{y}\right)^y, x^y\right\}$. This shows that indeed there exists a desired colouring of K_m.

To complete the proof, we now take a blow-up of the above colouring of K_m where we replace every vertex with a set of $\lfloor \frac{n}{m} \rfloor$ or $\lceil \frac{n}{m} \rceil$ vertices to obtain an n-vertex graph. Let $V(K_m) = \{v_1, \ldots, v_m\}$, and we denote the corresponding sets by $V_1, \ldots, V_m$. For $1 \leq i < j \leq m$, we colour the edges between V_i and V_j with the same colour as the edge between v_i and v_j. For $1 \leq i \leq m$, we colour the edges inside V_i with an arbitrary colour appearing on an edge incident to v_i. This shows that

$$f(n, r, s) < (s+1) \log r \cdot \left\lceil \frac{n}{m} \right\rceil \leq (s+1) \log r \cdot \left\lceil \frac{n}{r/(7s)} \right\rceil \leq \frac{8s(s+1) \log r}{r} \cdot n,$$

completing the proof. □

4. Concluding remarks and open problems

In this paper, we explored a natural 'power of many colours' generalisation of two classical problems, in the sense first explored by Erdős, Hajnal and Rado over 60 years ago in the case of the classical Ramsey questions. While it has predominantly been studied within various instances of the classical Ramsey problem, we believe that extending this approach to encompass other classical colouring problems holds considerable promise and potential for intriguing exploration.

The first question we considered asks how many vertices can we guarantee to connect using edges of up to s colours in any r-edge colouring of K_n, and we denoted the answer to this question by $f(n, r, s)$. The following theorem summarises what is known for this function.

Theorem 4.1. *Given integers $1 \leq s \leq r$ and n large enough compared to r, we have*

$$f(n, r, s) = \begin{cases} \Theta\left(\frac{2^s}{r}\right) \cdot n, & \text{for } s \leq (1 - o_r(1)) \log \log r; \\ \widetilde{\Theta}_r\left(\frac{s^2}{r}\right) \cdot n, & \text{for } s \leq \sqrt{r}; \\ \Theta(n), & \text{for } s = \Theta(\sqrt{r}); \\ \left(1 - e^{-\Theta(s^2/r)}\right) \cdot n, & \text{for } s \geq \sqrt{r \log r}; \\ \lceil(1 - 1/\binom{r}{s})n\rceil, & \text{for } s = \frac{r-1}{2} \text{ and } r \text{ odd}; \\ n, & \text{for } s \geq \frac{r}{2}. \end{cases}$$

Our new contributions here are the lower bound for the first and the upper bound for the second and fourth part of the above theorem, while the remainder of the theorem follows from the results of Liu, Morris and Prince [36]. We note that the same picture holds for the more general function $f(n, r, s, k)$, which requires a k-connected subgraph in the same setting, and is the answer to the question of Bollobás, apart from slightly less precise results in the last two parts, provided n is large enough compared to r and k. This holds since our new lower bound in the first part holds in this more general setting as shown in Theorem 3.5, and our new upper bounds also hold since $f(n, r, s, k) \leq f(n, r, s)$. Additionally, the matching lower bounds have been established in the more general setting in [36].

We note that also for $s = 1, 2$, tight results are known under certain divisibility assumptions as mentioned in the introduction. Our bound giving the first part above is tight up to an absolute constant factor between 4 and 8 depending on divisibility properties of r. With slightly more care, under appropriate divisibility assumptions, it can be improved to be tight up to a factor of $2 + \varepsilon$ for any $\varepsilon > 0$,

at a cost of the result only applying for progressively smaller values of s in terms of r. This gap of about two has its origins in the fact that our proof uses the case $s = 1$ as its starting point and grows by about a factor of two in each step, whereas the aforementioned results show a jump by a factor of 4 when going to the $s = 2$ case. This can possibly be rectified by starting from the $s = 2$ case. But due to the far more complicated nature of that proof, the fact that we need a stronger result than the existence of just a single desired subgraph, as well as the fact that we would still be left with a constant factor gap without the divisibility assumptions, we leave this open as a potentially good student project.

