INDUCED SUBGRAPH DENSITY. VII. THE FIVE-VERTEX PATH

TUNG NGUYEN, ALEX SCOTT, AND PAUL SEYMOUR

ABSTRACT. We prove the Erdős-Hajnal conjecture for the five-vertex path P_5; that is, there exists $c > 0$ such that every n-vertex graph with no induced P_5 has a clique or stable set of size at least n^c. This completes the verification of the Erdős-Hajnal property of all five-vertex graphs. Indeed, we show a stronger statement, that P_5 satisfies the polynomial version of a theorem of Rödl. To achieve this, we combine simple probabilistic and structural ideas with the iterative sparsification framework introduced in the series.

1. INTRODUCTION

All graphs in this paper are finite and with no loops or parallel edges. For graphs G, H, a *copy* of H in G is an injective map $\varphi\colon V(H) \to V(G)$ satisfying $uv \in E(H)$ if and only if $\varphi(u)\varphi(v) \in E(G)$, for all $u, v \in V(H)$; and G is H*-free* if there is no copy of H in G. Let $\overline{G}$ denote the complement of G. We say that H has the *Erdős-Hajnal property* if there exists $c > 0$ such that every n-vertex H-free graph has a clique or stable set of size at least n^c. Thus, H has the Erdős-Hajnal property if and only if $\overline{H}$ does. A conjecture of Erdős and Hajnal [10, 11] says:

Conjecture 1.1. *Every graph H has the Erdős-Hajnal property.*

Over 25 years ago, Gyárfás [13] suggested proving Conjecture 1.1 for every five-vertex graph H; and since then, this problem has been reiterated in [5, 9, 20]. By a theorem of Alon, Pach, and Solymosi [1] that the class of graphs with the Erdős-Hajnal property is closed under vertex-substitution, the problem reduces to showing Conjecture 1.1 for three graphs with five vertices: the bull (obtained from the four-vertex path by adding a new vertex adjacent to the two middle vertices), the five-cycle C_5, and the five-vertex path P_5 (or equivalently, the *house* $\overline{P_5}$). Chudnovsky and Safra [6] showed the Erdős-Hajnal property for the bull (see [8, 15] for two new proofs using different methods); and more recently Chudnovsky, Scott, Seymour, and Spirkl [8] showed it for C_5, but the P_5 case has remained open. There has been a sequence of successively stronger partial results for P_5. Let G be P_5-free, with n vertices, and let m be the size of its largest clique or stable set. Then there exists $c > 0$ such that:

- $m \geq 2^{c(\log n)^{1/2}}$, by a general theorem of Erdős and Hajnal [11] (the result is not special to P_5; the same holds with any excluded induced subgraph);
- $m \geq 2^{c(\log n \log\log n)^{1/2}}$, and again this is true with any excluded induced subgraph [4];
- $m \geq 2^{c(\log n)^{2/3}}$, by a result of P. Blanco and M. Bucić [2];
- $m \geq 2^{(\log n)^{1-o(1)}}$, and this is true with P_5 replaced by any path [16].

But finally we can prove the full conjecture for P_5:

Theorem 1.2. *P_5 has the Erdős-Hajnal property.*

As in some previous papers of this series, our main result is more general and says that P_5 actually satisfies the polynomial form of a theorem of Rödl, but to discuss this we need some further definitions and results. For a graph G, $|G|$ denotes the number of vertices of G. For $\varepsilon > 0$, we say that G is ε*-sparse* if its maximum degree is at most $\varepsilon|G|$, and ε*-restricted* if one of $G, \overline{G}$ is ε-sparse. We also say $S \subseteq V(G)$ is ε*-restricted* if $G[S]$ is ε*-restricted*. Rödl's theorem [19] then states that:

Date: October 25, 2023; revised January 5, 2024.
The first and the third authors were partially supported by AFOSR grant FA9550-22-1-0234 and NSF grant DMS-2154169. The second author was supported by EPSRC grant EP/X013642/1.

Theorem 1.3. *For every $\varepsilon \in (0, \frac{1}{2})$ and every graph H, there exists $\delta > 0$ such that every H-free graph G has an ε-restricted induced subgraph with at least $\delta|G|$ vertices.*

The original proof of Rödl used the regularity lemma and gave tower-type dependence of δ on ε. Fox and Sudakov [12] provided a proof that produces the bound $\delta = 2^{-d(\log \frac{1}{\varepsilon})^2}$ for Theorem 1.3 (here $d > 0$ is some constant depending on H only); and currently the best known bound for this theorem is $\delta = 2^{-d(\log \frac{1}{\varepsilon})^2/\log\log\frac{1}{\varepsilon}}$, obtained in [4]. Fox and Sudakov [12] also conjectured that in Theorem 1.3, δ can be taken to be a power of ε. More exactly, we say that a graph H has the *polynomial Rödl property* if there exists $d > 0$ such that for every $\varepsilon \in (0, \frac{1}{2})$, every H-free graph G has an ε-restricted induced subgraph with at least $\varepsilon^d|G|$ vertices. It is not hard to check that H has the Erdős-Hajnal property if it has the polynomial Rödl property. The Fox–Sudakov conjecture is then the following:

Conjecture 1.4. *Every graph H has the polynomial Rödl property.*

It is unknown whether Conjecture 1.1 implies Conjecture 1.4. As mentioned above, the main result of this paper says that Conjecture 1.4 holds for $H = P_5$, which contains Theorem 1.2:

Theorem 1.5. *P_5 has the polynomial Rödl property.*

2. Blockades, and some proof ideas

If $k \geq 0$ is an integer, we define $[k] := \{1, 2, \dots, k\}$. If G is a graph, and $A, B \subseteq V(G)$ are disjoint, we say that (A, B) is *anticomplete* in G (or A is *anticomplete* to B in G) if there is no edge between A, B; and we say that (A, B) is *complete* in G (or A is *complete* to B in G) if (A, B) is anticomplete in $\overline{G}$. Also, a vertex $v \in V(G) \setminus A$ is *mixed* on A if it has a neighbour and a nonneighbour in A. A *blockade* in G is a sequence $\mathcal{B} = (B_1, \dots, B_k)$ of disjoint (and possibly empty) subsets of $V(G)$; its *length* is k and its *width* is $\min_{i\in[k]}|B_i|$. For $\ell, w \geq 0$, $\mathcal{B}$ is an (ℓ, w)-*blockade* if it has length at least ℓ and width at least w. We say that $\mathcal{B}$ is *pure* if (B_i, B_j) is complete or anticomplete for all distinct $i, j \in [k]$, *complete* in G if (B_i, B_j) is complete in G for all distinct $i, j \in [k]$, and *anticomplete* in G if (B_i, B_j) is anticomplete in G for all distinct $i, j \in [k]$.

For $x > 0$ and disjoint $A, B \subseteq V(G)$, we say that B is *x-sparse* to A in G if every vertex in B has at most $x|A|$ neighbours in A. For $A, B \neq \emptyset$, the *edge density* between A, B in G is the number of edges between A, B in G divided by $|A||B|$; and we say that (A, B) is *weakly x-sparse* in G if the edge density between A, B in G is at most x. A blockade $\mathcal{B} = (B_1, \dots, B_k)$ in G is *x-sparse* in G if B_j is x-sparse to B_i in G for all $i, j \in [k]$ with $i < j$.

The proof of Theorem 1.5 is in two parts; first we prove a lemma, and then we use the lemma to prove the main theorem. The lemma is of some interest in its own right, so let us discuss it here before we go on. The following was proved in [7]:

Theorem 2.1. *If H is a forest, there exists $c > 0$ such that if G is an H-free, c-sparse graph with $|G| \geq 2$, then there exist disjoint $A, B \subseteq V(G)$ with $|A|, |B| \geq c|G|$ such that A, B are anticomplete. If H is not a forest, there is no such c.*

It is straightforward, using Theorem 1.3 and Theorem 2.1 (applied in the complement), to deduce the following, a version of Theorem 2.1 without the sparsity hypothesis:

Theorem 2.2. *If H is a forest, then for all d with $0 < d \leq 1/2$ there exists $c > 0$ such that if G is an $\overline{H}$-free graph with $|G| \geq 2$, then there exist disjoint $A, B \subseteq V(G)$ with $|A|, |B| \geq c|G|$ such that either A, B are complete, or A, B are weakly d-sparse to each other. If H is not a forest, then for all d with $0 < d \leq 1/2$, there is no such c.*

Let H be a graph: let us say H is *nice* (for lack of a better word) if there exist $a, b > 0$ such that for every $\overline{H}$-free graph G and every ε with $0 < \varepsilon \leq 1/2$, there is an $(\varepsilon^{-1}, \lfloor\varepsilon^a|G|\rfloor)$-blockade $(B_1, \dots, B_\ell)$ in

G, such that for all distinct $i, j \in [\ell]$, (B_i, B_j) is either complete or weakly ε^b-sparse in G. A key lemma of this paper is that P_5 is nice; but before we go on to its proof, let us consider niceness in general. Which graphs are nice? By taking $\varepsilon = 1/2$, Theorem 2.2 implies that every nice graph is a forest; but perhaps all forests are nice. We have not been able to decide that, but we would like to make three points:

