Sparse graphs with no polynomial-sized anticomplete pairs

Maria Chudnovsky*
Princeton University, Princeton, NJ 08544

Jacob Fox
Stanford University, Stanford, CA 94305-2125

Alex Scott†
Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK

Paul Seymour‡
Princeton University, Princeton, NJ 08544

Sophie Spirkl
Princeton University, Princeton, NJ 08544

December 31, 2017; revised June 16, 2018

Abstract

A graph is *H-free* if it has no induced subgraph isomorphic to H. An old conjecture of Conlon, Sudakov and the second author asserts that:

- For every graph H, there exists $\varepsilon > 0$ such that in every H-free graph with $n > 1$ vertices there are two disjoint sets of vertices, of sizes at least $\varepsilon n^{\varepsilon}$ and εn, complete or anticomplete to each other.

This is equivalent to:

- The "sparse linear conjecture": For every graph H, there exists $\varepsilon > 0$ such that in every H-free graph with $n > 1$ vertices, either some vertex has degree at least εn, or there are two disjoint sets of vertices, of sizes at least $\varepsilon n^{\varepsilon}$ and εn, anticomplete to each other.

We prove a number of partial results towards the sparse linear conjecture. In particular, we prove it holds for a large class of graphs H, and we prove that something like it holds for all graphs H. More exactly, say H is "almost-bipartite" if H is triangle-free and $V(H)$ can be partitioned into a stable set and a set inducing a graph of maximum degree at most one. (This includes all graphs that arise from another graph by subdividing every edge at least once.) Our main result is:

- The sparse linear conjecture holds for all almost-bipartite graphs H.

*Supported by NSF grant DMS-1550991. This material is based upon work supported in part by the U. S. Army Research Laboratory and the U. S. Army Research Office under grant number W911NF-16-1-0404.

†Supported by a Leverhulme Research Fellowship

‡Supported by ONR grant N00014-14-1-0084 and NSF grant DMS-1265563.

(It remains open when H is the triangle K_3.) There is also a stronger theorem:

- For every almost-bipartite graph H, there exist $\varepsilon, t > 0$ such that for every graph G with $n > 1$ vertices and maximum degree less than εn, and for every c with $0 < c \leq 1$, either G contains $\varepsilon c^t n^{|H|}$ induced copies of H, or there are two disjoint sets $A, B \subseteq V(G)$ with $|A| \geq \varepsilon c^t n$, $|B| \geq \varepsilon n$, and with at most $c|A| \cdot |B|$ edges between them.

We also prove some variations on the sparse linear conjecture, such as:

- For every graph H, there exists $\varepsilon > 0$ such that in every H-free graph with $n > 1$ vertices, either some vertex has degree at least εn, or there are two disjoint sets A, B of vertices with $|A| \cdot |B| \geq \varepsilon n^{1+\varepsilon}$, anticomplete to each other.

1 Introduction

All graphs in this paper are finite and have no loops or parallel edges. We denote the number of vertices of G by $|G|$, and $G[X]$ denotes the subgraph induced on $X \subseteq V(G)$. If G, H are graphs, we say that G *contains* H if some induced subgraph of G is isomorphic to H, and G is *H-free* otherwise. If $A, B \subseteq V(G)$ are disjoint, we say A is *complete* to B if every vertex in A is adjacent to every vertex in B, and *anticomplete* to B if there is no edge between A and B.

Erdős, Hajnal and Pach [7] proved:

1.1 *For every graph H there exists $\varepsilon > 0$, such that for every H-free graph G with $n > 1$ vertices, there are disjoint $A, B \subseteq V(G)$, complete or anticomplete, with $|A|, |B| \geq \varepsilon n^{\varepsilon}$.*

The goal of this paper is to strengthen 1.1. For instance, we shall prove in section 6 that:

1.2 *For every graph H there exists $\varepsilon > 0$ such that in every H-free graph G with $n > 1$ vertices, there are disjoint $A, B \subseteq V(G)$, complete or anticomplete, with $|A| \cdot |B| \geq \varepsilon n^{1+\varepsilon}$.*

We will also prove a number of other strengthenings of 1.1. Here are some ways in which we could try to modify it:

- Make $|B| \geq \varepsilon n$; conjecture 1.3 below says this is always possible. We cannot ask for both A, B to be linear, however; a random graph construction shows that this can only be true when both H and its complement $\overline{H}$ are forests, that is, H is an induced subgraph of a four-vertex path.
- Get more than two sets.
- Replace "H-free" by a weaker hypothesis, that there are not many copies of H in G.
- Generalize "complete or anticomplete" to "$(1 - c)$-dense or c-sparse" (these limit the number of edges between A, B).
- Assuming that G is "ε-bounded" (that is, its maximum degree is less than $\varepsilon|G| - 1$), eliminate the "complete" or "$(1 - c)$-dense" outcome.

We will prove various combinations of these. For instance, in 3.2 we satisfy the second, third and fourth bullets, and also the fifth in 5.4. Our main result, 7.6, satisfies the first, third, fourth and fifth bullets, but only when H is "almost-bipartite".

We need a few definitions. Let G be a graph. For a vertex $v \in V(G)$, we use $N(v)$ to denote its set of neighbours, and we define $N[v] = N(v) \cup \{v\}$. For $\varepsilon > 0$, a graph G is *ε-bounded* if $|N[v]| < \varepsilon|G|$ for all $v \in V(G)$. A pair (A, B) of subsets of $V(G)$ is an *(x, y)-pair* for $x, y \geq 0$, if $A \cap B = \emptyset$ and $|A| \geq x$ and $|B| \geq y$.

There is a conjecture of Conlon, Sudakov and the second author [4] that more than 1.2 is true, that we can make the larger of A, B linear in n:

1.3 Conjecture: *For every graph H there exists $\varepsilon > 0$ such that every H-free graph G with $n > 1$ vertices contains a complete or anticomplete $(\varepsilon n^{\varepsilon}, \varepsilon n)$-pair.*

If we restrict attention to ε-bounded graphs, then there cannot exist A, B as in 1.3 complete to each other; so 1.3 would imply:

1.4 Conjecture: *For every graph H there exists $\varepsilon > 0$ such that in every H-free ε-bounded graph G with $n > 1$ vertices, there is an anticomplete $(\varepsilon n^{\varepsilon}, \varepsilon n)$-pair.*

A theorem of Rödl [10] shows that a graph H satisfies 1.3 if and only both H and $\overline{H}$ satisfy 1.4. Thus 1.3 (for all H) is equivalent to 1.4 (for all H). On the other hand, for certain graphs H, 1.4 turns out to be much more tractable than 1.3.

We say H is *almost-bipartite* if it is triangle-free and its vertex set can be partitioned into a stable set and a set that induces a subgraph with maximum degree at most one. A consequence of our main result is:

1.5 *All almost-bipartite graphs H satisfy 1.4.*

(It remains open when H is a triangle, however.) The full conjecture 1.3 has not been proved for many graphs H. In [1], the authors prove that 1.3 holds for a five-cycle, but otherwise it has only been proved so far for graphs H that are induced subgraphs of a four-vertex path. A consequence of our results is that two more graphs satisfy 1.3, namely a four-cycle and its complement.

2 Density theorems

If $A, B \subseteq V(G)$ are disjoint, $E(A, B)$ denotes the set of edges of G with one end in A and one in B. For $c \geq 0$, a pair (A, B) of subsets of $V(G)$ is

- *c-sparse* if $A \cap B = \emptyset$ and $|E(A, B)| \leq c|A| \cdot |B|$; and
- *c-dense* if $A \cap B = \emptyset$ and $|E(A, B)| \geq c|A| \cdot |B|$.

The best general bound for the Erdős-Hajnal conjecture [5] to date was proved by Erdős and Hajnal in [6], namely:

2.1 *For every graph H, there exists $\varepsilon > 0$ such that for every H-free graph G with $n > 0$ vertices, some clique or stable set of G has cardinality at least $2^{\varepsilon\sqrt{(\log n)}}$.*

One of the key steps in proving this was to prove the following:

2.2 *For every graph H, there exist $\varepsilon, s > 0$ such that for every H-free graph G with $n > 1$ vertices, and every c with $0 \le c \le 1$, G contains either a c-sparse or a $(1-c)$-dense $(\varepsilon c^s n, \varepsilon c^s n)$-pair.*

Conlon, Sudakov and the second author [4] asked whether one of the sets A, B could always be chosen of linear size, independent of c: that is,

2.3 Conjecture: *For every graph H there exist $\varepsilon, s > 0$ such that for every H-free graph G on $n > 1$ vertices, and all c with $0 \le c \le 1$, G contains either a c-sparse or a $(1-c)$-dense $(\varepsilon c^s n, \varepsilon n)$-pair.*

For ε-bounded graphs this becomes:

2.4 Conjecture: *For every graph H there exist $\varepsilon, s > 0$ such that for every H-free ε-bounded graph G on $n > 1$ vertices and all c with $0 \le c \le 1$, there is a c-sparse $(\varepsilon c^s n, \varepsilon n)$-pair in G.*

There are implications between these conjectures; in fact 2.3 $\Rightarrow$ 1.3 $\Rightarrow$ 1.4 and 2.3 $\Rightarrow$ 2.4 $\Rightarrow$ 1.4. We have seen that 1.3 implies 1.4.

Proof of 2.4, assuming 2.3. Let ε' and s satisfy 2.3, and let $\varepsilon = \varepsilon'/2$. Let G be H-free, and let $0 < c \le 1$. We may assume $n \ge 2$. By 2.3, there exist disjoint A, B with $|A| \ge \varepsilon' c^s n$ and $|B| \ge \varepsilon' n$, such that (A, B) is either c-sparse or $(1-c)$-dense. If (A, B) is a c-sparse pair, then (A, B) satisfies 2.4 as required, because $\varepsilon' \ge \varepsilon$. If (A, B) is $(1-c)$-dense and not c-sparse, and so $\max(c, 1-c)$-dense and therefore at least $(1/2)$-dense, then some vertex in A has degree at least $|B|/2 \ge \varepsilon n$ and again 2.4 holds. ∎

Proof of 1.4, assuming 2.4. Let ε' and s satisfy 2.4; and let

$$\varepsilon = \min\left(\varepsilon'/2, 1/(s+1), 1/4\right).$$

We claim that ε satisfies 1.4.

Let G be ε-bounded and H-free with $n > 1$ vertices, and let $x = \varepsilon n^{\varepsilon}$. Choose c such that $c^s n = n^{\varepsilon}$, that is, $c = n^{-(1-\varepsilon)/s}$.

(1) *We may assume that $x \ge 1$, and $x \le \varepsilon' c^s n$, and $cx \le 1/4$.*

Let $v \in V(G)$. Since $|N[v]| < \varepsilon n$, and $\varepsilon \le 1/2$, it follows that v has at least εn non-neighbours; and since we may assume that v and its non-neighbours do not form an anticomplete $(\varepsilon n^{\varepsilon}, \varepsilon n)$-pair, it follows that $x \ge 1$. The second claim holds since $c^s n = n^{\varepsilon}$; and the third since $cx = n^{-(1-\varepsilon)/s}(\varepsilon n^{\varepsilon})$, and $\varepsilon - (1-\varepsilon)/s \le 0$ and $\varepsilon \le 1/4$. This proves (1).

