"In general, we don't know much about these Ramsey numbers."

Jacob Fox, 2018

1 Ramsey's Theorem

We begin with the following innocuous-looking theorem from nearly 90 years ago.

Theorem 1 (Ramsey, 1930)**.** *For any $k \in \mathbb{N}$, there is some $N \in \mathbb{N}$ so that, no matter how we color the edges of the complete graph K_N in red or blue, there will be a monochromatic K_k.*

If we let $r(k)$ denote the minimal such N, then this theorem simply asserts that $r(k) < \infty$. In order to prove this, we will actually introduce a slightly more general notion: for $k, \ell \in \mathbb{N}$, let $r(k, \ell)$ denote the minimal N so that any two-coloring of the edges of K_N contains either a red K_k or a blue K_ℓ. With this notation, we have that $r(k) = r(k, k)$.

Proof, due to Erdős and Szekeres. We will prove that $r(k, \ell) < \infty$ for all k, ℓ, by induction on $k+\ell$. As a base case, observe that $r(k, 1) = 1$ for all k. Indeed, in any two-coloring of the edges of K_1 (of which there are none), we will always find a (trivial) blue K_1. In particular, $r(k, 1) < \infty$ for all k.

For the inductive step, we claim that

$$r(k, \ell) \leq r(k-1, \ell) + r(k, \ell-1).$$

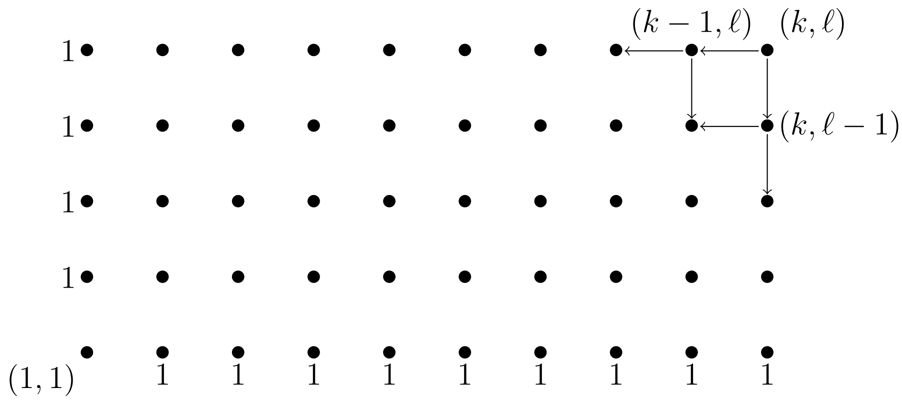
Indeed, let $N = r(k-1, \ell) + r(k, \ell-1)$, and fix some vertex $v \in K_N$. Consider any two-coloring of the edges of K_N, and look at the number of red and blue edges incident with v. Since there are $N-1$ such edges in total, by the pigeonhole principle, either there are at least $r(k-1, \ell)$ red edges or at least $r(k, \ell-1)$ blue edges. In the first case, by the definition of $r(k-1, \ell)$, we can find among the red neighbors a red K_{k-1} or a blue K_ℓ. If we find a blue K_ℓ, we are done, and if we find a red K_{k-1}, we can add to it the vertex v to form a red K_k, since all edges from v into this red K_{k-1} are red. The analogous argument shows that $r(k, \ell-1)$ blue edges out of v also suffice. □

It is natural to ask how large $r(k, \ell)$ is. In fact, even though he was primarily concerned with logic and set theory, Ramsey asked this very question in the paper where he proved Ramsey's Theorem, and was disappointed in how weak the bounds he found were. From the above argument, we can read off the following upper bound.

Corollary 1 (Erdős–Szekeres)**.**

$$r(k, \ell) \leq \binom{k+\ell-2}{k-1}.$$

Proof. One way to prove this is to observe that the recurrence $r(k,\ell) \le r(k-1,\ell)+r(k,\ell-1)$ looks like the recurrence defining Pascal's triangle. In my opinion, a more satisfying proof is to consider the following picture.