- Perhaps niceness is a halfway point towards proving Theorem 1.5 for forests, because every forest H with the polynomial Rödl property is nice. To see this, suppose H has the polynomial Rödl property; then we have some $d > 0$ such that for every $\varepsilon \in (0, \frac{1}{2})$ and every $\overline{H}$-free graph G, there exists an ε^{2d}-restricted $S \subseteq V(G)$ with $|S| \geq \varepsilon^{2d^2}|G|$. If $G[S]$ is ε^{2d}-sparse then it is easy to get a weakly ε^d-sparse $(\varepsilon^{-1}, \lfloor \varepsilon^{10d^2}|G| \rfloor)$-blockade in $G[S]$. If $\overline{G}[S]$ is ε^{2d}-sparse then we can increase d if necessary and iterate Theorem 2.1 to get a complete $(\varepsilon^{-1}, \lfloor \varepsilon^{10d^2}|G| \rfloor)$-blockade in $G[S]$ (we omit the details).
- The niceness of a forest H by itself does not seem enough to prove the Erdős-Hajnal (or polynomial Rödl) property for H directly. Niceness gives us a blockade in which all the pairs are sparse or complete. We can make a graph with a vertex for each block, with an edge for each complete pair of blocks, and we would know that this "pattern graph" is $\overline{H}$-free; but we know nothing else about it. If we apply induction to it, we prove just the "near-polynomial Rödl" property of H (that is, δ can be taken as $2^{-(\log \frac{1}{\varepsilon})^{1+o(1)}}$ in Theorem 1.3), which implies the "near-Erdős-Hajnal" property ($2^{(\log n)^{1-o(1)}}$ in place of n^c).
- Let us say H is *strongly nice* if it satisfies the niceness condition with "weakly ε^d-sparse" changed to "anticomplete". This is too strong to be interesting, because when ε is a constant that would mean every H-free graph contains a linear pure pair, which is not true unless $|H| \leq 4$ (see [7]). In the other direction, let us say H is *weakly nice* if it satisfies the niceness condition with "complete" changed to "weakly ε^b-sparse in $\overline{G}$". This *is* still an interesting property. We don't know that being weakly nice is equivalent to either the polynomial Rödl property or the near-polynomial Rödl property, but it is somewhere between them: every graph H with the polynomial Rödl property is weakly nice (not just forests); and every weakly nice graph has the near-polynomial Rödl property. We have nothing else to say about it in this paper.

Returning to P_5: the first half of the proof of Theorem 1.5 is to prove a lemma that says that P_5 is nice. We will in fact prove:

Lemma 2.3. *There exists $d \geq 40$ for which the following holds. Let $\varepsilon \in (0, \frac{1}{2})$, and let G be a $\overline{P_5}$-free graph. Then there is an $(\varepsilon^{-1}, \lfloor \varepsilon^{10d^2}|G| \rfloor)$-blockade $(B_1, \ldots, B_\ell)$ in G such that for all distinct $i, j \in [\ell]$, (B_i, B_j) is either complete or weakly ε^d-sparse in G.*

Both the proof of Lemma 2.3, and its application to prove Theorem 1.5, use what we call "iterative sparsification", which can be summarized as follows. We start with a graph G, that is H-free for some fixed H, and we are given x with $0 < x \leq 1/2$. In order to prove the polynomial Rödl property for H, we need to show that G contains an x-restricted induced subgraph with at least $\text{poly}(x)|G|$ vertices, where the polynomial depends on H but not in G. We can assume that x is at most any positive constant that is convenient. For the method to work, there needs to be a lemma that says that for any value of $y \geq x$, if we have an induced subgraph F of G that is y-restricted, then either

- there is an induced subgraph F' of F with $|F'| \geq \text{poly}(y')|F|$ that is y'-restricted, where $y^D \leq y' \leq y^d$ for some fixed $D \geq d > 1$; or
- some other good thing happens.

To use the lemma, we choose a subgraph F of linear size that is y-restricted for y some small constant (we can do this, for instance by applying Rödl's theorem). Now we apply the lemma to F, and, if the "other good thing" does not happen, we find F' and y'. Repeat, and if the "other good thing" never happens, we recursively generate a nested sequence of induced subgraphs that are y-restricted for smaller and smaller values of y, and with size at least some polynomial in (the current value of) y times $|G|$. If y becomes smaller than the target x, then the first time it does so, it is not *much* smaller than x (because it is not much smaller than the previous value of y), and then we have the x-restricted induced subgraph that we wanted. So we can assume that at some stage the "other good thing" happens.

3. Some preliminaries

In this section we gather several basic results. A graph G is *anticonnected* if $\overline{G}$ is connected; and an induced subgraph F of G is an *anticonnected component* of G if $\overline{F}$ is a connected component of $\overline{G}$. The following fact says that graphs without large anticonnected components contain long and wide complete blockades.

Lemma 3.1. *Let $k \geq 2$ be an integer, and let G be a graph whose anticonnected components have size less than $|G|/k$. Then there is a complete $(k, |G|/k^2)$-blockade in G.*

Proof. By the hypothesis, there exists $n \geq 0$ minimal for which there is a partition $S_0 \cup S_1 \cup \cdots \cup S_n = V(G)$ such that $(S_0, S_1, \ldots, S_n)$ is a complete blockade in G with $|S_i| < |G|/k$ for all $i \in [n] \cup \{0\}$. In particular $n + 1 > k$ and so $n \geq k$. We may assume $|S_0| \leq |S_1| \leq \ldots \leq |S_n|$. If there exists $i \in [n]$ with $|S_i| < |G|/(2k)$, then $|S_{i-1} \cup S_i| < |G|/k$ and so $(S_0, \ldots, S_{i-2}, S_{i-1} \cup S_i, S_{i+1}, \ldots, S_n)$ would contradict the minimality of n. Hence $|S_i| \geq |G|/(2k) \geq |G|/k^2$ for all $i \in [n]$; and so $(S_1, \ldots, S_n)$ is a complete $(k, |G|/k^2)$-blockade in G. This proves Lemma 3.1. ■

The following simple probabilistic lemma will be useful in Section 4.

Lemma 3.2. *Let $x \in (0, \frac{1}{2})$. Let G be a bipartite graph with bipartition (A, B) where every vertex in B has at least $x|A|$ neighbours in A. Then there exists $A' \subseteq A$ such that $|A'| \leq 1/x$ and there are at least $\frac{1}{2}|B|$ vertices in B with a neighbour in A'.*

Proof. Let $k := \lfloor 1/x \rfloor$; we may assume that $|A| \geq k$. Choose $s_1, \ldots, s_k \in A$ uniformly and independently at random, and let $S = \{s_1, \ldots, s_k\}$. For each $v \in B$, since v has at least $x|A|$ neighbours in A, the probability that none of $s_1, \ldots, s_k$ is such a neighbour is at most

$$\left(\frac{|A| - x|A|}{|A|}\right)^k = (1 - x)^{\lfloor 1/x \rfloor}.$$

If $x > 1/3$, then $(1 - x)^{\lfloor 1/x \rfloor} = (1 - x)^2 \leq 4/9 \leq 1/2$. If $x \leq 1/3$, then $x\lfloor 1/x \rfloor \geq 3/4$, and so

$$(1 - x)^{\lfloor 1/x \rfloor} \leq e^{-x\lfloor 1/x \rfloor} \leq e^{-3/4} \leq 1/2.$$

So, in either case, the expected number of vertices in B with no neighbour in S is at most $|B|/2$; and hence there is a choice of $A' \subseteq A$ with the desired property. This proves Lemma 3.2. ■

For $\ell, w \geq 0$ and a graph G, an *(ℓ, w)-comb* in G is a sequence of pairs $((a_i, B_i) : i \in [k])$ where

- $(B_1, \ldots, B_\ell)$ is an (ℓ, w)-blockade in G;
- $a_1, \ldots, a_k$ are pairwise distinct, and $\{a_1, \ldots, a_k\}, B_1, \ldots, B_k$ are pairwise disjoint subsets of $V(G)$; and
- for all distinct $i, j \in [k]$, a_i is adjacent to every vertex of B_i in G and nonadjacent to every vertex of B_j in G.

We call $a_1, \ldots, a_k$ the *apexes* of the comb.

To prove Lemma 2.3, we need a special case of the "comb" lemma from [8].

Lemma 3.3. *Let G be a graph and let $A, B \subseteq V(G)$ be nonempty and disjoint, such that each vertex in A has at most $\Delta > 0$ neighbours in B. Then either:*

- *at most $20\sqrt{|B|\Delta}$ vertices in B have a neighbour in A; or*

- *for some integer $k \ge 1$, there is a $(k, |B|/k^2)$-comb $((a_i, B_i) : i \in [k])$ in G where $a_i \in A$ and $B_i \subseteq B$ for all $i \in [k]$.*

The final ingredient we need is a well-known result for sparse P_5-free graphs [3], a special case of Theorem 2.1. We include a short proof here for completeness.

