THE TYPICAL STRUCTURE OF GRAPHS WITH NO LARGE CLIQUES

JÓZSEF BALOGH, NEAL BUSHAW, MAURÍCIO COLLARES NETO, HONG LIU, ROBERT MORRIS, AND MARYAM SHARIFZADEH

ABSTRACT. In 1987, Kolaitis, Prömel and Rothschild proved that, for every fixed $r \in \mathbb{N}$, almost every n-vertex K_{r+1}-free graph is r-partite. In this paper we extend this result to all functions $r = r(n)$ with $r \leqslant (\log n)^{1/4}$. The proof combines a new (close to sharp) supersaturation version of the Erdős–Simonovits stability theorem, the hypergraph container method, and a counting technique developed by Balogh, Bollobás and Simonovits.

1. Introduction

Determining the extremal properties of graphs which avoid a clique of a given size is one of the oldest problems in combinatorics, going back to the early paper of Mantel [18] and the groundbreaking work of Ramsey [22], Erdős and Szekeres [14] and Turán [25] over 70 years ago. The study of the *typical* properties of such graphs was initiated by Erdős, Kleitman and Rothschild [12], who proved in 1976 that almost all triangle-free graphs on n vertices are bipartite[1]. This result was extended to K_{r+1}-free graphs, for every fixed $r \in \mathbb{N}$, ten years later by Kolaitis, Prömel and Rothschild [16], who showed that almost all such graphs are r-partite. Various extensions of this theorem have since been obtained, see for example [6, 21] for work on other forbidden subgraphs, and [8, 20] for a sparse analogue.

In this paper we extend the result of Kolaitis, Prömel and Rothschild in a different direction, to K_{r+1}-free graphs where $r = r(n)$ is a function which is allowed to grow with n. More precisely, we prove the following theorem.

Theorem 1.1. *Let $r = r(n) \in \mathbb{N}_0$ be a function satisfying $r \leqslant (\log n)^{1/4}$ for every $n \in \mathbb{N}$. Then almost all K_{r+1}-free graphs on n vertices are r-partite.*

Note that if $r \geqslant 2\log_2 n$ then almost all graphs are K_{r+1}-free (and almost none are r-partite if $r \ll n/\log n$), so the bound on r in Theorem 1.1 is not far from being best possible. It would be extremely interesting (and likely very difficult) to determine the largest $\alpha \in [1/4, 1]$ such that the theorem holds for some function $r = (\log n)^{\alpha+o(1)}$. It may well be the case that this supremum is equal to 1, though we are not prepared to state this as a conjecture.

Research supported in part by a Simons Fellowship, NSF CAREER Grant DMS-0745185, Marie Curie FP7-PEOPLE-2012-IIF 327763, Arnold O. Beckman Research Award (UIUC Campus Research Board 13039) (JB), CAPES bolsa Proex (MCN), a CNPq bolsa PDJ (NB) and a CNPq bolsa de Produtividade em Pesquisa (RM).

[1] That is, the proportion of n-vertex triangle-free graphs that are not bipartite goes to zero as $n \to \infty$.

Theorem 1.1 improves a recent result of Mousset, Nenadov and Steger [19], who showed that, for the same[2] family of functions $r = r(n)$, the number of n-vertex K_{r+1}-free graphs is

$$2^{t_r(n)+o(n^2/r)}, \tag{1}$$

where $t_r(n) = \mathbf{ex}(n, K_{r+1})$ denotes the number of edges of the Turán graph, the r-partite graph on n vertices with the maximum possible number of edges. A bound of this type for fixed $r \in \mathbb{N}$ was originally proved in [12], and extended to an arbitrary (fixed) forbidden graph H in [11]. The problem for H-free graphs with $v(H) \to \infty$ as $n \to \infty$ was first studied by Bollobás and Nikiforov [9], who proved bounds corresponding to (1) whenever $v(H) = o(\log n)$ and $\chi(H_n) = r + 1$ is fixed. For more precise bounds for a fixed forbidden graph H, see [5], and for similar bounds in the hereditary (i.e., induced-H-free) setting, see [1, 4, 10] and the references therein.

The proof of Theorem 1.1 has three main ingredients. The first is the so-called 'hypergraph container method', which was recently developed by Balogh, Morris and Samotij [7], and independently by Saxton and Thomason [24]. This method was used by Mousset, Nenadov and Steger to prove Theorem 3.2, below, from which they deduced the bound (1) using a supersaturation theorem of Lovász and Simonovits [17].

In order to obtain the much more precise result stated in Theorem 1.1, we will use the method of Balogh, Bollobás and Simonovits [5, 6], who determined the structure of almost all H-free graphs for every fixed graph H. This powerful technique (see Sections 4 and 5) allows one to compare the number of K_{r+1}-free graphs that are 'close' to being r-partite, with the total number of K_{r+1}-free graphs.

The missing ingredient is the main new contribution of this paper. In order to deduce from Theorem 3.2 a bound on the number of K_{r+1}-free graphs that are 'far' from being r-partite, we will need an analogue of the Lovász–Simonovits supersaturation result, mentioned above, for the well-known stability theorem of Erdős and Simonovits [13]. Although a weak such analogue can easily be obtained via the regularity lemma, this gives bounds which are far from sufficient for our purposes. Instead we will adapt a recent argument due to Füredi [15] in order to prove the following close-to-best-possible such result. We say that a graph G is *t-far from being r-partite*[3] if $\chi(G') > r$ for every subgraph $G' \subset G$ with $e(G') > e(G) - t$.