We've placed the value 1 at each position $(k,1)$ or $(1,\ell)$ in this grid, corresponding to the base case $r(k,1) = r(1,\ell) = 1$ of our induction. At the point (k,ℓ) in the grid, we know that the value $r(k,\ell)$ at that point is upper-bounded by the sum of the values immediately below and immediately to the left. Applying this same recurrence to these adjacent nodes, we see that every left/down path from (k,ℓ) to the boundary will contribute 1 in the final sum (corresponding to the value 1 at the boundary points). Thus, $r(k,\ell)$ is upper-bounded by the number of left/down paths to the boundary, which is in turn equal to the number of left/down paths from (k,ℓ) to $(1,1)$, which is exactly $\binom{k+\ell-2}{k-1}$. ☐

In particular, Stirling's approximation tells us that

$$r(k,k) \le \binom{2k-2}{k-1} \approx \frac{C}{\sqrt{k}}4^k$$

for some constant C. For a lower bound, the simplest way to lower-bound $r(k,k)$ is to exhibit a coloring of the edges of K_N with no monochromatic K_k, which implies $r(k,k) \ge N+1$. For instance, one can take $k-1$ clusters, each of size $k-1$, and color each cluster red internally and blue between clusters. Then this will certainly have no monochromatic K_k, and yields the bound $r(k,k) \ge (k-1)^2+1$. Though not nothing, this quadratic bound is miles away from the exponential upper bound above. Still today, we don't know how to explicitly construct an example demonstrating an exponential lower bound. Nevertheless, Erdős proved the following theorem, which is remarkable both for the result and the proof technique.

Theorem 2 (Erdős, 1947)**.**

$$r(k,k) \ge ck\sqrt{2}^k$$

for some constant $c > 0$.

Proof. We randomly color each edge of K_N either red or blue, each with probability 1/2 and independently. Then it is simple to check that if $N = ck\sqrt{2}^k$ for c sufficiently small, then the probability of seeing a monochromatic K_k is less than 1. Therefore, there must exist *some* coloring with no monochromatic K_k, and even though we don't know what it is, we can deduce that $r(k,k) \geq ck\sqrt{2}^k$. □

Thus, we see that $r(k,k)$ grows exponentially in k, and the base of the exponent is between $\sqrt{2}$ and 4. To this day, more than 70 years later, no improvement has been made to these constants. In fact, despite a lot of attention, very little improvement has happened at all. For the lower bound, Spencer was able to use a more sophisticated probabilistic approach to improve the constant c by a factor of 2, but no one has improved anything but the constant. For the upper bound, there have been some more substantial improvements, which are the topic of the rest of this talk, but they have only affected the lower-order terms, and the exponential constant of 4 remains unchanged.

2 Quasirandomness

Since our only known exponential lower bounds for the Ramsey numbers $r(k,k)$ come from random constructions, many people have asked whether this is necessary, and whether all colorings with no small monochromatic cliques exhibit various random-like properties. The most important random-like notion is that of *quasirandomness*, originally defined and explored by Thomason and Chung–Graham–Wilson. For disjoint vertex sets X, Y in a graph G, let $e(X,Y)$ denote the number of edges between X and Y.

Definition 1. Let $\varepsilon > 0$ and $p \in (0,1)$. A graph G on n vertices is called (p,ε)-*quasirandom* if, for any pair of disjoint vertex sets X, Y,

$$|e(X,Y) - p|X||Y|| < \varepsilon n^2. \qquad (P_1)$$

Note that if G is a random graph where each edge exists independently with probability p, then $p|X||Y|$ is exactly the expected value of $e(X,Y)$. Moreover, standard concentration results like the Chernoff bound imply that if n is large compared to ε (namely $n \gg 1/\varepsilon$), then a random graph will indeed be quasirandom with exponentially high probability. Note too that in a red/blue coloring of K_N, the colors are complementary, so the red graph will be (p,ε)-quasirandom if and only if the blue graph is $(1-p,\varepsilon)$-quasirandom. Thus, we can extend this definition and speak of quasirandom colorings.

This definition may feel somewhat artificial. There are many other properties satisfied by random graphs, and it is strange to isolate this one as the one to define quasirandomness. However, the astonishing fact, due to Chung–Graham–Wilson (and, in some special cases, to Thomason and others) is that this definition is equivalent to many other properties shared by random graphs. Literally dozens of equivalent properties are now known, but here are a few of the important ones. As above, $\varepsilon > 0$ is some parameter and n is the number of vertices in G. For two vertices $u, v \in V(G)$, let $\operatorname{codeg}(u,v)$ denote their codegree, namely the number of vertices w such that $u \sim w$ and $v \sim w$. Finally, for a fixed graph H on s vertices, let $N_H(G)$ be the number of labeled copies of H in G (i.e. the number of times H appears as a subgraph of G, scaled by the number of automorphisms of H). Then we say that G has property (P_2) if

$$\sum_{u,v \in V} \left|\operatorname{codeg}(u,v) - p^2 n\right| < \varepsilon n^3. \tag{P_2}$$