Now $c \le 1$, so by 2.4, there is a c-sparse $(\varepsilon' c^s n, \varepsilon' n)$-pair (A, B). By (1) and 2.5 below, there is an anticomplete $(x, |B|/2)$-pair (A', B') with $A' \subseteq A$ and $B' \subseteq B$. But then (A', B') satisfies 1.4. This proves 1.4. ∎

(The proof that 2.3 implies 1.3 is similar and we omit it.) We just used a lemma that produces anticomplete pairs from c-sparse pairs:

2.5 *Let (A, B) be a c-sparse pair in a graph G. If $x \ge 1/2$ (not necessarily an integer), and $x \le |A|$ and $cx \le 1/4$, there is an anticomplete $(x, |B|/2)$-pair (A', B') with $A' \subseteq A$ and $B' \subseteq B$.*

Proof. Let $d = \lceil x \rceil$; then since $x \geq 1/2$ it follows that $d \leq 2x$. Since $x \leq |A|$ and hence $d \leq |A|$, there is a subset of A with cardinality d. By averaging over all such subsets, it follows that there exists $A' \subseteq A$ with $|A'| = d$ such that (A', B) is c-sparse. In particular, there are at most $cd|B| \leq 2cx|B| \leq |B|/2$ vertices in B with a neighbour in A'; let B' be the other vertices in B, and then the theorem holds. This proves 2.5. ∎

3 Saturation

For a graph H and a graph G, a *copy of H in G* is an isomorphism ϕ between H and an induced subgraph of G. (Thus, there are six copies of K_3 in K_3.) In particular, G contains H if and only if there is a copy of H in G. For $\alpha \geq 0$, we say that a graph G is *(α, H)-saturated* if there are at least $\alpha|G|^{|H|}$ copies of H in G.

With all these result and conjectures, one can try replacing "H-free" by "not (α, H)-saturated" for the appropriate choice of α. For instance, we mentioned earlier a theorem of Rödl [10]; it says that for all H and $\varepsilon > 0$, there exists $\delta > 0$ such that if G is H-free, there is an induced subgraph J with $|J| \geq \delta|G|$ such that one of $|E(J)|, |E(\overline{J})|$ is at most $\varepsilon|J|(|J| - 1)/2$. There is a saturation version of this, the following, due to Sudakov and the second author [8]:

3.1 *Let H be a graph, and let $\varepsilon > 0$. Then there exist $\alpha, \delta > 0$ such that for every graph G, if G is not (α, H)-saturated, then G contains an induced subgraph J with $|J| \geq \delta|G|$, such that one of $|E(J)|, |E(\overline{J})|$ is at most $\varepsilon|J|(|J| - 1)/2$.*

Similarly, one can strengthen 2.2. In fact we will prove the following in section 6; it strengthens 2.2 in two ways, replacing "H-free" by "not (α, H)-saturated" and producing k sets instead of two. ($\mathbb{N}$ denotes the set of non-negative integers.)

3.2 *For every graph H and $k \in \mathbb{N}$, there exist $\varepsilon, s, K > 0$ such that for every graph G with $n > K$ vertices, and every c with $0 \leq c \leq 1$, if G is not $(\varepsilon c^s, H)$-saturated, then there are pairwise disjoint subsets $A_1, \dots, A_k \subseteq V(G)$ such that either:*

- *(A_i, A_j) is a c-sparse $(\varepsilon c^s n, \varepsilon c^s n)$-pair for all distinct $i, j \in \{1, \dots, k\}$; or*
- *(A_i, A_j) is a $(1 - c)$-dense $(\varepsilon c^s n, \varepsilon c^s n)$-pair for all distinct $i, j \in \{1, \dots, k\}$.*

As usual, as we will prove in 5.4, if we require that G is ε-bounded, then we can omit the second outcome.

In light of this, one might try the saturation strengthenings of the two conjectures from the previous section. 2.3 could be strengthened to:

3.3 Conjecture: *For every graph H there exist $\varepsilon, s > 0$ such that for every graph G on $n > 1$ vertices, and all c with $0 \leq c \leq 1$, either G is $(\varepsilon c^s, H)$-saturated, or there is a c-sparse or a $(1 - c)$-dense $(\varepsilon c^s n, \varepsilon n)$-pair in G.*

Similarly, 2.4 could be strengthened to:

3.4 Conjecture: *For every graph H there exist $\varepsilon, s > 0$ such that for every ε-bounded graph G on n vertices and all c with $0 \leq c \leq 1$, either G is $(\varepsilon c^s, H)$-saturated, or there is a c-sparse $(\varepsilon c^s n, \varepsilon n)$-pair in G.*

As in section 2, we have the implication 3.3 ⇒ 3.4, and clearly 3.3 ⇒ 2.3, and 3.4 ⇒ 2.4. Moreover, 3.1 shows that H satisfies 3.3 if and only if both H and $\overline{H}$ satisfy 3.4.

We will prove 3.4 when H is almost-bipartite. Our main theorem (proved in section 7) says:

3.5 *For every almost-bipartite graph H there exist $\varepsilon, s > 0$ such that for every ε-bounded graph G on n vertices and all c with $0 \le c \le 1$, either G is $(\varepsilon c^s, H)$-saturated, or there is a c-sparse $(\varepsilon c^s n, \varepsilon n)$-pair in G.*

The graphs H such that both H and $\overline{H}$ are almost-bipartite are the five-cycle, the four-cycle, and its complement, as well as all induced subgraphs of these graphs. Therefore, our results imply that 3.3 holds for these graphs.

Finally, we remark that we cannot do better than "$(\varepsilon c^s, H)$-saturated" in 3.4, that is, 3.4 becomes false if we replace "$(\varepsilon c^s, H)$-saturated" by "(ε, H)-saturated". This can be seen by letting $H = K_2$. Let $\varepsilon, s > 0$; we will show they do not satisfy the modified 3.4. Let $n \in \mathbb{N}$, $\delta = 1/(2s+2)$, and $p = n^{-\delta/2}$, and let G be an n-vertex random graph in which every edge is present independently with probability p. It follows that G has $\approx \frac{1}{2}n^{2-\delta/2}$ edges in expectation, so for n sufficiently large, with high probability G is not (ε, H)-saturated. Also the probability that there is an anticomplete $(n^\delta, \frac{1}{2}\varepsilon n)$-pair in G is at most

$$3^n(1-p)^{\frac{1}{2}\varepsilon n^{1+\delta}} \le 3^n e^{-\frac{1}{2}\varepsilon n^{1+\delta/2}} \to 0$$

as $n \to \infty$; so for n large, with high probability, G has no anticomplete $(n^\delta, \frac{1}{2}\varepsilon n)$-pair.

Let $c = n^{-\delta}/4$. Since $n^\delta \ge 1$, and $n^\delta \le \varepsilon c^s n$ (for large n), and $cn^\delta = 1/4$, it follows from 2.5 (with $x = n^\delta$) that, if there is no anticomplete $(n^\delta, \frac{1}{2}\varepsilon n)$-pair in G, then there is also no c-sparse $(\varepsilon c^s n, \varepsilon n)$-pair in G. So with high probability, G is not (ε, H)-saturated and has no c-sparse $(\varepsilon c^k n, \varepsilon n)$-pair.

4 A game on a graph

Let H be a graph, and let $T \subseteq V(H)$ be a stable set. A graph H' is a *T-successor of H* if

- $V(H) = V(H')$ and H is a proper subgraph of H'; and
- every edge in $E(H') \setminus E(H)$ has both ends in T.

Let H be a graph. For $k \ge 2$ and $m \ge k$, the *k-tuple game for H* on m vertices is the following game between two players, A and B. Let G_0 be a graph with m vertices and no edges. Rounds of the game will add edges to G_0, making a sequence of graphs $G_1, G_2, \cdots$, all with the same vertex set and each a proper subgraph of the next. In round i, player A selects a stable set T of cardinality k in G_{i-1}, and player B choose a T-successor G_i of G_{i-1}. Player A wins if at some stage there is an induced subgraph isomorphic to H.

More precisely, during the ith round (starting with $i = 1$), the following moves take place in this order:

- if G_{i-1} contains H, player A has won; otherwise, if G_{i-1} has no stable set of cardinality k, then player B has won;

- if neither of these, player A chooses a k-vertex stable subset T of G_{i-1};
- player B chooses a T-successor G_i of G_{i-1}.

Then a new round commences. Since at least one edge is added in every round, this game terminates after a finite number of rounds.

For a graph H, we say that H is *(m, k)-forcible* if there is a strategy for player A to play the k-tuple game for H on m vertices and always win, that is, reach a graph G that contains H. We say that such a strategy *forces H*. The main result of this section is that for every H and k, there exists $m \geq 0$ such that H is (m, k)-forcible. We begin by proving the base cases:

4.1 *Every graph H is $(|H|, 2)$-forcible. If furthermore $|E(H)| = 0$, then H is $(|H|, k)$-forcible for all $k \geq 2$.*

Proof. The second statement holds since G_0 is isomorphic to H if H has no edges. For the first statement, player A picks a bijection f between $V(H)$ and $V(G_0)$, and player A ensures that in every round i, G_{i-1} is isomorphic to a (not necessarily induced) subgraph of H. Therefore either H is isomorphic to G_{i-1} or there is an edge $uv \in E(H)$ such that $f(u), f(v)$ are not adjacent in G_{i-1}. In the first case, player A stops the game; in the second, player A picks the set $\{f(u), f(v)\}$. This forces player B to add the edge $f(u)f(v)$. After $|E(H)|$ rounds, G_i has $|E(H)|$ edges and hence is isomorphic to H. ∎

4.2 *Let H_1, H_2 be graphs, and let $k \geq 2$ and $m_1, m_2 \in \mathbb{N}$ such that for all $i \in \{1, 2\}$, H_i is (m_i, k)-forcible. Then the disjoint union of H_1 and H_2 is $(m_1 + m_2, k)$-forcible.*

Proof. Let G_0 have $m_1 + m_2$ vertices, partitioned into V_1, V_2 with $|V_i| = m_i$ for $i \in \{1, 2\}$. Player A first plays according to the k-tuple game for H_1 on $G_0[V_1]$. Since H_1 is (m_1, k)-forcible, this game stops after a finite number s_1 of rounds and $G_{s_1-1}[V_1]$ contains H_1. Since every k-tuple picked by player A is contained in V_1, it follows that V_2 is stable and anticomplete to V_1 in G_{s_1-1}. Instead of stopping in round s_1, player A now plays according to the k-tuple game for H_2 on $G_{s_1-1}[V_2]$. After a finite number s_2 of rounds, $G_{s_1+s_2-1}[V_2]$ contains H_2, and V_1 remains anticomplete to V_2. But this implies that $G_{s_1+s_2-1}$ contains the disjoint union of H_1 and H_2. ∎

4.3 *Let H be a graph and $k \geq 2$. Then there exists $m \geq k$ such that H is (m, k)-forcible.*

Proof. We prove this by induction on $k + |E(H)|$. The statement holds in the base cases, when either $k = 2$ or $|E(H)| = 0$, by 4.1. Now let $k > 2$ and $|E(H)| > 0$; let $e = uv \in E(H)$ and suppose that $H \setminus \{e\}$ is (m_1, k)-forcible and that H is $(m_2, k-1)$-forcible. Let $m = m_1^{m_2}$. We claim that H is (m, k)-forcible.

For an integer $s \geq 1$, we say an *s-star* is a graph J with $s(|H| - 1) + 1$ vertices partitioned into sets $V_1, \ldots, V_s, \{w\}$ such that

- $|V_i| = |H| - 1$ for all $i \in \{1, \ldots, s\}$;
- V_i is anticomplete to V_j for all $i, j \in \{1, \ldots, s\}$ with $i \neq j$; and

- $J[V_i \cup \{w\}]$ is isomorphic to $H \setminus \{e\}$ and w maps to u under this isomorphism, for all $i \in \{1, \dots, s\}$.

The vertex w is called the *centre* of J.