Theorem 1.2. *For every $n, r, t \in \mathbb{N}$, the following holds. Every graph G on n vertices which is t-far from being r-partite contains at least*

$$\frac{n^{r-1}}{e^{2r} \cdot r!} \left(e(G) + t - \left(1 - \frac{1}{r}\right) \frac{n^2}{2} \right)$$

copies of K_{r+1}.

[2]In fact, a very slightly weaker theorem was stated in [19], but a little additional case analysis easily gives the result for all $r \leqslant (\log n)^{1/4}$.

[3]Similarly, we say that G is t-close to being r-partite if it is not t-far from being r-partite.

Note that the graph obtained by adding t edges to the Turán graph $T_r(n)$ is t-far from being r-partite and has roughly $t \cdot (n/r)^{r-1}$ copies of K_{r+1}, so Theorem 1.2 is sharp to within a factor of roughly e^r. We prove this supersaturated stability theorem in Section 2, and use it in Section 3 to count the K_{r+1}-free graphs that are n^{2-1/r^2}-far from being r-partite. We prove various simple properties of almost all K_{r+1}-free graphs in Section 4, and finally, in Section 5, we use the Balogh–Bollobás–Simonovits method to deduce Theorem 1.1.

2. A supersaturated Erdős-Simonovits stability theorem

In this section, we prove our 'supersaturated stability theorem' for K_{r+1}-free graphs. As noted in the Introduction, we do so by adapting a proof of Füredi [15].

Given a graph G, a vertex $v \in V(G)$ and an integer $m \in \mathbb{N}$, let us write $K_m(G)$ for the number of m-cliques in G, and $K_m(v)$ for the number of such m-cliques containing v.

Proof of Theorem 1.2. We will prove by induction on r that

$$K_{r+1}(G) \geqslant \frac{n^{r-1}}{c(r)} \left(e(G) + t - \left(1 - \frac{1}{r}\right) \frac{n^2}{2} \right), \tag{2}$$

where $c(r) := 2(r+1)^{r-1} r^{r-1}/r!$, for every graph G on n vertices which is t-far from being r-partite. Since $c(r) \leqslant e^{2r} r!$, the theorem follows from (2).

Note first that the theorem holds in the case $r = 1$, since a graph is t-far from being 1-partite if and only if $e(G) \geqslant t$, and hence G has more than $\frac{e(G)+t}{2}$ copies of K_2, as required. So let $r \geqslant 2$ and assume that the result holds for $r-1$. Let $n, t \in \mathbb{N}$, and let G be a graph that is t-far from being r-partite.

First, for each $v \in V(G)$, set $B_v = N(v)$ (the set of neighbors of v in G) and $A_v = V(G) \setminus B_v$, and observe that

$$\sum_{u \in A_v} d(u) = e(G) + e(A_v) - e(B_v), \tag{3}$$

where $e(X)$ denotes the number of edges in the graph $G[X]$. Now, the graph $G[B_v]$ is $\big(t - e(A_v)\big)$-far from being $(r-1)$-partite, and so, by the induction hypothesis,

$$K_{r+1}(v) \geqslant \frac{|B_v|^{r-2}}{c(r-1)} \left(e(B_v) + t - e(A_v) - \left(1 - \frac{1}{r-1}\right) \frac{|B_v|^2}{2} \right), \tag{4}$$

since each copy of K_r in $G[B_v]$ corresponds to a copy of K_{r+1} in G that contains v.

Combining (3) and (4), noting that $|B_v| = d(v)$, and summing over v, it follows that

$$(r+1) \cdot K_{r+1}(G) \geqslant \sum_{v \in V(G)} \frac{d(v)^{r-2}}{c(r-1)} \left(e(G) + t - \sum_{u \in A_v} d(u) - \left(1 - \frac{1}{r-1}\right) \frac{d(v)^2}{2} \right).$$

We claim that

$$\sum_{v \in V(G)} \sum_{u \in A_v} d(u) d(v)^{r-2} \leqslant \sum_{v \in V(G)} \sum_{u \in A_v} d(v)^{r-1} = \sum_{v \in V(G)} d(v)^{r-1} \big(n - d(v)\big). \tag{5}$$

Indeed, let $X = \big\{(v,u) : v \in V(G),\, u \in A_v\big\}$ denote the set of ordered pairs in the sum above, and note that $(v,u) \in X$ if and only if $uv \notin E(G)$. Since X is symmetric, the inequality in (5) now follows immediately for $r = 2$, and by the Cauchy-Schwarz inequality

$$\sum_{(v,u)\in X} d(u)d(v) \leqslant \bigg(\sum_{(v,u)\in X} d(u)^2 \bigg)^{1/2} \bigg(\sum_{(v,u)\in X} d(v)^2 \bigg)^{1/2}$$

for $r = 3$. For $r \geqslant 4$, applying Hölder's inequality with $p = r-2$ and $q = (r-2)/(r-3)$ gives

$$\sum_{(v,u)\in X} d(u)d(v)^{r-2} \leqslant \bigg(\sum_{(v,u)\in X} d(u)^{r-2}d(v) \bigg)^{1/p} \bigg(\sum_{(v,u)\in X} d(v)^{r-1} \bigg)^{1/q},$$

since $\big(r-2-\frac{1}{r-2}\big)\frac{r-2}{r-3} = \frac{r^2-4r+3}{r-3} = r-1$. Once again using the symmetry of X, and noting that $1 - 1/p = 1/q$, the claimed inequality (5) follows.

