Clique supersaturation

Quentin Dubroff, Benjamin Gunby, Bhargav Narayanan, and Sam Spiro

Abstract. We study how many copies of a graph F that another graph G with a given number of cliques is guaranteed to have. For example, one of our main results states that for all $t \geq 2$, if G is an n vertex graph with $kn^{3/2}$ triangles and k is sufficiently large in terms of t, then G contains at least

$$\Omega\left(\min\left\{k^t n^{3/2}, k^{\frac{2t^2}{3t-1}} n^{\frac{5t-2}{3t-1}}\right\}\right)$$

copies of $K_{2,t}$, and furthermore, we show these bounds are essentially best-possible provided either $k \geq n^{1/2t}$ or if certain bipartite-analogues of well known conjectures for Turán numbers hold.

1. Introduction

In this paper, we shall study a generalised supersaturation problem. Broadly speaking, 'extremal problems' ask for the largest 'size' N that a combinatorial object can have before containing at least one structure F, and in turn, 'supersaturation problems' ask about how many copies of F are guaranteed to exist if a combinatorial object has 'size' substantially larger than the extremal value N. In addition to being natural refinements of extremal problems in their own right, supersaturation problems also arise often in a number of other contexts. For example, supersaturation results were used by Erdős and Simonovits [7] to obtain upper bounds on the Turán number of the hypercube. More recently, supersaturation has proven to be a key ingredient for various asymptotic enumeration results proved using the method of hypergraph containers developed independently by Balogh, Morris, and Samotij [2] and by Saxton and Thomasson [19].

Here, we investigate the following problem: how many copies of a graph F can we guarantee in a graph G with a specified number of copies of another graph H? More precisely, given two graphs F and G, we define

$$\mathcal{N}(F,G) = \#\,\text{subgraphs of } G \text{ isomorphic to } F,$$

and our aim then is to prove lower bounds on $\mathcal{N}(F,G)$ as a function of $\mathcal{N}(H,G)$. We will informally refer to problems of this form as *generalized supersaturation problems.*

An immediate obstruction to this problem is the existence of F-free graphs G which contain a large number of copies of H. To this end, we define the *generalized Turán* *number*, first introduced by Alon and Shikhelman [1], by

$$\mathrm{ex}(n, F, H) = \max\{\mathcal{N}(H, G) : G \text{ is an } F\text{-free graph on } n \text{ vertices}\},$$

and we write $\mathrm{ex}(n, F) = \mathrm{ex}(n, F, K_2)$ to denote the (classical) *Turán number of* F. Observe that, by definition, if G is an n vertex graph then

$$\mathcal{N}(H, G) > \mathrm{ex}(n, F, H) \implies \mathcal{N}(F, G) > 0, \tag{1}$$

and this is best possible since there exist F-free graphs with $\mathcal{N}(H, G) \le \mathrm{ex}(n, F, H)$.

We are interested in quantitative versions of the trivial bound (1). For example, Halfpap and Palmer [10] proved that if $\chi(F) > \chi(H)$ and $\varepsilon > 0$, then any n vertex graph G with $\mathcal{N}(H, G) \ge (1 + \varepsilon)\mathrm{ex}(n, F, H)$ has $\mathcal{N}(F, G) = \Omega_\varepsilon(n^{v(F)})$, i.e., G contains a constant proportion of the copies of F in K_n. Hence, the central interest in the case $\chi(F) > \chi(H)$ is in proving asymptotically tight bounds for $\mathcal{N}(F, G)$ as a function of ε. For example, work of Razborov [17] completely solves this asymptotic problem of minimizing the number of K_3's in a graph with a given number of edges, and this was later generalized by Reiher [18] to handle general cliques K_r in place of the triangle K_3.

In this paper, we focus on the 'degenerate' setting $\chi(F) \le \chi(H)$ where the focus for supersaturation centers around proving coarse (i.e., order of magnitude) bounds on $\mathcal{N}(F, G)$. One classical example in this setting is the following conjecture of Erdős and Simonovits [7].

Conjecture 1.1 ([7]). *Let F be a graph with $\mathrm{ex}(n, F) = O(n^\alpha)$. If G is an n-vertex graph with $e(G) = kn^\alpha$ and $k \ge k_0(F)$, then*

$$\mathcal{N}(F, G) = \Omega(k^{e(F)} n^{v(F)-(2-\alpha)e(F)}).$$

We note that this conjecture, if true, would be best possible by considering G to be the random graph with kn^α edges. Conjecture 1.1 is known to hold (possibly with non-optimal values of α) for a large number of graphs, such as even cycles [15] and all graphs which satisfy Sidorenko's conjecture [20, Theorem 9]. One result of particular importance to us will be the following which confirms Conjecture 1.1 for complete bipartite graphs when $\alpha = 2 - 1/s$.

Proposition 1.2 ([8]). *For all $s \le t$, there exists a constant $C = C(s, t)$ such that if G is an n-vertex graph with $e(G) = kn^{2-1/s}$ and $k \ge C$, then $\mathcal{N}(K_{s,t}, G) = \Omega(k^{st} n^s)$.*

Other generalized supersaturation results in the degenerate setting include work of Cutler, Nir, and Radcliffe [6] who studied the case when F, H are each either cliques or stars; as well as Gerbner, Nagy, and Vizer [9] who initiated the systematic study of generalized supersaturation results and who proved a number of results when F, H are both bipartite.

1.1. **Our results.** In this paper, we focus on generalized supersaturation problems when $H = K_r$. That is, we ask how many copies of a given graph F is guaranteed in another graph G if G has N copies of K_r, and we informally refer to this as the *clique supersaturation problem.* For example, we have the following.

Lemma 1.3. *If F is a graph with $v(F) \leq r$ and if G is a graph with $\mathcal{N}(K_r, G) = N$, then $\mathcal{N}(F, G) = \Omega(N^{v(F)/r})$. Moreover, the graph G consisting of a clique of size $N^{1/r}$ satisfies $\mathcal{N}(K_r, G) = \Omega(N)$ and $\mathcal{N}(F, G) = O(N^{v(F)/r})$.*

Lemma 1.3 follows immediately from the Kruskal-Katona theorem, which implies that any graph with N copies of K_r has at least $\Omega(N^{v(F)/r})$ copies of $K_{v(F)}$. Due to Lemma 1.3, we will only consider F with $v(F) > r$ throughout this paper.

Returning to the general problem: when $r = 2$, Conjecture 1.1 predicts that the solution to the clique supersaturation problem is always achieved by the random graph $G_{n,p}$ with N copies of K_2, i.e. when $p = Nn^{-2}$. For larger r, it again makes sense to look at what happens for $G_{n,p}$. To this end, if we want $G_{n,p}$ to have on the order of N copies of K_r, then we should take $p = (Nn^{-r})^{1/\binom{r}{2}}$, which gives

$$\mathbb{E}[\mathcal{N}(F, G_{n,p})] = \Theta\left((Nn^{-r})^{e(F)/\binom{r}{2}} n^{v(F)}\right). \tag{2}$$

Our first main result significantly improves upon this trivial bound for a wide range of F and $r \geq 3$. For this result, we recall that a graph F is 2-*balanced* if for all $F' \subseteq F$ with $v(F') \geq 3$, we have

$$\frac{e(F') - 1}{v(F') - 2} \leq \frac{e(F) - 1}{v(F) - 2}.$$

Theorem 1.4. *Let F be a 2-balanced graph with $e(F) \geq 2$ and let $2 \leq r < v(F)$ be an integer. For all $1 \leq N \leq \binom{n}{r}$, there exists an n-vertex graph G with $\mathcal{N}(K_r, G) = \Omega(N)$ and with*

$$\mathcal{N}(F, G) = O\left((Nn^{-r})^{e(F)\beta_r(F)} n^{v(F)}\right),$$

where

$$\beta_r(F) = \frac{v(F) - 2}{(r-2)(e(F) - 1) + v(F) - 2}.$$

Note that this bound is strictly smaller than the bound (2) from $G_{n,p}$ whenever $\beta_r(F) > \binom{r}{2}^{-1}$, and this is equivalent to having

$$\frac{r+1}{2} > \frac{e(F) - 1}{v(F) - 2} \quad \text{for } r, v(F) \geq 3.$$

For example, this inequality holds if $F = K_{2,t}$ when $t \geq 2$ and $r \geq 3$. More generally, Theorem 1.4 implies the same result holds when K_r is replaced by any r-vertex graph

H. In this case, the bound does better than the corresponding bound coming from $G_{n,p}$ precisely when

$$\frac{e(H)-1}{v(H)-2} > \frac{e(F)-1}{v(F)-2}$$

provided $v(H), v(F) \geq 3$.

A crucial part of our proof of Theorem 1.4 will be the following general lower bound on $\mathrm{ex}(n, K_r, F)$, which may be of independent interest.