Perhaps the most immediate question is to determine whether the $\log r$ term is necessary in the second part of Theorem 4.1 since this is the least precise result we have. Given that our lower bound for the first part of Theorem 4.1 shows that a $(\log r)^{1-o_r(1)}$ factor is necessary when $s \approx \log \log r$, this leads us to the following conjecture:

Conjecture 4.2. *For any $1 \leq s \leq \frac{\sqrt{r}}{\log r}$ and n large enough, we have $f(n, r, s) = \frac{s^2 (\log r)^{1-o_r(1)}}{r} \cdot n$.*

This also highlights the very natural question of how exactly the transition from the answer being $o_r(1) \cdot n$ to $\Theta(1) \cdot n$ happens. See [36] for a more detailed discussion of this particular question.

Our results show that the two functions f and g behave differently for $s \ll \frac{\sqrt{r}}{\log r}$ and behave similarly when $s \geq \sqrt{r \log r}$. It would be interesting to determine when precisely the functions start to exhibit different behaviour. Let us mention an explicit question in this direction which could be interesting.

Question 4.3. *Given n and r, what is the smallest value of s for which $f(n, r, s) = g(n, r, s)$?*

We know this holds for $s \geq \frac{r-1}{2}$ by the final part of Theorem 4.1, and it seems already nontrivial to decide it for $s = \frac{r}{2} - 1$ (and r even). However, for n large compared to r, our lower bound of $g(n, r, s) \geq \frac{(1+o_r(1))s}{\sqrt{r}} n$ for $s \ll \sqrt{r}$ compared to the upper bound of $f(n, r, s) \leq \frac{(1+o_r(1))s}{\sqrt{2r}} \cdot n$, from [36], shows there is a gap of at least about a factor of $\sqrt{2}$ when $s \ll \sqrt{r}$.

Let us now turn to the question of how many vertices will be touched by edges of some s colours in any r-edge colouring of K_n, the answer to which we denoted by $g(n, r, s)$. We determine $g(n, r, s)$ up to a constant factor for the whole range, provided n is large enough, and we even determine it up to lower order terms for the vast majority of the range. With this in mind, the most glaring open problem is to determine how exactly the transition from the answer being $\frac{(1+o_r(1))s}{\sqrt{r}}$ for $s \ll \sqrt{r}$ to being $(1 - o_r(1))n$ for $s \gg \sqrt{r}$ occurs. We do know the behaviour of the $o_r(1)$ term when $s \geq \sqrt{r \log r}$, but we believe the same behaviour – namely, Corollary 2.8 – should hold already for $s \gg \sqrt{r}$. We note, however, that one cannot hope to improve Proposition 2.7 directly since for $s \ll \sqrt{r \log r}$, the hypergraph with $e^{O(s^2/r)} \ll \sqrt{r} \leq s$ edges can certainly be covered by s vertices. However, one might hope to have a hypergraph with more edges but with the property that all vertex subsets of size $r - s$ contain many edges as we sketched in the remark following Corollary 2.8, where we explain how one can obtain our best upper bound of $g(n, r, s) \leq \left(1 - e^{-O\left(s\sqrt{\frac{\log r}{r}}\right)}\right)n$, for the range $\sqrt{r} \ll s \ll \sqrt{r \log r}$.

Let us also highlight the following natural problem, extending a classical question of Erdős and Lovász [18] from their famous 1975 paper introducing the Local Lemma, which asks for the minimum number of edges in an s-uniform intersecting hypergraph with cover number equal to s. This question was famously solved by Kahn in 1994 [33] and was one of Erdős's three favourite combinatorial problems (see, for example, [14]).

Question 4.4. *Let $r \geq 2s - 1 \geq 3$ be integers. What is the minimum number of edges in an r-vertex s-uniform intersecting hypergraph with cover number equal to s?*

