Induced subgraph density. III. The pentagon and the bull

Tung Nguyen[1]
Princeton University,
Princeton, NJ 08544, USA

Alex Scott[2]
Oxford University,
Oxford, UK

Paul Seymour[3]
Princeton University,
Princeton, NJ 08544, USA

July 4, 2022; revised August 2, 2023

[1]Supported by AFOSR grants A9550-19-1-0187 and FA9550-22-1-0234, and by NSF grants DMS-1800053 and DMS-2154169.

[2]Supported by EPSRC grant EP/X013642/1

[3]Supported by AFOSR grants A9550-19-1-0187 and FA9550-22-1-0234, and by NSF grants DMS-1800053 and DMS-2154169.

Abstract

A theorem of Rödl says that for every graph H, and every $\varepsilon > 0$, there exists $\delta > 0$ such that if G is a graph that has no induced subgraph isomorphic to H, then there exists $X \subseteq V(G)$ with $|X| \geq \delta|G|$ such that one of $G[X], \overline{G}[X]$ has at most $\varepsilon\binom{|X|}{2}$ edges. But for fixed H, how does δ depends on ε?

If the dependence is polynomial, then H satisfies the Erdős-Hajnal conjecture; and Fox and Sudakov conjectured that the dependence is polynomial for *every* graph H. This conjecture is substantially stronger than the Erdős-Hajnal conjecture itself, and until recently it was not known to be true for any non-trivial graphs H. The preceding paper of this series showed that it is true for P_4, and all graphs obtainable from P_4 by vertex-substitution.

Here we will show that the Fox-Sudakov conjecture is true for all the graphs H that are currently known to satisfy the Erdős-Hajnal conjecture. In other words, we will show that it is true for the bull, and the 5-cycle, and induced subgraphs of them, and all graphs that can be obtained from these by vertex-substitution.

There is a strengthening of Rödl's theorem due to Nikiforov, that replaces the hypothesis that G has no induced subgraph isomorphic to H, with the weaker hypothesis that the density of induced copies of H in G is small. We will prove the corresponding "polynomial" strengthening of Nikiforov's theorem for the same class of graphs H.

1 Introduction

Some terminology and notation: $G[X]$ denotes the induced subgraph with vertex set X of a graph G; $|G|$ denotes the number of vertices of G; $\overline{G}$ is the complement graph of G; and a graph is *H-free* if it has no induced subgraph isomorphic to H. The *edge-density* of a graph G is its number of edges divided by $\binom{|G|}{2}$.

An important theorem of Rödl [15] says:

1.1 *For every graph H and every $\varepsilon > 0$, there exists $\delta > 0$ such that for every H-free graph G, there exists $X \subseteq V(G)$ with $|X| \geq \delta|G|$ such that one of $G[X], \overline{G}[X]$ has edge-density at most ε.*

How does δ depend on ε, for a given graph H? Fox and Sudakov [10] studied this, and proposed the conjecture (conjecture 7.1 in their paper) that the dependence is polynomial, or more exactly:

1.2 Conjecture: *For every graph H there exists $c > 0$ such that for every ε with $0 < \varepsilon \leq 1/2$ and every H-free graph G, there exists $X \subseteq V(G)$ with $|X| \geq \varepsilon^c|G|$ such that one of $G[X], \overline{G}[X]$ has edge-density at most ε.*

With Jacob Fox, we showed in [11] that 1.2 holds for all graphs H that can be obtained by vertex-substitution starting from graphs with at most four vertices. But which other graphs H satisfy 1.2? This is closely connected with the Erdős-Hajnal conjecture [8, 9], that:

1.3 Conjecture: *For every graph H there exists $c > 0$ such that for every H-free graph G, there exist a clique or stable set X of G with with $|X| \geq |G|^c$.*

Fox and Sudakov showed that every graph H satisfying their conjecture 1.2 also satisfies the Erdős-Hajnal conjecture 1.3, but what about the converse? Can we show that all graphs H that satisfy 1.3 also satisfy 1.2? In general, no, we have no idea how to do this. But for all the graphs H that are *currently known* to satisfy 1.3, we will show that they also satisfy 1.2.

We need to define "vertex-substitution" before we go on. Let H_1, H_2 be graphs, let $v \in V(H_1)$, and let N be the set of all neighbours of v in H_1. Let H be obtained from the disjoint union of $H_1 \setminus \{v\}$ and H_2 by making every vertex of H_2 adjacent to every vertex in N. Then H is obtained by *substituting H_2 for the vertex v of H_1*, and this operation is called *vertex-substitution*.