Theorem 1.5. *If F is a 2-balanced graph with $e(F) \geq 2$, then for all $2 \leq r < v(F)$ we have*

$$\mathrm{ex}(n, K_r, F) = \Omega(n^{2-\frac{v(F)-2}{e(F)-1}}).$$

Theorem 1.5 recovers the classic result $\mathrm{ex}(n, F) = \Omega(n^{2-\frac{v(F)-2}{e(F)-1}})$ for 2-balanced graphs, though we emphasize that the proof for $r > 2$ is somewhat more involved than the easy deletion argument which proves the classic $r = 2$ case. We also note that Theorem 1.5 can be close to best possible. For instance, it is known that $\mathrm{ex}(n, K_r, K_{2,t}) = \Theta_t(n^{3/2})$ for t sufficiently large in terms of r [1, 21], and in this case, Theorem 1.5 gives a lower bound of $\Omega(n^{3/2-\frac{1}{4t-2}})$, which is quite close to tight.

We next turn to bounds for specific choices of F. For this, it will not make sense to consider an arbitrary number of cliques N, as no copies of F will be guaranteed if $N \leq \mathrm{ex}(n, K_r, F)$. As such, we will normalize the N in our results by replacing N with kn^α whenever[1] $\mathrm{ex}(n, K_r, F) = O(n^\alpha)$.

Perhaps the most natural case of F to consider for the clique supersaturation problem is when $F = K_t$ is itself a clique. The case $t \leq r$ is completely solved by the Kruskal-Katona theorem, and the case $t > r$ is solved up to order of magnitude by the result of Halfpap and Palmer [10].

After cliques, the next simplest case is when F is a tree. This too is relatively easy to solve.

Proposition 1.6. *For all trees T and integers $2 \leq r < v(F)$, there exists a constant $k_0 = k_0(T)$ such that if G is an n-vertex graph with $\mathcal{N}(K_r, G) = kn$ and $k \geq k_0$, then*

$$\mathcal{N}(T, G) = \Omega(k^{(v(T)-1)/(r-1)} n).$$

Moreover, the graph G consisting of the disjoint union of $k^{-1/(r-1)}n$ cliques of size $k^{1/(r-1)}$ satisfies $\mathcal{N}(K_r, G) = \Omega(kn)$ and $\mathcal{N}(T, G) = O(k^{(v(T)-1)/(r-1)} n)$.

---

[1] We will specifically normalize our results based off the following bounds: (1) $\mathrm{ex}(n, K_r, T) = \Theta(n)$ whenever T is a tree, (2) $\mathrm{ex}(n, K_r, K_{2,t}) = O(n^{3/2})$, with this result being tight if t is sufficiently large in terms of r, (3) $\mathrm{ex}(n, K_r, K_{s,t}) = O(n^{r-\binom{r}{2}/s})$ whenever $r \leq s$, with this result being tight if t is sufficiently large in terms of r; see [1, 21].

Alternatively, Theorem 1.4 can be used in Proposition 1.6 instead of the disjoint union of cliques to get the same bound. Sharper bounds for $F = K_{1,s}$ were obtained by Cutler, Nir and Radcliffe [6, Theorem 1.9] when $\mathcal{N}(K_r, G) = (1+\varepsilon)\mathrm{ex}(n, K_{1,s}, K_r)$. This and the Kruskal-Katona theorem are the only results we are aware of studying degenerate clique supersaturation problems prior to this work.

We next look at complete bipartite graphs $K_{s,t}$. This is a natural case to study given that Proposition 1.2 is relatively easy to prove and essentially solves the case of $r = 2$. However, this problem becomes significantly more complex for $r > 2$, even in the simplest (non-tree) case of $s = 2$.

Theorem 1.7. *For all integers $t \geq 2$ and $2 < r < 2 + t$, there exists a constant $k_0 = k_0(t)$ such that if G is an n-vertex graph with $\mathcal{N}(K_r, G) = kn^{3/2}$ and $k \geq k_0$, then*

$$\mathcal{N}(K_{2,t}, G) = \Omega(\min\{k^{\frac{t}{r-2}} n^{3/2}, k^{\frac{2t^2}{(2t-1)(r-2)+t}} n^{\frac{(3t-2)(r-2)+2t}{(2t-1)(r-2)+t}}\}).$$

Moreover, if $2 < r < 2 + t$, there exists an n vertex graph G with $\mathcal{N}(K_r, G) = \Omega(kn^{3/2})$ and

$$\mathcal{N}(K_{2,t}, G) = O(k^{\frac{2t^2}{(2t-1)(r-2)+t}} n^{\frac{(3t-2)(r-2)+2t}{(2t-1)(r-2)+t}}).$$

Also, conditional on the existence of a bipartite graph B on $U \cup V$ with $|U| = k^{-2/(r-2)}n^{3/2}$, $|V| = n$, $e(B) = \Omega(k^{-1/(r-2)}n^{3/2})$, and such that every pair of vertices in V has fewer than t paths of length at most 4 between them; there exists an n vertex graph G with $\mathcal{N}(K_r, G) = \Omega(kn^{3/2})$ and

$$\mathcal{N}(K_{2,t}, G) = O(\min\{k^{\frac{t}{r-2}} n^{3/2}, k^{\frac{2t^2}{(2t-1)(r-2)+t}} n^{\frac{(3t-2)(r-2)+2t}{(2t-1)(r-2)+t}}\}).$$

The bipartite graph B in the conditional portion of this theorem is closely related to a bipartite analogue of the infamous problem of determining $\mathrm{ex}(n, C_8)$; see the concluding remarks for more on this.

One might hope that the methods of Theorem 1.7 could be extended to $K_{s,t}$ in general, but there are fundamental obstacles to this. Indeed, every construction in this paper will turn out to be the union of nearly-disjoint cliques of roughly the same size, and one can essentially show[2] that any construction of this form will fail to do better than the random graph $G_{n,p}$ for $K_{s,t}$ when $r \leq s$; see Lemma 4.3 for an exact statement and the concluding remarks for further discussions.

2. Supersaturation

In this section we prove our supersaturation results, i.e. the lower bounds of Proposition 1.6 and Theorem 1.7. We begin with two preliminary results that will be useful

[2]Of course, every graph G is the disjoint union of K_2's, so this exact claim is not quite true. However, Lemma 4.3 will imply that this claim is true if we further impose that a large portion of the K_r's in G lie within the cliques used in its union.

in our proofs. First, the following shows that if a set of vertices is contained in many copies of K_r, then they must have many neighbors.

Lemma 2.1. *If G is a graph and $S \subseteq V(G)$ is a set of vertices which is contained in at least ℓ cliques of size $r > |S|$, then there are at least $\ell^{1/(r-|S|)}$ vertices adjacent to every vertex of S.*

Proof. Let $N(S)$ denote the set of vertices adjacent to every vertex of S. Observe that every K_r containing S can be identified by choosing $r - |S|$ vertices from $N(S)$. Thus we must have

$$\ell \leq \binom{|N(S)|}{r-|S|} \leq |N(S)|^{r-|S|},$$

from which the result follows. □

To motivate our next lemma, we note that in proving Proposition 1.6, it would be useful to work with a subgraph $G' \subseteq G$ which has large minimum degree. Unfortunately, the standard lemma saying that we can find a $G' \subseteq G$ of minimum degree comparable to the average degree of G will be too weak for our purposes, as it could be the case that G' has very few vertices (in which case we may not be able to find many copies of T in G'). We get around this with the following lemma from [14, Lemma 2.5], which gives substantially stronger bounds on $\delta(G')$ if $v(G')$ is small. We will in fact need a slight generalization of this result to hypergraphs, which can be proven with an identical argument.

Lemma 2.2 ([14])**.** *Let H be an n-vertex hypergraph with $\emptyset \notin E(H)$. For all real $b \geq 1$, there exists a subgraph $H' \subseteq H$ with $v(H') > 0$ and minimum degree at least*

$$2^{-b} \left(\frac{v(H')}{n}\right)^{1/b} \frac{e(H)}{v(H')}.$$

With this we can prove our supersaturation result for trees.

Proof of Proposition 1.6. Recall that we wish to show for all trees T and $2 \leq r < v(T)$, if T is a tree and G is an n-vertex graph with kn copies of K_r, then G contains at least $\Omega(k^{(v(T)-1)/(r-1)})$ copies of T provided k is sufficiently large in terms of T.

Let H by the r-uniform hypergraph with $V(H) = V(G)$ whose hyperedges are copies of K_r in G. Let $H' \subseteq H$ be the subhypergraph guaranteed by Lemma 2.2 with $b = \frac{v(T)-1}{v(T)-r} > 1$, and let $G' \subseteq G$ by the induced subgraph with $V(G') = V(H')$. By unwinding the definitions, we see that every vertex of G' is contained in at least

$$\ell = 2^{-\frac{v(T)-1}{v(T)-r}} \left(\frac{v(G')}{n}\right)^{(v(T)-r)/(v(T)-1)} \frac{kn}{v(G')} = 2^{-\frac{v(T)-1}{v(T)-r}} kn^{\frac{r-1}{v(T)-1}} v(G')^{\frac{1-r}{v(T)-1}}$$

copies of K_r, which by Lemma 2.1 implies every vertex of G' has degree at least $\ell^{1/(r-1)}$. Observe that $\ell \geq 2^{-\frac{v(T)-1}{v(T)-r}} k$, so by taking k sufficiently large, we may assume $\ell^{1/(r-1)} \geq 2v(T)$.