Lemma 3.4. *Let $\eta = 2^{-5}$; then for every η-sparse P_5-free graph G, there is an anticomplete $(2, \lfloor \eta|G| \rfloor)$-blockade in G.*

Proof. Let G be η-sparse and P_5-free; and suppose that there is no anticomplete $(2, \lfloor \eta|G| \rfloor)$-blockade in G. Then $|G| \ge \eta^{-1}$; and by Lemma 3.1 with $k = 2$, G has a connected component F with $|F| \ge \frac{1}{2}|G|$. Let $v \in V(F)$ and A be the set of neighbours of v in F; then $A \neq \emptyset$. Let $F' := F \setminus (A \cup \{v\})$. Since $|F'| \ge |F| - |A| - 1 \ge (\frac{1}{2} - 2\eta)|G| \ge \frac{1}{3}|G|$, and therefore $\frac{1}{4}|F'| \ge \eta|G|$, Lemma 3.1 gives a connected component J of F' with $|J| \ge \frac{1}{2}|F'| \ge \frac{1}{6}|G|$. Since F is connected, there are $u \in A, w \in V(J)$ with $uw \in E(F)$. Let B be the set of vertices in J adjacent to u in F; then $w \in B$ and $|B| \le \eta|G|$. Thus $|J \setminus B| \ge \frac{1}{6}|G| - \eta|G| \ge \frac{1}{8}|G| = 4\eta|G|$. Again, by Lemma 3.1 with $k = 2$, $J \setminus B$ has a connected component J' with $|J'| \ge \frac{1}{2}|J \setminus B| \ge 2\eta|G|$. Hence, since J is connected and w has degree at most $\eta|G| < |J'|$, w is mixed on $V(J')$ in J; and so there are $z, z' \in V(J')$ with $wz \in E(J), wz' \notin E(J)$. But then $\{u, v, w, z, z'\}$ forms a copy of P_5 in G, a contradiction. This proves Lemma 3.4. ∎

4. Using a comb

We will obtain Lemma 2.3 as a consequence of the followng:

Lemma 4.1. *There exists $d \ge 40$ for which the following holds. For every $x \in (0, 2^{-d})$ and every $\overline{P_5}$-free graph G with $|G| \ge x^{-d}$, there exist $k \in [2, 1/x]$ and a pure or x-sparse $(k, |G|/k^d)$-blockade in G.*

And the first step of the proof of Lemma 4.1 is Lemma 4.2 below; let us sketch the proof of that. Let $x \le y$ be sufficiently small positive variables, and let G be a y-sparse $\overline{P_5}$-free graph. If G is actually $y/2$-sparse then G is already (much) sparser than what we knew about it; so let us assume that there is a vertex v of degree at least $(y/2)|G|$ in G. We will apply Lemma 3.3 to obtain a comb between the neighbourhood B of v and the rest of the graph; but instead of taking a comb with apexes in B that expands into the rest of G (as was done in [8]), we will build an "upside-down" comb $((a_i, B_i) : i \in [k])$ (for some $k \ge 1$), with apexes in $V(G) \setminus B$ that goes from the rest of G back into B (in other words, v is nonadjacent to $a_1, \dots, a_k$ and adjacent to every vertex in $B_1 \cup \dots \cup B_k$; see Fig. 1). Such a comb is potentially useful, because if we can arrange for every $G[B_i]$ to be anticonnected (Lemma 3.1), then the blockade $\mathcal{B} = (B_1, \dots, B_k)$ has to be pure: whenever there is a vertex from some B_j mixed on another block B_i, the anticonnectivity of $G[B_i]$ would then give a copy of the house $\overline{P_5}$ in G that contains v and a_i (Fig. 1).

Thus, $\mathcal{B}$ is pure; but to satisfy the lemma, it must have the right length and width. First, we need its width to be at least $\text{poly}(1/k)|G|$ where k is its length. The blocks $B_1, \dots, B_k$ are subsets of B; and the application of Lemma 3.3 tells us that $\mathcal{B}$ is a $(k, |B|/O(k^2))$-blockade in $G[B]$, and so a $(k, (y/2)|G|/O(k^2))$-blockade in G, but it gives us no lower bound on k. To ensure that the width of $\mathcal{B}$ is at least $\text{poly}(1/k)|G|$, we need k to be at least some small power of y^{-1}. But we can arrange this as follows. Let us choose the comb to that it contains no vertices outside B that see at least a $y^{1/2}$ fraction of B. There are not many such vertices (at most $O(y^{1/2})|G|$), because $|B| \ge y|G|$ and everyone in B sees at most $y|G|$ vertices outside. In other words, by letting A be the set of vertices with at most $y^{1/2}|B|$ neighbours in B, we have $|A| \ge (1 - O(y^{1/2}))|G|$; so let us choose the comb with every apex a_i in A. Then the width of the comb is at least $|B|/O(k^2)$ and at most $y^{1/2}|B|$, and this ensures that $k \ge \Omega(y^{-1/4})$, as we wanted. Consequently we can arrange that $\mathcal{B}$ is a pure $(k, |G|/O(k^6))$-comb in G.

Another thing we need, for $\mathcal{B}$ to satisfy the lemma, is a good *upper* bound on its length k. We can arrange that $k \le \operatorname{poly}(1/x)$ (or another good thing happens), by putting a further restriction on how we choose the comb. Given A, B as above, if there are too many vertices of B (at least half, say) seeing fewer than $x^2|A|$ vertices in A, then it is easy to obtain subsets $A' \subseteq A$, $B' \subseteq B$ with $|A'| \ge (1-O(x))|A| \ge (1-O(y^{1/2}))|G|$ and $|B'| \ge \frac{1}{2}|B| \ge \Omega(y)|G|$ such that A' is x-sparse to B'. This is another desirable outcome for us, since we can iterate inside A', and if we keep getting this outcome, we will produce an x-sparse $(\Omega(y^{-1/2}), \Omega(y)|G|)$-blockade. So we may assume there are at least $\frac{1}{2}|B|$ vertices of B with at least $x^2|A|$ neighbours in A; and then Lemma 3.2 gives us some subset S of A of size at most x^{-2} that "covers" a constant fraction of B. By Lemma 3.3, the apexes $a_1, \dots, a_k$ of the comb can be taken from S, and so $k \le x^{-2}$ as a consequence.

That was a sketch of the proof of Lemma 4.2. Next we will write it out, with cosmetic adjustments in the constant factors and exponents.

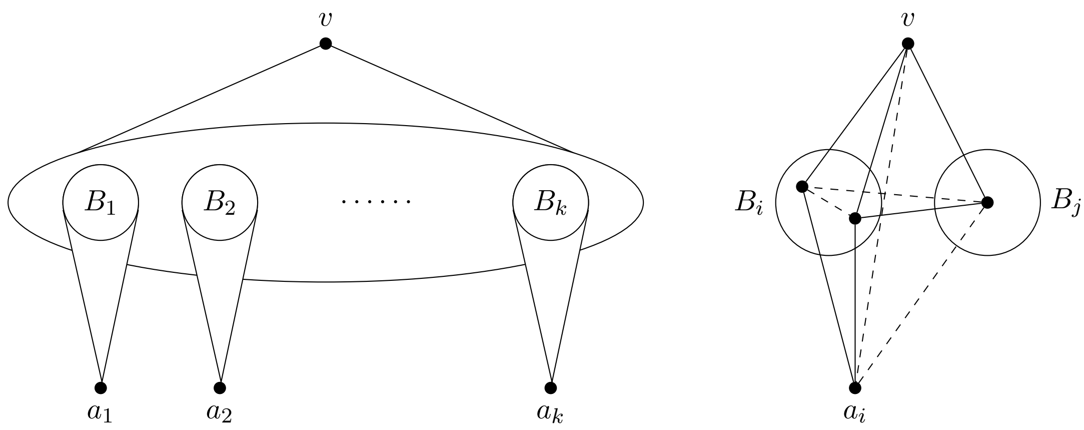

FIGURE 1. Making a house from an upside-down comb with anticonnected blocks.

We begin with the following lemma:

Lemma 4.2. *Let $x, y > 0$ with $x \le y \le 2^{-8}$, and let G be a y^3-sparse $\overline{P_5}$-free graph. Then either:*

- *G is $2y^4$-sparse;*
- *there exist $k \in [y^{-1/4}, 1/x]$ and a pure $(k, |G|/k^{26})$-blockade in G; or*
- *there are disjoint $X, Y \subseteq V(G)$ such that $|X| \ge \lfloor y^4|G| \rfloor$, $|Y| \ge (1-4y)|G|$, and Y is x-sparse to X.*

Proof. Assume that the first and third outcomes do not hold; then $y^4|G| \ge 1$. Since the first outcome does not hold, G has a vertex v of degree at least $2y^4|G|$. Let N be its set of neighbours.

Claim 4.3. *There exist $A \subseteq V(G) \setminus (N \cup \{v\})$ and $B \subseteq N$ such that*

- *$|B| \ge y^4|G|$ and $|A| \ge (1-3y)|G|$; and*
- *every vertex in B has at least $x^2|A|$ neighbours in A and A is y^2-sparse to B.*

Subproof. We have $|N| \ge 2y^4|G|$. Let A' be the set of vertices in $V(G) \setminus (N \cup \{v\})$ with at least $\frac{1}{2}y^2|N|$ neighbours in N. By averaging, there is a vertex in N with at least $\frac{1}{2}y^2|A'|$ neighbours in A'; and so $\frac{1}{2}y^2|A'| \le y^3|G|$, which yields that $|A'| \le 2y|G|$. Let $A := V(G) \setminus (N \cup A' \cup \{v\})$; then since $1 + y^2|G| \le y|G|$, we have

$$|A| \ge |G| - (1 + y^2|G| + 2y|G|) \ge (1-3y)|G|.$$

Let N' be the set of vertices in N with at most $x^2|A|$ neighbours in A, and let $B := N \setminus N'$. There are at most $x|A|$ vertices in A with more than $x|N'|$ neighbours in N', since there are at most $x^2|A| \cdot |N'|$ edges between A and N'; so there are at least

$$|A| - x|A| \ge (1-3y)|G| - x|G| \ge (1-4y)|G|$$

vertices in A with at most $x|N'|$ neighbours in N'. Thus, $|N'| \le y^4|G| \le \frac{1}{2}|N|$, since the third outcome of the lemma does not hold, and so

$$|B| = |N| - |N'| \ge \tfrac{1}{2}|N| \ge y^4|G|.$$

Since A is $\frac{1}{2}y^2$-sparse to N, it is y^2-sparse to B. This proves Claim 4.3. □

Let A, B be given by Claim 4.3; then Lemma 3.2 (with x^2 in place of x) gives $S \subseteq A$ with $|S| \le x^{-2}$ such that there are at least $\frac{1}{2}|B|$ vertices in B with a neighbour in S. Since $y < \frac{1}{40}$, more than $20\sqrt{|B|\Delta} = 20y|B|$ vertices in B have a neighbour in S. So by Lemma 3.3 with $\Delta = y^2|B|$ for some integer $\ell \ge 1$, there is an $(\ell, |B|/\ell^2)$-comb $((a_i, B_i) : i \in [\ell])$ in G where $a_i \in S$ and $B_i \subseteq B$ for all $i \in [\ell]$.