Combining the inequalities above, we obtain

$$(r+1)\cdot K_{r+1}(G) \;\geqslant\; \sum_{v\in V(G)} \frac{d(v)^{r-2}}{c(r-1)} \bigg(e(G) + t - d(v)n + \bigg(1 + \frac{1}{r-1}\bigg)\frac{d(v)^2}{2} \bigg).$$

Since the factor in parentheses is minimized when $d(v) = \frac{r-1}{r}\cdot n$, it follows that

$$(r+1)\cdot K_{r+1}(G) \;\geqslant\; \sum_{v\in V(G)} \frac{d(v)^{r-2}}{c(r-1)} \bigg(e(G) + t - \bigg(1 - \frac{1}{r}\bigg)\frac{n^2}{2} \bigg).$$

Finally, note that every graph G is $\big(e(G)/r\big)$-close to being r-partite (take a random partition), and hence we may assume that $\big(1+\frac{1}{r}\big)e(G) \geqslant \big(1-\frac{1}{r}\big)\frac{n^2}{2}$, since otherwise the theorem is trivial. Thus, by the convexity of x^{r-2},

$$\sum_{v\in V(G)} d(v)^{r-2} \geqslant n \cdot \bigg(\frac{2e(G)}{n}\bigg)^{r-2} \geqslant \bigg(\frac{r-1}{r+1}\bigg)^{r-2} n^{r-1},$$

and so, since $c(r-1)\cdot(r+1)^{r-1} = c(r)\cdot(r-1)^{r-2}$, it follows that

$$K_{r+1}(G) \;\geqslant\; \frac{n^{r-1}}{c(r)} \bigg(e(G) + t - \bigg(1 - \frac{1}{r}\bigg)\frac{n^2}{2} \bigg),$$

as claimed. □

3. An approximate structural result

In this section we will prove the following approximate version of Theorem 1.1.

Theorem 3.1. *Let $r = r(n) \in \mathbb{N}$ be a function satisfying $r \leqslant (\log n)^{1/4}$ for each $n \in \mathbb{N}$. Then almost all K_{r+1}-free graphs on n vertices are n^{2-1/r^2}-close to being r-partite.*

Theorem 3.1 is a straightforward consequence of Theorem 1.2 and the following theorem proved by Mousset, Nenadov and Steger [19], which was proved using the hypergraph container method of Balogh, Morris and Samotij [7] and Saxton and Thomason [24]. The following theorem is slightly stronger than the result stated in [19], but follows easily from essentially the same proof.

Theorem 3.2. *Let $r = r(n) \in \mathbb{N}$ be a function satisfying $r \leqslant (\log n)^{1/4}$ for each sufficiently large $n \in \mathbb{N}$. There exists a collection $\mathcal{C}$ of graphs such that the following hold:*

- (a) *every K_{r+1}-free graph on n vertices is a subgraph of some $G \in \mathcal{C}_n$,*
- (b) *$K_{r+1}(G) \leqslant n^{r+1-2/r^2}$ for every $G \in \mathcal{C}_n$, and*
- (c) *$|\mathcal{C}_n| \leqslant \exp\big(n^{2-2/r^2}\big)$,*

where $\mathcal{C}_n = \big\{G \in \mathcal{C} : v(G) = n\big\}$.

Deducing Theorem 3.1 from Theorems 1.2 and 3.2 is straightforward.

Proof of Theorem 3.1. For each $t \in \mathbb{N}$, set

$$\mathcal{F}_t = \left\{G : e(G) \geqslant \left(1 - \frac{1}{r}\right)\frac{|G|^2}{2} - \frac{t}{2} \text{ and } G \text{ is } t\text{-far from being } r\text{-partite}\right\},$$

and observe that if $G \in \mathcal{F}_t$, then

$$K_{r+1}(G) \geqslant \frac{|G|^{r-1} \cdot t}{e^{2r+1} \cdot r!},$$

by Theorem 1.2. Therefore, letting $\mathcal{C}$ be the collection of graphs given by Theorem 3.2, and setting $t = n^{2-1/r^2}$, it follows from property (b) and the bound $r \leqslant (\log n)^{1/4}$ that $\mathcal{C}_n \cap \mathcal{F}_t = \emptyset$.

Now, for each K_{r+1}-free graph G on n vertices that is n^{2-1/r^2}-far from being r-partite, we have $G \in C$ for some $C \in \mathcal{C}_n$, and by the observations above and the definition of $\mathcal{F}_t$, it follows that

$$e(C) \leqslant \left(1 - \frac{1}{r}\right)\frac{n^2}{2} - \frac{t}{2}.$$

Therefore, summing over all such containers, the number of such graphs is at most

$$\exp\big(n^{2-2/r^2}\big) \cdot 2^{t_r(n) - t/2} \ll 2^{t_r(n) - t/4},$$

which is clearly smaller than the number of K_{r+1}-free graphs on n vertices, as required. □

4. Some properties of a typical K_{r+1}-free graph

In this section we will prove some useful structural properties of almost all K_{r+1}-free graphs. These structural properties will allow us (in Section 5) to count the K_{r+1}-free graphs that are close to being r-partite, and hence to complete the proof of Theorem 1.1. We emphasize that the lemmas in this section were all proved for fixed $r \in \mathbb{N}$ in [5], and no extra ideas are required in order to extend their proofs to our more general setting.