Note that in a random graph, any two vertices will have roughly $p^2 n$ vertices in common with high probability. For a fixed $s \geq 4$, we say that G has property $(P_3(s))$ if, for any graph H on s vertices,

$$\left|N_H(G) - n^s p^{\mathsf{e}(H)}\right| < \varepsilon n^s, \tag{$P_3(s)$}$$

where $\mathsf{e}(H)$ is the number of edges in H. Note that in a random graph, every ordered collection of s vertices has a probability $p^{\mathsf{e}(H)}$ of producing a copy of H, and there are roughly n^s such ordered sets, if n is large. Finally, we say that G has property (P_4) if

$$\left|\mathsf{e}(G) - p\binom{n}{2}\right| < \varepsilon n^2 \qquad \text{and} \qquad \left|N_{C_4}(G) - p^4 n^4\right| < \varepsilon n^4, \tag{P_4}$$

where C_4 is the cycle graph on four vertices. Note that $(P_3(s))$ says that the counts of *all* graphs on s vertices are roughly the same as they would be in a random graph, whereas property (P_4) says that only the counts of edges and four-cycles are roughly correct.

Then the astonishing fact mentioned above is that all these properties are equivalent, in the following sense. Fix $p \in (0,1)$. For every $\varepsilon > 0$, there is some $\varepsilon' > 0$ such that property (P_i) with parameter ε' implies property (P_j) with parameter ε, for any i, j. The explicit dependence on the parameter ε is often ignored, and we simply say that a graph satisfying any of these properties is quasirandom, with the understanding that it also satisfies all the others, with some parameter; nevertheless, in many instances, the precise dependence of ε' on ε is very important, and there are still several cases where the correct growth order is unknown. Probably the most surprising of these is implications is the fact that simply correctly counting the number of edges and C_4s immediately implies the correct counts for all fixed graphs.

Before continuing, let me mention one conjecture connecting Ramsey's theorem with quasirandomness.

Conjecture 1 (Sós)**.** *For every $\varepsilon > 0$, there exists some $k_0 \in \mathbb{N}$ so that the following holds. If we let $k \geq k_0$ and $N = r(k,k)-1$, then every red/blue coloring of K_N with no monochromatic K_k (i.e. a maximally large coloring with no monochromatic K_k) is $(\frac{1}{2}, \varepsilon)$-quasirandom.*

This conjecture, in some sense, would explain why proving explicit (i.e. non-random) lower bounds on Ramsey numbers is so hard; it asserts that every maximal Ramsey coloring is random-like in a precise sense. Note however that the converse of this conjecture is very false, because quasirandomness is a sort of "global" property that cannot by itself detect "local" properties such as small monochromatic cliques. For instance, if we fix a graph G, we can construct its *blow-up* by replacing every vertex by t vertices and every edge by a complete bipartite graph $K_{t,t}$; then it is simple to check that if G was quasirandom, then its blow-up will be as well, whereas this blow-up will contain some independent sets of size t. In particular, if we blow up a quasirandom coloring, we will obtain a new quasirandom coloring, but it will have enormous cliques in one of the two colors, and in particular will be very far from a maximal Ramsey coloring.

Though this conjecture is still open (and potentially quite difficult), other connections between Ramsey theory and quasirandomness are known, and quasirandomness is used to prove the strongest upper bound on $r(k,k)$ to date. Recall that the Erdős–Szekeres argument gave an upper bound of $r(k,k) = O(\frac{4^k}{\sqrt{k}})$; the following theorem remains the only known super-polynomial improvement on this bound.

Theorem 3 (Conlon, 2009)**.** *There is a constant $c > 0$ so that*

$$r(k,k) \leq k^{-\frac{c \log k}{\log \log k}} 4^k.$$

This result builds on and improves previous results of Rödl, who showed that $r(k,k) = o(4^k/\sqrt{k})$, and Thomason, who gave the polynomial improvement $r(k,k) = O(4^k/k)$; the main additional idea in Conlon's proof is the introduction of quasirandomness techniques.

Proof idea. The basic idea in Conlon's proof is to suppose that the Erdős–Szekeres bound were close to true, obtain strong structural information about extremal colorings, and ultimately derive a contradiction.