Since A is y^2-sparse to B, $|B|/\ell^2 \le y^2|B|$ and so $\ell \in [y^{-1}, x^{-2}]$. Let $k := \lceil \ell^{1/4} \rceil \in [y^{-1/4}, 1/x]$; then $|B| \ge y^4|G| \ge |G|/\ell^4 \ge |G|/k^{16}$ and $(B_1, \dots, B_k)$ is a $(k, |B|/k^8)$-blockade (note that $k \le \sqrt{\ell} \le x^{-1/2}$). Let $I := [k]$.

Claim 4.4. *There is a pure $(k, |B|/k^{10})$-blockade in $G[B]$.*

Subproof. For each $i \in I$, if $\overline{G}[B_i]$ has no anticonnected component of size at least $|B_i|/k$, then Lemma 3.1 gives a complete $(k, |B_i|/k^2)$-blockade in $G[B_i]$ (note that $k \ge y^{-1/4} \ge 4$); and this satisfies the claim since $|B_i|/k^2 \ge |B|/k^{10}$. Hence, we may assume each $G[B_i]$ has an anticonnected component D_i with

$$|D_i| \ge |B_i|/k^2 \ge |B|/k^{10}.$$

For distinct $i, j \in I$, if there exists some $u \in D_j$ mixed on D_i, then u would have a neighbour $w \in D_i$ and a nonneighbour $z \in D_i$ such that $wz \notin E(G)$ since D_i is anticonnected; and so $\{v, u, w, z, a_i\}$ would form a copy of $\overline{P_5}$ in G (see Fig. 1), a contradiction. Thus $(D_i : i \in I)$ is a pure blockade in $G[B]$ of length k and width at least $|B|/k^{10}$. This proves Claim 4.4. □

Since $|B|/k^{10} \ge |G|/k^{26}$, Claim 4.4 gives a pure $(k, |G|/k^{26})$-blockade in G, which is the second outcome of the lemma. This proves Lemma 4.2. ■

Next, we iterate Lemma 4.2 to turn its third outcome (an x-sparse pair) into an x-sparse blockade outcome, as follows.

Lemma 4.5. *Let $c := 2^{-8}$. Let $x, y > 0$ with $x \le y \le c$, and let G be a cy^3-sparse $\overline{P_5}$-free graph. Then either:*

- *there exists $S \subseteq V(G)$ such that $|S| \ge c|G|$ and $G[S]$ is $2y^4$-sparse;*
- *there exist $k \in [y^{-1/4}, 1/x]$ and a pure $(k, |G|/k^{30})$-blockade in G; or*
- *there is an x-sparse $(y^{-1}, \lfloor y^6|G| \rfloor)$-blockade in G.*

Proof. Suppose that none of the outcomes holds; then $y^6|G| \ge 1$. Thus there exists $n \ge 0$ maximal such that there is an x-sparse blockade $(B_0, B_1, \dots, B_n)$ with $|B_{i-1}| \ge \lfloor y^6|G| \rfloor$ for all $i \in [n]$ and $|B_n| \ge (1-4y)^n|G|$. Since the third outcome does not hold, $n < y^{-1}$; and so by the inequality $1 - t \ge 4^{-t}$ for all $t \in [0, \frac{1}{2}]$,

$$|B_n| \ge (1-4y)^n|G| \ge 4^{-4yn}|G| > 4^{-4}|G| = c|G| \ge y|G| \ge x|G|.$$

Hence $G[B_n]$ has maximum degree at most $cy^3|G| < y^3|B_n|$; and since the first outcome does not hold, $G[B_n]$ is not $2y^4$-sparse. Therefore, by Lemma 4.2, either:

- there exist $k \in [y^{-1/4}, 1/x]$ and a pure $(k, |B_n|/k^{26})$-blockade in G; or
- there are disjoint $X, Y \subseteq B_n$ such that $|X| \ge \lfloor y^6|B_n| \rfloor$, $|Y| \ge (1-4y)|B_n|$, and Y is x-sparse to X.

The first bullet cannot hold since $|B_n|/k^{26} \ge y|G|/k^{26} \ge |G|/k^{30}$ and the second outcome of the lemma does not hold. Thus the second bullet holds; but then $(B_0, B_1, \dots, B_{n-1}, X, Y)$ would contradict the maximality of n since $|X| \ge \lfloor y^5|B_n| \rfloor \ge \lfloor y^6|G| \rfloor$. This proves Lemma 4.5. ■

The next result contains the "iterative sparsification" step of the proof. It allows us to replace the the cy^3-sparsity hypothesis of Lemma 4.5 with a "sparsity a small constant" hypothesis and still deduce (essentially) the same conclusion.

Lemma 4.6. *Let $c := 2^{-8}$. Let $x \in (0, c^5)$, and let G be a c^{16}-sparse $\overline{P_5}$-free graph. Then either:*

- *for some $k \in [1/c, 1/x]$, there is a pure $(k, |G|/k^{34})$-blockade in G; or*
- *for some $y \in [x, c^5]$, there is an x-sparse $(y^{-1}, \lfloor y^7|G|\rfloor)$-blockade in G.*

Proof. Suppose that neither of the two outcomes holds. Let $y \in [cx, c^5]$ be minimal such that G has a cy^3-sparse induced subgraph F with $|F| \geq y|G|$. (This is possible, since taking $y = c^5$ has the property.) Suppose that $y < x$; then F is x^3-sparse with $|F| \geq y|G| \geq cx|G| \geq x^2|G| \geq x^{-5}$. Because $\lceil x^{-1}\rceil \cdot \lceil \frac{1}{4}x|F|\rceil \leq 2x^{-1} \cdot \frac{1}{2}x|F| = |F|$, there is an $(x^{-1}, \frac{1}{4}x|F|)$-blockade in F, which is then x-sparse since $\frac{1}{4}x \geq x^2$. Thus, since $\frac{1}{4}x|F| \geq \frac{1}{4}cx^2 \geq x^3|G|$, this would be an x-sparse $(x^{-1}, x^3|G|)$-blockade in G, a contradiction.

Consequently $y \geq x$. By Lemma 4.5 applied to F, either:

- F has a $2y^4$-sparse induced subgraph with at least $c|F| \geq cy|G|$ vertices;
- there exist $k \in [y^{-1/4}, 1/x] \subseteq [1/c, 1/x]$ and a pure $(k, |F|/k^{30})$-blockade in F; or
- there is an x-sparse $(y^{-1}, \lfloor y^6|F|\rfloor)$-blockade in F.

The first bullet would give a $2y^4$-sparse induced subgraph of F (and so of G) with at least $cy|G|$ vertices, which contradicts the minimality of y since $2y^4 \leq c^4y^3 = c(cy)^3$. If the second bullet holds, then since $|F|/k^{30} \geq y|G|/k^{30} \geq |G|/k^{34}$, there would be a pure $(k, |G|/k^{34})$-blockade in G, a contradiction. If the third bullet holds, then since $y^6|F| \geq y^7|G|$, there would be an x-sparse $(y^{-1}, \lfloor y^7|G|\rfloor)$-blockade in G, a contradiction. This proves Lemma 4.7. ■

Next, by applying Rödl's theorem 1.3, we remove the sparsity hypothesis in Lemma 4.6 completely, and prove Lemma 4.1, which we restate:

Lemma 4.7. *There exists $d \geq 40$ for which the following holds. For every $x \in (0, 2^{-d})$ and every $\overline{P_5}$-free graph G with $|G| \geq x^{-d}$, there exist $k \in [2, 1/x]$ and a pure or x-sparse $(k, |G|/k^d)$-blockade in G.*

Proof. Let $c := 2^{-8}$, and $\eta := 2^{-5}$, and let $\xi := c^{16}$). By Theorem 1.3, there exists $\theta \in (0,1)$ such that every $\overline{P_5}$-free graph G contains a ξ-restricted induced subgraph with at least $\theta|G|$ vertices. We shall prove that every $d \geq 40$ with $2^{d-1} \geq (\eta\theta)^{-1}$ satisfies the lemma. To show this, let $x \in (0, 2^{-d})$, and let G be $\overline{P_5}$-free with $|G| \geq x^{-d}$. We must show that there exists $k \in [2, 1/x]$ such that there is a pure or x-sparse $(k, |G|/k^d)$-blockade in G. By the choice of θ, G has a ξ-restricted induced subgraph F with $|F| \geq \theta|G|$. If $\overline{F}$ is ξ-sparse, then since $\overline{F}$ is P_5-free, Lemma 3.4 gives an anticomplete $(2, \lfloor \eta|F|\rfloor)$-blockade in $\overline{F}$; and we are done since $\lfloor \eta|F|\rfloor \geq \lfloor \eta\theta|G|\rfloor \geq \lfloor |G|/2^{d-1}\rfloor \geq |G|/2^d$ by the choice of d. Hence, we may assume that F is ξ-sparse (and so is c^{16}-sparse). Since $x \in (0, 2^{-d}) \subseteq (0, c^5)$, Lemma 4.6 implies that either:

- for some $k \in [1/c, 1/x]$, there is a pure $(k, |S|/k^{34})$-blockade in F; or
- for some $y \in [x, c^5]$, there is an x-sparse $(y^{-1}, \lfloor y^7|F|\rfloor)$-blockade in F.