(1) *For every $s \geq 1$, the s-star is (m_1^s, k)-forcible.*

We prove this by induction on s. For $s = 1$, this follows since $H \setminus \{e\}$ is (m_1, k)-forcible and is isomorphic to the 1-star. Now let $s > 1$. By induction, the $(s-1)$-star is (m_1^{s-1}, k)-forcible, and so by 4.2, the disjoint union of m_1 graphs, each an $(s-1)$-star, is (m_1^s, k)-forcible. Starting with a graph G_0 with m_1^s vertices and no edges, player A uses the strategy that forces this disjoint union, until at the end of round p, say, G_p has an induced subgraph with m_1 components $H_1, \dots, H_m$, each an $(s-1)$-star. For $i \in \{1, \dots, m_1\}$, let u_i denote the centre of H_i, and let $U = \{u_1, \dots, u_{m_1}\}$. It follows that $|U| = m_1$ and U is stable in G_p. By applying the strategy for $H \setminus \{e\}$ to U starting at round $p+1$, it follows that there is a strategy for player A that produces at the end of some round q a graph G_q containing an induced subgraph that consists of m_1 disjoint $(s-1)$-stars, pairwise anticomplete except for edges with both ends in a set U as defined above, and a subset $U' \subseteq U$ such that $G_q[U']$ is isomorphic to $H \setminus \{e\}$. Let $i \in \{1, \dots, m_1\}$ such that $u_i \in U'$ maps to u under the isomorphism between $G_p[U']$ and the graph $H \setminus \{e\}$. It follows that $U' \setminus \{u_i\}$ is anticomplete to $V(H_i) \setminus \{u_i\}$, and therefore $G_p[U' \cup V(H_i)]$ is an s-star. This proves (1).

By (1), it follows that an m_2-star is $(m_1^{m_2}, k)$-forcible. This implies that in the k-tuple game on m vertices, player A can guarantee that in some round s, G_s contains an m_2-star H' with vertex set $V_1 \cup \dots \cup V_{m_2} \cup \{w\}$ (with notation as before). Let w be the centre of H', and for $i \in \{1, \dots, m_2\}$, let v_i be the vertex corresponding to v in the isomorphism between $G_s[V_i \cup \{w\}]$ and $H \setminus \{e\}$. Let $V = \{v_1, \dots, v_{m_2}\}$. By the definition of an m_2-star, it follows that V is a stable set in G_s, and $|V| = m_2$. Now player A uses the winning strategy of the $(k-1)$-tuple game for H on m_2 vertices by starting with $G_s[V]$; except in every round, player A picks $T \cup \{w\}$ instead of the set $T \subseteq V$ of $(k-1)$ vertices that the strategy for the $(k-1)$-tuple game produces. If in round $s' > s$, player B adds an edge incident with w, say wv_i, then $G_{s'}[V_i \cup \{w\}]$ is isomorphic to H, and the result follows. Therefore, we may assume that in every round $s' > s$, player A picks a set $T \cup \{w\}$, and player B adds at least one edge with both ends in T. Since H is $(m_2, k-1)$-forcible, it follows that for some $s' > s$, $G_{s'}[V]$ contains H. This proves 4.3. ∎

5 Sparse k-tuples

In this section we prove a lemma that is used for all the difficult results of the paper, 5.4 below. Its proof needs the game from the previous section, and some other preliminaries.

5.1 *Let G, H be graphs, and let $\alpha > 0$ such that G is (α, H)-saturated. Let H' be an induced subgraph of H. Then G is (α, H')-saturated.*

Proof. Let $n = |G|$, and let $\mathcal{T}$ be the set of copies of H' in G. For every copy ϕ of H in G, $\phi|_{V(H')}$ is a copy of H' in G; we say that ϕ *came from* $\phi|_{V(H')}$. For every $\phi' \in \mathcal{T}$, there are at most $n^{|H|-|H'|}$

copies of H that came from ϕ' (since there are at most $n^{|H|-|H'|}$ ways to extend ϕ' from a function from $V(H')$ to $V(G)$, to a function from $V(H)$ to $V(G)$). It follows that

$$|\mathcal{T}| \geq n^{|H'|-|H|}\alpha n^{|H|} = \alpha n^{|H'|},$$

and so G is (α, H')-saturated. This proves 5.1. ∎

Let G be a graph with n vertices, and let $A_1, \ldots, A_k \subseteq V(G)$ be pairwise disjoint such that

- for all $i \in \{1, \ldots, k\}$, $|A_i| \geq \alpha n$; and
- for all $i, j \in \{1, \ldots, k\}$ with $i \neq j$, (A_i, A_j) is a c-sparse pair.

Then we call $A_1, \ldots, A_k$ a *c-sparse (α, k)-tuple* in G. Thus a c-sparse (x, x)-pair is a c-sparse $(x/n, 2)$-tuple.

5.2 *Let G, H be graphs. Let $S \subseteq V(H)$ be a stable set with $|S| = k \geq 2$; and let $0 \leq \alpha, c \leq 1$ such that G is (α, H)-saturated. Then either G contains a c-sparse $\left(\frac{\alpha}{2k^k}, k\right)$-tuple, or there is an S-successor H' of H such that G is $\left(\frac{c\alpha}{2^{k^2}k^k}, H'\right)$-saturated.*

Proof. Let $\ell = |H|$ and $n = |G|$. For a copy ϕ of H, we say that $\phi|_{H \setminus S}$ is the *anchor* of ϕ. We say that a copy ψ of $H \setminus S$ is *weighty* if there are at least $\frac{1}{2}\alpha n^k$ copies of H in G with anchor ψ. Let $\mathcal{T}$ be the set of weighty copies of $H \setminus S$. Since there are at most $n^{\ell-k}$ copies of $H \setminus S$, there are at most $\frac{1}{2}\alpha n^\ell$ copies of H whose anchors are not weighty. Consequently, there are at least $\frac{1}{2}\alpha n^\ell$ copies of H whose anchors are weighty. For $\psi \in \mathcal{T}$, let $\mathcal{A}(\psi)$ be the set of copies of H with anchor ψ. It follows that

$$\sum_{\psi \in \mathcal{T}} |\mathcal{A}(\psi)| \geq \alpha n^\ell / 2.$$

Let $S = \{s_1, \ldots, s_k\}$. For $i \in \{1, \ldots, k\}$, we let $U_i = \{\phi(s_i) : \phi \in \mathcal{A}(\psi)\}$. Now let $V_1, \ldots, V_k$ be a random partition of $U = U_1 \cup \cdots \cup U_k$ in which, for all $i \in \{1, \ldots, k\}$, every vertex of U is in V_i with probability $1/k$ independently. Let $\phi \in \mathcal{A}(\psi)$. It follows that the probability that $\phi(s_i) \in V_i$ for all $i \in \{1, \ldots, k\}$ is $1/k^k$. Therefore there is a choice of $V_1, \ldots, V_k$ such that

$$|\{\phi \in \mathcal{A}(\psi) : \phi(s_i) \in V_i \text{ for all } i \in \{1, \ldots, k\}\}| \geq |\mathcal{A}(\psi)|/k^k.$$

Fix such a choice of $V_1, \ldots, V_k$, and let $W_i = U_i \cap V_i$ for all $i \in \{1, \ldots, k\}$. It follows that

$$|W_1| \cdots |W_k| \geq \{\phi \in \mathcal{A}(\psi) : \phi(s_i) \in W_i \text{ for all } i \in \{1, \ldots, k\}\}| \geq |\mathcal{A}(\psi)|/k^k,$$

since $\phi(s_i) \in U_i$ for all $\phi \in \mathcal{A}(\psi)$ and $i \in \{1, \ldots, k\}$. Since $|\mathcal{A}(\psi)|/k^k \geq \alpha n^k k^{-k}/2$, and $|W_1| \cdots |W_k| \leq n^{k-1}|W_i|$, it follows that $|W_i| \geq \frac{\alpha}{2k^k} n$ for all $i \in \{1, \ldots, k\}$. If for all distinct $i, j \in \{1, \ldots, k\}$, (W_i, W_j) is a c-sparse pair, then $W_1, \ldots, W_k$ is a c-sparse $\left(\frac{c\alpha}{2k^k}, k\right)$-tuple, and 5.2 follows. Therefore, we may assume that there exist distinct $i, j \in \{1, \ldots, k\}$ such that (W_i, W_j) is not c-sparse.

Let $\mathcal{W}(\psi)$ be the set of all $w = (w_1, \ldots, w_k)$ such that $w_h \in W_h$ for all $h \in \{1, \ldots, k\}$, and w_i, w_j are adjacent. For $w = (w_1, \ldots, w_k) \in \mathcal{W}(\psi)$, let $\phi_w(v) = \psi(v)$ if $v \in V(H) \setminus S$, and $\phi_w(s_h) = w_h$ for all $h \in \{1, \ldots, k\}$. Since $w_h \in U_h$ for each i, and w_i, w_j are adjacent, it follows that ϕ_w is a copy of an S-successor of H.

Since the pair (W_i, W_j) is not c-sparse,

$$|\mathcal{W}(\psi)| \geq c|W_1| \cdots |W_k| \geq c|\mathcal{A}(\psi)|/k^k.$$

But $\sum_{\psi \in \mathcal{T}} |\mathcal{A}(\psi)| \geq \frac{1}{2}\alpha n^\ell$, and so

$$\sum_{\psi \in \mathcal{T}} |\mathcal{W}(\psi)| \geq \frac{c\alpha}{2k^k} n^\ell.$$

This implies that G contains at least $\frac{c\alpha}{2k^k} n^\ell$ copies of S-successors of H. Now H has at most 2^{k^2-1} distinct S-successors, since that bounds the number of distinct graphs on k vertices; and therefore, there is an S-successor H' of H such that G contains at least $2^{-k^2} k^{-k} \alpha c n^\ell$ copies of H'. It follows that G is $\left(\frac{c\alpha}{2^{k^2} k^k}, H'\right)$-saturated. This proves 5.2. ∎

5.3 *Let H be a graph with $|E(H)| = 0$, and let $\varepsilon > 0$ with $\varepsilon|H| \leq 1/2$. Let G be an ε-bounded graph. Then G is $\left(2^{-|H|}, H\right)$-saturated.*

Proof. We prove this by induction on $|H|$. For $|H| = 0$, H is the null graph, and so G is trivially $(1, H)$-saturated. Now let $|H| > 0$, and let $v \in V(H)$. From the inductive hypothesis, it follows that G is $\left(2^{-|H|+1}, H \setminus \{v\}\right)$-saturated. Let ϕ be a copy of $H \setminus \{v\}$ in G. Since G is ε-bounded, it follows that there are at most $\varepsilon(|H| - 1)|G| \leq |G|/2$ vertices of G that are equal to or adjacent to a vertex in the image of ϕ; and so there are at least $|G|/2$ that are not. Consequently there are at least $|G|/2$ vertices w such that the function ϕ_w is a copy of H, where $\phi_w(x) = \phi(x)$ if $x \in V(H) \setminus \{v\}$, and $\phi(v) = w$. Summing over ϕ, it follows that there are at least $2^{-|H|}|G|^{|H|}$ copies of H in G, and so G is $\left(2^{-|H|}, H\right)$-saturated. This proves 5.3. ∎

5.4 *Let $k \geq 2$ and $t \geq 0$ be integers and let H be a graph. Then there exist $S, \varepsilon > 0$ such that for every ε-bounded graph G, and for all c with $0 < c \leq 1$, either G is $\left(\varepsilon c^S, H\right)$-saturated, or G contains a c^{s+t}-sparse $(\varepsilon c^s, k)$-tuple for some $s \in \{0, \ldots, S\}$.*

Proof. Let $\ell = |H|$; we may assume that $\ell > 0$. By 4.3, it follows that there exists $L \in \mathbb{N}$ such that H is (L, k)-forcible, and so $L \geq \ell$. Let $r = \binom{L}{2} + 1$. For $i \in \{0, \ldots, r\}$, let $\delta_i = \left(\frac{1}{k^k 2^{k^2}}\right)^i 2^{-L}$; and let $\varepsilon = \min(1/(2L), \delta_r)$. Then $\delta_{i+1} \leq \frac{1}{k^k 2^{k^2}} \delta_i$ for each i. Let $S = 2^{r+t}$ and $s_i = 2^{i+t} - t$ for each i. Let G be ε-bounded.

Let H_0 be an L-vertex graph with no edges. By 5.3, it follows that G is (δ_0, H_0)-saturated. We will play the k-tuple game for H on L vertices starting with the graph H_0; and the graph passed to player A at the start of round i is denoted by H_i. Since every round (except the last) adds an edge, it follows that there are at most s rounds. Player A will use an optimal strategy, one that will guarantee that at some round, H_i will contain H. We will guide player B depending on the graph G, and when the game terminates we will obtain information about G.