What sort of structural information can one obtain? As a toy model, let's suppose the Erdős–Szekeres bound were exactly correct for all k, ℓ, and let's fix a maximal coloring (on $N = r(k,\ell) - 1$ vertices) with no red K_k or blue K_ℓ. Then the observation that drives the Erdős–Szekeres argument is that in this coloring, every vertex must have red degree at most $r(k-1,\ell) - 1$ and blue degree at most $r(k,\ell-1) - 1$. However, since the Erdős–Szekeres bound is assumed to be tight, there are precisely $r(k-1,\ell) + r(k,\ell-1) - 2$ vertices other than v, so both these upper bounds must be tight. Thus, under this assumption, we see that every vertex must have the same red degree (namely $r(k-1,\ell) - 1$) and the same blue degree. Let $p = (r(k-1,\ell) - 1)/N$, so that every vertex has red degree pN. Then by our assumption that the Erdős–Szekeres bound is always tight, we see that

$$p \approx \frac{\binom{k+\ell-3}{k-2}}{\binom{k+\ell-2}{k-1}} = \frac{k-1}{k+\ell-2} \approx \frac{k}{k+\ell}.$$

In fact, we can generalize this observation to get stronger structural results. Namely, let Q be some red K_t in the coloring, for some $t < k$. Then we claim that Q lies in fewer than $r(k-t,\ell)$ red K_{t+1}s. Said differently, there are fewer than $r(k-t,\ell)$ vertices all of which have a red edge to every vertex of Q. Indeed, suppose there were at least $r(k-t,\ell)$ such vertices. Then by the definition of $r(k-t,\ell)$, they would contain among them either some blue K_ℓ, contradicting our assumption, or some red K_{k-t}. But we can then take the union of this red K_{k-t} with Q to obtain a red K_k, using the fact that all the edges between Q and the K_{k-t} would be blue. By the same argument, we can conclude that every blue K_t lies in fewer than $r(k, \ell - t)$ blue K_{t+1}s.

Now, let's first apply this observation in the case when $t = 2$. Then this says that for every red edge, say the edge uv, there are fewer than $r(k-2, \ell)$ other vertices w such that uw and vw are both red. In other words, the red codegree of u and v is less than $r(k-2, \ell)$. However, by our assumption that the Erdős–Szekeres bound is tight, we have that

$$\operatorname{codeg}_R(u, v) < r(k-2, \ell) = \binom{k+\ell-4}{k-3} = \frac{(k-3)(k-2)}{(k+\ell-2)(k+\ell-3)}\binom{k+\ell-2}{k-1} \approx p^2 N.$$

We can in fact obtain a matching lower bound, as follows. Note that the number of vertices w so that uw is red while vw is blue is less than $r(k-1, \ell-1) \approx p(1-p)N$, by the same argument as before. By interchanging the roles of u and v, we obtain the same bound for the number of w with uw blue and vw red. However, by applying the inclusion-exclusion principle to the red neighborhoods of u and v, this implies that

$$\operatorname{codeg}_R(u, v) \geq r(k-1, \ell) - r(k-1, \ell-1) \approx p^2 N.$$

Finally, again by inclusion-exclusion, we see that this also implies that the blue codegree of u and v is approximately $(1-p)^2 N$. At this point, we can interchange the roles of red and blue, and we find that for any pair of vertices, their red codegree is approximately $p^2 N$ and their blue codegree is approximately $(1-p)^2 N$. Thus, our coloring is quasirandom, by property (P_2).

However, once we have this quasirandomness, we can use property (P_3) to count the number of monochromatic cliques in our coloring. For instance, there are roughly $n^t p^{\binom{t}{2}}$ red K_ts, and roughly $n^{t+1} p^{\binom{t+1}{2}}$ red K_{t+1}s. However, we also know that each red K_t can lie in at most $r(k-t, \ell)$ red K_{t+1}s. If our counts are sufficiently precise and hold for sufficiently many values of t, these bounds contradict each other, which implies that the Erdős–Szekeres bound is not tight.