With this in mind, we claim that

$$\mathcal{N}(T, G') \geq v(G') \cdot (\ell^{1/(r-1)} - v(T))^{v(T)-1}/v(T)!.$$

Indeed, because T is a tree, we can order its vertices $x_1, \ldots, x_{v(T)}$ in such a way that every x_i with $i > 1$ has a (unique) neighbor x_j with $j < i$. With this ordering, we build our copies of T greedily by selecting any vertex of G' and label it y_1, and then iteratively given that we have chosen $y_1, \ldots, y_{i-1}$ and that x_i is adjacent to x_j with $j < i$, we choose y_i to be any neighbor of y_j that is not equal to any of the already selected vertices $y_1, \ldots, y_{i-1}$. It is not difficult to check that this procedure terminates with a set of vertices $y_1, \ldots, y_{v(T)}$ which forms a copy of T in $G' \subseteq G$, and that the number of ways of going through this procedure is at least $v(G') \cdot (\ell^{1/(r-1)} - v(T))^{v(T)-1}$. The same tree can be generated at most $v(T)!$ times by this algorithm, giving the bound above.

Using $\ell^{1/(r-1)} \geq 2v(T)$ and the definition of ℓ, we obtain

$$\mathcal{N}(T, G) \geq \mathcal{N}(T, G') = \Omega(v(G')\ell^{(v(T)-1)/(r-1)}) = \Omega(k^{(v(T)-1)/(r-1)} n),$$

giving the desired result. □

We next prove our supersaturation result for $K_{2,t}$.

Proof of the lower bound in Theorem 1.7. Recall that we wish to show that for all $t \geq 2$ and $2 \leq r < 2 + t$ that if G is an n-vertex graph with $kn^{3/2}$ copies of K_r, then G contains at least

$$\Omega\left(\min\left\{k^{\frac{t}{r-2}} n^{3/2}, k^{\frac{2t^2}{(2t-1)(r-2)+t}} n^{\frac{(3t-2)(r-2)+2t}{(2t-1)(r-2)+t}}\right\}\right)$$

copies of $K_{2,t}$ provided k is sufficiently large in terms of t.

The idea behind our proof is the following: either our graph G has many edges (in which case it will contain many copies of $K_{2,t}$ by Proposition 1.2), or we can assume every pair of adjacent vertices has many common neighbors (in which case we can build copies of $K_{2,t}$ greedily). More precisely, Proposition 1.2 implies there exists a constant $C = C(t)$ such that the following holds:

$$\text{if } e(G) \geq Cn^{3/2}, \text{ then } \mathcal{N}(K_{2,t}, G) = \Omega(e(G)^{2t} n^{2-3t}). \tag{3}$$

We use the following when $e(G)$ is small.

Claim 2.3. *If $e(G) \leq \max\{C, k^{1/2}\} n^{3/2}$ and k is sufficiently large in terms of C, then*

$$\mathcal{N}(K_{2,t}, G) = \Omega\left(k^{\frac{t}{r-2}} n^{\frac{3t}{2(r-2)}} e(G)^{\frac{r-2-t}{r-2}}\right). \tag{4}$$

Proof. We form copies of $K_{2,t}$ by starting with an edge $e = uv$ and then choosing any t of the common neighbors of u, v. Note that this process generates each $K_{2,t}$ in at most 2 ways (and in at most 1 way if $t > 2$).

To estimate the number of copies of $K_{2,t}$ formed in this way, let $\deg(e)$ denote the number of K_r's containing the edge e. By Lemma 2.1, the two vertices of e have at least $\deg(e)^{1/(r-2)}$ common neighbors. Thus

$$\mathcal{N}(K_{2,t}, G) \geq \frac{1}{2} \sum_{e \in E(G)} \binom{\deg(e)^{1/(r-2)}}{t} \geq \frac{1}{2} \sum_{e \in E(G)} \left(\frac{\deg(e)^{t/(r-2)}}{t^t} - 1 \right),$$

where this last step used the inequality $\binom{x}{t} \geq x^t/t^t - 1$ valid for all x.

Since $t > r-2$ by hypothesis, the function $x^{t/(r-2)}/t^t - 1$ is convex, and hence the expression above is minimized when each $\deg(e)$ is equal to the average value

$$\ell = \binom{r}{2} k n^{3/2} e(G)^{-1},$$

so we find

$$\mathcal{N}(K_{2,t}, G) \geq \frac{1}{2} e(G) \left(\frac{\ell^{t/(r-2)}}{t^t} - 1 \right).$$

Note that if $e(G) \leq \max(C, k^{1/2}) n^{3/2}$ then $\ell \geq 2t^t$ for k sufficiently large, meaning $\frac{\ell^{t/(r-2)}}{t^t} - 1 = \Omega\left(\ell^{t/(r-2)}\right)$. In total then, we find

$$\mathcal{N}(K_{2,t}, G) = \Omega(e(G)\ell^{t/(r-2)}) = \Omega\left(k^{\frac{t}{r-2}} n^{\frac{3t}{2(r-2)}} e(G)^{\frac{r-2-t}{r-2}} \right)$$

as desired. ∎

We now split up our analysis based off of the value of k. Recalling the value of C defined before (3), we first consider the case that $k \leq C^{\frac{(2t-1)(r-2)+t}{t}} n^{\frac{r-2}{2t}}$. If $e(G) \geq Cn^{3/2}$, then by (3) we have

$$\mathcal{N}(K_{2,t}, G) = \Omega(n^2) = \Omega\left(k^{\frac{t}{r-2}} n^{3/2} \right),$$

with the last step using our assumption on k, proving the result. If instead $e(G) \leq Cn^{3/2}$, then (4) gives a lower bound of $\Omega(k^{\frac{t}{r-2}} n^{\frac{3}{2}})$ as desired.

From now on we assume $k \geq C^{\frac{(2t-1)(r-2)+t}{t}} n^{\frac{r-2}{2t}}$. First consider the case

$$e(G) \geq k^{\frac{t}{(2t-1)(r-2)+t}} n^{\frac{(6t-4)(r-2)+3t}{(4t-2)(r-2)+2t}} \geq Cn^{3/2},$$

where this last inequality holds by our assumption on k. By (3) we find

$$\mathcal{N}(K_{2,t}, G) = \Omega\left(k^{\frac{2t^2}{(2t-1)(r-2)+t}} n^{\frac{(6t^2-4t)(r-2)+3t^2}{(2t-1)(r-2)+t}} \cdot n^{2-3t} \right) = \Omega\left(k^{\frac{2t^2}{(2t-1)(r-2)+t}} n^{\frac{(3t-2)(r-2)+2t}{(2t-1)(r-2)+t}} \right),$$

giving the desired bound. If instead $e(G) \le k^{\frac{t}{(2t-1)(r-2)+t}} n^{\frac{(6t-4)(r-2)+3t}{(4t-2)(r-2)+2t}} \le k^{1/2}n^{3/2}$, then by (4) we have

$$\begin{aligned}\mathcal{N}(K_{2,t},G) &= \Omega\left(\left[k^t n^{\frac{3t}{2}} \cdot k^{\frac{(r-2-t)t}{(2t-1)(r-2)+t}} n^{\frac{(r-2-t)((6t-4)(r-2)+3t)}{(4t-2)(r-2)+2t}}\right]^{1/(r-2)}\right)\\ &= \Omega\left(\left[k^{\frac{2t^2(r-2)}{(2t-1)(r-2)+t}} n^{\frac{((6t-4)(r-2)+4t)(r-2)}{(4t-2)(r-2)+2t}}\right]^{1/(r-2)}\right)\\ &= \Omega\left(k^{\frac{2t^2}{(2t-1)(r-2)+t}} n^{\frac{(3t-2)(r-2)+2t}{(2t-1)(r-2)+t}}\right),\end{aligned}$$

again giving the desired bound. □

3. Constructions

In this section, we construct graphs G with many K_r's but few copies of $K_{2,t}$. Motivated by Proposition 1.6 and Lemma 1.3 whose extremal constructions were unions of cliques, it is perhaps reasonable to consider G which are union of cliques. And indeed, our constructions will consist of two different G of this form: one coming from the union of random cliques, the other from the union of cliques which avoids certain structures.

3.1. Uniform Random Cliques.

In this subsection we prove Theorem 1.4 by constructing graphs G which contain many copies of K_r and few copies of F when F is 2-balanced. Intuitively, the graph G will be formed by taking the union of roughly u cliques of size m chosen uniformly at random.