We wish to arrange that for each round i of the game, G is $(\delta_{i-1} c^{s_{i-1}}, H_{i-1})$-saturated, and we will guide player B to arrange this. Suppose that this is true for $i - 1$. If player A stops the game in this round, then H_{i-1} contains H, and G is $(\delta_{i-1} c^{s_{i-1}}, H_{i-1})$-saturated. Since H is an induced subgraph of H_{i-1}, it follows that G is $\left(\varepsilon c^S, H\right)$-saturated by 5.1, and 5.4 follows. Therefore, we may assume that player A selects a stable subset T of $V(H_{i-1})$ of size k. We now apply 5.2 with $\alpha = \delta c^r$, $m = t + s_{i-1}$, $\delta = \delta_{i-1}$, $c = c^m$, and $s = s_{i-1}$, and deduce that either G has a $c^{t+s_{i-1}}$-sparse $(\delta_i c^{s_{i-1}}, k)$-tuple, and we are done, or, since $s_i = 2s_{i-1} + t$, there is an S-successor H_i of H_{i-1} such that G is $(\delta_i c^{s_i}, H')$-saturated, and player B returns H_i and the game continues. Since the k-tuple game terminates in at most r rounds, and since player A is using a strategy that forces H, it follows that when this k-tuple game terminates, in round $i \le r$ say, either G is $(\delta_i c^{s_i}, H')$-saturated, or G has a $c^{t+s_{i-1}}$-sparse $(\delta_i c^{s_{i-1}}, k)$-tuple. This proves 5.4. ∎

We remark that in the case of $k = 2$, we can bound s by $|E(H)|$:

5.5 *Let $k \in \mathbb{N}$ and let H be a graph with s edges. Then there exists $\varepsilon > 0$ such that for every ε-bounded graph G, and for all c with $0 \le c \le 1$, either G is $(\varepsilon c^s, H)$-saturated, or G contains a c-sparse $(\varepsilon c^s, 2)$-tuple.*

Proof. This follows by changing slightly the strategy of player B in the proof of 5.4: just maintain that G is $\left(\delta_0 c^i, H_i\right)$-saturated. The result follows since H is $(|H|, 2)$-forcible and player A has a strategy that forces H in s rounds. ∎

A special case of 5.4 is of interest and worth stating separately:

5.6 *Let $k \ge 2$ be an integer, and let H be a graph. Then there exist $s \in \mathbb{N}$ and $\varepsilon > 0$ such that for every ε-bounded graph G, and for all c with $0 \le c \le 1$, either G is $(\varepsilon c^s, H)$-saturated, or G contains a c-sparse $(\varepsilon c^s, k)$-tuple.*

Proof. By 5.4 with $t = 0$, there exist $S, \varepsilon > 0$ such that for every ε-bounded graph G, and for all c with $0 \le c \le 1$, either G is $\left(\varepsilon c^S, H\right)$-saturated, or G contains a c^r-sparse $(\varepsilon c^r, k)$-tuple for some $r \in \{0, \dots, S\}$. In the second case G contains a c-sparse $(\varepsilon c^S, k)$-tuple, so in both cases the theorem holds taking $s = S$. ∎

6 Excluding general graphs

The results of the previous section can be applied to deduce several results about excluding general graphs, that we obtain in this section. We need first:

6.1 *Let G be a graph with n vertices and at most $\varepsilon n(n-1)/2$ edges with $n \ge 2\varepsilon^{-1}$. Then there is an induced subgraph J with $|J| \ge n/2$ such that J is 2ε-bounded.*

Proof. Choose distinct $v_1, \dots, v_k \in V(G)$ with k maximum such that for $1 \le i \le k$, v_i has at least $2\varepsilon(n-i+1) - 1$ neighbours in $V(G) \setminus \{v_1, \dots, v_i\}$. Let $m = \lceil n/2 \rceil$. If $k \ge m$, then there are at least

$$\sum_{1 \le i \le m} (2\varepsilon(n-i+1) - 1) = 2\varepsilon m(n - (m-1)/2) - m \ge 3\varepsilon mn/2 - m \ge (3\varepsilon n/2 - 1)n/2$$

edges in G with an end in $\{v_1, \dots, v_m\}$. Consequently $(3\varepsilon n/2 - 1)n/2 \le \varepsilon n(n-1)/2$, contradicting that $n \ge 2\varepsilon^{-1}$. So $k \le n/2$. But from the maximality of k, $G[V(G) \setminus \{v_1, \dots, v_k\}]$ is 2ε-bounded, and therefore satisfies the theorem. ∎

We use 6.1 to prove a consequence of 3.1.

6.2 *For every graph H and every $\varepsilon > 0$ there exist $\alpha, \delta > 0$ such that for every graph G, if G is not (α, H)-saturated, then G either satisfies $|G| \le 4(\delta\varepsilon)^{-1}$ or contains an induced subgraph J with $|J| \ge \delta|G|$ and such that either J or $\overline{J}$ is ε-bounded.*

Proof. Let H be a graph, and let $\varepsilon > 0$. Let $\alpha, \delta > 0$ satisfy 3.1, with ε, δ replaced by $\varepsilon/2, 2\delta$ respectively. Now let G be a graph that is not (α, H)-saturated. By 3.1, G contains an induced subgraph J with $|J| \ge 2\delta|G|$ and such that either $|E(J)| \le \varepsilon|J|(|J|-1)/4$ or $|E(\overline{J})| \le \varepsilon|J|(|J|-1)/4$. In the first case, by 6.1, either $|J| < 4\varepsilon^{-1}$ and hence $|G| < 4(\varepsilon\delta)^{-1}$, or G contains an ε-bounded induced subgraph with at least $|J|/2 \ge \delta|G|$ vertices. In the second case we use the same argument in the complement. This proves 6.2. ∎

This is used to prove 3.2, which we restate:

6.3 *For every graph H and $k \in \mathbb{N}$, there exist $\varepsilon, s, K > 0$ such that for every graph G with $n > K$ vertices, and every c with $0 \le c \le 1$, if G is not $(\varepsilon c^s, H)$-saturated, then there are pairwise disjoint subsets $A_1, \dots, A_k \subseteq V(G)$ such that either:*

- *(A_i, A_j) is a c-sparse $(\varepsilon c^s n, \varepsilon c^s n)$-pair for all distinct $i, j \in \{1, \dots, k\}$; or*
- *(A_i, A_j) is a $(1-c)$-dense $(\varepsilon c^s n, \varepsilon c^s n)$-pair for all distinct $i, j \in \{1, \dots, k\}$.*

Proof. Let ε', s satisfy 5.6 both for H and for $\overline{H}$. Let $\alpha, \delta \le 1$ be as in 6.2 for H and ε'. Let $K = 4(\delta\varepsilon')^{-1}$. Let $\varepsilon = \min\left(\alpha, \varepsilon'\delta^{|H|}\right)$. We claim that s, K, ε satisfy the theorem.

Let G be a graph. By 6.2, it follows that either G is (α, H)-saturated (and thus (ε, H)-saturated), or $|G| \le 4(\delta\varepsilon')^{-1} = K$, or G contains an induced subgraph J with $|J| \ge \delta|G|$ such that either J or $\overline{J}$ is ε'-bounded. We may assume the third of these holds. Let $0 \le c \le 1$.

Suppose first that J is ε'-bounded. By 5.6, it follows that either J is $(\varepsilon' c^s, H)$-saturated (and so G is $\left(\varepsilon'\delta^{|H|}c^s, H\right)$-saturated), or J contains a c-sparse $(\varepsilon' c^s, k)$-tuple. We may assume the latter; but then G contains a c-sparse $(\varepsilon'\delta^{|H|}c^s, k)$-tuple, and 6.3 follows. In the case when $\overline{J}$ is ε'-bounded, we apply the same argument in the complement, using $\overline{H}$ instead of H. This proves 6.3. ∎

Before the next result we need two easy lemmas:

6.4 *Let $k \in \mathbb{N}$, let G be a graph, and let $P_1, \dots, P_k$ be pairwise disjoint subsets of $V(G)$, where $|P_i| = p_i$ for $1 \le i \le k$. Let $d_1, \dots, d_k$ be such that for all distinct $i, j \in \{1, \dots, k\}$, and every $v \in P_i$, v has at most d_j neighbours in P_j. Let $q \in \mathbb{N}$ with $((k-1)d_i + 1)q \le p_i$ for $1 \le i \le k$. Then there are subsets $Q_i \subseteq P_i$ for $1 \le i \le k$, each of cardinality q, and pairwise anticomplete.*

Proof. For $k = 0$ the result is vacuously true, so we assume that $k \ge 1$ and that the result holds for $k-1$. Choose $Q_k \subseteq P_k$ of cardinality q (this is possible since $q \le ((k-1)d_i + 1)q \le p_i$), and for $1 \le i \le k-1$ let P_i' be the set of vertices in P_i with no neighbour in Q_k. Thus $|P_i'| \ge p_i - qd_i \ge ((k-2)d_i + 1)q$ for $1 \le i \le k-1$, so from the inductive hypothesis there exist $Q_i \subseteq P_i'$ of cardinality q for $1 \le i \le k-1$, pairwise anticomplete; and they are all anticomplete to Q_k. This proves 6.4. ∎

This extends to:

6.5 *Let $k \in \mathbb{N}$, let G be a graph, let $0 \le c \le 1$, and let $P_1, \dots, P_k$ be disjoint subsets of $V(G)$, pairwise c-sparse. Let $|P_i| = p_i$ for $1 \le i \le k$. Let $q \in \mathbb{N}$ such that $2q(2(k-1)^2 cp_i + 1) \le p_i$ for $1 \le i \le k$. Then there exist $Q_i \subseteq P_i$ for $1 \le i \le k$, each of cardinality q, and pairwise anticomplete.*

Proof. We may assume that $c > 0$. For all distinct $i, j \in \{1, \dots, k\}$, let $B_{i,j}$ be the set of vertices $v \in P_i$ with more than $2(k-1)cp_j$ neighbours in B_j. Since there are at most cp_ip_j edges between P_i and P_j, it follows that $|B_{i,j}| \le cp_ip_j/(2(k-1)cp_j) = p_i/(2(k-1))$. By taking the union of $B_{i,j}$ for all $j \ne i$, we deduce that there are at most $p_i/2$ vertices in P_i that have more than $2(k-1)cp_j$ neighbours in B_j for some $j \ne i$; and so there are at least $p_i/2$ vertices in P_i that have at most $2(k-1)cp_j$ neighbours in P_j for each $j \ne i$. By 6.4, there are subsets $Q_i \subseteq P_i$ for $1 \le i \le k$, each of cardinality q, and pairwise anticomplete. This proves 6.5. ∎

The following result shows that there are k sets, each of size $\varepsilon n^{\varepsilon}$, and pairwise anticomplete, if we exclude a graph H as an induced subgraph of an ε-bounded graph. This is similar to 1.1, except that we assume sparsity and guarantee anticomplete sets; and we get more than two anticomplete sets.

6.6 *Let H be a graph and $k \ge 2$ be an integer. Then there exists $\varepsilon > 0$ such that in every ε-bounded H-free graph G with $|G| = n \ge 2$, there are k disjoint subsets of $V(G)$, pairwise anticomplete and each of cardinality at least $\varepsilon n^{\varepsilon}$.*

Proof. Let $\varepsilon', s > 0$ be as in 5.6. Let

$$\varepsilon = \min\left(\varepsilon'/8, \frac{1}{2|H|}, \frac{1}{16k^2}, \frac{1}{s+1}\right),$$

and let G be ε-bounded and H-free, and let $n = |G|$. By 5.3, it follows that G contains a stable set of size k; therefore, we may assume that $\varepsilon n^{\varepsilon} > 1$. Let $c = n^{-1/(s+1)}$.