If the first bullet holds, then $|G| \geq x^{-d} \geq k^d$, $k \geq 1/c = 2^8$, and $d \geq 40$ which together imply

$$|F|/k^{34} \geq \theta|F|/k^{34} \geq 2^{-d}|F|/k^{34} \geq k^{-d/8}|G|/k^{34} \geq |F|/k^d;$$

and so there would be a pure $(k, |F|/k^d)$-blockade in G and we are done. If the second bullet holds, then since

$$\lfloor y^7|F|\rfloor \geq \lfloor \theta y^7|G|\rfloor \geq \lfloor 2^{1-d}y^7|G|\rfloor \geq \lfloor 2y^{d/8+7}|G|\rfloor \geq \lfloor 2y^d|G|\rfloor \geq y^d|G|,$$

there would be an x-sparse $(y^{-1}, y^d|G|)$-blockade in G and we are done. This proves Lemma 4.7. ■

5. The proof of Lemma 2.3

Next we will deduce Lemma 2.3 from Lemma 4.7. If we take x to be a power of ε^d, then Lemma 4.7 already gives us something like what we want for Lemma 2.3, but the blockade we obtain might have length too small. If so, then it still has very large blocks, and we can apply Lemma 4.7 to each block to get a longer blockade, and repeat. This idea is formalized in the following general theorem (with no $\overline{P_5}$-free condition), which is a slight modification of a theorem of [15].

Theorem 5.1. *Let G be a graph, and let $\varepsilon \in (0, \frac{1}{2})$ and $d \geq 1$. Let $x := \varepsilon^{5d}$. Assume that for every induced subgraph F of G with $|F| \geq \varepsilon^d|G|$, there exists $k \in [2, 1/x]$ such that there is a pure or x-sparse $(k, |F|/k^d)$-blockade in F. Then there is an $(\varepsilon^{-1}, \lfloor x^{2d}|G|\rfloor)$-blockade $(B_1, \ldots, B_\ell)$ in G, such that for all distinct $i, j \in [\ell]$, (B_i, B_j) is either complete or weakly ε^d-sparse in G.*

Proof. We may assume that $|G| \geq x^{-2d}$. Let J be a graph; and for each $j \in V(J)$ let A_j be a nonempty subset of $V(G)$, pairwise disjoint, such that for all distinct $i, j \in J$, A_i is complete to A_j whenever i, j are adjacent in J. We call $\mathcal{L} = (J, (A_j : j \in V(J)))$ a *layout*. A pair $\{u, v\}$ of distinct vertices of G is *undecided* for a layout $(J, (A_j : j \in V(J)))$ if there exists $j \in V(J)$ with $u, v \in A_j$; and *decided* otherwise. A decided pair $\{u, v\}$ is *wrong* for $(J, (A_j : j \in V(J)))$ if there are distinct $i, j \in V(J)$ such that $u \in A_i$, $v \in A_j$, and u, v are adjacent in G while i, j are nonadjacent in J. We are interested in layouts in which the number of wrong pairs is only a small fraction of the number of decided pairs. Choose a layout $\mathcal{L} = (J, (A_j : j \in V(J)))$ satisfying the following:

- $|A_j| \geq \varepsilon^{2d}|G|$ for each $j \in V(J)$;
- $\sum_{j \in V(J)} |A_j|^{1/d} \geq |G|^{1/d}$;
- the number of wrong pairs is at most x times the number of decided pairs; and
- subject to these three conditions, $|J|$ is maximum.

(This is possible since we may take $|J| = 1$ and $A_1 = V(G)$ to satisfy the first three conditions.)

Claim 5.2. *We may assume that $|J| \leq \varepsilon^{-1}$.*

Subproof. Assume that $|J| \geq \varepsilon^{-1}$. Since the number of wrong pairs is at most x times the number of decided pairs and so at most $x|G|^2$, for every distinct $i, j \in V(J)$ that are nonadjacent in J, the number of edges between A_i, A_j is at most $x|G|^2 \leq x\varepsilon^{-4d}|A_i||A_j| = \varepsilon^d|A_i||A_j|$; that is, (A_i, A_j) is weakly ε^d-sparse. Since $|A_i| \geq \varepsilon^{2d}|G| \geq x^{2d}|G|$ for each $j \in V(J)$, $(A_j : j \in V(J))$ is thus a blockade satisfying the theorem. This proves Claim 5.2. $\square$

Let $A \in \{A_j : j \in V(J)\}$ satisfy $|A| = \max_{j \in V(J)} |A_j|$. Since $\sum_{j \in V(J)} |A_j|^{1/d} \geq |G|^{1/d}$, and $|J| \leq \varepsilon^{-1}$ by Claim 5.2, it follows that $|A|^{1/d} \geq \varepsilon|G|^{1/d}$, that is, $|A| \geq \varepsilon^d|G|$. By applying the hypothesis to $G[A]$, we obtain a pure or x-sparse $(k, |A|/k^d)$-blockade $(B_1, \ldots, B_\ell)$ in $G[A]$, for some $k \in [2, 1/x]$. Let K be the graph with vertex set $[\ell]$, such that for all distinct $p, q \in [\ell]$, p is adjacent to q in K if and only if B_p is complete to B_q in $G[A]$; in particular K is edgeless if $(B_1, \ldots, B_\ell)$ is x-sparse in $G[A]$.

Claim 5.3. *$k \geq \varepsilon^{-1}$.*

Subproof. Suppose that $k \leq \varepsilon^{-1}$. Then each of the sets $B_1, \ldots, B_\ell$ has size at least $|A|/k^d \geq \varepsilon^d|A|$. By substituting K for the vertex of J corresponding to A, and replacing A by $B_1, \ldots, B_\ell$, we obtain a new layout $\mathcal{L}' = (J', (A'_j : j \in V(J')))$ say, where $|J'| > |J|$. We claim that this violates the choice of $\mathcal{L}$; and so we must verify that $\mathcal{L}'$ satisfies the first three bullets in the definition of $\mathcal{L}$. To see this, observe that each B_p satisfies $|B_p| \geq \varepsilon^d|A| \geq \varepsilon^{2d}|G|$, and so the first bullet is satisfied. For the second bullet, since $B_1, \ldots, B_\ell$ all have size at least $|A|/k^d$, it follows that

$$|B_1|^{1/d} + \cdots + |B_\ell|^{1/d} \geq |A|^{1/d},$$

and so $\sum_{j\in V(J')}|A'_j|^{1/d}\ge|G|^{1/d}$. For the third bullet, let P be the set of all decided pairs for $\mathcal{L}$, and $Q\subseteq P$ the set of wrong pairs for $\mathcal{L}$; and define P',Q' similarly for $\mathcal{L}'$. Then $P\subseteq P'$ and $|Q|\le x|P|\le x|P'|$. Let R be the set of all pairs $\{u,v\}$ with $u,v\in A$ such that u,v belong to different blocks of $(B_1,\dots,B_\ell)$. Then $R\subseteq P'\setminus P$ and $Q'\setminus Q\subseteq R$. If $(B_1,\dots,B_\ell)$ is pure in $G[A]$ then $|Q'|\le|Q|\le x|P'|$; and if $(B_1,\dots,B_\ell)$ is x-sparse in $G[A]$, then $|Q'\setminus Q|\le x|R|$ which yields $|Q'\setminus Q|\le x|P'\setminus P|$, and so

$$|Q'|\le|Q|+|Q'\setminus Q|\le x|P|+x|P'\setminus P|=x|P'|.$$

This contradicts the choice of $\mathcal{L}$, and so proves Claim 5.3. □

Since $k\le 1/x$ and $|A|\ge\varepsilon^d|G|\ge x^d|G|$, we have $|B_p|\ge|A|/k^d\ge x^d|A|\ge x^{2d}|G|$ for each $p\in[\ell]$; and for all distinct $p,q\in[\ell]$, (B_p,B_q) is either complete or weakly ε^d-sparse since $x=\varepsilon^{5d}\le\varepsilon^d$. Hence $(B_1,\dots,B_\ell)$ satisfies the theorem. This proves Theorem 5.1. ■

By combining Lemma 4.7 and Theorem 5.1, we prove Lemma 2.3, which we restate:

Lemma 5.4. *There exists $d\ge 40$ for which the following holds. Let $\varepsilon\in(0,\frac12)$, and let G be a $\overline{P_5}$-free graph. Then there is an $(\varepsilon^{-1},\lfloor\varepsilon^{10d^2}|G|\rfloor)$-blockade $(B_1,\dots,B_\ell)$ in G, such that for all distinct $i,j\in[\ell]$, (B_i,B_j) is either complete or weakly ε^d-sparse in G.*

Proof. We claim that $d\ge 40$ given by Lemma 4.7 satisfies the lemma. Let $x:=\varepsilon^{5d}\in(0,2^{-d})$; and we may assume that $|G|\ge\varepsilon^{-10d^2}=x^{-2d}$. For every induced subgraph F of G with $|F|\ge\varepsilon^d|G|$, we have $|F|\ge\varepsilon^dx^{-2d}\ge x^{-d}$; and so by the choice of d, there exists $k\in[2,1/x]$ such that there is a pure or x-sparse $(k,|F|/k^d)$-blockade in F. Theorem 5.1 now gives an $(\varepsilon^{-1},\lfloor x^{2d}|G|\rfloor)$-blockade $(B_1,\dots,B_\ell)$ in G, such that for all distinct $i,j\in[\ell]$, (B_i,B_j) is either complete or weakly ε^d-sparse in G. Since $x^{2d}=\varepsilon^{10d^2}$, this proves Lemma 5.4. ■

This completes the first half of the proof of Theorem 1.5.

6. Deducing Theorem 1.5

In this section we complete the proof of Theorem 1.5. Let us make one point which might clarify why we need two rounds of iterative sparsification. Lemma 5.4 gives us blockades with the property that every pair of blocks is complete or weakly sparse: let us call them "semisparse" for this discussion. Lemma 4.5 tells us essentially that:

- If G is $\overline{P_5}$-free and $O(y^3)$-sparse, then either we can sparsify further or there is a semisparse blockade of length at least $(1/y)^{1/4}$ and at most $1/x$.