More precisely, given a real number u and integers m, n, define the random clique graph $G_{u,m,n}$ as follows. Let $C_1, \ldots, C_{\binom{n}{m}}$ be an enumeration of the m-element subsets of $[n]$ and let $B_1, \ldots, B_{\binom{n}{m}}$ be i.i.d. Bernoulli random variables with probability of success $u\binom{n}{m}^{-1}$. Define $G_{u,m,n}$ to be the graph on $[n]$ with edge set

$$\bigcup_{i:B_i=1} \binom{C_i}{2}. \tag{5}$$

We will prove two properties about the random clique graph $G_{u,m,n}$: that it contains a relatively large number of K_r's, and that it contains few copies of other F (provided u and m are chosen appropriately). We begin with the clique estimate.

Lemma 3.1. *For all $r \ge 2$, there exists $\delta = \delta(r) > 0$ such that if $u \ge 1, m \ge 2r$ and $um^r \le n^r$, then*

$$\Pr(\mathcal{N}(K_r, G_{u,m,n}) > \delta u m^r) > \delta.$$

Proof. We record the following binomial tail bound that will be needed in the proof, see for example [11, Theorem 2.1]: if $X \sim \mathrm{Bin}(n,p)$, then

$$\Pr(X \geq np + t) \leq \exp[-t^2/(2np + 2t/3)]. \tag{6}$$

Returning to the main proof, let $X = \binom{m}{r}|\{i : B_i = 1\}|$ (which counts the number of r-cliques in G *with multiplicity*), so $X \sim \binom{m}{r}\mathrm{Bin}(\binom{n}{m}, u\binom{n}{m}^{-1})$. Using the general fact that $\Pr(\mathrm{Bin}(M,p) \geq Mp/2) \geq 1/2$ if $Mp \geq 1$ (which follows from e.g. [12]), together with $u \geq 1$ and $m \geq 2r$ gives

$$\Pr\left(X \geq \frac{um^r}{2^{r+1}r!}\right) \geq \Pr\left(X \geq \frac{1}{2}u\binom{m}{r}\right) \geq \frac{1}{2}. \tag{7}$$

Similarly, if $\lambda(A) = |\{i : B_i = 1, A \subset C_i\}|$ counts the multiplicity of a fixed r-clique A, then $\lambda \sim \mathrm{Bin}(\binom{n-r}{m-r}, u\binom{n}{m}^{-1})$. This random variable has expectation

$$u(m)_r/(n)_r \lesssim um^r/n^r =: \mu.$$

and this is at most 1 by hypothesis. Note that (6) says that for $t \geq 6$,

$$\Pr(\lambda(A) \geq 1 + t) \leq \exp[-t^2/(2 + 2t/3)] \leq \exp[-t].$$

Thus, if $Y_i = |\{A : 2^i \leq \lambda(A) < 2^{i+1}\}|$, we find that $\mathbb{E}[Y_i] \leq \binom{n}{r}\mu\exp[-2^{i-1}]$ for, say $i \geq 10$. By Markov's inequality, we have

$$\Pr\left(Y_i \geq \binom{n}{r}\mu\exp[-2^{i-2}]\right) < \exp[-2^{i-2}] \leq \exp[-i] \tag{8}$$

for large enough i. Letting L be a large constant to be determined, we have

$$X = \sum_A \lambda(A) < \sum_i 2^{i+1}Y_i \leq \sum_{i=1}^{L} 2^{i+1}Y_i + \sum_{i>L} 2^{i+1}Y_i.$$

By (8), the last sum is at most $2\binom{n}{r}\mu\sum_{i>L+1}(2/e)^i \leq 4(2e^{-1})^L um^r/r!$ with probability at least $1 - \sum_{i>L}\exp[-i]$. Thus, using (7) and taking L sufficiently large, we have with probability at least 1/3 that

$$\frac{um^r}{2^{r+2}r!} \leq \sum_{i=1}^{L} 2^{i+1}Y_i \leq 2^{L+1}\sum_i Y_i = 2^{L+1}\mathcal{N}(K_r, G),$$

which completes the proof. □

We next aim to prove the random clique construction contains few copies of F for certain ranges of parameters.

Lemma 3.2. *Let F be a 2-balanced graph with $e(F) \geq 2$. If u, m, n with $m \leq n$ are such that $um^2 < n^2$ and $u(m/n)^{2-\frac{v(F)-2}{e(F)-1}} > 1$, then*

$$\mathbb{E}[\mathcal{N}(F, G_{u,m,n})] = O\left((um^2n^{-2})^{e(F)}n^{v(F)}\right).$$

The condition $um^2 < n^2$ intuitively means the cliques of $G_{u,m,n}$ will be close to edge disjoint. The condition $u(m/n)^{2-\frac{v(F)-2}{e(F)-1}} > 1$ is best possible for the conclusion to hold, as otherwise the count $u\binom{m}{v(F)}$ coming from copies of F within a single clique will be larger.

We need a few technical definitions to prove Lemma 3.2. These definitions are based off the observation that for a given copy $\tilde{F}$ of F to be present in $G_{u,m,n}$, there must exist some (minimal) set of m-subsets $\{C_{i_1}, \ldots, C_{i_s}\}$ which cover the edges of $\tilde{F}$ and which are all present as cliques in $G_{u,m,n}$.

With this in mind, given a graph $\tilde{F} \subseteq K_n$, we say that a family $\mathcal{C}$ of m-subsets of K_n is an *$\tilde{F}$-covering* if $E(\tilde{F}) \subseteq \bigcup_{C \in \mathcal{C}} \binom{C}{2}$, if each $C \in \mathcal{C}$ contains at least one edge of $\tilde{F}$, and if $C \cap V(\tilde{F}) \neq C' \cap V(\tilde{F})$ for all distinct $C, C' \in \mathcal{C}$. Given $G_{u,n,n}$ and a family of m-subsets $\mathcal{C} = \{C_{i_1}, \ldots, C_{i_s}\}$ of K_n, we let $\mathcal{B}(\mathcal{C})$ denote the event that $B_{i_j} = 1$ for all $1 \leq j \leq s$ (that is, this is the event that each of the C_{i_j} appear in the union of (5)). Given a graph F, we let $Z(F, G_{u,m,n})$ denote the set of pairs $(\tilde{F}, \mathcal{C})$ such that $\tilde{F} \subseteq K_n$ and $\mathcal{C}$ is an $\tilde{F}$-covering with $\mathcal{B}(\mathcal{C})$ occurring.

The crucial observation is the following.

Lemma 3.3. *For all graphs F, we have $\mathcal{N}(F, G_{u,m,n}) \leq Z(F, G_{u,m,n})$.*

Proof. Observe that if $\tilde{F} \subseteq G_{u,m,n}$ is isomorphic to F, then there exists an $\tilde{F}$-covering $\mathcal{C}$ such that $\mathcal{B}(\mathcal{C})$ occurs; namely by taking a minimal set of m-subsets C_{i_j} with $B_{i_j} = 1$ that contain the set of edges of $\tilde{F}$ (which must exist if $\tilde{F} \subseteq G_{u,m,n}$). Thus for each subgraph $\tilde{F} \subseteq G_{u,m,n}$ counted by $\mathcal{N}(F, G_{u,m,n})$, there exists at least one pair $(\tilde{F}, \mathcal{C})$ counted by $Z(F, G_{u,m,n})$, proving the bound. □

With Lemma 3.3 in hand, we see that Lemma 3.2 will immediately be implied by the following result.

Lemma 3.4. *Let F be a 2-balanced graph with $e(F) \geq 2$. If u, m, n with $m \leq n$ are such that $um^2 < n^2$ and $u(m/n)^{2-\frac{v(F)-2}{e(F)-1}} > 1$, then*

$$\mathbb{E}[Z(F, G_{u,m,n})] = O\left((um^2 n^{-2})^{e(F)} n^{v(F)}\right).$$

Proof. Fix any $\tilde{F} \subseteq K_n$ isomorphic to F. Since there are at most $n^{v(F)}$ choices for $\tilde{F}$, we see that it suffices to prove that the expected number of $\tilde{F}$-coverings $\mathcal{C}$ for which $\mathcal{B}(\mathcal{C})$ occurs is at most $O((um^2 n^{-2})^{e(F)})$. For this we need a few more definitions.