Let us fix throughout this section a function $2 \leqslant r = r(n) \leqslant (\log n)^{1/4}$, and let us denote by $\mathcal{G}$ the collection of K_{r+1}-free graphs on n vertices that are n^{2-1/r^2}-close to being r-partite. We begin with two simple definitions.

Definition 4.1 (Optimal partitions)**.** An r-partition $(U_1, \ldots, U_r)$ of the vertex set of a graph G is called *optimal* if the number of interior edges, $\sum_{i=1}^r e(U_i)$, is minimized.

Definition 4.2 (Uniformly dense graphs)**.** We say that a graph G is *uniformly dense* if for every optimal r-partition $(U_1, \ldots, U_r)$ and every pair $\{i, j\} \subset [r]$, we have

$$e(A, B) > \frac{|A||B|}{32} \tag{6}$$

for every $A \subset U_i$ and $B \subset U_j$ with $|A| = |B| \geqslant 2^{-8r}n$.

Lemma 4.3. *The number of graphs in $\mathcal{G}$ that are not uniformly dense is at most*

$$2^{t_r(n) - 2^{-17r}n^2},$$

and therefore almost all K_{r+1}-free graphs are uniformly dense.

Proof. In order to count such graphs, we first choose the optimal partition $\mathcal{U} = (U_1, \ldots, U_r)$, the pair $\{i, j\} \subset [r]$, and the sets $A \subset U_i$ and $B \subset U_j$ for which (6) fails. We then choose the edges between A and B, and finally the remaining edges. Note first that we have at most r^n choices for $\mathcal{U}$, at most r^2 choices for $\{i, j\}$, and at most 2^{2n} choices for the pair (A, B).

Now, the number of choices for the edges between A and B is at most

$$\sum_{k=0}^{|A||B|/32} \binom{|A||B|}{k} \leqslant n^2 (32e)^{|A||B|/32} \leqslant 2^{|A||B|/4},$$

and the number of choices for the remaining edges is at most

$$2^{t_r(n) - |A||B|} \binom{n^2}{n^{2-1/r^2}} \leqslant 2^{t_r(n) - |A||B|} \exp\Big(n^{2-1/r^2} \log n \Big) \leqslant 2^{t_r(n) - |A||B|/2},$$

since $\mathcal{U}$ is optimal, $|A||B| \geqslant 2^{-16r}n^2$, and each $G \in \mathcal{G}$ is n^{2-1/r^2}-close to being r-partite.

It follows that the number of graphs in $\mathcal{G}$ that are not uniformly dense is at most

$$r^{n+2} \cdot 2^{2n} \cdot 2^{t_r(n) - |A||B|/4} \leqslant 2^{t_r(n) - 2^{-17r}n^2},$$

as claimed. ∎

Our next definition controls the maximum degree inside the parts of an optimal partition.

Definition 4.4 (Internally sparse graphs)**.** A graph G is said to be *internally sparse* if, for every optimal partition $\mathcal{U} = (U_1, \ldots, U_r)$ of G, we have

$$\Delta\big(G[U_i]\big) \leqslant 2^{-3r}n. \tag{7}$$

for every $1 \leqslant i \leqslant r$. Otherwise we say that G is *internally dense.*

Lemma 4.5. *If $G \in \mathcal{G}$ is internally dense then it is not uniformly dense.*

We will prove Lemma 4.5 using the following embedding lemma[4] from [2].

Lemma 4.6. *Let $\alpha > 0$, let G be a graph, and let $W_1, \dots, W_r \subset V(G)$ be disjoint sets of vertices. Suppose that for every pair $\{i, j\} \subset [r]$ and every pair of sets $A \subset W_i$ and $B \subset W_j$ with $|A| \geqslant \alpha^r |W_i|$ and $|B| \geqslant \alpha^r |W_j|$, we have $e(A, B) > \alpha |A||B|$.*

Then G contains a copy of K_r with one vertex in each set W_j.

Proof of Lemma 4.5. Suppose for a contradiction that $G \in \mathcal{G}$ is both internally dense and uniformly dense. Let $\mathcal{U} = (U_1, \dots, U_r)$ be the optimal partition given by Definition 4.4, and suppose that $v \in U_1$ has degree at least $2^{-3r} n$ in $G[U_1]$. For each $i \in [r]$, let $W_i = N(v) \cap U_i$, and observe that $|W_i| \geqslant 2^{-3r} n$, since $\mathcal{U}$ is optimal.

Observe that $W_1, \dots, W_r$ satisfy the conditions of Lemma 4.6 with $\alpha = 1/32$, since G is uniformly dense, so $e(A, B) > |A||B|/32$ for every pair $\{i, j\} \subset [r]$, and every $A \subset U_i$ and $B \subset U_j$ with $|A| = |B| \geqslant 2^{-8r} n$. Thus, by Lemma 4.6, there exists a copy of K_r in the neighborhood of v, which (including v) gives a copy of K_{r+1} in G. But this is a contradiction, since our graph is K_{r+1}-free, and so every internally dense graph $G \in \mathcal{G}$ is not uniformly dense, as claimed. □

Our final definition controls the sizes of the parts in an optimal partition.