Similarly, if we instead assume that the Erdős–Szekeres bound is "close to" tight, then we will deduce that all vertices have roughly the same red degree and roughly the same blue degree, with the approximation error here depending on how close to true the Erdős–Szekeres bound is assumed to be. Similarly, we can get that the red codegrees all roughly equal their expected values, where the quality of the approximation again depends on the assumption we make. Thus, we see that as we try to prove stronger and stronger upper bounds via this approach (namely assuming that the bound is less and less close to true), we will obtain weaker and weaker structural information, since all our approximations will become less precise. In particular, at some point we will stop getting sufficiently strong quasirandomness in order to gain precise counts of K_ts, which means we will no longer be able to obtain the desired contradiction. □

3 Ramsey numbers of books

We begin with an important definition, which generalizes the Ramsey numbers studied above.

Definition 2. For a fixed graph H, its *Ramsey number* $r(H)$ is the minimal N such that any red/blue coloring of K_N contains a monochromatic copy of H. Note that since every graph H is contained in some complete graph, Ramsey's theorem implies that $r(H)$ exists and is finite for all graphs H.

The basic question in the field of *graph Ramsey theory* is to understand how the function $r(H)$ depends on the graph H. Of course, this question includes the classical question on the growth of the Ramsey numbers $r(k,k)$, since $r(K_k) = r(k,k)$. However, it turns out that for many different classes of graphs, the behavior of $r(H)$ is markedly different from the behavior of $r(k,k)$, and an extremely rich theory has developed around graph Ramsey numbers.

Recall that the key observation that allowed Conlon to both derive the quasirandomness property of extremal colorings and to use it to obtain a contradiction was the following generalization of the Erdős–Szekeres observation: in a coloring of K_N with no monochromatic K_k, every monochromatic K_t has fewer than $n = r(k-t,k)$ extensions into a monochromatic K_{t+1}, for every $1 \le t \le k-1$. Indeed, if some monochromatic (say red) K_k had at least n extensions into red K_{k+1}s, then those extending vertices would contain among them either a blue K_k or a red K_{k-t}, which could be added to the original K_t to obtain a red K_k. This motivates the following definition.

Definition 3. The *book graph* $B_n^{(t)}$ is the graph on $t+n$ vertices consisting of n copies of K_{t+1} glued along a common K_t. The terminology comes from the case $t = 2$, where we have n triangles glued along a common edge; we picture the triangles as "pages" and the edge as the "spine" of a book.

With this language, the observation above is that for any choice parameters $1 \le t \le k-1$, we have that a coloring of K_N with no monochromatic K_k also contains no monochromatic $B_n^{(t)}$, where $n = r(k-t,k)$. In other words, if we set $n = r(k-t,k)$, we obtain the bound

$$r(k,k) \le r(B_n^{(t)}).$$

This suggests that understanding the Ramsey numbers of book graphs could be used to improve the upper bound on $r(k,k)$. Notice too that in this regime, n is generally much larger than t, so we will often limit ourselves to the case when t is fixed and $n \to \infty$.

To understand $r(B_n^{(t)})$, let's begin with lower bounds. As is often the case in Ramsey theory, a random bound is a good starting point. Indeed, let's suppose we color each edge of K_N red or blue independently with probability $1/2$. Then if Q is a fixed monochromatic K_t and v is a vertex not in Q, the probability that v forms a monochromatic extension of Q is precisely 2^{-t}, since there are t edges from Q to v and they all need to receive the same color as Q. Thus, by linearity of expectation, Q will have $2^{-t}(N-t)$ monochromatic extensions in expectation. By using standard concentration results, we can show that with high probability the number of extensions will not greatly exceed its expectation, and we find that for t fixed,

$$r(B_n^{(t)}) \ge 2^t n - o(n).$$

Thomason conjectured that this was asymptotically tight. In fact, he conjectured the following very precise conjecture, which seems hopelessly out of reach with present techniques.

Conjecture 2 (Thomason)**.**

$$r(B_n^{(t)}) \leq 2^t(t+n-2)+2.$$

Notice that $B_1^{(t)}$ is just the clique K_{t+1}, so plugging in $n=1$ to this conjecture would imply the bound $r(t+1,t+1) \leq t2^t$, which would already be a tremendous improvement on the current best bound. This suggests one reason why this conjecture seems unlikely to be resolved soon.

Nevertheless, a weaker conjecture is just that the random construction is asymptotically tight. Indeed, this is true.