The only difference compared to the classical question is that we in addition restrict the number of vertices one may use. So the main new feature of the above question is what happens if we are 'space restricted' and are not allowed to use as many vertices as we please when building the hypergraph. The $r \geq 2s - 1$ condition is here since such a graph trivially cannot exist if $r \leq 2s - 2$ since any set of $s - 1$

vertices would be a cover. However, for $r = 2s - 1$, the answer is easily seen to be $\binom{2s-1}{s}$ since the only way to achieve the properties is to take the complete hypergraph. For larger values of r, there is always a trivial lower bound of s edges, and the result of Kahn shows that for r large enough in terms of s, this is close to tight and the answer is linear in s. In fact, the usual double counting argument tells us that we need at least

$$\binom{r}{s-1}\bigg/\binom{r-s}{s-1} \geq \left(\frac{r}{r-s}\right)^{s-1} = \left(1 + \frac{s}{r-s}\right)^{s-1} \geq e^{\frac{s(s-1)}{2(r-s)}}$$

edges, where in the final inequality, we used $r \geq 2s$ and $1 + x \geq e^{x/2}$ for any real $x \leq 1$. This shows that for $s \gg \sqrt{r \log r}$, the answer is growing substantially faster than just linearly in s. Our Proposition 3.6 essentially solves a weakening of this question where we allow the uniformity to be somewhat larger than the lower bound on the cover number, which suffices to translate to a rough bound for $g(n, r, s)$. A good understanding of the answer to the above question would translate to a more precise result here as well.

In light of Lemma 2.2, the following natural extremal question arises.

Question 4.5. What is the minimum number of edges in an r-vertex intersecting hypergraph if any subset of m vertices needs to contain at least t edges?

In light of the discussion following Lemma 2.2, if we allow multi-hypergraphs in the above question, it would become essentially equivalent to determining $g(n, r, s)$, and in particular, most of our results on the $g(n, r, s)$ function came from results we proved for the above question. What turned out to be the most relevant instance for us is if we set $m = r - s$ and if one further restricts the hypergraph to be $(s + 1)$-uniform (note that it cannot have an edge of size at most s). Indeed, in this case, any edge is contained in $\binom{r-s-1}{s}$ sets of size $r - s$, so if we can achieve that every set of $r - s$ vertices contains the same number of edges, then every set of $r - s$ vertices would contain $|E(H)|\binom{r-s-1}{s}/\binom{r}{r-s}$ sets of size $s + 1$. Provided we can do so and ensure our hypergraph is intersecting, via Lemma 2.2, we would get an upper bound of $g(n, r, s) \leq \left(1 - \binom{r-s-1}{s}/\binom{r}{s}\right)n$, which, given some divisibility conditions. is tight via Lemma 2.1 with $d = s$. Indeed, this is exactly how our precise result at the end of the regime Proposition 2.11 proceeds, where we simply use a complete hypergraph. In general, this has a certain design-like flavour and seems to be asking whether a somewhat pseudorandom, intersecting hypergraph exists, a common theme in a number of still open questions from the aforementioned paper of Erdős and Lovász [18]. In our setting, approximate results in this direction also lead to approximate results on the $g(n, r, s)$ function and are behind most of our more precise results on this function.

Question 4.6. For which values of r, s does there exist an r-vertex, $(s + 1)$-uniform intersecting hypergraph in which all sets of size $r - s$ contain the same positive number of edges?

Another potentially interesting direction for future work is to consider what happens if one allows base graphs other than the complete one. In particular, one can ask a sparse random analogue, where we replace K_n with the binomial random graph $\mathcal{G}(n, p)$ and let p vary. The famous examples of results of this type are the works of Conlon and Gowers [11] and Schacht [44]. Both our questions for $s = 1$, the largest monochromatic connected component question (as well as several of its extensions) and the clique covering problem have also been considered in this direction. Various instances have also been considered for expander, pseudorandom and arbitrary large minimum degree base graphs. See, for example, [5, 26] for more details.

References

[1] E. Aigner-Horev, D. Hefetz and M. Krivelevich, 'Cycle lengths in randomly perturbed graphs', *Random Structures Algorithms* **65**(4) (2023), 867–884.

[2] N. Alon, P. Erdős, D. S. Gunderson and M. Molloy, 'A Ramsey-type problem and the Turán numbers', *J. Graph Theory* **40**(2) (2002), 120–129. MR 1899117

[3] N. Alon and J. H. Spencer, *The Probabilistic Method* fourth edn. (John Wiley & Sons, 2016).