Definition 4.7 (Balanced graphs)**.** A graph G is said to be *balanced* if, for every optimal partition $\mathcal{U} = (U_1, \dots, U_r)$ of G, we have

$$\frac{n}{r} - 2^{-3r} n \leqslant |U_i| \leqslant \frac{n}{r} + 2^{-3r} n \tag{8}$$

for every $1 \leqslant i \leqslant r$. Otherwise we say that G is *unbalanced*.

Lemma 4.8. *The number of unbalanced graphs in $\mathcal{G}$ is at most*

$$2^{t_r(n) - 2^{-8r} n^2},$$

and therefore almost all K_{r+1}-free graphs are balanced.

Proof. Let $G \in \mathcal{G}$ be an unbalanced graph, and let $\mathcal{U} = (U_1, \dots, U_r)$ be an optimal partition of G for which (8) fails. Note that

$$\sum_{i=1}^{r-1} \sum_{j=i+1}^{r} |U_i||U_j| \leqslant t_r(n) - 2^{-7r} n^2,$$

[4] In fact, the version stated here is slightly more general than [2, Lemma 3.1], but follows from exactly the same proof.

since moving a vertex from a set of size at least $n/r+a$ to one of size $n/r-b$ creates at least $a+b$ new potential cross edges. The number of such graphs $G \in \mathcal{G}$ is therefore at most

$$r^n \cdot 2^{t_r(n) - 2^{-7r} n^2} \cdot \binom{n^2}{n^{2-1/r^2}} \leqslant 2^{t_r(n) - 2^{-8r} n^2},$$

as claimed. □

5. The proof of Theorem 1.1

In this section we will deduce Theorem 1.1 from Theorem 3.1, using the method of Balogh, Bollobás and Simonovits [5, 6]. Recall from the previous section that almost all K_{r+1}-free graphs are uniformly dense, internally sparse and balanced.

Let us fix throughout this section a function $2 \leqslant r = r(n) \leqslant (\log n)^{1/4}$.

Definition 5.1. Let $\mathcal{Q}(n,r)$ denote the collection of K_{r+1}-free graphs on n vertices that are not r-partite, but are n^{2-1/r^2}-close to being r-partite, and are moreover uniformly dense, internally sparse and balanced.

Let $\mathcal{K}(n,r)$ denote the collection of K_{r+1}-free graphs on n vertices. We will prove the following proposition, which completes the proof of Theorem 1.1.

Proposition 5.2. *For every sufficiently large $n \in \mathbb{N}$,*

$$|\mathcal{Q}(n,r)| \leqslant 2^{-2^{-5r} n} \cdot |\mathcal{K}(n,r)|.$$

The idea of the proof is as follows. We will define a collection of bipartite graphs F_m (see Definition 5.8) with parts $\mathcal{Q}(n,r,m)$ and $\mathcal{K}(n,r)$, where the sets $\mathcal{Q}(n,r,m)$ form a partition of $\mathcal{Q}(n,r)$ (see Definitions 5.4 and 5.5). These bipartite graphs will have the following property: the degree in F_m of each $G \in \mathcal{Q}(n,r,m)$ will be significantly larger than the degree of each $G \in \mathcal{K}(n,r)$ (see Lemmas 5.10 and 5.12). The result will then follow by double counting the edges of each F_m and summing over m.

In order to define $\mathcal{Q}(n,r,m)$ and F_m, we will need the following simple concept.

Definition 5.3 (Bad sets)**.** Let G be a graph and let $U \subset V(G)$. A set of r vertices $R \subset V(G) \setminus U$ is said to be *bad* towards U if it has no common neighbor in U.

In the following definition we may choose the partition $\mathcal{U}$ and the sets $X^{(1)}, \ldots, X^{(r)}$ arbitrarily, subject to the given conditions.

Definition 5.4. For each $G \in \mathcal{Q}(n,r)$, fix an optimal partition $\mathcal{U} = (U_1, \ldots, U_r)$ of $V(G)$, and for each $j \in [r]$ choose a maximal collection of vertex-disjoint sets $X^{(j)} = \big\{ R_1^{(j)}, \ldots, R_{\ell(j)}^{(j)} \big\}$ such that $R_i^{(j)}$ is bad towards U_j for each $i \in [\ell(j)]$. We define

$$m(G) := \max \big\{ \ell(j) : j \in [r] \big\},$$

let $j(G)$ denote the smallest j for which this maximum is attained, and set

$$X(G) \,:=\, R_1^{(j(G))} \cup \cdots \cup R_{\ell(j(G))}^{(j(G))}.$$

With this definition in place, it is natural to partition $\mathcal{Q}(n,r)$ by the size of $m(G)$.

Definition 5.5. For each $m \in \mathbb{N}$, we define

$$\mathcal{Q}(n,r,m) \,=\, \big\{ G \in \mathcal{Q}(n,r) \,:\, m(G) = m \big\}.$$

Before continuing, let us note a simple but key fact.