That result passed through the machinery of iterative sparsification, and was converted to Lemma 5.4. The latter works in any $\overline{P_5}$-free graph, with no sparsity condition, and we can specify the length of the blockade it gives us, by choose $1/\varepsilon$ appropriately. In particular, we can apply it in a y-sparse graph, choosing ε to be some huge power of y; and we deduce that:

- If G is $\overline{P_5}$-free and y-sparse, then either we can sparsify further or there is a semisparse blockade of length a huge power of $1/y$.

So this is a much more powerful version of Lemma 4.5, and the length of this blockade gives rise to a new way to sparsify, that is the key to the remainder of the proof of Theorem 1.5.

After Lemma 5.4, the next step in the proof of Theorem 1.5 is to prove Lemma 6.1 below, and that is where we use the blockades given by Lemma 2.3. Let us sketch its proof. Let y be a small positive variable, and let G be a y-sparse $\overline{P_5}$-free graph. Again, we try to do sparsification; if we can find a slightly smaller value y' such that there is a y'-sparse induced subgraph of size $\mathrm{poly}(y'/y)|G|$, we will take that as an outcome. We apply Lemma 5.4 with $\varepsilon=y^d$ to get a $(y^{-d},\lfloor y^{10d^3}|G|\rfloor)$ blockade $\mathcal{B}=(B_1,\dots,B_\ell)$ in G (where $\ell=\lceil y^{-d}\rceil$) such that every pair (B_i,B_j) is either complete or weakly y^{d^2}-sparse. Here we can arrange each B_i to be anticonnected in G and of size about $y^{10d^3}|G|$ (up to minor changes in their sizes and the density between them). How does the rest of G attach to $\mathcal{B}$? Let v be some vertex not in any of the blocks of $\mathcal{B}$. Then v is anticomplete to some of the blocks, complete to others, and mixed on the remainder. If there is some v outside of $\mathcal{B}$ that is mixed on at least $y\ell$ blocks, then no two of these blocks are complete to each other; for otherwise there would be a copy of $\overline{P_5}$; this is where the complete property is crucial (see Fig. 2). Hence, these $y\ell$ blocks are pairwise weakly y^{d^2}-sparse; and so their union has edge density about $O((y\ell)^{-1}) = O(y^{d-1})$ and size at least $y^{10d^3}|G|$, which is a desirable sparsification outcome. So we assume that there is no such v. It follows that there is some B_i with at most $O(y)|G|$ vertices of G mixed on it. But only a few vertices are complete to B_i since G is y-sparse; so almost all are anticomplete to B_i. More exactly, B_i is anticomplete to a vertex subset of size $(1 - O(y))|G|$, which is another desirable outcome since $|B_i|$ is about $y^{10d^3}|G|$. (This type of argument also appears in [17] where we show that graphs of bounded VC-dimension have polynomial-sized cliques or stable sets.)

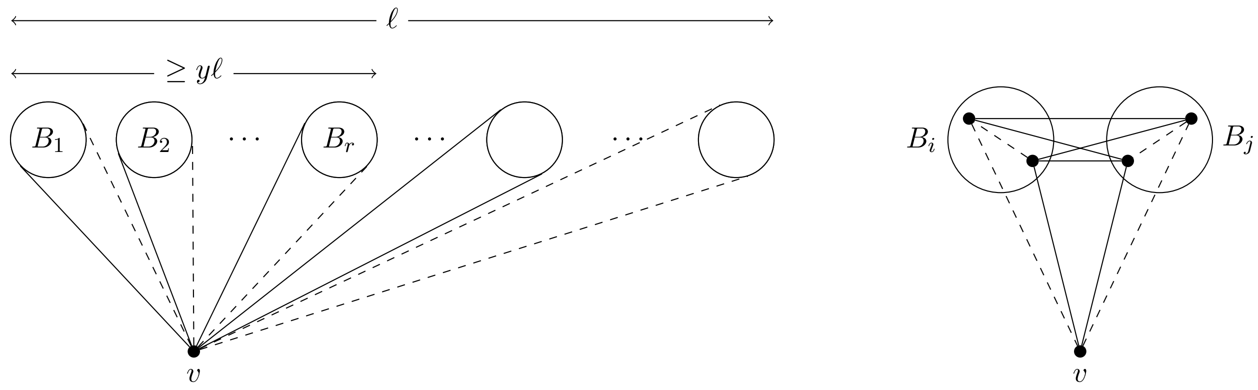

FIGURE 2. Using a really long semisparse blockade.

Lemma 6.1. *There exists $d \geq 40$ such that the following holds. Let $y \in (0, \frac{1}{2})$, and let G be a y-sparse $\overline{P_5}$-free graph. Then either:*

- *there exists $S \subseteq V(G)$ with $|S| \geq y^{30d^3}|G|$ such that $G[S]$ is y^{2d}-sparse;*
- *there is a complete $(y^{-1}, y^{33d^3}|G|)$-blockade in G; or*
- *there are disjoint $X, Y \subseteq V(G)$ such that $|X| \geq y^{33d^3}|G|$, $|Y| \geq (1-3y)|G|$, and Y is anticomplete to X in G.*

Proof. We claim that $d \geq 40$ given by Lemma 5.4 satisfies the lemma. To show this, let y, G be as in the lemma statement; and assume that the first two outcomes do not hold. In particular $|G| \geq y^{-30d^3}$ since the first outcome does not hold. Let $\varepsilon := y^{3d} \in (0, 2^{-3d})$; then $|G| \geq y^{-30d^3} = \varepsilon^{-10d^2}$. Let $\ell := \lceil \varepsilon^{-1} \rceil$ and $m := \lfloor \varepsilon^{10d^2}|G| \rfloor \geq \frac{1}{2}\varepsilon^{10d^2}|G|$.

Claim 6.2. *There is a blockade $(B_1, \ldots, B_\ell)$ in G such that:*

- *for all $i \in [\ell]$, B_i is anticonnected in G and $|B_i| = \lceil \varepsilon^2 m \rceil$; and*
- *for all distinct $i, j \in [\ell]$, (B_i, B_j) is either complete or ε^{d-8}-sparse to each other in G.*

Subproof. By Lemma 5.4, there is an $(\varepsilon^{-1}, \lfloor \varepsilon^{10d^2}|G| \rfloor)$-blockade $(A_1, \ldots, A_\ell)$ in G, where $\ell = \lceil \varepsilon^{-1} \rceil \leq 2\varepsilon^{-1}$, such that for all distinct $i, j \in [\ell]$, (A_i, A_j) is complete or weakly ε^d-sparse in G. Let J be the graph with vertex set $[\ell]$ where distinct $i, j \in V(J)$ are adjacent in J if and only if A_i is complete to A_j in G.

For each $i \in [\ell]$, let X_i be a uniformly random subset of A_i of size $m = \lfloor \varepsilon^{10d^2}|G| \rfloor$. For all distinct $i, j \in [\ell]$ with $ij \notin E(J)$, the expected number of edges between X_i, X_j in G is at most $\varepsilon^d |X_i||X_j|$; and so, since $\frac{1}{2}\ell^2 = \frac{1}{2}\lceil \varepsilon^{-1} \rceil^2 \leq \varepsilon^{-2}$, with positive probability (X_i, X_j) is weakly ε^{d-2}-sparse for all distinct $i, j \in [\ell]$ with $ij \notin E(J)$.

For $i = 1, 2, \ldots, \ell$ in turn, define a subset B_i of X_i as follows. Assume that $B_1, \ldots, B_{i-1}$ have been defined, such that $|B_p| = \lceil \varepsilon^2 m \rceil$ for all $1 \leq p < q \leq \ell$ with $pq \notin E(J)$ and $p < i$,

- B_p is ε^{d-6}-sparse to B_q and B_q is ε^{d-8}-sparse to B_p if $q < i$; and
- B_p is ε^{d-4}-sparse to X_q if $q \ge i$.

For each $p \in [\ell] \setminus \{i\}$ with $pi \notin E(J)$, let C_p be the set of vertices in X_i with at least $\varepsilon^{d-8}|B_p|$ neighbours in B_p if $p < i$, and let C_p be the set of vertices in X_i with at least $\varepsilon^{d-4}|X_p|$ neighbours in X_p if $p > i$; then $|C_p| \le \varepsilon^2|X_i|$ for all $p \in [\ell] \setminus \{i\}$. Let $D_i := X_i \setminus (\bigcup_{p\in[\ell]\setminus\{i\},pi\notin E(J)} C_p)$; then $|D_i| \ge (1-\varepsilon^2\ell)|X_i| \ge (1-2\varepsilon)|X_i| \ge \frac{1}{2}m$. If $G[D_i]$ has no anticonnected component of size at least $|D_i|/\ell$, then Lemma 3.1 (with $k = \ell$) would give a complete $(\ell, |D_i|/\ell^2)$-blockade in $G[D_i]$; but this satisfies the second outcome of the lemma since $|D_i|/\ell^2 \ge \frac{1}{8}\varepsilon^2 m \ge \frac{1}{16}\varepsilon^{2+10d^2}|G| \ge \varepsilon^{11d^2}|G| = y^{33d^3}|G|$ and $\ell \ge \varepsilon^{-1} \ge y^{-1}$, a contradiction. Thus, $G[D_i]$ has an anticonnected component B_i with $|B_i| \ge |D_i|/\ell \ge \frac{1}{4}\varepsilon m \ge \varepsilon^2 m$. By removing vertices from B_i if necessary, we may assume that $|B_i| = \lceil\varepsilon^2 m\rceil$. For every $1 \le p < i$ with $pi \notin E(J)$, since B_p is ε^{d-4}-sparse to X_i, it follows that B_p is ε^{d-6}-sparse to B_i; and B_i is ε^{d-8}-sparse to B_p by definition.