Until very recently, the only graphs that were known to satisfy 1.3 were the *bull* [4] (that is, the graph obtained from a four-vertex path by adding a fifth vertex adjacent to the two middle vertices of the path); the five-vertex cycle C_5 [5]; all induced subgraphs of these two; and all graphs that can be obtained from these by repeated vertex-substitution [2]. We will show that all these graphs satisfy 1.2, that is:

1.4 *If H can be constructed from induced subgraphs of the bull and C_5 by repeated vertex-substitution, then H satisfies 1.2.*

It is no longer true that the graphs of 1.4 are the only graphs known to satisfy the Erdős-Hajnal conjecture. In a forthcoming paper [12], we will give infinitely more graphs that are "prime" (that is, cannot be obtained by vertex-substitution from smaller graphs), and that satisfy 1.3. But we show in that paper that these new graphs also satisfy 1.2.

There is a proof given in [5] that shows that both the bull and C_5 satisfy 1.3, and we will modify it to show that they satisfy 1.2. To finish the proof of 1.4, it would then be enough to show that the class of graphs that satisfy 1.2 is closed under vertex-substitution. We do not know how to do that, but all is not lost: we will prove that the graphs in 1.4 have a property even stronger stronger than 1.2, which *is* closed under vertex-substitution. Let us explain.

A *copy* of H in G is an isomorphism from H to an induced subgraph of G. Let $\operatorname{ind}_H G$ be the number of copies of H in G. There is a theorem of Nikiforov [14], strengthening Rödl's theorem:

1.5 *For every graph H and all $\varepsilon > 0$, there exists $\delta > 0$ such that for every graph G, if $\operatorname{ind}_H(G) \leq \delta|G|^{|H|}$, then there exists $S \subseteq V(G)$ with $|S| \geq \delta|G|$ such that one of $G[S], \overline{G}[S]$ has edge-density at most ε.*

Again, one could ask how δ depends on ε. Let us say that H is *viral* if there exists $d > 0$ such that for every graph G and every ε with $0 < \varepsilon \leq 1/2$, either

- there exists $X \subseteq V(G)$ with $|X| \geq \varepsilon^d|G|$ such that one of $G[X], \overline{G}[X]$ has edge-density at most ε; or
- $\operatorname{ind}_H(G) \geq \varepsilon^d|G|^{|H|}$.

Thus, the dependence of δ on ε in 1.5 is polynomial, if and only if H is viral. So, now we are considering three successively stronger properties of a graph H:

- H satisfies the Erdős-Hajnal conjecture;
- H satisfies the Fox-Sudakov conjecture;
- H is viral.

The third implies the second, and the second implies the first, but we do not know either of the converse implications. To appreciate the relative strength of these conjectures, it is instructive to think about the case when H is the four-vertex path P_4. Proving that P_4 satisfies the Erdős-Hajnal conjecture is easy; but proving that it satisfies the Fox-Sudakov conjecture is non-trivial, and this was proved in [11]. Proving that P_4 is viral is highly non-trivial. In [11] it was proved via a "polynomial removal lemma" for P_4, proved by Alon and Fox [1]. (It also is a special case of the results of this paper.)

On the other hand, perhaps all graphs H have all three of the properties above. In this paper we show that any graph currently known to have any one of these properties has all three (*pace* the results of [12]):

1.6 *If H can be constructed from induced subgraphs of the bull and C_5 by repeated vertex-substitution, then H is viral.*

There are two advantages of proving 1.6 rather than 1.4: we prove something stronger, but more importantly, viral graphs *are* closed under vertex-substitution, and all their induced subgraphs are also viral. The following was proved in [11]:

1.7 *If H_1, H_2 are viral and H is obtained by substituting H_2 for a vertex of H_1, then H is viral.*

So to prove 1.6 and hence 1.4, it suffices to show the following two results:

1.8 *The bull and C_5 are viral.*

1.9 *If H is viral then so are all its induced subgraphs.*

1.9 is a special case of 2.1, proved in section 2. The proof of 1.8 occupies most of the paper. We will approach it as follows. Let us say a set $\mathcal{H}$ of graphs is *viral* if there exists $d > 0$ such that for every x with $0 < x \le 1/2$, and for every graph G with $\mathrm{ind}_H(G) \le x^d|G|^{|H|}$ for each $H \in \mathcal{H}$, there exists $S \subseteq V(G)$ with $|S| \ge x^d|G|$ such that one of $G[S], \overline{G}[S]$ has edge-density at most x.