[4] R. C. Baker, G. Harman and J. Pintz, 'The difference between consecutive primes. II', *Proc. London Math. Soc. (3)* **83**(3) (2001), 532–562. MR 1851081

[5] D. Bal and L. DeBiasio, 'Partitioning random graphs into monochromatic components', *Electron. J. Combin.* **24**(1) (2017), Paper No. 1.18, 25. MR 3609188

[6] T. Bohman, A. Frieze, M. Krivelevich and R. Martin, 'Adding random edges to dense graphs', *Random Structures Algorithms* **24**(2) (2004), 105–117. MR 2035870

[7] B. Bollobás and A. Gyárfás, 'Highly connected monochromatic subgraphs', *Discrete Math.* **308**(9) (2008), 1722–1725. MR 2392611

[8] M. Bucić and A. Khamseh, 'The $(t − 1)$-chromatic Ramsey number for paths', *J. Graph Theory* **102**(4) (2023), 784–796. MR 4563218

[9] S. Bustamante and M. Stein, 'Monochromatic tree covers and Ramsey numbers for set-coloured graphs', *Discrete Math.* **341**(1) (2018), 266–276. MR 3713408

[10] K. M. Chung and C. L. Liu, 'A generalization of Ramsey theory for graphs', *Discrete Math.* **21**(2) (1978), 117–127. MR 523059

[11] D. Conlon and W. T. Gowers, 'Combinatorial theorems in sparse random sets', *Ann. of Math. (2)* **184**(2) (2016), 367–454. MR 3548529

[12] D. Conlon, J. Fox, X. He, D. Mubayi, A. Suk and J. Verstraëte, 'Set-coloring Ramsey numbers via codes', *Studia Scientiarum Mathematicarum Hungarica* **61**(1) (2024), 1-15.

[13] R. Diestel, *Graph Theory* (Graduate Texts in Mathematics) fifth edn., vol. 173 (Springer, Berlin, 2018). MR 3822066

[14] P. Erdős, 'On the combinatorial problems which I would most like to see solved', *Combinatorica* **1**(1) (1981), 25–42. MR 602413

[15] P. Erdős and R. L. Graham, 'On partition theorems for finite graphs', in *Infinite and Finite Sets (Colloq., Keszthely, 1973; Dedicated to P. Erdős on His 60th Birthday), Vol. I* (Colloq. Math. Soc.) vol. 10 (János Bolyai, 1975), 515–527. MR 373959

[16] P. Erdős, A. Hajnal and R. Rado, 'Partition relations for cardinal numbers, *Acta Math. Acad. Sci. Hungar.* **16** (1965), 93–196. MR 202613

[17] P. Erdős and H. Hanani, 'On a limit theorem in combinatorial analysis', *Publ. Math. Debrecen* **10** (1963), 10–13. MR 166116

[18] P. Erdős and L. Lovász, 'Problems and results on 3-chromatic hypergraphs and some related questions', in *Infinite and Finite Sets (Colloq., Keszthely, 1973; Dedicated to P. Erdős on His 60th birthday), Vol. II* (Colloq. Math. Soc.) vol. 10 (János Bolyai, 1975), 609–627. MR 382050

[19] P. Erdős and L. Rényi, 'On some combinatorical problems', *Publ. Math. Debrecen* **4** (1956), 398–405. MR 78938

[20] M. K. Fort Jr. and G. A. Hedlund, 'Minimal coverings of pairs by triples', *Pacific J. Math.* **8** (1958), 709–719. MR 103831

[21] S. Fujita, H. Liiu and C. Magnant, 'Monochromatic structures in edge-coloured graphs and hypergraphs-a survey', *Int. J. Graph Theory Appl.* **1**(1) (2015), 3–56.

[22] Z. Füredi, 'Covering pairs by $q^2 + q + 1$ sets', *J. Combin. Theory Ser. A* **54**(2) (1990), 248–271. MR 1059999

[23] Z. Füredi, 'Maximum degree and fractional matchings in uniform hypergraphs', *Combinatorica* **1**(2) (1981), 155–162. MR 625548

[24] Z. Füredi, 'Matchings and covers in hypergraphs', *Graphs Combin.* **4**(2) (1988), 115–206. MR 943753

[25] L. Gerencsér and A. Gyárfás, 'On Ramsey-type problems', *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **10** (1967), 167–170. MR 239997

[26] L. Gishboliner, M. Krivelevich and P. Michaeli, 'Discrepancies of spanning trees and Hamilton cycles', *J. Combin. Theory Ser. B* **154** (2022), 262–291. MR 4374842

[27] A. Gyárfás, 'Partition coverings and blocking sets in hypergraphs', *Communications of the Computer and Automation Institute of the Hungarian Academy of Sciences* **71** (1977), 62.