By 5.6, it follows that either G is $(\varepsilon' c^s, H)$-saturated or G contains a c-sparse $(\varepsilon' c^s, k)$-tuple $A_1, \dots, A_k$, and since G is H-free, the latter holds. Let $q = \lceil \varepsilon n^{\varepsilon} \rceil$. By 6.5 it suffices to show that $2q(2(k-1)^2 cp_i + 1) \le p_i$ for $1 \le i \le k$, where $p_i = |A_i|$. Thus, it suffices to check that $2q(2(k-1)^2 cp_i) \le p_i/2$ and $2q \le p_i/2$, that is, $8qc(k-1)^2 \le 1$ and $4q \le \varepsilon' c^s n$. Since $\varepsilon n^{\varepsilon} > 1$, it follows that $q \le 2\varepsilon n^{\varepsilon}$; so it suffices to show that $16\varepsilon n^{\varepsilon} c(k-1)^2 \le 1$ and $8\varepsilon n^{\varepsilon} \le \varepsilon' c^s n$.

For the first, since $c = n^{-1/(s+1)}$, we must show that $16\varepsilon n^{\varepsilon} n^{-1/(s+1)}(k-1)^2 \le 1$, and this is true since $\varepsilon \le 1/(s+1)$ and $16\varepsilon(k-1)^2 \le 1$. For the second, we must show that $8\varepsilon n^{\varepsilon} \le \varepsilon' n^{1-s/(s+1)}$, and this is true since $\varepsilon \le 1/(s+1)$ and $\varepsilon \le \varepsilon'/8$. This proves 6.6. ∎

The next result is an improvement of 1.1 in the ε-bounded case.

6.7 *Let H be a graph. Then there exists $\varepsilon > 0$ such that if G is H-free and ε-bounded, then G has an anticomplete pair (A, B) with $|A| \cdot |B| \ge \varepsilon n^{1+\varepsilon}$.*

Proof. Let S, ε' be as in 5.4, setting $k = 2$ and $t = 1$. We may assume that $\varepsilon' \le 1/4$. Let

$$\varepsilon = \min\left((\varepsilon')^2/2, \frac{1}{2S+1}\right);$$

we claim that ε satisfies the theorem. Let G be ε-bounded, and let $|G| = n$. Let $c = n^{-1/(2S+1)}$. It follows from 5.4 that either G is $\left(\varepsilon' c^S, H\right)$-saturated, or G contains a c^{s+1}-sparse $(\varepsilon' c^s, 2)$-tuple for some $s \in \{0, \dots, S\}$. Since G is H-free, the latter holds, and so G contains a c^{s+1}-sparse $(\varepsilon' c^s n, \varepsilon' c^s n)$-pair (A, B) for some $s \in \{0, \dots, S\}$. Let $t = (s+1)/(2S+1)$, and $m = \lceil \varepsilon' n^t \rceil$.

(1) *We may assume that $\varepsilon' n^t \geq 1/2$, and $\varepsilon' n^t \leq |A|$, and $\varepsilon' n^t c^{s+1} \leq 1/4$.*

Let $v \in V(G)$; then we may assume that $|\{v\}| \cdot |V(G) \setminus N[v]| < \varepsilon n^{(1+\varepsilon)}$, for otherwise the theorem holds. But v has at least $(1-\varepsilon)n$ non-neighbours, so $(1-\varepsilon)n < \varepsilon n^{1+\varepsilon}$, and hence

$$n^\varepsilon > 1/\varepsilon - 1 \geq 1/\varepsilon' - 1 \geq 1/(2\varepsilon').$$

Consequently $\varepsilon' n^t \geq \varepsilon' n^\varepsilon \geq 1/2$, so the first holds. The second holds since $n^t \leq c^s n$, and the third since $n^t c^{s+1} = 1$. This proves (1).

By 2.5, taking $x = \varepsilon' n^t$ and with c replaced by c^{s+1}, there is an anticomplete $(\varepsilon' n^t, |B|/2)$ pair (A', B') with $A' \subseteq A$ and $B' \subseteq B$. Then

$$|A'| \cdot |B'| \geq \left(\varepsilon' n^t\right)\left(\frac{1}{2}\varepsilon' c^s n\right) = \frac{1}{2}\varepsilon'^2 n^{1+1/(2S+1)} \geq \varepsilon n^{1+\varepsilon}.$$

This proves 6.7. ∎

This implies 1.2, which we restate.

6.8 *For every graph H there exists $\varepsilon > 0$ such that in every H-free graph G with $n > 1$ vertices, there exist disjoint sets $A, B \subseteq V(G)$ that are complete or anticomplete to one another, with $|A| \cdot |B| \geq \varepsilon n^{1+\varepsilon}$.*

Proof. Let ε' satisfy 6.7 for both H and $\overline{H}$. Let δ satisfy 6.2 for ε' and H. Let

$$\varepsilon = \min\left(\varepsilon' \delta^{1+\varepsilon'}, (\delta\varepsilon')^2/16\right).$$

Let G be H-free, and let $n = |G| \geq 2$. By 6.2, it follows that either $n \leq 4/(\delta\varepsilon')$, or G contains an induced subgraph J with at least δn vertices such that one of J, $\overline{J}$ is ε'-bounded. If $n \leq 4/(\delta\varepsilon')$, then $n^{1+\varepsilon} \leq n^2 \leq \varepsilon^{-1}$ from the definition of ε. Choose distinct $u, v \in V(G)$, and then $\{u\}, \{v\}$ is an complete or anticomplete pair (A, B) with $|A| \cdot |B| = 1 \geq \varepsilon n^{1+\varepsilon}$, and 6.8 holds. Therefore, we may assume that G contains an induced subgraph J with at least δn vertices such that one of J, $\overline{J}$ is ε'-bounded. If J is ε'-bounded, then 6.7 implies that J contains an anticomplete pair (A, B) with

$$|A| \cdot |B| \geq \varepsilon' |J|^{1+\varepsilon'} \geq \varepsilon n^{1+\varepsilon}.$$

If $\overline{J}$ is ε'-bounded, we apply the same argument in the complement, using $\overline{H}$, obtaining a complete pair in G. This proves 6.8. ∎

We have given a long and complicated proof for 6.8, since it is a consequence of other results that we needed anyway; but 6.8 can be proved directly, much more easily, using a minor variant of the original proof of 1.1 by Erdős, Hajnal and Pach [7], as follows. We use the following lemma. For $k \ge 1$, define $e_k = 1 - 2^{1-k}$.

6.9 *Let H be a graph with $k \ge 1$ vertices $h_1, \dots, h_k$, and let $t \ge 5^{2^{k-2}}$ be a real number. Let G be a k-partite graph, with parts $V_1, \dots, V_k$, each of cardinality at least $5t^{e_k}$. Then either*

- *for $1 \le i \le k$ there exists $v_i \in V_i$ such that for $1 \le i < j \le k$, v_i, v_j are adjacent in G if and only if h_i, h_j are adjacent in H; or*
- *there exist i, j with $1 \le i < j \le k$ and subsets $A \subseteq V_i$ and $B \subseteq V_j$, such that A, B are complete or anticomplete to each other, and $|A| \cdot |B| \ge t$.*

Proof. We may assume that H is a complete graph, by replacing all edges between V_i, V_j by the bipartite complement if h_i, h_j are nonadjacent. If $k = 1$ the result is trivial. We assume $k > 1$ and proceed by induction on k.

Define $n = 5t^{e_k}$ and $d = 5t^{e_{k-1}}$. If there exists $v_1 \in V_1$ such that v_1 has at least d neighbours in each of $V_2, \dots, V_k$, then the result follows by induction (applied to $H \setminus \{h_1\}$ and the sets $N[v_1] \cap V_i$ $(2 \le i \le k)$), since $t^{2^{2-k}} \ge t^{2^{1-k}} \ge 5$.

So we may assume that each vertex in V_1 has fewer than d neighbours in one of $V_2, \dots, V_k$; and so we may assume that at least $n/(k-1)$ vertices in V_1 have fewer than d neighbours in V_2. Now since $t \ge 5^{2^{k-2}}$ by hypothesis, it follows that

$$t^{e_{k-1}} \ge 5^{2^{k-2}(1-2^{2-k})} = 5^{2^{k-2}-1} \ge k-1.$$

Consequently $n/(k-1) \ge n/(2d)$. Let x be an integer with $|x - n/(2d)| \le 1/2$; say $x = n/(2d) + p$, where $-1/2 \le p \le 1/2$. Choose a set $A \subseteq V_1$ with $|A| = x$, such that all its members have at most d neighbours in V_2. Let B be the set of vertices in V_2 with no neighbour in A; then $|B| \ge n - dx$. Now

$$|A| \cdot |B| \ge x(n - dx) = (n/(2d) + p)(n - d(n/(2d) + p)) = n^2/(4d) - p^2 d = 5t/4 - p^2 d.$$

But $p^2 d \le t/4$ since $|p| \le 1/2$ and $d \le t$ from the hypothesis, and so $|A| \cdot |B| \ge t$. This proves 6.9. ∎

We deduce 6.8, slightly strengthened to the following:

6.10 *Let H be a graph with $k \ge 2$ vertices. Define $\sigma = 1/(2^{k-1} - 1)$. If G is an H-free graph with $n > 1$ vertices, there exist sets $A, B \subseteq V(G)$, complete or anticomplete to each other, with $|A| \cdot |B| \ge \frac{1}{45} n^{1+\sigma}$.*

Proof. Since $n \ge 1$, there is a vertex either with at least $\lfloor n/2 \rfloor$ neighbours or at least $\lfloor n/2 \rfloor$ non-neighbours, and so we may assume that $\lfloor n/2 \rfloor < \frac{1}{45} n^{1+\sigma}$. Now $n/3 \le \lfloor n/2 \rfloor$, since $n \ge 2$, and so $1/3 < \frac{1}{45} n^{\sigma}$, that is, $n > 15^{1/\sigma}$. Let $t = \frac{1}{45} n^{1+\sigma}$. If $t \le 5^{2^{k-2}}$ then

$$\frac{1}{45} n^{1+\sigma} \le 5^{2^{k-2}} \le 15^{1/\sigma}/3$$

and so $n^{1+\sigma} \le 15^{1/\sigma+1}$, contradicting that $n > 15^{1/\sigma}$. Thus $t > 5^{2^{k-2}}$.

If $n \le 5kt^{e_k} + k$, then since $n/(6k) \le (n-k)/(5k)$ (because $n > 15^{1/\sigma}$ and $15 \ge (6k)^\sigma$), it follows that $n/(6k) \le t^{e_k} = 45^{-e_k} n^{(1+\sigma)e_k} = 45^{-e_k} n$, so $45^{-e_k} \ge 1/6$, a contradiction. Thus $n > 5kt^{e_k} + k$, and so we can divide the vertex set of G into k sets $V_1, \dots, V_k$ each of cardinality at least $5t^{e_k}$.

From 6.9 applied to the corresponding k-partite graph, there are sets $A, B \subseteq V(G)$, complete or anticomplete to each other, with $|A| \cdot |B| \ge t$, as required. ∎

7 Excluding almost-bipartite graphs

We recall that a graph H is *almost-bipartite* if H is triangle-free and there is a partition of $V(H)$ into A, B such that A is a stable set and $H[B]$ is a graph with maximum degree one. We call such a pair (A, B) an *almost-bipartition.*

In this section we prove the main theorem of the paper. It says that for every almost-bipartite graph H, there are $s, \varepsilon > 0$ such that for every ε-bounded graph G, and all $0 < c \le 1$, if G is not $(\varepsilon c^s, H)$-saturated then G has a c-sparse $(\varepsilon c^s n, \varepsilon n)$-pair. The difference with 5.5 is that the latter only tells us that G contains a c-sparse $(\varepsilon c^s n, \varepsilon c^s n)$-pair; and so far we only know how to prove the stronger statement for almost-bipartite graphs. Before we begin on the proof (which is elaborate), it might be helpful if we sketch the main ideas.