Let $\widehat{C_5}$ be the graph obtained from C_5 by adding a new vertex of degree two, adjacent to two adjacent vertices of C_5. We will prove that:

1.10 *The set $\{\widehat{C_5}, \overline{\widehat{C_5}}\}$ is viral.*

Since both of $\widehat{C_5}, \overline{\widehat{C_5}}$ contain both the bull and C_5 as induced subgraphs, this will imply 1.8. We will give a sketch of the proof of 1.10 in section 3.

For $k \ge 3$, C_k denotes the k-vertex cycle. It was shown in [5] that for $k = 6, 7$ there exists $c > 0$ such that for every graph G that is both C_k-free and $\overline{C_k}$-free, there exists a clique or stable set X of G with with $|X| \ge |G|^c$. When $k = 6$, this can be strengthened: the proof of 1.10 with minor modifications shows that:

1.11 *$\{C_6, \overline{C_6}\}$ is viral.*

We sketch the proof at the end of section 10. We have not yet been able to show that $\{C_7, \overline{C_7}\}$ is viral; but in a forthcoming paper [13], we shall prove that $\{C_k, \overline{C_\ell}\}$ is "near-viral" (that is, δ can be chosen as $2^{(\log \frac{1}{\varepsilon})^{1+o(1)}}$ in 1.5) for *any* two integers $k, \ell \ge 3$.

2 Viral sets

For every finite set $\mathcal{H}$ of non-null graphs, every $x > 0$, and every graph G, define

$$\mu_{\mathcal{H}}(x, G) := \max_{H \in \mathcal{H}} \frac{\mathrm{ind}_H(G)}{(x|G|)^{|H|}}.$$

We say $\mathcal{H}$ is *viral* if there exist $d > 0$ such that for every x with $0 < x \le 1/2$, and for every graph G with $\mu_{\mathcal{H}}(x^d, G) \le 1$ there exists $S \subseteq V(G)$ with $|S| \ge x^d|G|$ such that one of $G[S], \overline{G}[S]$ has edge-density at most x. It is easy to check that a graph H is viral if and only if $\{H\}$ is viral. We call d a *viral exponent* for $\mathcal{H}$. Here is a useful lemma for manipulating viral sets.

2.1 *Let $\mathcal{H}, \mathcal{J}$ be finite sets of non-null graphs, and suppose that each member of $\mathcal{J}$ has an induced subgraph isomorphic to a member of $\mathcal{H}$. If $\mathcal{J}$ is viral then $\mathcal{H}$ is viral.*

Proof. Let d be a viral exponent for $\mathcal{J}$. Choose p such that each $J \in \mathcal{J}$ has an induced subgraph isomorphic to some $H \in \mathcal{H}$ with $|J| \le p|H|$. We will show that $d' = pd$ is a viral exponent for $\mathcal{H}$. Let G be a graph and let $0 < x \le 1/2$ such that $\mathrm{ind}_H(G) \le (x^{d'}|G|)^{H|}$ for each $H \in \mathcal{H}$. Thus $\mathrm{ind}_H(G) \le x^{d'|H|}|G|^{|H|}$ for each $H \in \mathcal{H}$.

Let $J \in \mathcal{J}$. Then there exists $H \in \mathcal{H}$ isomorphic to an induced subgraph of J. Each copy of J in G is an extension of a copy of H in G, and each copy of H in G extends to at most $|G|^{|J|-|H|}$ copies of J; so

$$\mathrm{ind}_J(G) \le |G|^{|J|-|H|}\,\mathrm{ind}_H(G) \le |G|^{|J|-|H|}x^{d'|H|}|G|^{|H|} = x^{pd|H|}|G|^{|J|} \le x^{d|J|}|G|^{|J|}.$$

Since d is a viral exponent for $\mathcal{J}$, there exists $S \subseteq V(G)$ with $|S| \ge x^d|G| \ge x^{d'}|G|$ such that one of $G[S], \overline{G}[S]$ has edge-density at most x. This proves 2.1. ∎