This completes the inductive definition of $B_1, \dots, B_\ell$; and it is not hard to check that $(B_1, \dots, B_\ell)$ is a blockade of G satisfying the claim. This proves Claim 6.2. □

Let $B := V(G) \setminus (B_1 \cup \cdots \cup B_\ell)$; then since $\varepsilon \le y^2$, we have

$$|B| \ge |G| - \ell\lceil\varepsilon^2 m\rceil \ge |G| - 2\ell\varepsilon^2 m \ge |G| - 4\varepsilon m \ge |G| - m \ge (1-\varepsilon)|G| \ge (1-y^2)|G|.$$

Claim 6.3. *No vertex in B is mixed on at least $y\ell$ blocks among $(B_1, \dots, B_\ell)$.*

Subproof. Suppose there is such a vertex $v \in B$; and assume that it is mixed on $B_1, \dots, B_r$, where $r \ge y\ell \ge y^{2d+1}$. If there are distinct $i, j \in [r]$ such that B_i is complete to B_j in G, then since B_i, B_j are anticonnected in G, there would be $u_i, w_i \in B_i$ and $u_j, w_j \in B_j$ such that $u_iv, u_jv \in E(G)$ and $w_iv, w_jv \notin E(G)$; but then $\{v, u_i, u_j, v_i, v_j\}$ would form a copy of $\overline{P_5}$ in G (see Fig. 2), a contradiction. Thus, B_i is ε^{d-8}-sparse to B_j for all distinct $i, j \in [r]$. Let $S := \bigcup_{i\in[r]} B_i$; then $|S| = rm$ and $G[S]$ has maximum degree at most

$$m + r\varepsilon^{d-8}m \le (y^{2d+1} + \varepsilon^{d-8})rm \le 2y^{2d+1}rm \le y^{2d}rm = y^{2d}|S|$$

where the penultimate inequality holds since $\varepsilon^{d-8} = y^{3d(d-8)} \le y^{3d} \le y^{2d+1}$ (note that $d \ge 40$). Thus $G[S]$ is y^{2d}-sparse; but then S satisfies the first outcome of the lemma since $|S| = rm \ge \varepsilon^{10d^2}|G| = y^{30d^2}|G|$, a contradiction. This proves Claim 6.3. □

Claim 6.3 says that every vertex in B is mixed on fewer than $y\ell$ blocks among $(B_1, \dots, B_\ell)$; and so there exists $i \in [\ell]$ such that there are fewer than $y|B|$ vertices in B mixed on B_i. Thus, since G is y-sparse, there are at most $y|G| + y|B|$ vertices in B with a neighbour in B_i. Let Y be the set of vertices in B with no neighbour in B_i; then, because $|B| \ge (1-y^2)|G|$, we have

$$|Y| \ge (1-y)|B| - y|G| \ge (1-y)(1-y^2)|G| - y|G| \ge (1-3y)|G|$$

and the third outcome of the lemma holds since $|B_i| \ge \varepsilon^2 m \ge \frac{1}{2}\varepsilon^{2+10d^2}|G| \ge \varepsilon^{11d^2}|G| = y^{33d^2}|G|$. This proves Lemma 6.1. ■

Let us now turn the third outcome of Lemma 6.1 into an anticomplete blockade outcome.

Lemma 6.4. *There exists $d \ge 40$ such that the following holds. Let $y \in (0, 4^{-6}]$, and let G be a y-sparse $\overline{P_5}$-free graph. Then either:*

- *there exists $S \subseteq V(G)$ with $|S| \ge y^{16d^3}|G|$ such that $G[S]$ is y^d-sparse; or*
- *there is a complete or anticomplete $(y^{-1/2}, y^{18d^3}|G|)$-blockade in G.*

Proof. We claim that $d \ge 40$ given by Lemma 6.1 satisfies the lemma. We may assume $|G| \ge y^{-16d^3}$, for otherwise the first outcome trivially holds. Let $n \ge 0$ be maximal such that there is an anticomplete blockade $(B_0, B_1, \dots, B_n)$ of G with $|B_n| \ge (1-2y^{1/2})^n|G|$ and $|B_{i-1}| \ge y^{18d^3}|G|$ for all $i \in [n]$. If

$n \ge y^{-1/2}$ then the second outcome of the lemma holds; and so we may assume $n < y^{-1/2}$. Then since $y \le 4^{-6}$,

$$|B_n| \ge (1-3y^{1/2})^n|G| \ge 4^{-3y^{1/2}n}|G| \ge 4^{-3}|G| \ge y|G| \ge y^{-15d^3} = (y^{-1/2})^{30d^3}$$

and so $G[B_n]$ has maximum degree at most $y|G| \le 4^3y|B_n| \le y^{1/2}|B_n|$ since $y \le 4^{-6}$. Thus, by Lemma 6.1 (with $y^{1/2}$ in place of y), either:

- there exists $S \subseteq B_n$ with $|S| \ge y^{15d^3}|B_n|$ such that $G[S]$ is y^d-sparse;
- there is a complete $(y^{-1/2}, y^{17d^3}|B_n|)$-blockade in $G[B_n]$; or
- there are disjoint $X, Y \subseteq B_n$ such that $|X| \ge y^{17d^3}|B_n|$, $|Y| \ge (1-2y^{1/2})|B_n|$, and Y is anticomplete to X in G.

If the first bullet holds, then $|S| \ge y^{15d^3}|B_n| \ge y^{16d^3}|G|$ and the first outcome of the lemma holds. If the second bullet holds, then since $y^{17d^3}|B_n| \ge y^{18d^3}|G|$, the second outcome of the lemma holds. If the third bullet holds, then since $|X| \ge y^{17d^3}|B_n| \ge y^{18d^3}|G|$ and $|Y| \ge (1-2y^{1/2})|B_n| \ge (1-2y^{1/2})^{n+1}|G|$, $(B_0, B_1, \dots, B_{n-1}, X, Y)$ would contradict the maximality of n. This proves Lemma 6.4. ■

Next we eliminate the sparsity hypothesis of Lemma 6.4, by means of Rödl's theorem 1.3 and iterative sparsification.

Lemma 6.5. *There exists $a \ge 1$ such that the following holds. For every $x \in (0, \frac{1}{2})$ and every $\overline{P_5}$-free graph G, either:*

- *G has an x-restricted induced subgraph with at least $x^a|G|$ vertices; or*
- *there is a complete or anticomplete $(k, |G|/k^a)$-blockade in G, for some $k \in [2, 1/x]$.*

Proof. Let $c := 4^{-6}$ and $\eta = 2^{-5}$. Let $d \ge 40$ be given by Lemma 6.4. By Theorem 1.3, there exists $t \ge 36d^2$ such that for every $\overline{P_5}$-free graph G, there exists $S \subseteq V(G)$ with $|S| \ge c^t|G|$ such that $G[S]$ is c-restricted. We shall prove that every $a \ge 2dt$ with $2^{a-1} \ge (\eta c^t)^{-1}$ satisfies the lemma. To show this, let $x \in (0, c)$, and let G be $\overline{P_5}$-free with $|G| \ge x^{-a}$. Assume that the second outcome of the lemma does not hold; that is, there is no $k \in [2, 1/x]$ such that there is a complete or anticomplete $(k, |G|/k^a)$-blockade in G. By the choice of θ, there is a c-restricted $S \subseteq V(G)$ with $|S| \ge \theta|G|$. If $\overline{G}[S]$ is c-sparse, then since $\overline{G}[S]$ is P_5-free, Lemma 3.4 gives an anticomplete $(2, \lfloor\eta|S|\rfloor)$-blockade in $\overline{G}[S]$, a contradiction since $\lfloor\eta|S|\rfloor \ge \lfloor\eta c^t|G|\rfloor \ge \lfloor|G|/2^{a-1}\rfloor \ge |G|/2^a$ by the choice of a. Hence, $G[S]$ is c-sparse. Thus, there exists $y \in [x^d, c]$ minimal (note that $x^d < 2^{-d} < 2^{-12} = c$) such that G has a y-sparse induced subgraph F with $|F| \ge y^t|G|$.

Claim 6.6. *$y < x$.*

Subproof. Suppose not. By Lemma 6.4, either:

- F has a y^d-sparse induced subgraph with at least $y^{16d^3}|F|$ vertices; or
- there is a complete or anticomplete $(y^{-1/2}, y^{18d^3}|F|)$-blockade in F.

Note that $18d^3 + t \le dt \le \frac{1}{2}a$ since $d \ge 2$, $t \ge 36d^2 \ge \frac{18d^3}{d-1}$, and $a \ge 2dt$. Thus, if the first bullet holds, then G would have a y^d-sparse induced subgraph with at least $y^{16d^3}|F| \ge y^{16d^3+t} \ge y^{dt}|G|$ vertices, which contradicts the minimality of y since $y^d \ge x^d$. If the second bullet holds, then since $y^{18d^3}|F| \ge y^{18d^3+t}|G| \ge y^{dt}|G| \ge y^{a/2}|G|$, there would be a complete or anticomplete $(y^{-1/2}, y^{a/2}|G|)$-blockade in G, which satisfies the second outcome of the lemma (with $k = y^{-1/2}$) because $x \le y^{1/2} \le c^{1/2} \le \frac{1}{2}$, a contradiction. This proves Claim 6.6. □

Since $x^d \le y < x$, we have that F is x-sparse and $|F| \ge y^t|G| \ge x^{dt}|G| \ge x^a|G|$. Thus the first outcome of the lemma holds, proving Lemma 6.5. ■

We are now ready to deduce the polynomial Rödl property of P_5. The proof method holds under a more general setting, and is similar to and simpler in part than that of Theorem 5.1.