Let us see how to do it if H is actually bipartite rather than just almost-bipartite. Let A, B be a bipartition, and choose $\varepsilon > 0$ very small and s very large, in terms of H. Now let G be ε-bounded, and let $0 \le c \le 1$, and assume G has no c-sparse $(\varepsilon c^s n, \varepsilon n)$-pair; we need to prove G is $(\varepsilon c^s, H)$-saturated. From 5.6, we can arrange the constants such that G contains a c-sparse $(\delta c^{s_1}, |A|)$-tuple (for some constant s_1 much smaller than s). So, take such an a-tuple $C_1, \dots, C_a$ say, where $a = |A|$. From now on we will only count copies of H where for each i, the vertex representing the ith vertex of A is contained in C_i, and hope this will already give us enough copies. This a-tuple is c-sparse, so if we pick a vertex from each at random, then with high probability the transversal we generate is stable. This property is crucial. However, we are going to need to shrink the sets to a small fraction of their original size (scaled by powers of c) and these shrunken sets may be very dense to one another, and we might lose the crucial property that transversals are mostly stable. We can avoid this by choosing the original sets more carefully, using 5.4 with some large value of t, instead of just 5.6; so let us do that instead. Now the edges between the C_i's will give us no further trouble.

Most of the vertices of G lie in none of $C_1, \dots, C_a$ (we can prove that all the C_i's have cardinality at most εn); and each vertex in C_i is only adjacent to at most εn of the outside vertices. Consequently most of the outside vertices are only adjacent to at most $2\varepsilon |C_i|$ vertices in C_i; discard the others. Actually, just discard those adjacent to more than $2a\varepsilon |C_i|$ vertices in C_i; we can afford to do this for each i and still keep a good fraction of the outside vertices.

If D is the set of surviving vertices outside $C_1, \dots, C_a$, the pair (C_i, D) is not a c-sparse $(\varepsilon c^s n, \varepsilon n)$-pair, so most vertices outside have at least $c|C_i|/(2a)$ neighbours in C_i; discard those that do not, for each i. Thus D still contains a constant fraction of the original vertices of G, where the constant depends on H (actually, just on a) but not on c. Pick any one of those vertices, decide it is going to represent the first vertex b_1 say of B, and shrink all the sets C_i so that v is complete to some of them and anticomplete to the others, according to the vertices in A that b_1 is adjacent to in H. Now repeat for b_2, and so on; the sets C_i are shrinking by factors of c at each stage, but the number of choices for the next vertex in B remains linear in n independent of c. This would prove that G is

$(\varepsilon c^s, H)$-saturated when H is bipartite, since the shrunken sets still have the property that random transversals are mostly stable.

How can we modify the proof to work when H is almost-bipartite? Let (A, B) be the almost-bipartition. We start almost the same, applying 5.4 to get a c^{s_1+t}-sparse $(\delta c^{s_1}, 6|A|)$-tuple, $C_1, \ldots, C_{6a}$. (Note the 6.) Previously we filled in the vertices of B one at a time, proving there were linearly many choices at each step. Now we fill in the edges of $H[B]$ one at a time, that is, we will add the vertices of B two at a time in adjacent pairs. (We can assume that $H[B]$ is a perfect matching.) As before, we can arrange that random transversals of the C_i's are mostly stable, even after shrinking the C_i's by factors of c; and that every vertex outside has a decent number of neighbours in each C_i (not too large and not too small). So any vertex outside can play the role of any one vertex of B, but how do we get an edge of outside vertices to represent an edge of H? If we hope an edge uv of outside vertices could represent an edge b_1b_2 of H, we need u to have many neighbours nonadjacent to v in certain of the C_i's (because b_1 has certain neighbours nonadjacent to b_2 in H), and vice versa; and it need not have any such neighbours. We don't know how to control things directly in this way.

On the other hand, we can get many ($\Omega(n^2)$, regarding c as a constant) edges uv of outside vertices that "disagree" in this way for at least one-sixth of the values of i. The bad news is, we can't control which one-sixth of the values this is. But there is also good news; on that one-sixth of the values, we can shrink the sets C_i to make the adjacency to uv whatever we want (except, no triangles). Slightly more exactly, we can arrange that for many edges uv of outside vertices, there exists $I \subseteq \{1, \ldots, 6a\}$ with $|I| = a$, such that for each $i \in I$ there are many vertices in C_i adjacent to u and not v, *and* many adjacent to v and not u. (We won't need both sets for a given value of i, because H is triangle-free; but we don't yet know which set we will need.) So, pick one of these pairs uv; for the five-sixth of the values of i not in I, we shrink C_i to make it anticomplete to both u, v; and for each $i \in I$, we can shrink the sets C_i to make C_i complete to u and anticomplete to v, or vice versa, or anticomplete to both; whichever we want.

Which should we choose? We are given the power to add an edge with any adjacency we like to I, but with no control over the set I. We can make this work as follows. If we can turn the a sets corresponding to I into a (blowup of) a copy of H by using the edge uv and shrinking the C_i's for $i \in I$ appropriately (and using some of the edges added at earlier steps), do so; and if not, make I closer to being part of a blowup of H. After adding a bounded number of edges (making sure all the added edges are anticomplete to one another, and adding each to be as useful as possible in this way), there must be a blowup of H, because at every step, some a-subset of $\{1, \ldots, 6a\}$ gets closer to being in a blowup of H, and there are only $\binom{6a}{a}$ such subsets. That is the proof; we would have shown that there are $\Omega(n^{2k})$ induced k-edge matchings in G (for some constant k), each including a submatching of size $b/2$ that for some $I \subseteq \{1, \ldots, 6a\}$ with $|I| = a$, has the correct adjacency to a big subset of C_i for each $i \in I$, and which therefore extends to $\Omega(n^a)$ copies of H.

Let us say this carefully. Let G be a graph, and let $C, D \subseteq V(G)$ be disjoint. A pair (R, S) of disjoint subsets of D is C-*split* if $|N(u) \cap C| \leq |N(v) \cap C|$ for each $u \in R$ and $v \in S$.

7.1 *Let G be a graph and $d, k \in \mathbb{N}$. Let $C_1, \ldots, C_k, D \subseteq V(G)$ with $|D| \geq 2^k d$. Then there is a set $I \subseteq \{1, \ldots, k\}$ with $|I| \geq k/2$, and disjoint subsets A, B of D with $|A|, |B| = d$ such that (A, B) is C_i-split for all $i \in I$.*

Proof. We start by proving the following claim.

(1) *Let G be a graph, and $d, k \geq 1$ be integers. Let $C_1, \ldots, C_k, D \subseteq V(G)$ with $|D| \geq 2^k d$. Then there are disjoint subsets A, B of D with $|A|, |B| = d$ such that for all $i \in \{1, \ldots, k\}$, either (A, B) or (B, A) is C_i-split.*

We prove this by induction on k. For $k = 1$, we have $|D| \geq 2d$; choose distinct $v_1, \ldots, v_{2d} \in D$, numbered such that $|C_1 \cap N[v_j]| \geq |C_1 \cap N[v_i]|$ for all i, j with $1 \leq i < j \leq 2d$. Then $A = \{v_1, \ldots, v_d\}$ and $B = \{v_{d+1}, \ldots, v_{2d}\}$ is the desired partition.

Now let $k > 1$. From the inductive hypothesis, with d replaced by $2d$, it follows that there are disjoint subsets A, B of D with $|A| = |B| = 2d$, such that for all $i \in \{1, \ldots, k-1\}$, either (A, B) or (B, A) is C_i-split. Choose $j \in \mathbb{N}$ maximum such that

- the number of vertices in A with at most $j - 1$ neighbours in C_k is at most d, and
- the number of vertices in B with at most $j - 1$ neighbours in C_k is at most d.

(This is possible, because for $j = 0$ both bullets are true, and for $j \geq 2d$ both are false, since $d \geq 1$.) By exchanging A, B if necessary, we may assume that the number of vertices in A with at most j neighbours in C_k is more than d. Consequently there is a subset $A' \subseteq A$ with $|A'| = d$ such that each vertex in A' has at most j neighbours in C_k; and a subset $B' \subseteq B$ with $|B'| = d$ such that each vertex in B' has at least j neighbours in C_k. Thus (A', B') is C_k-split, and for $1 \leq i \leq k-1$, either (A', B') or (B', A') is C_i-split. This proves (1).

Now we apply (1). We may assume that $d, k \geq 1$; we apply (1) to $C_1, \ldots, C_k$ and obtain A, B as in (1). Then one of $(A, B), (B, A)$ satisfy the theorem. This proves 7.1. ∎

7.2 *Let $0 < \varepsilon \leq 1/4$, let $s \in \mathbb{N}$ with $k \geq 1$, let $0 \leq c \leq 1$, and let G be a ε-bounded graph with $n > 0$ vertices, and with no c-sparse $(\varepsilon c^s n, \varepsilon n)$-pair. Then $n > \varepsilon^{-1}$, and $c < 2\varepsilon$.*

Proof. Let v be a vertex of G. Then $1 \leq |N[v]| < \varepsilon n$ since G is ε-bounded, and the first statement follows.

For the second statement, since $\varepsilon \leq 1/4$, it follows that $n > 4$. Let $A \subseteq V(G)$ with $|A| = \lfloor n/2 \rfloor$. Since $|A| \geq n/4 \geq \varepsilon c^s n$, it follows that $(A, V(G) \setminus A)$ is not a c-sparse pair, and so $|E(A, V(G) \setminus A)| > c|A| \cdot |V(G) \setminus A|$. Therefore, there is a vertex v in A with at least $c|V(G) \setminus A|$ neighbours in $V(G) \setminus A$, so v has degree at least $cn/2$. Since every vertex has degree less than εn, it follows that $cn/2 < \varepsilon n$, and the second statement follows. This proves 7.2. ∎

This is used to prove the main lemma, that we can get many edges that disagree on one-sixth of the values of i.

7.3 *Let $k, s \geq 1$ be integers, let $0 < \varepsilon \leq k^{-1} 2^{-k-5}$, and let G be an ε-bounded graph with n vertices. Let $0 < c \leq 1$, such that G does not have a c-sparse $(\varepsilon c^s n, \varepsilon n)$-pair. Let $C_1, \ldots, C_k, D$ be pairwise disjoint subsets of $V(G)$, such that $|C_i| \geq 2\varepsilon c^s n$ for $1 \leq i \leq k$, and $|D| \geq 3n/4$. Let E^* be the set of edges uv of $G[D]$ such that for at least $k/6$ values of $i \in \{1, \ldots, k\}$, there are at least $c^2|C_i|$ vertices in C_i adjacent to u and not to v, and at least $c^2|C_i|$ adjacent to v and not to u. Then $|E^*| \geq c^2 2^{-2k-9} n^2$.*

Proof. For $1 \le i \le k$, let D_i be the set of vertices in D with at most $c|C_i|$ neighbours in C_i; and let F_i be the set with at least $8k\varepsilon|C_i|$ neighbours in C_i. Then $|D_i| \le \varepsilon n \le n/(8k)$, since (C_i, D_i) is not a c-sparse $(\varepsilon c^s n, \varepsilon n)$-pair; and $|F_i| \le n/(8k)$, since

$$8k\varepsilon|C_i| \cdot |F_i| \le |E(C_i, D)| \le \varepsilon|C_i|n.$$

It follows that $|D_i \cup F_i| \le n/(4k)$ for $1 \le i \le k$, and so

$$D' = D \setminus \bigcup_{i\in\{1,\dots,k\}} (D_i \cup F_i)$$

satisfies $|D'| \ge |D| - n/4 \ge n/2$. Let $d = \lfloor 2^{-k-2}n \rfloor$. Since $2^{-k-2}n \ge \varepsilon n > 1$ by 7.2, it follows that $d \ge 2^{-k-3}n$.

Since $|D'| \ge 2^{k+1}d$, 7.1 implies there exist disjoint subsets $A, B \subseteq D'$ with $|A| = |B| = 2d$, and $I \subseteq \{1, \dots, k\}$ with $|I| \ge k/2$, such that (B, A) is C_i-split for all $i \in I$.