3 Blockades, and a sketch of the main proof

If $A, B \subseteq V(G)$ are disjoint, and $0 \le x \le 1$, we say that B is *x-sparse* to A if every vertex in B has at most $x|A|$ neighbours in A; and B is *x-dense* to A if every vertex in B has at least $x|A|$ neighbours in A. Let us say a *blockade* in G is a sequence $\mathcal{B} = (B_1, \dots, B_k)$ of pairwise disjoint subsets of $V(G)$, and we call $B_1, \dots, B_k$ its *blocks*. (In some earlier papers, the blocks of a blockade must be nonempty, but here it is convenient to allow empty blocks.) The *length* of the blockade $\mathcal{B} = (B_1, \dots, B_k)$ is k, and its *width* is the minimum of the cardinalities of its blocks. If its length is at least ℓ and width at least w we call it an *(ℓ, w)-blockade*. For $\varepsilon > 0$, the blockade $\mathcal{B} = (B_1, \dots, B_k)$ is *ε-sparse* if $B_{i+1} \cup \dots \cup B_k$ is ε-sparse to B_i for all i with $1 \le i \le k$; and similarly $\mathcal{B}$ is *$(1-\varepsilon)$-dense* if $B_{i+1} \cup \dots \cup B_k$ is $(1-\varepsilon)$-dense to B_i for all i with $1 \le i \le k$.

Using blockades is critical to our approach; let us give here a sketch of the remainder of the paper. We need to prove that $\mathcal{H} = \{\widehat{C_5}, \overline{\widehat{C_5}}\}$ is viral; so we have a graph G in which there are only a small number of copies of $\widehat{C_5}$ and $\overline{\widehat{C_5}}$ (at most $x^{6d}|G|^6$ of each, say, where we can choose d for convenience, though it must be independent of x and G), and we need to prove that there is a subset S with $|S| \ge x^d|G|$ that has edge-density, or edge-density in the complement, at most x. The next section proves a crucial result: that we can replace finding our target set S with finding a long and wide blockade that is either sparse or dense. There is a tradeoff between length and width: we accept length two if the width is large enough. So, more exactly, it is enough if for some d independent of G, x, we can prove that there is an x-sparse or $(1-x)$-dense blockade of length some k between 2 and $1/x$, and width at least $\lfloor |G|/k^d \rfloor$. This reduction, from finding a large set with small edge-density, in G or $\overline{G}$, to finding a sparse blockade in G or $\overline{G}$, works for any set $\mathcal{H}$; there is nothing special about $\widehat{C_5}, \overline{\widehat{C_5}}$ here. There is a further, easier, reduction, a simple application of Nikiforov's theorem: it suffices to find such a blockade in a graph G with maximum degree at most $|G|/d$.

Now our goal is to prove that for some d, and all $0 < x \le 1/d$, if G is a graph in which there are only $x^{6d}|G|^6$ copies of each of $\widehat{C_5}$ and $\overline{\widehat{C_5}}$, and with maximum degree at most $|G|/d$, then there is an x-sparse or $(1-x)$-dense blockade of length some k between 2 and $1/x$, and width at least $\lfloor |G|/k^d \rfloor$. Assume there is no such blockade. If we can halve the maximum degree by deleting only a small set of vertices, do so, and repeat. If this succeeds in driving down the maximum degree to at most $x|G'|$ where G' is the graph that remains, while still $|G'| \ge |G|/2$, then our task is easy: any blockade in G' of the required length and width is x-sparse. So we may assume that this process stops quite soon: the graph is still large, but now T is large, where T is the set of vertices with degree at least half the maximum. For $v \in T$, let $A(v)$ be the neighbour set of v, and let $B(v)$ be the vertices in the remainder of the graph with many neighbours in $A(v)$ (at least $x|A(v)|$ such neighbours). There are two kinds of vertices $v \in T$: $|B(v)|$ may be large or small. We can get rid of all $v \in T$ with $|B(v)|$ small; because if v is such a vertex, delete $v, A(v)$ and $B(v)$, and repeat. This cannot happen many times, since the sequence of $A(v)$'s we produce is an x-sparse blockade. So we end up with a graph, G' say, still large, with many vertices with degree close to the maximum, and they all have $B(v)$ large. (This is the content of 6.1.) How can we use that?