Lemma 5.6. *$m(G) \geqslant 1$ for every $G \in \mathcal{Q}(n,r)$.*

Proof. This follows from the fact that G is not r-partite. Indeed, suppose that $m(G) = 0$ and let $x_0 x_1 \in E(G[U_1])$ be an 'interior' edge of G with respect to $\mathcal{U}$. Since there are no bad r-sets towards U_j for any $j \in [r]$, we can recursively choose vertices $x_j \in U_j$ such that $\{x_0, \ldots, x_j\}$ forms a clique. But this is a contradiction, since G is K_{r+1}-free. □

In order to establish an upper bound on those m which we need to consider, we count those graphs in $\mathcal{Q}(n,r)$ for which $m(G)$ is large.

Lemma 5.7. *If $m \geqslant 2^{-6r} n$, then*

$$|\mathcal{Q}(n,r,m)| \,\leqslant\, 2^{t_r(n) - mn/2^{3r}}.$$

Proof. Let $m \geqslant 2^{-6r} n$, and consider the number of ways of constructing a graph $G \in \mathcal{Q}(n,r,m)$. We have at most r^n choices for the partition $\mathcal{U}$, at most $\binom{n}{r}^m$ choices for the set $X(G)$, and r choices for $j = j(G)$. Moreover, we have at most

$$2^{t_r(n) - |U_j||X(G)|} \big(2^r - 1 \big)^{|U_j||X(G)|/r} \,\leqslant\, 2^{t_r(n) - mn/2^{2r}}$$

choices for the edges between different parts of $\mathcal{U}$, since $X(G)$ is composed of r-sets that are bad towards U_j, and G is balanced. Finally, we have at most $n^{O(n^{2-1/r^2})}$ choices for the edges inside parts of $\mathcal{U}$, since G is n^{2-1/r^2}-close to being r-partite.

It follows that

$$|\mathcal{Q}(n,r,m)| \,\leqslant\, r^n \cdot \binom{n}{r}^m \cdot r \cdot n^{O(n^{2-1/r^2})} \cdot 2^{t_r(n) - mn/2^{2r}} \,\leqslant\, 2^{t_r(n) - mn/2^{3r}}$$

as required, since $m \geqslant 2^{-6r} n$, $n \log r, mr \log n \ll mn/2^{9r}$ and $n^{-1/r^2} \log n \ll 2^{-9r}$. □

From now on, let us fix a function $1 \leqslant m = m(n) \leqslant 2^{-6r} n$. We are ready to define the bipartite graph F_m.

Definition 5.8. Define a map $\Phi_m \colon \mathcal{Q}(n,r,m) \to 2^{\mathcal{K}(n,r)}$ by placing $H \in \Phi_m(G)$ if and only if H can be constructed from G by first removing all edges of G that are incident to $X(G)$, and then adding an arbitrary subset of the edges between $X(G)$ and $V(G) \setminus \big(X(G) \cup U_{j(G)} \big)$.

Let F_m be the bipartite graph with edge set $\{(G,H) : H \in \Phi_m(G)\}$.

We first observe that the map Φ_m is well-defined.

Lemma 5.9. *If $G \in \mathcal{Q}(n,r,m)$ and $H \in \Phi_m(G)$, then H is K_{r+1}-free.*

Proof. This follows easily from the fact that G is K_{r+1}-free, and the maximality of $X(G)$. Indeed, if there exists a copy of K_{r+1} in H, then it must contain a vertex of $X(G)$, and therefore it must contain no other vertices of $X(G) \cup U_{j(G)}$. Hence it contains exactly r vertices of $V(G) \setminus \big(X(G) \cup U_{j(G)}\big)$, and by the maximality of $X(G)$ these have a common neighbor in $U_{j(G)}$. But this contradicts our assumption that G is K_{r+1}-free, as required. $\square$

We are now ready to prove our first bound on the degrees in F_m.

Lemma 5.10. *For every $G \in \mathcal{Q}(n,r,m)$,*

$$\log_2 |\Phi_m(G)| \geqslant \left(1 - \frac{1}{r} - \frac{1}{2^{3r}} - \frac{mr}{n}\right) mnr.$$

Proof. This follows immediately from the fact that G is balanced. Indeed, we have two choices for each of the

$$|X(G)| \cdot \big|V(G) \setminus \big(X(G) \cup U_{j(G)}\big)\big| \geqslant mr \cdot \left(1 - \frac{1}{r} - \frac{1}{2^{3r}} - \frac{mr}{n}\right) n \tag{9}$$

potential edges between $X(G)$ and $V(G) \setminus \big(X(G) \cup U_{j(G)}\big)$. $\square$

In order to bound the degrees in F_m of vertices in $\mathcal{K}(n,r)$, we will need the following lemma, which counts the optimal partitions in the neighborhood of such a vertex. We note that here, the upper bound on m from Lemma 5.7 is crucial.

Lemma 5.11. *For each $H \in \mathcal{K}(n,r)$, there are at most $2^{n/2^{4r}}$ distinct partitions $\mathcal{U}$ of $V(H)$ such that $\mathcal{U}$ is an optimal partition of some graph $G \in \Phi_m^{-1}(H)$.*

Proof. We will use the fact that each $G \in \Phi_m^{-1}(H)$ is uniformly dense and n^{2-1/r^2}-close to being r-partite to show that the optimal partitions in question must be 'close' to one another.