[28] A. Gyárfás, 'Large monochromatic components in edge colorings of graphs: a survey', in *Ramsey Theory (Progr. Math.)* vol. 285 (Birkhäuser/Springer, New York, 2011), 77–96. MR 2759045

[29] A. Gyárfás and G. N. Sárközy, 'Matchings with few colors in colored complete graphs and hypergraphs', *Discrete Math.* **343**(5) (2020), 111831, 8. MR 4054970

[30] J. R. Henderson, 'Permutation decomposition of (0, 1)-matrices and decomposition transversals', PhD dissertation, 1971, California Institute of Technology.

[31] P. Horák and N. Sauer, 'Covering complete graphs by cliques', *Ars Combin.* **33** (1992), 279–288. MR 1174851

[32] S. Janson, 'Poisson approximation for large deviations', *Random Structures Algorithms* **1**(2) (1990), 221–229. MR 1138428

[33] J. Kahn, 'On a problem of Erdős and Lovász. II. $n(r) = O(r)$', *J. Amer. Math. Soc.* **7**(1) (1994), 125–143. MR 1224593

[34] X. Li and L. Wang, 'Forbidden rainbow subgraphs that force large monochromatic or multicolored k-connected subgraphs', *Discrete Appl. Math.* **285** (2020), 18–29. MR 4110091

[35] H. Liu, 'Recent extremal problems in combinatorics', PhD dissertation, 2006, The University of Memphis. MR 2708624

[36] H. Liu, R. Morris and N. Prince, 'Highly connected multicoloured subgraphs of multicoloured graphs', *Discrete Math.* **308**(22) (2008), 5096–5121. MR 2450447

[37] H. Liu, R. Morris and N. Prince, 'Highly connected monochromatic subgraphs of multicolored graphs', *J. Graph Theory* **61**(1) (2009), 22–44. MR 2514097

[38] H. Liu and Y. Person, 'Highly connected coloured subgraphs via the regularity lemma', *Discrete Math.* **309**(21) (2009), 6277–6287. MR 2551981

[39] C. Lo, H. Wu, and Q. Xie, 'A Ramsey type problem for highly connected subgraphs', Preprint, 2023, arXiv:2008.09001v4.

[40] T. Łuczak, 'Highly connected monochromatic subgraphs of two-colored complete graphs', *J. Combin. Theory Ser. B* **117** (2016), 88–92. MR 3437613

[41] W. H. Mills, 'Covering designs. I. Coverings by a small number of subsets', *Ars Combin.* **8** (1979), 199–315. MR 557075

[42] W. H. Mills, 'A covering of pairs by quintuples', *Ars Combin.* **18** (1984), 21–31. MR 823828

[43] N. J. Pullman, 'Clique ccverings of graphs—a survey', in *Combinatorial Mathematics, X (Adelaide, 1982)* (Lecture Notes in Math.) vol. 1036 (Springer, Berlin, 1983), 72–85. MR 731572

[44] M. Schacht, 'Extremal results for random discrete structures', *Ann. of Math. (2)* **184**(2) (2016), 333–365. MR 3548528

[45] Y. Shiloach, U. Vishkin and S. Zaks, 'Golden ratios in a pairs covering problem', *Discrete Math.* **41**(1) (1982), 7–65. MR 676863

[46] D. T. Todorov, 'Some coverings derived from finite planes', in *Finite and Infinite Sets, Vol. I, II (Eger, 1981)* (Colloq. Math. Soc.) vol. 37 (János Bolyai) (North-Holland, Amsterdam, 1984), 697–710. MR 818272

[47] D. T. Todorov, 'On the covering of pairs by 13 blocks', *C. R. Acad. Bulgare Sci.* **38**(6) (1985), 691–694. MR 805444