Given a family $\mathcal{A}$ of subsets of $V(\tilde{F})$, we say that an $\tilde{F}$-covering $\mathcal{C}$ has *type* $\mathcal{A}$ if

$$\{C \cap V(\tilde{F}) : C \in \mathcal{C}\} = \mathcal{A}.$$

We say that a family $\mathcal{A}$ is *valid* if $\tilde{F} \subseteq \bigcup_{A \in \mathcal{A}} \binom{A}{2}$ and if each $A \in \mathcal{A}$ contains at least one edge of $\tilde{F}$. Observe that by definition, if $\mathcal{C}$ is an $\tilde{F}$-covering, then $\mathcal{C}$ is of type $\mathcal{A}$ for some valid $\mathcal{A}$ with $|\mathcal{A}| = |\mathcal{C}|$. Given a valid $\mathcal{A}$, we define

$$w(\mathcal{A}) = u^{|\mathcal{A}|}(m/n)^{\sum_{A\in\mathcal{A}}|A|}.$$

Claim 3.5. *For any family $\mathcal{A}$, let $\mathcal{T}(\mathcal{A})$ denote the number of $\tilde{F}$-coverings $\mathcal{C}$ of type $\mathcal{A}$ such that $\mathcal{B}(\mathcal{C})$ occurs. Then*

$$\mathbb{E}[\mathcal{T}(\mathcal{A})] \le w(\mathcal{A}). \tag{9}$$

Proof. Let $\mathcal{A} = \{A_1, \ldots, A_s\}$, and let $\mathcal{S}$ be the set of $\tilde{F}$-coverings $\mathcal{C}$ of type $\mathcal{A}$. Since $|\mathcal{S}|$ is at most the number of tuples $(C_{i_1}, \ldots, C_{i_s})$ such that each C_{i_j} is an m-subset containing A_j, we find that

$$|\mathcal{S}| \le \prod_{j=1}^{s} \binom{n-|A_j|}{m-|A_j|} = \prod_{j=1}^{s} \binom{n}{m}\binom{m}{|A_j|}\Big/\binom{n}{|A_j|} \le \prod_{j=1}^{s} \binom{n}{m}(m/n)^{|A_j|},$$

Now, any fixed set $\mathcal{C} = \{C_{i_1}, \ldots, C_{i_s}\}$ of distinct m-subsets has $\mathcal{B}(\mathcal{C})$ occurring with probability $u^s\binom{n}{m}^{-s}$, so by a union bound we see that

$$\mathbb{E}[\mathcal{T}(\mathcal{A})] \le u^s\binom{n}{m}^{-s}|\mathcal{S}| \le \prod_{j=1}^{s} u(m/n)^{|A_j|} = w(\mathcal{A}),$$

proving the claim. ☐

With this claim, we see that the expected number of $\tilde{F}$-coverings for which $\mathcal{B}(\mathcal{C})$ occurs is at most

$$\sum_{\mathcal{A}} \mathbb{E}[\mathcal{T}(\mathcal{A})] \le 2^{2^{v(F)}} \cdot \max_{\mathcal{A}} w(\mathcal{A}),$$

where the sum and the maximum range over all valid families $\mathcal{A}$. Thus to prove the result, it suffices to show

$$\max_{\mathcal{A}} w(\mathcal{A}) \le (um^2n^{-2})^{e(F)}, \tag{10}$$

where the maximum ranges over all valid families $\mathcal{A}$. We will call any $\mathcal{A}$ achieving the maximum in (10) a *maximizer*, and our goal will be to show that the only maximizers are those with $A \in E(F)$ for all $A \in \mathcal{A}$. We do this through the following two claims.

Claim 3.6. *Every maximizer $\mathcal{A}$ has $|A \cap B| \le 1$ for all distinct $A, B \in \mathcal{A}$.*

Proof. Let $\mathcal{A}$ be a maximizer and assume for contradiction that $|A \cap B| \ge 2$ for some distinct $A, B \in \mathcal{A}$. In this case, the set $\mathcal{A}' = (\mathcal{A} \setminus \{A, B\}) \cup \{A \cup B\}$ is a valid family which satisfies $|\mathcal{A}'| = |\mathcal{A}| - 1$ and

$$\sum_{D\in\mathcal{A}'} |D| = -|A \cap B| + \sum_{D\in\mathcal{A}} |D|,$$

since $|A \cup B| = |A| + |B| - |A \cap B|$. Thus

$$w(\mathcal{A}') = u^{-1}(m/n)^{-|A\cap B|}w(\mathcal{A}) > w(\mathcal{A}),$$

where the inequality used $|A \cap B| \geq 2$ together with $um^2 < n^2$ and $m \leq n$ applied $|A \cap B| - 2$ times. This contradicts $\mathcal{A}$ being a maximizer, giving the result. $\square$

Claim 3.7. *Every maximizer $\mathcal{A}$ has $A \in E(F)$ for all $A \in \mathcal{A}$.*

Proof. Assume for contradiction that $\mathcal{A}$ is a maximizer with $|A| > 2$ for some $A \in \mathcal{A}$. Let $\mathcal{E}$ denote the set of edges of $F' = F[A]$, noting that $\sum_{e \in \mathcal{E}} |e| = 2e(F')$ and $|A| = v(F')$. Let $\mathcal{A}' = (\mathcal{A} \setminus \{A\}) \cup \mathcal{E}$. Observe that $\mathcal{A}'$ is also a valid family with $|\mathcal{A}'| = |\mathcal{A}| + e(F') - 1$ (noting that we have $\mathcal{E} \cap \mathcal{A} = \emptyset$ by the previous claim since $A \in \mathcal{A}$ and $\mathcal{A}$ is a maximizer). Thus,

$$w(\mathcal{A}') = u^{e(F')-1}(m/n)^{2e(F')-v(F')}w(\mathcal{A}) > w(\mathcal{A}),$$

where the last step used

$$u(m/n)^{2-\frac{v(F')-2}{e(F')-1}} \geq u(m/n)^{2-\frac{v(F)-2}{e(F)-1}} > 1,$$

with the first inequality using that F is 2-balanced (and $m \leq n$) and the last inequality using the hypothesis of the lemma. This contradicts $\mathcal{A}$ being a maximizer, so we must have $|A| \leq 2$ for all $A \in \mathcal{A}$. Moreover, each $A \in \mathcal{A}$ must contain an edge of F by definition of $\mathcal{A}$ being valid, giving the result. $\square$

The claims above imply $\mathcal{A} = E(F)$ is the only maximizer, in which case $w(\mathcal{A}) = u^{e(F)}(m/n)^{2e(F)}$, giving (10) and hence the result. $\square$

In addition to proving Lemma 3.2, Lemma 3.4 can be used to derive our general lower bound on $\mathrm{ex}(n, K_r, F)$.

Proof of Theorem 1.5. Recall that we aim to prove that if F is a 2-balanced graph with $e(F) \geq 2$ and $2 \leq r < v(F)$ an integer, then

$$\mathrm{ex}(n, F, K_r) = \Omega(n^{2-\frac{v(F)-2}{e(F)-1}}).$$

Let $m = r$ and $u = 2(n/m)^{2-\frac{v(F)-2}{e(F)-1}} = \Omega(n^{2-\frac{v(F)-2}{e(F)-1}})$. Given a constant $\alpha > 0$, we let $G_\alpha = G_{\alpha u,m,n}$. Note that the conditions of Lemma 3.4 apply to G_1 (though not necessarily for G_α). We aim to show that for small α, we can alter G_α to make it F-free by removing a small portion of its K_r's.

Let $Z_\alpha = Z(F, G_\alpha)$. Since $v(F) > r = m$, any $\tilde{F}$-covering $\mathcal{C} = \{C_{i_1}, \ldots, C_{i_s}\}$ of some $\tilde{F} \cong F$ must have $s \geq 2$, so by definition of Z we find $\mathbb{E}[Z_\alpha] \leq \alpha^2 \mathbb{E}[Z_1]$. By Lemma 3.4 we see

$$\mathbb{E}[Z_1] = O((um^2n^{-2})^{e(F)}n^{v(F)}) = O(n^{2-\frac{v(F)-2}{e(F)-1}}) = O(u),$$

and thus

$$\mathbb{E}[Z_\alpha] = O(\alpha^2 u).$$

Recall that $B_1, B_2, \ldots$ are the Bernoulli random variables associated to G_α such that the ith clique C_i is included in the union (5) for G_α if $B_i = 1$. Let $Y(G_\alpha)$ denote the number of i such that $B_i = 1$, noting that $\mathcal{N}(K_r, G_\alpha) \geq Y(G_\alpha)$ and that $\mathbb{E}[Y(G_\alpha)] = \alpha u$. Thus we find $\mathbb{E}[Y(G_\alpha)] - \mathbb{E}[Z_\alpha] > \alpha u - O(\alpha^2 u)$, so for all α there is a realization G'_α of G_α with

$$Y(G'_\alpha) - Z(G'_\alpha) > \alpha u - O(\alpha^2 u).$$

Taking α sufficiently small and removing one clique from each pair $(\tilde{F}, \{C_{i_1}, \ldots\})$ in G'_α counted by $Z(F, G'_\alpha)$ (meaning we remove the clique from the union (5), which does not necessarily remove any edges from G'_α) results in an F-free graph G''_α with

$$\mathcal{N}(K_r, G''_\alpha) \geq Y(G''_\alpha) = \Omega(u) = \Omega(n^{2 - \frac{v(F)-2}{e(F)-1}}). \qquad \square$$

With this all established, we can now prove our main result for this subsection.