To be precise, let $G_1, G_2 \in \Phi_m^{-1}(H)$, and let $\mathcal{U} = (U_1, \ldots, U_r)$ be an optimal partition of G_1 and $\mathcal{V} = (V_1, \ldots, V_r)$ be an optimal partition of G_2. We claim that

$$\big|\big\{ j \in [r] \,:\, |U_i \cap V_j| > 2^{-6r} n + 2mr \big\}\big| \leqslant 1$$

for every $i \in [r]$. Indeed, suppose that

$$\big|U_i \cap V_j\big| > 2^{-6r} n + 2mr \qquad \text{and} \qquad \big|U_i \cap V_{j'}\big| > 2^{-6r} n + 2mr,$$

set $A = \big(U_i \cap V_j\big) \setminus \big(X(G_1) \cup X(G_2)\big)$ and $B = (U_i \cap V_{j'}) \setminus \big(X(G_1) \cup X(G_2)\big)$, and note that, since G_2 is uniformly dense, we have $e_{G_2}(A,B) > |A||B|/32 > 2^{-12r-5} n^2$. But these edges are all contained in U_i, so this contradicts the fact that G_1 is n^{2-1/r^2}-close to being r-partite, as required.

It follows that (by renumbering the parts if necessary) we have

$$\big|U_i \setminus V_i\big| \,\leqslant\, r \cdot \big(2^{-6r} n + 2mr\big) \,\leqslant\, 2^{-5r} n$$

for every $i \in [r]$, where second inequality follows since $m \leqslant 2^{-6r}n$. Set $D_i = U_i \setminus V_i$, and observe that the partition $\mathcal{U}$ and the collection $(D_1, \ldots, D_r)$ together determine $\mathcal{V}$. It follows that the number of optimal partitions is at most

$$\left(\sum_{k=0}^{2^{-5r}n} \binom{n}{k} \right)^r \leqslant\, n^r \cdot \binom{n}{2^{-5r}n}^r \,\leqslant\, 2^{r \log n} \cdot \big(e2^{5r}\big)^{r2^{-5r}n} \,\leqslant\, 2^{n/2^{4r}},$$

as required. □

We can now bound the degrees on the right. Recall than in Definition 5.4 we chose a 'canonical' optimal partition for each graph $G \in \mathcal{Q}(n,r)$.

Lemma 5.12. *We have*

$$\log_2 \big|\Phi_m^{-1}(H)\big| \,\leqslant\, \left(1 - \frac{1}{r} - \frac{1}{2^{2r}}\right) rmn$$

for every $H \in \mathcal{K}(n,r)$.

Proof. First let us fix a partition $\mathcal{U} = (U_1, \ldots, U_r)$, and count the number of graphs $G \in \mathcal{Q}(n,r,m)$ with $\Phi_m(G) = H$ whose optimal partition is $\mathcal{U}$. To do so, first note that we have $\binom{n}{r}^m$ choices for $X(G)$, and at most r choices for $j = j(G)$. Now, since G is internally sparse and balanced, each vertex $v \in X(G)$ has at most $2^{-3r}n$ neighbors in its own part of $\mathcal{U}$, and $\big||U_j| - n/r\big| \leqslant n/2^{3r}$. Thus we have at most

$$\binom{n}{2^{-3r}n} \cdot 2^{(1-2/r+1/2^{3r})n} \,\leqslant\, 2^{(1-2/r+2/2^{3r})n}$$

choices for the edges between each vertex $v \in X(G)$ and $V(G) \setminus (X(G) \cup U_j)$. Finally, by the definition of bad sets, and since G is balanced, we have at most

$$(2^r - 1)^{(1/r+1/2^{3r})mn} \,\leqslant\, 2^{(1/r-3/2^{2r})mnr}$$

choices for the edges between $X(G)$ and U_j.

Since, by Lemma 5.11, we have at most $2^{n/2^{4r}}$ choices for the partition $\mathcal{U}$, it follows that

$$\begin{aligned} \log_2 \big|\Phi_m^{-1}(H)\big| \;&\leqslant\; mr \log n + \log r + \left(1 - \frac{2}{r} + \frac{1}{2^{2r}} + \frac{1}{r} - \frac{3}{2^{2r}} + \frac{1}{2^{4r-1}}\right) mnr \\ &\leqslant\; \left(1 - \frac{1}{r} - \frac{1}{2^{2r}}\right) mnr, \end{aligned}$$

as claimed. □

Finally we put the pieces together and prove Proposition 5.2.

Proof of Proposition 5.2. We claim first that

$$|\mathcal{Q}(n,r,m)| \;\leqslant\; 2^{-2^{-3r}mnr} \cdot |\mathcal{K}(n,r)|. \tag{10}$$

To prove this, we simply double count the edges of F_m, using Lemmas 5.10 and 5.12. Indeed, we have

$$\log_2\left(\frac{|\mathcal{Q}(n,r,m)|}{|\mathcal{K}(n,r)|}\right) \;\leqslant\; \left(1-\frac{1}{r}-\frac{1}{2^{2r}}\right)mnr - \left(1-\frac{1}{r}-\frac{1}{2^{3r}}-\frac{mr}{n}\right)mnr,$$

which implies (10) since $m \leqslant 2^{-3r}n$.