So far we have not used any special properties of $\widehat{C_5}$ and $\overline{\widehat{C_5}}$; the proof so far works for any set of graphs. But now we will. It turns out that whenever $|B(v)|$ is large, and there is no long, wide, x-sparse or $(1-x)$-dense blockade in G, there are poly$(x)|G|^5$ copies of $\widehat{C_5}$ or $\overline{\widehat{C_5}}$ containing v (where poly(x) means some polynomial in x), or poly$(x)|G|^6$ of them within $A(v)$. If we can prove that, we are done, because it will contradicts that there are not many copies of $\widehat{C_5}$ or $\overline{\widehat{C_5}}$ in total.

Proving this is the topic of section 7, 8, 9, 10, but here is a summary. Fix a vertex v with degree at least half the maximum degree, and with $|B(v)|$ large. First, certain helpful subgraphs (with two, three or six vertices) of $G[A(v)]$ will, together with some of their neighbours in $B(v)$, supply many copies of $\widehat{C_5}$ or $\overline{\widehat{C_5}}$, either using v or within $A(v)$; and so we may assume such subgraphs are rare. (This is in sections 8 and 10.) Consequently we may find a large subset A' of $A(v)$ including none of these helpful subgraphs, and still covering a great deal, say B', of $B(v)$. (This is a result of section 9.) Now we need the "comb" theorem of [5], which says, roughly, that in these circumstances one can find $a_1, \dots, a_k \in A'$, and subsets $B_1, \dots, B_k$ of B', such that each a_i is adjacent to all members of B_i and to no members of B_j when $j \neq i$; and k is large, and the sets B_i are all large (again, there is a tradeoff; we accept small k if the sets are very large). Since $G[A']$ is $\widehat{C_5}$-free and $\overline{\widehat{C_5}}$-free (because we counted copies of these as among the "helpful" subgraphs), A' includes a clique or stable set A'' of size polynomial in $|A'|$ (this was a theorem of [5]). So let us focus on A''. For $a_i \in A''$, we can find a large subset C_i of B_i (a "core") such that however it is divided into two large parts, there are many edges between the two parts; because if this were not the case, we could find a long, wide, x-sparse blockade, all within B_i, a contradiction. (This is in section 8.)

Now we look for "rainbow" copies of P_4: copies of P_4 with each vertex in a different set C_i where $a_i \in A''$. For each choice of $a_1, \dots, a_4 \in A''$, we can prove that there are not many rainbow copies of P_4 within the four sets $C_1, \dots, C_4$, since $\{a_1, \dots, a_4\}$ includes none of our helpful subgraphs and $C_1, \dots, C_4$ are cores. Finally, we use a theorem that says that, given a long wide blockade with not many rainbow P_4's, we can find a long, wide x-sparse or x-dense blockade; and that is the content of section 7.

4 Sparse blockades and sparse subsets

We begin with an important lemma, that says that if every large induced subgraph of G admits a sufficiently long and wide blockade that is sufficiently sparse or dense, then there is a large induced subgraph that has either small edge-density or small edge-density in the complement graph.

4.1 *Let G be a graph, and let $\varepsilon \in (0, \frac{1}{2})$ and $d \geq 1$. Let $x = \varepsilon^{12d}$. Suppose that for every induced subgraph F of G with $|F| \geq \varepsilon^{4d}|G|$, there is an x-sparse or $(1-x)$-dense blockade in F of length k and width at least $|F|/k^d$, for some $k \in [2, 1/x]$. Then there exists $S \subseteq V(G)$ with $|S| \geq x^{d+1}|G|$ such that one of $G[S], \overline{G}[S]$ has edge-density at most ε.*

Proof. We may assume that $|G| > x^{-d-1}$, since otherwise we may take $|S| \leq 1$ to satisfy the theorem. Let J be a cograph, and for each $j \in V(J)$ let $A_j \subseteq V(G)$, pairwise disjoint. 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 either

- u, v are adjacent in G and i, j are nonadjacent in J; or
- u, v are nonadjacent in G, and i, j are adjacent 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:

- $|A_j| \geq \varepsilon^{6d}|G|$ 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 $A_1 = G$ to satisfy the first three conditions.)

(1) *We may assume that* $|J| \leq 4\varepsilon^{-2}$.