Proof of Theorem 1.4. Recall that we wish to prove that if F is a 2-balanced graph with $e(F) \geq 2$ and $2 \leq r < v(F)$ is an integer, then for all $1 \leq N \leq \binom{n}{r}$, there exists an n-vertex graph G with $\mathcal{N}(K_r, G) = \Omega(N)$ and with

$$\mathcal{N}(F, G) = O\left((Nn^{-r})^{\frac{e(F)(v(F)-2)}{(r-2)(e(F)-1)+v(F)-2}} n^{v(F)} \right).$$

We first consider some trivial cases. If F is the disjoint union of K_2's, then one can check that the bound above is achieved by taking G to be a clique on $N^{1/r}$ vertices. If F has an isolated vertex x, then $F' = F - x$ has at least 3 vertices (since $e(F) \geq 2$) and $\frac{e(F')-1}{v(F')-2} > \frac{e(F)-1}{v(F)-2}$, contradicting F being 2-balanced. Thus we can assume F has no isolated vertices and at least one component which is not a K_2, from which it follows that

$$2e(F) > v(F) \geq 3,$$

where here we used that no isolated vertices implies $2e(C) \geq v(C)$ for all components C, and the component which is not a K_2 gives a strict inequality.

The result is also trivial if $N \leq \mathrm{ex}(n, K_r, F)$, as in this case there exist F-free graphs with the desired number of copies. Thus by Theorem 1.5 we can assume

$$N \geq cn^{2 - \frac{v(F)-2}{e(F)-1}}$$

for some $c \leq 1$. Similarly the result is trivial if $N = \Omega(n^r)$ by taking $G = K_n$, so we can assume N is at most a small constant times n^r (with this constant depending on F, r, c).

With all these assumptions above in mind, we set

$$C = 2rc^{-\frac{e(F)-1}{(r-2)(e(F)-1)+v(F)-2}},$$

$$u = 2(Nn^{-r})^{\frac{v(F)-2e(F)}{(r-2)(e(F)-1)+v(F)-2}} \text{ and}$$

$$m = C(Nn^{-r})^{\frac{e(F)-1}{(r-2)(e(F)-1)+v(F)-2}} \cdot n.$$

The lower bound $N \geq cn^{2-\frac{v(F)-2}{e(F)-1}}$ immediately gives $m \geq 2r$. Observe that $v(F) - 2e(F) < 0$ and $(r-2)(e(F)-1) + v(F) - 2 > 0$ due to the bound $2e(F) > v(F) \geq 3$ above and $r \geq 2$, which in particular implies $C \geq 1$ since $c \leq 1$. With this and our assumption that N is at most a small constant times n^r, we observe that $m < n$, that $u \geq 2$,

$$1 \leq um^r = 2CN \leq n^r,$$
$$u(m/n)^{2-\frac{v(F)-2}{e(F)-1}} = 2C^{2-\frac{v(F)-2}{e(F)-1}} > 1,$$

and

$$um^2n^{-2} = 2C^2(Nn^{-r})^{\frac{v(F)-2}{(r-2)(e(F)-1)+v(F)-2}} < 1.$$

Now consider $G = G_{u,m,n}$ and let δ be the constant from Lemma 3.1. By applying Markov's inequality to Lemma 3.2, we find that

$$\mathcal{N}(F,G) = O\left((Nn^{-r})^{\frac{e(F)(v(F)-2)}{(r-2)(e(F)-1)+v(F)-2}} n^{v(F)}\right)$$

with probability at least $1 - \delta/2$. By Lemma 3.1, we see $\mathcal{N}(K_r, G) = \Omega(N)$ with probability at least δ. In particular, G satisfies the desired properties with positive probability, showing such a graph exists. □

3.2. **Cliques from Bipartite Graphs.** Throughout this subsection we work with bipartite graphs B with ordered bipartitions (U, V).

The intuition for our construction is as follows. We again consider a graph G formed by taking the union of roughly u cliques of size m for some parameters u, m. We can not have m larger than what it was in the proof for Theorem 1.4, as otherwise the number of copies of $K_{2,t}$ contained within the m cliques will be too large. Thus we must take m to be smaller and u to be larger. If we put the u cliques down uniformly at random, then G would contain too many copies of $K_{2,t}$ which have each edge contained in a distinct clique (intuitively because G behaves locally like $G_{n,p}$). Ideally then, we want to place our cliques down so that there exists no $K_{2,t}$ with each edge contained in a distinct clique. For this the following will be useful.

Definition 3.8. *Given a bipartite graph B with ordered bipartition (U, V), we define the clique graph $K(B)$ to be the graph with vertex set V such that $v, v' \in V$ are adjacent if and only if v, v' have a common neighbor in U. Equivalently, $K(B)$ is formed by taking the union[3] of the cliques $N_B(u)$ with $u \in U$.*

Unwinding the intuition from above; we want to find a bipartite B which avoid subdivisions of $K_{2,t}$ (as these correspond to edges belonging to distinct cliques in $K(B)$), or equivalently, to avoid having t paths of length 4 between any two vertices of V in B. And indeed, the following shows that this is essentially all we need.

[3]A helpful mnemonic is that U is the set of cliques that we Union over.

Lemma 3.9. *If B is a bipartite graph with bipartition (U, V) such that every vertex of U has degree at most d and such that there are at most $\ell \geq 2$ distinct paths of length at most 4 between any two vertices of V, then for all $t > \ell$ we have*

$$\mathcal{N}(K_{2,t}, K(B)) \leq \ell^{2t} d^{2+t} |U|.$$

Note that the number of $K_{2,t}$'s within any of the $N_B(u)$ cliques of $K(B)$ is at most $d^{2+t}|U|$, so the lemma says that this trivial count is essentially correct.

Proof. We first claim that if $v_1, v_2, w_1, \ldots, w_t$ form a $K_{2,t}$ in $K(B)$ with v_1, v_2 adjacent to all of the w_j vertices, then $v_1v_2 \in E(K(B))$ (i.e. they have a common neighbor in B). Indeed, for each edge $v_iw_j \in E(K(B))$ there exists a $u_{i,j} \in U$ which is adjacent to both of these vertices. If $u_{1,j} = u_{2,j}$ for any j then we are done, so we assume this is not the case. Thus $v_1u_{1,j}w_ju_{2,j}v_2$ is a path of length 4 from v_1 to v_2 in B, and each of these $t > \ell$ paths are distinct since the w_j vertices are distinct. This is impossible by our condition on B, so the claim follows.

For any pair of vertices $v_1, v_2 \in K(B)$, we claim that there are at most $\ell(d+1)$ vertices w which are adjacent to both v_1, v_2 in $K(B)$. Indeed, for any such vertex w there must exist two (possibly non-distinct) vertices $u_1^w, u_2^w \in U$ such that $w, v_i \in N_B(u_i^w)$. If $u_1^w = u_2^w = u$, then in particular u is a common neighbor of v_1, v_2 in B. By assumption there are at most ℓ such vertices u, and each of them have at most d neighbors w. Thus the number of common neighbors with $u_1^w = u_2^w$ is at most ℓd. On the other hand, for each distinct w with $u_1^w \neq u_2^w$, there exists a distinct path $v_1u_1^wwu_2^wv_2$ of length 4 from v_1 to v_2 in B. By assumption, at most ℓ such vertices can exist, giving the claim.

By these two claims, every $K_{2,t}$ in $K(B)$ can be formed by first choosing v_1, v_2 adjacent in $K(B)$ (i.e. with a common neighbor in U) and then choosing some t of the at most $\ell(d+1) \leq 2\ell d$ common neighbors they have in $K(B)$. In total the number of ways of doing this is at most

$$d^2|U| \cdot \binom{2\ell d}{t} \leq \ell^{2t} d^{2+t} |U|. \qquad \square$$

It remains to find graphs B avoiding the structures in Lemma 3.9 which are "dense" (so that $K(B)$ will have many K_r's). To this end, we define $\mathrm{ex}(m, n, \mathcal{P}_{\leq 4}^{\ell+1})$ to be the maximum number of edges that a bipartite graph B with bipartition (U, V) and $|U| = m, |V| = n$ can have if there are at most ℓ distinct paths of length at most 4 between any two vertices of V, and in this case we say B *avoids* $\mathcal{P}_{\leq 4}^{\ell+1}$. It is relatively easy to upper bound this extremal number by adapting an argumenent of Bukh and Conlon [3, Lemma 1.1].

Lemma 3.10.

$$\mathrm{ex}(m, n, \mathcal{P}_{\leq 4}^{\ell+1}) < 4\ell^{1/4} n^{3/4} m^{1/2} + 10m + 10n.$$

Proof. Let $B = (U, V, E)$ be a bipartite graph with $|U| = m$, $|V| = n$, and $|E| = e$ which avoids $\mathcal{P}_{\leq 4}^{\ell+1}$. If $e < 10m$ or $e < 10n$ then the result follows, so assume this is not the case.