Summing (10) over m, and recalling that G is n^{2-1/r^2}-close to being r-partite, we obtain

$$|\mathcal{Q}\left(n,r\right)| \;\leqslant\; \sum_{m=1}^{2^{-3r}n} 2^{-2^{-3r}mnr} \cdot |\mathcal{K}(n,r)| \;+\; \sum_{m=2^{-3r}n}^{n} 2^{t_r(n)-mn/2^{2r}} \;\leqslant\; 2^{-2^{-5r}n} \cdot |\mathcal{K}(n,r)|,$$

by Lemmas 5.6 and 5.7, as required. □

Finally, let us deduce Theorem 1.1.

Proof of Theorem 1.1. By Theorem 3.1, almost all K_{r+1}-free graphs on n vertices are n^{2-1/r^2}-close to r-partite. We further showed in Lemmas 4.3, 4.5, and 4.8 that almost all of these graphs are either r-partite, or in $\mathcal{Q}(n,r)$. Since by Proposition 5.2, for sufficiently large n, the size of $\mathcal{Q}(n,r)$ is exponentially small compared to $\mathcal{K}(n,r)$, it follows that almost all K_{r+1}-free graphs are r-partite, as required. □

References

[1] N. Alon, J. Balogh, B. Bollobás and R. Morris, The structure of almost all graphs in a hereditary property, *J. Combin. Theory Ser. B*, **101** (2011), 85–110.

[2] N. Alon, J. Balogh, P. Keevash and B. Sudakov, The number of edge colorings with no monochromatic cliques, *J. London Math. Soc.*, **70** (2004), 273–288.

[3] N. Alon and J.H. Spencer, The probabilistic method, 3rd edition, *Wiley*, New York, (2008).

[4] J. Balogh and J. Butterfield, Excluding induced subgraphs: critical graphs, *Random Structures Algorithms*, **38** (2011), 100–120.

[5] J. Balogh, B. Bollobás and M. Simonovits. The number of graphs without forbidden subgraphs, *J. Combin. Theory Ser. B*, **91** (2004), 1–24.

[6] J. Balogh, B. Bollobás and M. Simonovits, The typical structure of graphs without given excluded subgraphs, *Random Structures Algorithms*, **34** (2009), 305–318.

[7] J. Balogh, R. Morris and W. Samotij, Independent sets in hypergraphs, to appear in *J. Amer. Math. Soc.*

[8] J. Balogh, R. Morris, W. Samotij and L. Warnke, The typical structure of sparse K_{r+1}-free graphs, to appear in *Trans. Amer. Math. Soc.*

[9] B. Bollobás and V. Nikiforov, The number of graphs with large forbidden subgraphs, *European J. Combin.*, **32** (2010), 1964–1968.

[10] B. Bollobás and A. Thomason, Threshold functions. *Combinatorica*, **7** (1987),35-38.

[11] P. Erdős, P. Frankl and V. Rödl, The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent, *Graphs Combin.*, **2** (1986), 113–121.

[12] P. Erdős, D.J. Kleitman and B.L. Rothschild, Asymptotic enumeration of K_n-free graphs, Colloquio Internazionale sulle Teorie Combinatorie (Rome, 1973), Tomo II, 19–27. Atti dei Convegni Lincei, No. 17, Accad. Naz. Lincei, Rome, 1976.
[13] P. Erdős and M. Simonovits, A limit theorem in graph theory, *Studia Sci. Math. Hungar.*, **1** (1966), 51–57.
[14] P. Erdős and G. Szekeres, A combinatorial problem in geometry, *Compos. Math.*, **2** (1935), 463–470.
[15] Z. Füredi, see http://www.renyi.hu/conferences/sze70/Talks/furedi.pdf
[16] P. G. Kolaitis, H. J. Prömel and B. L. Rothschild, $K_{\ell+1}$-free graphs: asymptotic structure and a 0-1 law, *Trans. Amer. Math. Soc.*, **303** (1987), 637–671.
[17] L. Lovász and M. Simonovits, On the number of complete subgraphs of a graph, II. *Studies in pure mathematics*, Birkhuser (1983), 459–495.
[18] W. Mantel, Problem 28, *Wiskundige Opgaven*, **10** (1907), 60–61.
[19] F. Mousset, R. Nenadov and A. Steger, On the number of graphs without large cliques, submitted.
[20] D. Osthus, H. J. Prömel and A. Taraz, For which densities are random triangle-free graphs almost surely bipartite?, *Combinatorica*, **23** (2003), 105–150.
[21] H. J. Prömel and A. Steger, The asymptotic number of graphs not containing a fixed color-critical subgraph, *Combinatorica*, **12** (1992), 463–473.
[22] F.P. Ramsey, On a problem of formal logic, *Proc. London Math. Soc.*, **30** (1930), 264–286.
[23] M. Simonovits, A method for solving extremal problems in graph theory, stability problems, *Theory of graphs (Proc. Colloq., Tihany, 1966)*, Academic press, New York, (1968), 279–319.
[24] D. Saxton and A. Thomason, Hypergraph containers, submitted.
[25] P. Turán, Eine Extremalaufgabe aus der Graphentheorie, *Mat. Fiz. Lapok*, **48** (1941), 436–452.

(JB, HL, MS): Department of Mathematical Sciences, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA
E-mail address: `jobal@math.uiuc.edu, hliu36|sharifz2@illinois.edu`

(NB): School of Mathematical and Statistical Sciences, Arizona State University, Tempe, AZ 85287 USA and IMPA, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ, Brasil
E-mail address: `neal@asu.edu`

(MCN, RM): IMPA, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ, Brasil
E-mail address: `collares|rob@impa.br`