Suppose that $|J| \geq 4\varepsilon^{-2}$. Since J is a cograph, it has a clique or stable set I of size at least $|J|^{1/2} \geq 2/\varepsilon$, and by taking complements if necessary, we may assume that I is a stable set. For each $i \in I$, choose $B_i \subseteq A_i$ with size $\lceil \varepsilon^{6d}|G| \rceil$, and let $S = \bigcup_{i \in I} B_i$. Thus $|S| \geq (2\varepsilon^{-1})\varepsilon^{6d}|G|$. We claim that $G[S]$ has edge-density at most ε. There are at most $|I|^{-1}\binom{|S|}{2}$ edges uv of $G[S]$ such that $u, v \in B_i$ for some $i \in I$; and the number of edges uv of $G[S]$ such that $u \in B_i$ and $v \in B_j$ for some distinct $i, j \in I$ is at most the number of wrong pairs of $\mathcal{L}$, and hence at most

$$x\binom{|G|}{2} \leq x|G|^2/2 \leq x(\varepsilon^{1-6d}|S|/2)^2/2 = x\varepsilon^{2-12d}|S|^2/8 \leq x\varepsilon^{2-12d}\binom{|S|}{2}/2.$$

Hence the number of edges of $G[S]$ is at most $(|I|^{-1} + x\varepsilon^{2-12d}/2)\binom{|S|}{2} \leq \varepsilon\binom{|S|}{2}$ since $|I|^{-1} \leq \varepsilon/2$ and $x\varepsilon^{2-12d}/2 \leq \varepsilon/2$. Moreover,

$$|S| \geq \varepsilon^{6d}|G| \geq x^{d+1}|G|,$$

and so the theorem is satisfied. This proves (1).

We may assume that $|A_1| \geq |A_j|$ for all $j \in V(J)$. Since $\sum_{j \in V(J)} |A_j|^{1/d} \geq |G|^{1/d}$, and $|J| \leq 4\varepsilon^{-2}$ by (1), it follows that $|A_1|^{1/d} \geq (\varepsilon^2/4)|G|^{1/d}$, that is,

$$|A_1| \geq \varepsilon^{2d}2^{-2d}|G| \geq \varepsilon^{4d}|G|.$$

By applying the hypothesis to $G[A_1]$, we deduce that there an x-sparse or $(1-x)$-dense blockade $(B_1, \ldots, B_k)$ in $G[A_1]$ where $k \in [2, 1/x]$, with width at least $|A_1|/k^d$. By taking complements, we may assume that $(B_1, \ldots, B_k)$ is x-sparse.

(2) $k \geq 2/\varepsilon$.

Suppose that $k \leq 2/\varepsilon \leq \varepsilon^{-2}$. Then each of the sets $B_1, \ldots, B_k$ has size at least $|A_1|/k^d \geq \varepsilon^{2d}|A_1|$. By substituting a k-vertex stable set for the vertex 1 in J, and replacing A_1 by $B_1, \ldots, B_n$, 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 $\mathcal{L}$. Each B_j satisfies

$$|B_j| \geq \varepsilon^{2d}|A_1| \geq \varepsilon^{6d}|G|,$$

and so the first bullet is satisfied. For the second bullet, since $B_1, \dots, B_k$ all have size at least $|A_1|/k^d$, it follows that

$$|B_1|^{1/d} + \cdots + |B_k|^{1/d} \geq |A_1|^{1/d},$$

and so $\sum_{j \in V(J)} |A'_j|^{1/d} \geq |G|^{1/d}$. 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 $|Q| \leq x|P|$. Let R be the set of all pairs $\{u, v\}$ with $u, v \in A_1$ such that u, v belong to different blocks of $(B_1, \dots, B_k)$. Then $R \subseteq P' \setminus P$; and $Q' \setminus Q \subseteq R$; and $|Q' \setminus Q| \leq x|R|$ since $(B_1, \dots, B_k)$ is x-sparse. Hence $|Q' \setminus Q| \leq x|P' \setminus P|$, and so

$$|Q'| \leq |Q| + |Q' \setminus Q| \leq x|P| + x|P' \setminus P| = x|P'|$$

since $P \subseteq P'$. This contradicts the choice of $\mathcal{L}$, and so proves (2).