Let $U' \subseteq U$ be the set of vertices with minimum degree at least $e/(2m)$. For $v \in V$, let $d'_v = |N(v) \cap U'|$, and note $\sum_{v \in V} d'_v \geq e/2$, as at most $e/2$ edges can be incident to vertices in $U - U'$. Let X denote the number of labelled copies of P_4 in $U' \cup V$ with endpoints in V. One can lower bound X by greedily embedding copies of P_4 starting with the middle vertex to get

$$X \geq \sum_{v \in V} \binom{d'_v}{2} (e/(2m) - 1)(e/(2m) - 2) \geq n \binom{e/(2n)}{2} \left(\frac{e}{4m}\right)^2 \geq \frac{e^4}{2^8 m^2 n},$$

where the second and third inequalities use that $e \geq 10m$ and $e \geq 10n$ respectively. On the other hand, there are $\binom{n}{2}$ choices for the endpoints of each path, and each pair may appear at most ℓ times as endpoints, so

$$X \leq \ell n^2/2.$$

Comparing the lower and upper bounds gives

$$e^4 \leq 2^7 \ell n^3 m^2,$$

which implies the lemma. ◻

The next result shows that whenever $\mathrm{ex}(m, n, \mathcal{P}_{\leq 4}^{\ell+1})$ is close to the upper bound of Lemma 3.10 we can find a graph with many K_r's and few $K_{2,t}$'s.

Lemma 3.11. *Fix $r \geq 3$, $\ell \geq 2$ and $t > \ell, r-2$. Assume there exists a $c > 0$ such that for all ε with $0 \leq \varepsilon \leq \frac{r-2}{2t}$ we have $\mathrm{ex}(n^{3/2-2\varepsilon}, n, \mathcal{P}_{\leq 4}^{\ell+1}) > 2cn^{3/2-\varepsilon}$. Then for all $0 \leq \varepsilon \leq \frac{r-2}{2t}$ there exists an n-vertex graph G with $\Omega_c(n^{3/2+(r-2)\varepsilon})$ r-cliques and $O_c(n^{3/2+t\varepsilon})$ $K_{2,t}$'s.*

Proof. We may assume n is sufficiently large in terms of c, as otherwise we can take $G = K_n$ to give the result. Let $B = (U, V, E)$ be a bipartite graph with $|U| = m = n^{3/2-2\varepsilon}$ and $|V| = n$ showing $\mathrm{ex}(n^{3/2-2\varepsilon}, n, \mathcal{P}_{\leq 4}^{\ell+1}) > 2cn^{3/2-\varepsilon}$. Let $W \subseteq U$ be the set of vertices w with $\deg_B(w) > D\ell n^{\varepsilon}$ for some large D to be determined.

Claim 3.12. $e(B[U \setminus W, V]) > e(B)/2$.

Proof. Consider the bipartite graph B_W induced by $W \cup V$. We are done if $e(B_W) \leq cn^{3/2-\varepsilon}$. Suppose this is not the case, and set $s = |W|$. Using n sufficiently large in terms of c and that $\varepsilon < 1/2$, we find

$$e(B_W) > cn^{3/2-\varepsilon} > 100s + 100n,$$

so Lemma 3.10 gives

$$s^{1/2} > \frac{e(B_W)}{5\ell^{1/4} n^{3/4}} > \frac{cn^{3/2-\varepsilon}}{5n^{3/4}\ell^{1/4}},$$

i.e. $s > c^2 n^{3/2-2\varepsilon}/(25\ell^{1/2})$. Therefore,

$$e(B_W) > s \cdot D\ell n^{\varepsilon} > \frac{n^{3/2-2\varepsilon}}{25\ell^{1/2}} \cdot D\ell n^{\varepsilon} = \ell^{1/2} n^{3/2-\varepsilon}(D/25),$$

which is more than $4\ell^{1/4}n^{3/4}m^{1/2} + 10(m+n)$ for large enough D, contradicting Lemma 3.10. $\square$

Let $U' = U \setminus W$, $B' = B[U' \cup V]$, and $|U'| = m'$. By the claim, $e(B') > e(B)/2$. Let $G = K(B')$ (recalling the definition of $K(B')$ above Lemma 3.9). We claim G satisfies the conclusion of the lemma.

Because B' avoids $\mathcal{P}_{\le 4}^{\ell+1}$, every r-set in V is contained in at most ℓ neighborhoods, so the number of K_r's in G is at least

$$\sum_{u \in U'} \binom{\deg_{B'}(u)}{r} \ell^{-1} \ge m' \binom{e(B)/(2m')}{r} \ell^{-1} \ge \frac{c^r n^{3/2+(r-2)\varepsilon}}{2^r r! \ell}$$

for sufficiently large n, with this last step using $m' \le m = n^{3/2-2\varepsilon}$. Note that Lemma 3.9 with the assumption that B' avoids $\mathcal{P}_{\le 4}^{\ell+1}$ implies G contains at most $O((2\ell)^{2t}(D\ell n^{\varepsilon})^{2+t} n^{3/2-2\varepsilon}) = O(n^{3/2+t\varepsilon})$ copies of $K_{2,t}$. $\square$

With this we can prove our upper bound result for $K_{2,t}$.

Proof of Theorem 1.7 Upper Bound. Recall that we wish to prove if $t \ge r-1 \ge 2$ and either $k \ge n^{\frac{r-2}{2t}}$ or if there exists a bipartite graph B on $U \cup V$ such that $|U| = k^{-2/(r-2)} n^{3/2}$, $|V| = n$, $e(B) = \Omega(k^{-1/(r-2)} n^{3/2})$, and such that every pair of vertices in V has less than t paths of length at most 4 between them; then there exists an n vertex graph G with $\mathcal{N}(K_r, G) = \Omega(kn^{3/2})$ and with

$$\mathcal{N}(K_{2,t}, G) = O(\min\{k^{\frac{t}{r-2}} n^{3/2+o(1)}, k^{\frac{2t^2}{(2t-1)(r-2)+t}} n^{\frac{(3t-2)(r-2)+2t}{(2t-1)(r-2)+t}}\}).$$

If $k \ge n^{\frac{r-2}{2t}}$ then the second term achieves the minimum, and in this case the general bound of Theorem 1.4 gives the result.

It remains to show that if $k \le n^{\frac{r-2}{2t}}$ and if bipartite graphs B as above exists, then there exist graphs with $\mathcal{N}(K_r, G) = \Omega(kn^{3/2})$ and $\mathcal{N}(K_{2,t}, G) = O(k^{\frac{t}{r-2}} n^{3/2})$. And indeed, this follows from the previous lemma. $\square$

As an aside, we note that the constructions considered here are similar to the constructions used in the recent breakthrough of Mattheus and Verstraëte [13] for off-diagonal Ramsey numbers. Indeed, as is made more explicit in [5], the construction of [13] comes from taking a bipartite graph B which avoids certain subgraphs, forming the clique graph $K(B)$, and then taking a random bipartition of each of its cliques.

4. Conclusion

In this paper we studied how many copies of a graph F another graph G is guaranteed to have if G contains a given number of K_r's. While our general result Theorem 1.4 gives effective bounds on many F, it would be desirable to get a better understanding of the problem for specific choices of F. Below we outline two such directions for problems, and for this we recall the notation $\mathcal{N}(F, G)$ which denotes the number of copies of F in G.

Tight Bounds for $K_{2,t}$. Theorem 1.7 solved the clique supersaturation problem when $F = K_{2,t}$ and G has $kn^{3/2}$ copies of K_r with $k \geq n^{(r-2)/2t}$. We believe Theorem 1.7 should give tight bounds even when $k < n^{(r-2)/2t}$, and in particular we conjecture the following.

Conjecture 4.1. *There exists t_0 such that for $t \geq t_0$ and $1 \leq k \leq n^{1/2t}$, there exists an n-vertex graph G with $\Omega(kn^{3/2})$ triangles and with*

$$\mathcal{N}(K_{2,t}, G) \leq k^t n^{3/2+o(1)}.$$

Note that the second half of Theorem 1.7 shows such G exists provided there exist bipartite graphs B with parts of sizes $k^{-2}n^{3/2}$ and n which have $e(B) = \Omega(k^{-1}n^{3/2})$, and which have fewer than t paths of length 4 between any two vertices in the part of size n. For example, if $t = 2$ and $k = n^{1/4}$, then this is equivalent to finding an n-vertex bipartite graph with $\Omega(n^{5/4})$ edges and which is C_4 and C_8-free, which is a notoriously open and difficult problem. More generally, the $t = 2$ case requires finding C_8-free unbalanced bipartite graphs of the largest possible density; see for example [16] for more on this.

The above suggests using Theorem 1.7 to solve Conjecture 4.1 with $t_0 = 2$ is quite difficult, but there is some hope that one can do this for t_0 sufficiently large. In particular, it is known at $k = n^{1/4}$ that (explicit) bipartite graphs B of this form exist for $t \geq 3$ due to an algebraic construction of Verstraëte and Williford [22], and it is plausible that one could modify their argument to construct B for additional values of k.