Theorem 4 (Conlon, 2018)**.** *For any fixed t,*

$$r(B_n^{(t)}) \leq 2^t n + o(n).$$

Proof sketch. The key tool used in Conlon's proof is Szemerédi's regularity lemma. For two vertex sets X and Y in a graph, we define their *density* to be the fraction of pairs in $X \times Y$ that are edges, namely

$$d(X,Y) = \frac{e(X,Y)}{|X||Y|}.$$

For $\varepsilon > 0$, we say that the pair (X,Y) is *ε-regular* if

$$|d(X,Y) - d(X',Y')| < \varepsilon$$

for any $X' \subseteq X, Y' \subseteq Y$ with $|X'| \geq \varepsilon|X|, |Y'| \geq \varepsilon|Y|$. This property basically says that the edges between X and Y are fairly uniformly distributed, and is closely reminiscent of the property (P_1) defining quasirandomness; indeed, there is a close connection between the two notions, and ε-quasirandomness is essentially equivalent to all pairs of "large" subsets being ε-regular.

The main reason we care about ε-regularity is that, much like property (P_3) of quasirandomness, it allows us to count subgraphs of a graph as though the edges were randomly distributed. Specifically, for any graph H with s vertices, and any collection of vertex sets $V_1, \ldots, V_s$ such that all pairs (V_i, V_j) are ε-regular, the *counting lemma* says that the number of copies of H with the ith vertex in V_i is roughly

$$\prod_{(i,j) \in E(H)} d(V_i, V_j) \prod_{i=1}^{s} |V_i|,$$

where the quality of the approximation depends on ε. Note that this quantity is the expected number of copies of H if the edges between the V_is were distributed randomly according to the densities $d(V_i, V_j)$, so this counting lemma is analogous to the implication $(P_1) \Longrightarrow (P_3(s))$.

Szemerédi's regularity lemma says that for any $\varepsilon > 0$, there exists some $M \in \mathbb{N}$, such that for every graph G on N vertices, we can find a partition $V(G) = V_1 \sqcup \cdots \sqcup V_m$ so that the following properties hold.

- $k \le M$, meaning that the complexity of the partition depends only on ε,
- $||V_i| - |V_j|| \le 1$ for all i, j, meaning that all parts have essentially the same size, and
- for all but $\varepsilon\binom{m}{2}$ pairs (i, j), we have that (V_i, V_j) is ε-regular.

In our simplification of Conlon's proof, Jacob Fox and I used a slight strengthening of Szemerédi's regularity lemma, that guarantees the following properties as well.

- For every i, the pair (V_i, V_i) is ε-regular, and
- there are at most εm values of j so that (V_i, V_j) is not ε-regular, meaning that the irregular pairs are "spread out."

The proof proceeds by starting with a coloring of K_N, with $N = (2^t + o(1))n$, and applying the regularity lemma to the red graph. Note that since the colors are complementary, a pair is ε-regular in red if and only if it's ε-regular in blue. Suppose that among the regularity partition, we can find a collection of parts $U_1, \ldots, U_t$ so that all pairs (U_i, U_i) and (U_i, U_j) are ε-regular, the red density inside each U_i is at least ε, and the blue density between each U_i and U_j is also at least ε; it is not hard to show that if no such configuration (or such a configuration with the colors flipped) exists, our coloring must have an extremely large monochromatic book, and in particular a monochromatic $B_n^{(t)}$.

In such a configuration, we can find many blue K_ks spanning the U_is, in addition to many red K_ks inside each U_i, by the counting lemma. Suppose we pick a random clique of one of these types, and some vertex v outside the configuration. Thanks to a technical analytic lemma (i.e. just a fact about real numbers), one can show that the probability that v yields a monochromatic extension of such a random clique is at least 2^{-t}. Specifically, the lemma says that for any real numbers $x_1, \ldots, x_t \in [0, 1]$,

$$\prod_{i=1}^{t} x_i + \frac{1}{t} \sum_{i=1}^{t} (1 - x_i)^t \ge 2^{1-t}.$$

If we let x_i be fraction of vertices in U_i that are blue to v, then the first term measures the probability that v extends a spanning blue clique, while the second term measures the average probability that v extends a red clique inside some V_i, averaged over all i. The lemma asserts that the average of these two values is at least 2^{-t}. By adding this fact up over all vertices v, we obtain at least $(2^{-t} - o(1))N = n$ extensions in expectation, so there must exist some clique with at least that many extensions, yielding our monochromatic $B_n^{(t)}$. $\square$