Theorem 6.7. *Let $\varepsilon \in (0, \frac{1}{2})$ and $a \geq 1$, and let G be a graph. Assume that for every induced subgraph F of G with $|F| \geq \varepsilon^{2a}|G|$, there exists $k \in [2, 1/\varepsilon]$ such that there is a complete or anticomplete $(k, |F|/k^a)$-blockade in F. Then G has an ε-restricted induced subgraph with at least $\varepsilon^{3a}|G|$ vertices.*

Proof. A *cograph* is a graph with no induced four-vertex path; and it is well known that every n-vertex cograph has a clique or stable set of size at least $\sqrt{n}$. Let $q \geq 1$ be a maximal integer such that there exist a cograph J with vertex set $[q]$ and a pure $(q, \varepsilon^{3a}|G|)$-blockade $(A_1, \dots, A_q)$ in G satisfying:

- for all distinct $i, j \in [q]$, (A_i, A_j) is complete in G if and only if $ij \in E(J)$; and
- $\sum_{j \in [q]} |A_j|^{1/a} \geq |G|^{1/a}$.

Claim 6.8. $q \geq \varepsilon^{-2}$.

Subproof. Suppose not. We may assume $|A_1| = \max_{j \in [q]} |A_j|$; then $q|A_1|^{1/a} \geq |G|^{1/a}$ which yields $|A_1| \geq |G|/q^a \geq \varepsilon^{2a}|G|$. Thus, the hypothesis gives $k \in [2, 1/\varepsilon]$ and a complete or anticomplete $(k, |A_1|/k^a)$-blockade $(B_1, \dots, B_\ell)$ in $G[A_1]$. Let J' be the graph obtained from J by substituting a complete or edgeless graph K for vertex 1 in J, such that $|K| = \ell$ and K is complete if and only if (B_1, B_2) is complete in $G[A_1]$. Then J' is a cograph with $|J'| > q$. Now $|B_i| \geq |A_1|/k^a \geq \varepsilon^a |A_1| \geq \varepsilon^{3a}|G|$ for all $i \in V(K)$; and $\sum_{i \in V(K)} |B_i|^{1/a} \geq k(|A_1|/k^a)^{1/a} = |A_1|^{1/a}$ which implies

$$\sum_{j \in [q] \setminus \{1\}} |A_j|^{1/a} + \sum_{i \in V(K)} |B_i|^{1/a} \geq \sum_{j \in [q]} |A_j|^{1/a} \geq |G|^{1/a}.$$

Consequently J' violates the maximality of q, a contradiction. This proves Claim 6.8. □

Since J is a cograph, it has a clique or stable set I with $|I| \geq \sqrt{q} \geq 1/\varepsilon$. For every $j \in I$, let $S_j \subseteq A_j$ with $|S_j| = \lceil \varepsilon^{3a}|G| \rceil$; and let $S := \bigcup_{j \in I} S_j$. Then $|S| = |I| \cdot |S_j| \geq \varepsilon^{3a}|G|$ for all $j \in I$. If I is a clique in J, then $\overline{G}[S]$ has maximum degree at most $|S|/|I| \leq \varepsilon|S|$; and if I is a stable set in J, then $G[S]$ has maximum degree at most $|S|/|I| \leq \varepsilon|S|$. Thus $G[S]$ is an ε-restricted induced subgraph of G with at least $\varepsilon^{3a}|G|$ vertices. This proves Theorem 6.7. ■

Proof of Theorem 1.5. Let $a \geq 1$ be given by Lemma 6.4. It suffices to show that for every $\varepsilon \in (0, \frac{1}{2})$, every $\overline{P_5}$-free graph G has an ε-restricted induced subgraph with at least $\varepsilon^{3a}|G|$ vertices. Suppose not. By Lemma 6.4 with $x = \varepsilon$, for every induced subgraph F of G with $|F| \geq \varepsilon^{2a}|G|$, either:

- F has an ε-restricted induced subgraph with at least $\varepsilon^a|F| \geq \varepsilon^{3a}|G|$ vertices; or
- there is a complete or anticomplete $(k, |F|/k^a)$-blockade in F for some $k \in [2, 1/x]$.

Since the first bullet cannot hold by our supposition, the second bullet holds for every such induced subgraph F. Then Theorem 6.7 implies that G has an ε-restricted induced subgraph with at least $\varepsilon^{3a}|G|$ vertices, contrary to the supposition. This proves Theorem 1.5. ■

7. The viral property

A graph H is *viral* if there exists $d > 0$ such that for every $\varepsilon \in (0, \frac{1}{2})$, every graph G with fewer than $(\varepsilon^d|G|)^{|H|}$ copies of H contains an ε-restricted induced subgraph with at least $\varepsilon^d|G|$ vertices; in other words, H is viral if and only if it satisfies the polynomial form of Nikiforov's theorem [18]. Thus, all viral graphs have the polynomial Rödl property and hence the Erdős-Hajnal property. It was shown in this series that, conversely, all graphs known to have the Erdős-Hajnal property are indeed viral (except for the five-vertex cycle; that will be shown to be viral in Tung Nguyen's thesis [14]). Indeed, we showed in [15] that various new graphs have the Erdős-Hajnal property, and in that paper it was essential for inductive purposes to prove the stronger result that they were viral.

What about P_5? One can bootstrap the Erdős-Hajnal property of P_5 into its viral property, by adapting the arguments in Section 4 to sparse graphs with few copies of $\overline{P_5}$ and using an extension of Lemma 3.4 to sparse graphs with few copies of P_5. We omit the details, which will appear in [14].

References

[1] N. Alon, J. Pach, and J. Solymosi. Ramsey-type theorems with forbidden subgraphs. *Combinatorica*, 21(2):155–170, 2001. Paul Erdős and his mathematics (Budapest, 1999). 1

[2] P. Blanco and M. Bucić. Towards the Erdős–Hajnal conjecture for P_5-free graphs. *Res. Math. Sci.*, to appear, arXiv:2210.10755, 2022. 1

[3] N. Bousquet, A. Lagoutte, and S. Thomassé. The Erdős–Hajnal conjecture for paths and antipaths. *J. Combin. Theory Ser. B*, 113:261–264, 2015. 5

[4] M. Bucić, T. Nguyen, A. Scott, and P. Seymour. Induced subgraph density. I. A $\log\log$ step towards Erdős–Hajnal, *submitted*, arXiv:2301.10147, 2023. 1, 2

[5] M. Chudnovsky. The Erdős–Hajnal conjecture—a survey. *J. Graph Theory*, 75(2):178–190, 2014. 1

[6] M. Chudnovsky and S. Safra. The Erdős–Hajnal conjecture for bull-free graphs. *J. Combin. Theory Ser. B*, 98(6):1301–1310, 2008. 1

[7] M. Chudnovsky, A. Scott, P. Seymour, and S. Spirkl. Pure pairs. I. Trees and linear anticomplete pairs. *Adv. Math.*, 375:107396, 20, 2020. 2, 3

[8] M. Chudnovsky, A. Scott, P. Seymour, and S. Spirkl. Erdős–Hajnal for graphs with no 5-hole. *Proc. Lond. Math. Soc. (3)*, 126(3):997–1014, 2023. 1, 4, 5

[9] D. Conlon, J. Fox, and B. Sudakov. Recent developments in graph Ramsey theory. In *Surveys in combinatorics 2015*, volume 424 of *London Math. Soc. Lecture Note Ser.*, pages 49–118. Cambridge Univ. Press, Cambridge, 2015. 1

[10] P. Erdős and A. Hajnal. On spanned subgraphs of graphs. In *Contributions To Graph Theory And Its Applications (Internat. Colloq., Oberhof, 1977) (German)*, pages 80–96. Tech. Hochschule Ilmenau, Ilmenau, 1977. 1

[11] P. Erdős and A. Hajnal. Ramsey-type theorems. *Discrete Appl. Math.*, 25(1-2):37–52, 1989. 1

[12] J. Fox and B. Sudakov. Induced Ramsey-type theorems. *Adv. Math.*, 219(6):1771–1800, 2008. 2

[13] A. Gyárfás. Reflections on a problem of Erdős and Hajnal. In *The Mathematics Of Paul Erdős, II*, volume 14 of *Algorithms Combin.*, pages 93–98. Springer, Berlin, 1997. 1

[14] T. Nguyen. *Induced Subgraph Density*, Ph.D. thesis, Princeton University, May 2025, in preparation. 14, 15

[15] T. Nguyen, A. Scott, and P. Seymour. Induced subgraph density. IV. New graphs with the Erdős–Hajnal property, arXiv:2307.06455, 2023. 1, 9, 14

[16] T. Nguyen, A. Scott, and P. Seymour. Induced subgraph density. V. All paths approach Erdős–Hajnal, arXiv:2307.15032, 2023. 1

[17] T. Nguyen, A. Scott, and P. Seymour. Induced subgraph density. VI. Bounded VC-dimension, *manuscript*, 2023. 11

[18] V. Nikiforov. Edge distribution of graphs with few copies of a given graph. *Combin. Probab. Comput.*, 15(6):895–902, 2006. 14

[19] V. Rödl. On universality of graphs with uniformly distributed edges. *Discrete Math.*, 59(1-2):125–134, 1986. 1

[20] A. Scott. Graphs of large chromatic number. In *Proceedings of the International Congress of Mathematicians 2022. Vol. VI.*, pages 4660–4681. EMS Press, 2023. 1

Princeton University, Princeton, NJ 08544, USA
Email address: tunghn@math.princeton.edu

Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK
Email address: scott@maths.ox.ac.uk

Princeton University, Princeton, NJ 08544, USA
Email address: pds@math.princeton.edu