Another potential avenue is through random polynomial graphs. This approach was used by Conlon [4] to show that at $k = n^{1/4}$, there is a t_0 such that bipartite graphs B of this form exist for $t \geq t_0$. By adapting his argument, it is possible to show that for any rational $0 \leq q \leq 1/4$, there exists t_q such that at $k = n^q$ bipartite graphs B of this form exist for $t \geq t_q$.

While the $k = n^q$ result above might be of independent interest, it does not suffice for our purposes. Indeed, the t_q we obtain will typically have $q > 1/2t_q$, and hence the range $k = n^q$ falls outside the scope of Conjecture 4.1. Still, it might be possible to prove Conjecture 4.1 with a more sophisticated approach using random polynomials.

General Graphs. We believe the bounds of Theorem 1.7 for $K_{2,t}$ should extend to all theta graphs F with a similar argument. However, the situation for general $K_{s,t}$ is entirely unclear, and we leave this as an open problem.

Problem 4.2. *Solve the K_r clique supersaturation problem for $K_{s,t}$ with $r, s, t \geq 3$.*

One can check that Theorem 1.4 gives better bounds for $K_{s,t}$ compared to the bound (2) coming from $G_{n,p}$ if and only if

$$r > \frac{2st - s - t}{s + t - 2} = s + \frac{(s-1)(t-s)}{s+t-2} = 2s - 1 - \frac{2(s-1)^2}{s+t-2}.$$

In particular, Theorem 1.4 never gives effective bounds when $r \leq s$, and for $r \geq 2s - 1$ it always gives a non-trivial bound. Theorem 1.4 used cliques placed uniformly at random, and one might hope that by placing cliques in a more careful way (say with the aid of a bipartite graph B which avoids certain structures), one could obtain bounds better than (2) for smaller values of r. Unfortunately, the following shows that this is essentially impossible.

Lemma 4.3. *Let $2 \leq s \leq t$ be integers and $r \leq \frac{2st-s-t}{s+t-2}$. There exists a constant $k_0 = k_0(r, s, t)$ such that if G is an n-vertex graph which is the union of u cliques of size m with $um^r = kn^{r-\binom{r}{2}/s}$ where $k \geq k_0$, and with every edge contained in at most $O(1)$ of the u cliques, then*

$$\mathcal{N}(K_{s,t}, G) = \Omega\left(k^{st/\binom{r}{2}} n^s\right).$$

We note that the quantity um^r is roughly the number of copies of K_r within one of the u cliques making up G.

Proof Sketch. If $m \ll k^{\frac{1}{r(r-1)}} n^{\frac{2s-r-1}{2s}}$, then a small computation together with the fact that each edge is in at most $O(1)$ cliques shows $e(G) \gg k^{1/\binom{r}{2}} n^{2-1/s}$, from which the result follows by Proposition 1.2. Otherwise, counting copies of $K_{s,t}$ within each of the u cliques gives

$$\mathcal{N}(K_{s,t}, G) = \Omega(um^{s+t}) = \Omega(kn^{r-\binom{r}{2}/s} \cdot m^{s+t-r}) = \Omega(k^{st/\binom{r}{2}} n^s),$$

where this last step implicitly uses the upper bound on r. □

Lemma 4.3 shows that any non-trivial construction for Problem 4.2 when $r \leq s$ must look substantially different from all the constructions used throughout this paper. We feel like such constructions should not exist, and as such we conjecture the following.

Conjecture 4.4. *If $2 \leq r \leq s \leq t$, then there exists a constant k_0 such that if G is an n-vertex graph with $\mathcal{N}(K_r, G) = kn^{r-\binom{r}{2}/s}$ and $k \geq k_0$, then*

$$\mathcal{N}(K_{s,t}, G) \geq k^{st/\binom{r}{2}} n^{s-o(1)}.$$

That is, if $r \leq s$, then we predict every graph G with a given number of K_r's contains at least as many $K_{s,t}$'s as the random graph with the same number of K_r's. It is perhaps natural to extend this conjecture to all $r \leq \frac{2s-s-t}{s+t-2}$ since this is the full range for Lemma 4.3, but we find this quantity too strange to make any statements with confidence.

Note that Proposition 1.2 implies Conjecture 4.4 for $r = 2$. Thus the next open case is $r = s = 3$, which we restate below.

Conjecture 4.5. *For all $t \geq 3$ there exists a constant k_0 such that if G is an n-vertex graph with kn^2 triangles and $k \geq k_0$, then*

$$\mathcal{N}(K_{3,t}, G) \geq k^t n^{3-o(1)}.$$

We believe we can adapt the argument of Theorem 1.7 for this problem to prove a lower bound of roughly $n^{7/3}$ when k is a large constant, but it seems like new ideas are needed to improve upon this.

Acknowledgements

The third author was supported by NSF grants DMS-2237138 and CCF-1814409, as well as a Sloan Research Fellowship. The fourth author was supported by an NSF postdoctoral research fellowship under grant DMS-2202730.

References

1. Noga Alon and Clara Shikhelman, *Many T copies in H-free graphs*, Journal of Combinatorial Theory, Series B **121** (2016), 146–172. 2, 4
2. József Balogh, Robert Morris, and Wojciech Samotij, *Independent sets in hypergraphs*, Journal of the American Mathematical Society **28** (2015), no. 3, 669–709. 1
3. Boris Bukh and David Conlon, *Rational exponents in extremal graph theory*, Journal of the European Mathematical Society **20** (2018), no. 7, 1747–1757. 16
4. David Conlon, *Graphs with few paths of prescribed length between any two vertices*, Bulletin of the London Mathematical Society **51** (2019), no. 6, 1015–1021. 19
5. David Conlon, Sam Mattheus, Dhruv Mubayi, and Jacques Verstraëte, *Ramsey numbers and the Zarankiewicz problem*, arXiv preprint arXiv:2307.08694 (2023). 18
6. Jonathan Cutler, JD Nir, and AJ Radcliffe, *Supersaturation for subgraph counts*, Graphs and Combinatorics **38** (2022), no. 3, 65. 2, 5
7. P Erdos and Miklós Simonovits, *Cube-supersaturated graphs and related problems*, Progress in graph theory (Waterloo, Ont., 1982) (1984), 203–218. 1, 2
8. Paul Erdős and Miklós Simonovits, *Supersaturated graphs and hypergraphs*, Combinatorica **3** (1983), 181–192. 2

9. Dániel Gerbner, Zoltán Lóránt Nagy, and Máté Vizer, *Unified approach to the generalized Turán problem and supersaturation*, Discrete Mathematics **345** (2022), no. 3, 112743. 2
10. Anastasia Halfpap and Cory Palmer, *On supersaturation and stability for generalized Turán problems*, Journal of Graph Theory **97** (2021), no. 2, 232–240. 2, 4
11. Svante Janson, Tomasz Łuczak, and Andrzej Rucinski, *Random graphs*, Wiley Publishing, 2000. 10
12. Nick Lord, *Binomial averages when the mean is an integer*, The Mathematical Gazette **94** (2010), no. 530, 331–332. 10
13. Sam Mattheus and Jacques Verstraete, *The asymptotics of* $r(4,t)$, arXiv preprint arXiv:2306.04007 (2023). 18
14. Gwen McKinley and Sam Spiro, *The random Turán problem for theta graphs*, arXiv preprint arXiv:2305.16550 (2023). 6
15. Robert Morris and David Saxton, *The number of* $C_{2\ell}$*-free graphs*, Advances in Mathematics **298** (2016), 534–580. 2
16. Assaf Naor and Jacques Verstraëte, *A note on bipartite graphs without* $2k$*-cycles*, Combinatorics, Probability and Computing **14** (2005), no. 5-6, 845–849. 19
17. Alexander A Razborov, *On the minimal density of triangles in graphs*, Combinatorics, Probability and Computing **17** (2008), no. 4, 603–618. 2
18. Christian Reiher, *The clique density theorem*, Annals of Mathematics (2016), 683–707. 2
19. David Saxton and Andrew Thomason, *Hypergraph containers*, Inventiones mathematicae **201** (2015), no. 3, 925–992. 1
20. Alexander F Sidorenko, *Inequalities for functionals generated by bipartite graphs*, Diskretnaya Matematika **3** (1991), no. 3, 50–65. 2
21. Sam Spiro, *Random polynomial graphs for random Turán problems*, arXiv preprint arXiv:2212.08050 (2022). 4
22. Jacques Verstraëte and Jason Williford, *Graphs without theta subgraphs*, Journal of Combinatorial Theory, Series B **134** (2019), 76–87. 19

Department of Mathematics, Rutgers University, Piscataway, NJ 08854, USA
Email address: qcd2@math.rutgers.edu

Department of Mathematics, Rutgers University, Piscataway, NJ 08854, USA
Email address: bg570@connect.rutgers.edu

Department of Mathematics, Rutgers University, Piscataway, NJ 08854, USA
Email address: narayanan@math.rutgers.edu

Department of Mathematics, Rutgers University, Piscataway, NJ 08854, USA
Email address: sas703@scarletmail.rutgers.edu