For $uv \in E(G)$ with $u, v \in D'$, and $i \in I$, we write $u \to_i v$ if $|(N(v) \setminus N(u)) \cap C_i| < c^2|C_i|$ (note that possibly $u \to_i v$ and $v \to_i u$ both hold). If neither of $u \to_i v$, $v \to_i u$ hold, we say uv is *i-incomparable*. If there are at least $|I|/3$ values of $i \in I$ such that $u \to_i v$, we write $u \to v$. (Again, possibly $u \to v$ and $v \to u$ both hold.) Thus, any edge uv that is i-incomparable for at least $|I|/3$ values of $i \in I$ belongs to E^*, and in particular, any edge uv for which $u \not\to v$ and $v \not\to u$ belongs to E^*.

(1) *We may assume that there is a vertex $u \in A$ such that the set $U = \{v \in A : u \to v\}$ satisfies $|U| \ge cd/4$.*

Choose disjoint subsets A_1, A_2 of A, both of cardinality d. Since $d \ge \varepsilon n$, and G has no c-sparse $(\varepsilon c^s n, \varepsilon n)$-pair, it follows that the pair (A_1, A_2) is not c-sparse. Hence $|E(A_1, A_2)| \ge cd^2$. For each edge uv with $u \in A_1$ and $v \in A_2$, either $u \to v$, or $v \to u$, or $uv \in E^*$; and if $|E^*| \ge c^2 2^{-2k-9}n^2$ we are done. So we may assume (exchanging A_1, A_2 if necessary) that there are at least $(cd^2 - c^2 2^{-2k-9}n^2)/2 \ge cd^2/4$ edges uv with $u \in A_1$ and $v \in A_2$, such that $u \to v$. Hence there exists $u \in A_1$ such that $u \to v$ for at least $cd/4$ values of $v \in A_2$. This proves (1).

Choose u and U as in (1). For $i \in I$, since $u \notin F_i$ and $i \in I$, it follows that $|N(u) \cap C_i| \le 8k\varepsilon|C_i|$; and consequently

$$|C_i \setminus N(u)| \ge (1 - 8k\varepsilon)|C_i| \ge \varepsilon c^s n.$$

For $i \in I$, let B_i be the set of vertices in B with at most $c|C_i \setminus N(u)|$ neighbours in $C_i \setminus N(u)$. Since $(C_i \setminus N(u), B_i)$ is a c-sparse pair, and $|C_i \setminus N(u)| \ge \varepsilon c^s n$, it follows that $|B_i| \le \varepsilon n$, for all $i \in I$. Let $B' = B \setminus \left(\bigcup_{i\in I} B_i\right)$; then $|B'| \ge |B| - k\varepsilon n \ge d/2$. Since $|U| \ge cd/4 \ge \varepsilon c^s n$ and $|B'| \ge \varepsilon n$, it follows that U, B' is not c-sparse, and so

$$|E(U, B')| \ge c|U| \cdot |B'| \ge c^2d^2/8 \ge c^2 2^{-2k-9}n^2.$$

We claim that $E(U, B') \subseteq E^*$, and the result will follow.

(2) $E(U, B') \subseteq E^*$.

Let vw be an edge with $v \in U$ and $w \in B'$; and let $I' = \{i \in I : u \to_i v\}$. We need to show that $vw \in E^*$; and to prove this, it suffices to show that for each $i \in I'$, $v \not\rightarrow_i w$ and $w \not\rightarrow_i v$, and so v, w are i-incomparable. Thus, let $i \in I'$. Now u has at most $8k\varepsilon|C_i| \le |C_i|/2$ neighbours in C_i, since $u \notin F_i$; and so $|C_i \setminus N(u)| \ge |C_i|/2$. But w has at least $c|C_i \setminus N(u)| \ge c|C_i|/2$ neighbours in $C_i \setminus N(u)$ since $w \in B'$, and v has at most $c^2|C_i|$ neighbours in $C_i \setminus N(u)$ since $u \rightarrow_i v$. Hence there are at least $c|C_i|/2 - c^2|C_i| \ge c^2|C_i|$ vertices in C_i that are adjacent to w and not to v, since $c < 2\varepsilon \le 1/4$ by 7.2, and it follows that $v \not\rightarrow_i w$. Moreover, we recall that (B, A) is C_i-split since $i \in I' \subseteq I$; and since $v \in A, w \in B$, it follows that $|N(w) \cap C_i| \le |N(v) \cap C_i|$. Consequently

$$|(N(v) \setminus N(w)) \cap C_i| \ge |(N(w) \setminus N(v)) \cap C_i| \ge c^2|C_i|,$$

and so $w \not\rightarrow_i v$. This proves (2).

From (2), $|E^*| \ge |E(U, B')| \ge c^2 2^{-2k-9} n^2$. This proves 7.3. ∎

An *induced matching* in a graph G is a subset $M \subseteq E(G)$, such that for all distinct $e, f \in M$, e, f have no common end and both ends of e are nonadjacent to both ends of f. We write $V(M)$ to denote the set of ends of members of M. A *blockade* in a graph G is a set $\mathcal{C} = \{C_1, \dots, C_k\}$ of pairwise disjoint subsets of $V(G)$. (We used the same term to mean something slightly different in [3].) We write $V(\mathcal{C}) = C_1 \cup \dots \cup C_k$. If $C'_i \subseteq C_i$ for $1 \le i \le k$, we call $\mathcal{C}' = \{C'_1, \dots, C'_k\}$ a *contraction* of $\mathcal{C}$; and for $\delta > 0$, if $|C'_i| \ge \delta|C_i|$ for $1 \le i \le k$. we call $\mathcal{C}'$ a *δ-contraction* of $\mathcal{C}$.

Now let M be an induced matching in $G \setminus V(\mathcal{C})$. We say $\mathcal{C}$ is *M-pure* if

- for $1 \le i \le k$, and for each $v \in V(M)$, v is either complete or anticomplete to C_i; and
- for each $e = uv \in M$ and $1 \le i \le k$, not both u, v are complete to C_i.

Let $\mathcal{C}$ be M-pure. For each $e = uv \in M$, let P, Q, R be respectively the sets of $i \in \{1, \dots, k\}$ such that u is complete to C_i, v is complete to C_i, and neither; then (P, Q, R) is a partition of $\{1, \dots, k\}$, and we call (P, Q, R) and (Q, P, R) the *supports* of $e = uv$. The number of distinct supports of edges in M is called the *richness* of M on $\mathcal{C}$. (More precisely, the richness is the number of partitions (P, Q, R) of M such that (P, Q, R) is a support of an edge of M.)

If $\mathcal{C} = \{C_1, \dots, C_k\}$ is a blockade and $I \subseteq \{1, \dots, k\}$, then $\{C_i : i \in I\}$ is called a *sub-blockade.* If $\mathcal{C}$ is M-pure, then so are its sub-blockades. We say M is *complete* on $\mathcal{C}$ if every partition of $\{1, \dots, k\}$ into three parts is the support of an edge of M.

Let $a \in \mathbb{N}$, and let $k = 6a$; and let $\mathcal{C} = \{C_1, \dots, C_k\}$ be a blockade. Then the *worth* of M on $\mathcal{C}'$ is ∞ if M is complete on some sub-blockade of cardinality a; and otherwise the worth of M is the sum, over all sub-blockades $\mathcal{C}'$ of cardinality a, of the richness of M on $\mathcal{C}'$. Thus, the worth of M is either less than $3^a\binom{6a}{a}$ or ∞.

Let us say an induced matching M in $G \setminus V(\mathcal{C})$ is *$\mathcal{C}$-successful* if there is an M-pure $c^{2|M|}$-contraction $\mathcal{C}'$ of $\mathcal{C}$, such that M has worth at least $|M|$ over $\mathcal{C}'$.

7.4 *Let $a \ge 1$ be an integer, let $k = 6a$, and let $R = 3^a\binom{k}{a}$; let $s \ge 2R$ be an integer, and let $0 < \varepsilon \le 6^{-k}$. Let G be an ε-bounded graph with n vertices. Let $0 < c \le 1$, such that G does not have a c-sparse $(\varepsilon c^s n, \varepsilon n)$-pair. Let $\mathcal{C} = \{C_1, \dots, C_k\}$ be a blockade in G, such that $2\varepsilon c^{s-2R} n \le |C_i| \le \varepsilon n$ for $1 \le i \le k$. Then for $0 \le m \le R$ there are at least $2^{-(2k+9)m} c^{2m} n^{2m}/m!$ $\mathcal{C}$-successful induced matchings in $G \setminus V(\mathcal{C})$ of cardinality m.*

Proof. We show first:

(1) *Let $0 < m \leq 3^a\binom{6a}{a}$, and let M be a $\mathcal{C}$-successful induced matching in $G \setminus V(\mathcal{C})$ of cardinality $m-1$. Then there are at least $c^2 2^{-2k-9} n^2$ $\mathcal{C}$-successful induced matchings in $G \setminus V(\mathcal{C})$ of cardinality m that include M.*

Since M is $\mathcal{C}$-successful, there is an M-pure $c^{2|M|}$-contraction $\mathcal{C}' = (C_1', \dots, C_k')$ of $\mathcal{C}$, such that M has worth at least $|M|$ on $\mathcal{C}'$. Let D be the set of all vertices in $V(G) \setminus V(\mathcal{C})$ that are anticomplete to $V(M)$. Thus $|D| \geq n - (k+2m)\varepsilon n$, since each $|C_i| \leq \varepsilon n$ and each vertex in $V(M)$ has at most εn neighbours; and so $|D| \geq 3n/4$, since $(k+2m)\varepsilon \leq 1/4$ because $m \leq 3^a\binom{6a}{a}$ and $\varepsilon \leq 6^{-k}$. Now each $|C_i'| \geq c^{2m-2}|C_i| \geq 2\varepsilon c^s n$, since $m \leq R$ and $s \geq 2R$. Since $(k+2m)\varepsilon n \leq n/4$, and $\varepsilon \leq 6^{-k} \leq k^{-1}2^{-k-5}$, we can apply 7.3 to $\mathcal{C}'$ and D. With E^* defined as in 7.3, we deduce that $|E^*| \geq c^2 2^{-2k-9} n^2$. Let $e = uv \in E^*$. We claim that $M \cup \{e\}$ is a $\mathcal{C}$-successful matching.

From the definition of E^*, there exists $I \subseteq \{1, \dots, k\}$ with $|I| = a$ such that for each $i \in I$, there are at least $c^2|C_i|$ vertices in C_i adjacent to u and not to v, and at least $c^2|C_i|$ adjacent to v and not to u. There are two cases:

- If M has worth ∞ on $\mathcal{C}'$, then for each $i \in \{1, \dots, k\}$, let C_i'' be the set of vertices in C_i' nonadjacent to both u, v, and $\mathcal{C}'' = \{C_1'', \dots, C_k''\}$. Then $|C_i''| \geq |C_i'| - 2\varepsilon n \geq c^2|C_i'|$ for each i, and $\mathcal{C}''$ is $(M \cup \{e\})$-pure; and $M \cup \{e\}$ has worth ∞ on $\mathcal{C}'' = \{C_1'', \dots, C_k''\}$.
- If M has finite worth on $\mathcal{C}'$, then in particular M is not complete on the sub-blockade $\{C_i' : i \in I\}$ of $\mathcal{C}'$. Choose a partition (P, Q, R) of I that is not a support of any edge in M; for each $i \in P$, let C_i'' be the set of vertices in C_i' adjacent to u and not to v; for each $i \in Q$, let C_i'' be the set of vertices in C_i' adjacent to v and not to u; and for each $i \in \{1, \dots, k\} \setminus (P \cup Q)$, let C_i'' be the set of vertices in C_i' nonadjacent to both u, v. Then again, for each i, $|C_i''| \geq c^2|C_i'|$, and $\mathcal{C}''$ is $(M \cup \{e\})$-pure. Moreover, the worth of $M \cup \{e\}$ on $\mathcal{C}'' = \{C_1'', \dots, C_k''\}$ is strictly more than the worth of M on $\mathcal{C}'$ (because the richness on the sub-blockade defined by I increased).