This theorem is an example of an instance where the magnitude of the error term really matters, for if we could get very strong control over this $o(n)$ term, we could dramatically improve the upper bound for $r(k, k)$. Indeed, this theorem implies that if n (and thus k) is sufficiently large, then $r(B_n^{(t)}) \le 2^{t+1}n$. Let's suppose we had sufficiently strong control on the error, so that this "sufficiently large" already kicks in when k is linear in t; namely, suppose that if $k \geq Ct$, then $r(B_n^{(t)}) \leq 2^{t+1}n$, for some $C > 1$, where $n = r(k-t,k)$. Then by our argument above, we know that

$$r(k,k) \leq r(B_n^{(t)}) \leq 2^{t+1} r(k-t,k) \leq 2^{t+1}\binom{2k-t-2}{k-1} \leq 2^{k/C}\binom{(2-1/C)k}{k} \leq (4-\delta)^k,$$

for some $\delta > 0$ depending on C. We applied the Erdős–Szekeres bound to $r(k-t,k)$, and then a standard bound on binomial coefficients in the last step; it says that any Hamming sphere of radius bounded away from $1/2$ contains an exponentially small fraction of the hypercube. The upshot is that if we could apply this bound for n roughly exponential in t (so that k is linear in t), we could obtain some exponential improvement to the upper bound for $r(k,k)$, which would be a major breakthrough.

So how good is the error term in Conlon's theorem? Unfortunately, it's very bad; his argument gives an upper bound of the form

$$r(B_n^{(t)}) \leq 2^t n + O\left(\frac{2^{t^{c_1}}}{(\log^* n)^{c_2}}\right) n$$

for some $c_1, c_2 > 0$, where $\log^*$ is the inverse of the tower function, and thus grows extremely slowly. Put another way, rather than obtaining the bound we need when n is exponential in t, Conlon's argument only kicks in when $n \geq 2^{2^{2^{\cdot^{\cdot^{\cdot}}}}}$, where the height of the tower is polynomial in t. This is because his proof uses the regularity lemma, which always yields such bounds. This is not nearly strong enough to obtain an exponential improvement on the upper bound for $r(k,k)$, but Conlon conjectured that this tower-type dependence is not necessary.

In recent work with Jacob Fox, we've tried to improve on Conlon's theorem. Our first result gives (somewhat) more reasonable control on the error term.

Theorem 5.

$$r(B_n^{(t)}) \leq 2^t n + O\left(\frac{2^{t^{c_1}}}{(\log\log\log n)^{c_2}}\right) n,$$

for some constants $c_1, c_2 > 0$.

Of course, this is also not nearly strong enough to give the desired exponential improvement; it kicks in "only" when $n \geq 2^{2^{2^{t^{c'}}}}$. Though our proof uses a weaker result than the regularity lemma in order to avoid the tower-type bounds, it is still closely related to Conlon's proof, and it seems unlikely that such techniques could be used to obtain strong enough bounds to improve the bound on $r(k,k)$.

Our second main result deals with the structure of large colorings with no monochromatic book. Recall that Sós conjectured that an extremal coloring with no monochromatic K_k must be quasirandom; we can prove an analogous result for books, in some sense explaining why the random bound is asymptotically correct for this problem.

Theorem 6. *For every $t \in \mathbb{N}$ and $\varepsilon > 0$, there exists some $\delta > 0$ and $N_0 \in \mathbb{N}$ so that the following holds. For any $N \geq N_0$, any red/blue coloring of K_N is either $(\frac{1}{2}, \varepsilon)$-quasirandom or else contains a monochromatic $B^{(t)}_{(2^{-t}+\delta)N}$. In other words, any maximal coloring that contains no monochromatic $B^{(t)}_n$ must be quasirandom.*

The $t = 2$ case was previously proven by Nikiforov, Rousseau, and Schelp, using different techniques. Our proof again utilizes the connection between regularity and quasirandomness, and a "nibbling" technique; rather than proving the coloring is quasirandom directly, we repeatedly find small parts that are regular to remainder, and by iteratively pulling them out, we obtain the global quasirandomness. Unlike Sós's conjecture for the classical Ramsey numbers, we are also able to prove a sort of converse to this theorem, which in particular shows that the "correct" size of a monochromatic book is another equivalent property to quasirandomness.

The general picture emerging from these results is that book Ramsey numbers share some similarities to classical Ramsey numbers, and there are concrete connections between them (e.g. the fact that classical Ramsey numbers can be upper-bounded by book Ramsey numbers), but it is possible to prove some results about book Ramsey numbers that are currently out of reach in the classical case.