In both cases, $\mathcal{C}''$ is a c^2-contraction of $\mathcal{C}'$ and hence a c^{2m}-contraction of $\mathcal{C}$; and the worth of $M \cup \{e\}$ on $\mathcal{C}''$ is either ∞, or at least one more than the worth of M on $\mathcal{C}'$ and so in either case $M \cup \{e\}$ has worth at least m on $\mathcal{C}''$. Consequently $M \cup \{e\}$ is a $\mathcal{C}$-successful matching, for at least $c^2 2^{-2k-9} n^2$ edges e. This proves (1).

For $m \geq 0$, let f_m denote the number of $\mathcal{C}$-successful induced matchings of cardinality m. We must show that $f_m \geq 2^{-(2k+9)m} c^{2m} n^{2m}/m!$. We proceed by induction on m; the claim is true if $m = 0$, so we assume $m > 0$ and the claim holds for $m-1$. Since every m-edge matching includes only m matchings of cardinality $m-1$, it follows from (1) that

$$mf_m \geq c^2 2^{-2k-9} n^2 f_{m-1} \geq c^2 2^{-2k-9} n^2 2^{-(2k+9)(m-1)} c^{2m-2} n^{2m-2}/(m-1)!,$$

and so $f_m \geq 2^{-(2k+9)m} c^{2m} n^{2m}/m!$. This proves 7.4. ∎

We also need a lemma about sparse k-tuples.

7.5 *Let G be a graph, let $c > 0$, and let $k \geq 2$ be an integer. Let $C_1, \dots, C_k$ be pairwise disjoint subsets of $V(G)$, pairwise c-sparse. Then there exist $B_i \subseteq C_i$ for $1 \leq i \leq k$ such that $|B_i| \geq |C_i|/2$ for $1 \leq i \leq k$, and for all distinct $i, j \in k$, every vertex in B_i has at most $4kc|B_j|$ neighbours in B_j.*

Proof. For each $j \neq i$, let $A_{i,j}$ be the set of vertices in C_i with at least $2kc|C_j|$ neighbours in C_j. Since there are only $c|C_i|\cdot|C_j|$ edges between C_i and C_j, it follows that $|A_{i,j}| \leq c|C_i|\cdot|C_j|/(2kc|C_j|) = |C_i|/(2k)$, and so the union of all the sets $A_{i,j} (j \neq i)$ has cardinality at most $|C_i|/2$. Let B_i be its complement in C_i. Thus $|B_i| \geq |C_i|/2$, and for all distinct i, j, each vertex in B_i has at most $2kc|C_j| \leq 4kc|B_j|$ neighbours in C_j, and therefore has at most that many in B_j. This proves 7.5. ▮

Now we are ready to prove our main theorem:

7.6 *Let H be almost-bipartite. Then there exist $s \in \mathbb{N}$ and $\varepsilon > 0$ such that for all c with $0 < c \leq 1$, if G is ε-bounded, then either G is $(\varepsilon c^s, H)$-saturated, or G has a c-sparse $(\varepsilon c^s n, \varepsilon n)$-pair.*

Proof. There is a "universal" almost-bipartite graph, defined as follows. Let $a > 0$ be an integer; and let A be a set of cardinality a. For each partition (P, Q, R) of A, where either $P = \emptyset$ or the least member of P is less than all members of Q, take an edge uv of new vertices, and make u adjacent to all $p \in P$ and v adjacent to all $q \in Q$. We call the result H_a; it is almost-bipartite, and it is easy to see that every almost-bipartite graph is an induced subgraph of H_a for some a. (The restriction on (P, Q, R) is to ensure that no two of the added edges have a common support.) Thus, to prove 7.6 in general, it suffices to prove it when $H = H_a$, for each a.

Let $k = 6a$, $R = 3^a\binom{k}{a}$, and $t = 2R + 1$. Let 5.4 hold with k, t, H, S, ε replaced by $k, t, H_a, S, \varepsilon'$ respectively. Define $p = 2^{-(2k+9)R}/R!$, and $q = (\varepsilon'/2)^k/2$. Let $s = 2R + k(2R + S)$ and

$$\varepsilon = \min\left(pq, \varepsilon'/2, 1/(8k^2a^2), 6^{-k}\right).$$

We claim that s, ε satisfy the theorem. For let $0 < c \leq 1$, and let G be ε-bounded, with $n > 0$ vertices.

We may assume that G has no c-sparse $(\varepsilon c^s n, \varepsilon n)$-pair; and now we must show that G is $(\varepsilon c^s, H)$-saturated. From 7.2, $c < 2\varepsilon$. We may assume that G is not $\left(\varepsilon' c^S, H\right)$-saturated, since $S \leq s$ and $\varepsilon' \geq \varepsilon$; so from 5.6, G contains a $c^{s'+t}$-sparse $(\varepsilon' c^{s'}, k)$-tuple $C_1, \ldots, C_k$ for some $s' \leq S$.

By 7.5, with c replaced by $c^{s'+t}$, there exists $B_i \subseteq C_i$ with $|B_i| \geq |C_i|/2$ for $1 \leq i \leq k$, such that for all distinct $i, j \in \{1, \ldots, k\}$, every vertex in B_i has at most $4kc^{s'+t}|B_j|$ neighbours in B_j. Let $\mathcal{B} = (B_1, \ldots, B_k)$. For all distinct $i, j \in \{1, \ldots, k\}$, since G has no c-sparse $(\varepsilon c^s n, \varepsilon n)$-pair, and the pair (B_1, B_2) is c-sparse, and $|B_i| \geq \varepsilon' c^{s'} n/2 \geq \varepsilon c^s n$ (since $\varepsilon \leq \varepsilon'/2$ and $s' \leq S \leq s$), it follows that $|B_j| \leq \varepsilon n$; and so $|B_1|, \ldots, |B_k| \leq \varepsilon n$. By 7.4, there are at least $pc^{2R}n^{2R}$ $\mathcal{B}$-successful induced matchings in $G \setminus V(\mathcal{B})$ of cardinality R.

(1) *For every $\mathcal{B}$-successful induced matching M in $G \setminus V(\mathcal{C})$ of cardinality R, there are at least $qc^{k(2R+S)}n^k$ choices of $(x_1, \ldots, x_k)$, such that $x_i \in C_i$ for $1 \leq i \leq k$, and the subgraph induced on*

$$V(M) \cup \{x_1, \ldots, x_k\}$$

contains H_a.

Let M be a $\mathcal{B}$-successful induced matching of cardinality R. Hence there is an M-pure c^{2R}-contraction $\mathcal{A} = (A_1, \ldots, A_k)$ of $\mathcal{B}$, such that M has worth at least $|M|$ on $\mathcal{A}$. Since for all distinct $i, j \in \{1, \ldots, k\}$, every vertex in B_i has at most $4kc^{s'+t}|B_j|$ neighbours in B_j, there are at most

$$4kc^{s'+t}|A_i| \cdot |B_j| \leq 4kc^{s'+t-2R}|A_i| \cdot |A_j| \leq 4kc|A_i| \cdot |A_j|$$

edges between A_i and A_j, since $s' + t - 2R \geq 1$; that is, $A_1, \dots, A_k$ is $4kc$-sparse.

Now since $|M| = R$, it follows that M has worth ∞ on $\mathcal{A}$, and so there exists $I \subseteq \{1, \dots, k\}$ with $|I| = a$ such that M is complete on the sub-blockade $(A_i : i \in I)$. Let N be the product of the cardinalities of $A_1, \dots, A_k$. There are N choices of a sequence $(x_1, \dots, x_k)$ such that $x_i \in A_i$ for $1 \leq i \leq k$. For all distinct $i, j \in I$, there are only $4kcN$ such choices in which x_i, x_j are adjacent, since (A_i, A_j) is $4kc$-sparse. Hence there are at least $(1 - 4ka^2c)N$ choices such that $\{x_i : i \in I\}$ is stable; and since $1 - 4ka^2c \geq 1/2$ (because $c < 2\varepsilon \leq 1/(8k^2a^2)$), this number is at least $N/2$. For each such choice of $(x_1, \dots, x_k)$, the subgraph induced on $V(M) \cup \{x_i : i \in I\}$ contains H_a, since M is complete on the sub-blockade $(A_i : i \in I)$. Since for each i, $|A_i| \geq c^{2R}|B_i| \geq c^{2R+s'}\varepsilon' n/2$, and $s' \leq S$, it follows that

$$N \geq c^{k(2R+S)}(\varepsilon'/2)^k n^k = 2qc^{k(2R+S)}n^k,$$

and this proves (1).

Multiplying the number of choices for M and the number of choices for $(x_1, \dots, x_k)$ (for each M), we deduce that altogether there are at least $pqc^{2R+k(2R+s')r}n^{2R+k}$ distinct induced subgraphs of G, each with $k + 2R$ vertices, and each containing H_a. Since $pq \geq \varepsilon$ and $2R + k(2R + s') \leq s$, it follows that there are at least $\varepsilon c^s n^{2R+k}$ such subgraphs. But each induced subgraph of G isomorphic to H_a is contained in at most $n^{k+2R-|H_a|}$ induced subgraphs of G with $k + 2R$ vertices, and so there are at least $\varepsilon c^s n^{|H_a|}$ distinct copies of H_a in G. This proves 7.6. ∎

Subdividing an edge uv of a graph H' means replacing uv by a path, whose internal vertices are not in $V(H')$ and have degree two in the new graph. A graph H is a *(≥ 1)-subdivision* of a graph H' if H arises from H' by subdividing each edge at least once; that is, replacing each edge (one at a time) by a path of length at least two. 7.6 implies the following.

7.7 *Let H' be a graph, and let H be a (≥ 1)-subdivision of H'. Then there exist $s \in \mathbb{N}$ and $\varepsilon > 0$ such that for all $c > 0$, if G is ε-bounded and H-free, then G has a c-sparse $(\varepsilon c^t n, \varepsilon n)$-pair.*

Proof. By 7.6, it suffices to prove that H is almost-bipartite. We first note that since H is a (≥ 1)-subdivision of some graph, it follows that H arises from some graph H' by subdividing every edge of H' either once or twice.

Now $V(H')$ is a stable set in H, and every vertex of $H \setminus V(H')$ has degree at most one in $H \setminus V(H')$ (since it is part of a path of length two or three with ends in $V(H')$). Since H is a (≥ 1)-subdivision of H', it follows that H is triangle-free, and therefore H is almost-bipartite. This proves 7.7. ∎

References

[1] M. Chudnovsky, J. Fox, A. Scott, P. Seymour and S. Spirkl. "Towards Erdős-Hajnal for graphs with no 5-hole", submitted for publication (manuscript March 2018).

[2] M. Chudnovsky, A. Scott, P. Seymour and S. Spirkl, "Sparse graphs without linear anticomplete pairs", submitted for publication (manuscript November 2017).

[3] M. Chudnovsky, A. Scott, P. Seymour and S. Spirkl, "Trees and linear anticomplete pairs", submitted for publication (manuscript March 2018).

[4] D. Conlon, J. Fox and B. Sudakov, "Recent developments in graph Ramsey theory", `https://arxiv.org/pdf/1501.02474.pdf`, problem 3.13.

[5] P. Erdős and A. Hajnal, "On spanned subgraphs of graphs", *Graphentheorie und Ihre Anwendungen* (Oberhof, 1977), www.renyi.hu/~p_erdos/1977-19.pdf.

[6] P. Erdős and A. Hajnal, "Ramsey-type theorems", *Discrete Applied Mathematics* **25** (1989), 37–52.

[7] P. Erdős, A. Hajnal and J. Pach, "A Ramsey-type theorem for bipartite graphs", *Geombinatorics* 10 (2000), 64–68.

[8] J. Fox and B. Sudakov, "Induced Ramsey-type theorems", *Advances in Mathematics*, 219(6) (2008), 1771–1800.

[9] A. Liebenau, M. Pilipczuk, P. Seymour and S. Spirkl, "Caterpillars in Erdős-Hajnal", submitted for publication (manuscript October 2017).

[10] V. Rödl, "On universality of graphs with uniformly distributed edges", *Discrete Math.* **59** (1986), 125–134.