Let $n = \lceil 2/\varepsilon \rceil$. For $1 \leq i \leq n$, choose $C_i \subseteq B_i$ with size $w := \lceil |A_1|/k^d \rceil$, uniformly at random. The probability that an edge between B_i, B_j has ends in C_i and C_j is $\frac{w^2}{|B_i| \cdot |B_j|}$, and since there are at most $x|B_i| \cdot |B_j|$ edges between B_i, B_j, the expected number of edges between C_i, C_j is at most xw^2. Consequently the probability that there are more than $xn^2w^2/2$ such edges is less than $2/n^2$. It follows that the probability that for all distinct $i, j \in \{1, \dots, n\}$, there are at most $xn^2w^2/2$ edges between C_i, C_j is positive, and so there is a choice of $C_1, \dots, C_n$ such that for all distinct i, j there are at most $xn^2w^2/2$ edges between C_i, C_j. Let $S = C_1 \cup \cdots \cup C_n$. The number of edges of $G[S]$ with ends in the same C_i is at most $(1/n)\binom{|S|}{2}$; and the number of edges of $G[S]$ with ends in distinct blocks C_i, C_j is at most $(xn^2w^2/2)(n^2/2) = xn^2|S|^2/4 \leq xn^2\binom{|S|}{2}$. Consequently $G[S]$ has at most $(1/n + xn^2)\binom{|S|}{2} \leq \varepsilon\binom{|S|}{2}$ edges, since $1/n \leq \varepsilon/2$ and $xn^2 \leq x(4/\varepsilon)^2 \leq \varepsilon/2$. Moreover,

$$|S| \geq w \geq |A_1|/k^d \geq \varepsilon^{4d}|G|/k^d \geq \varepsilon^{4d}x^d|G| \geq x^{d+1}|G|,$$

and hence S satisfies the theorem. This proves 4.1. ∎

5 Divisive sets

We recall that if $\mathcal{H}$ is a set of graphs, $x > 0$, and G is a graph, then

$$\mu_{\mathcal{H}}(x, G) := \max_{H \in \mathcal{H}} \frac{\text{ind}_H(G)}{(x|G|)^{|H|}}.$$

Thus $\mu_{\mathcal{H}}(\cdot, G)$ is nonincreasing for every G, and $\mu_{\mathcal{H}}(x, F) \leq \mu_{\mathcal{H}}(xy, G)$ for all $x, y > 0$ and every induced subgraph F of G with $|F| \geq y|G|$. We say that a finite set $\mathcal{H}$ of graphs is *divisive* if there exists $d > 2$ such that for every $x \in (0, d^{-1})$ and every graph G with $\mu_{\mathcal{H}}(x^d, G) \leq 1$, there exists $k \in [2, 1/x]$ such that there is an x-sparse or $(1-x)$-dense blockade in G of length at least k and width at least $\lfloor |G|/k^d \rfloor$. We say such a number d is an *exponent* for the divisiveness of $\mathcal{F}$. Note that if d is such an exponent, then so is every larger number, say f; because if $x \in (0, f^{-1})$ and $\mu_{\mathcal{H}}(x^f, G) \leq x$, then $x \leq f^{-1} \leq d^{-1}$, and $\mu_{\mathcal{H}}(x^d, G) \leq \mu_{\mathcal{H}}(x^f, G) \leq x$.

The next result is a variant of a theorem in [3]:

5.1 *If $\mathcal{H}$ is divisive then it is viral.*

Proof. Let $d > 1$ be an exponent for the divisiveness of $\mathcal{H}$, and let $c := 12(d+1)(d+2)$. We claim that c is a viral exponent for $\mathcal{H}$. To show this, let $\varepsilon \in (0, 1/2)$ and let G be a graph with $\mu_{\mathcal{H}}(\varepsilon^c, G) \le 1$. We must show that there exists $S \subseteq V(G)$ with $|S| \ge \varepsilon^c|G|$ such that one of $G[S], \overline{G}[S]$ has edge-density at most ε. We may assume that $|G| > \varepsilon^{-c}$, since otherwise we may take $|S| \le 1$ to satisfy the theorem. Let $d' = d + 1$, and $x = \varepsilon^{12d'}$. We claim that:

(1) *For every induced subgraph F of G with $|F| \ge \varepsilon^{4d'}|G|$, there is an x-sparse or $(1-x)$-dense blockade in F of length k and width at least $|F|/k^{d'}$, for some $k \in [2, 1/x]$.*

We observe that $x^d\varepsilon^{4d'} = \varepsilon^{12dd'+4d'} \ge \varepsilon^c$, and so $x^d|F| \ge \varepsilon^c|G|$. It follows that

$$\mu(x^d, F) \le \mu(\varepsilon^c, G) \le 1.$$

Since d is an exponent for the divisiveness of $\mathcal{H}$, there exists $k \in [2, 1/x]$ such that there is an x-sparse or $(1-x)$-dense blockade in F of length at least k and width at least $\lfloor |F|/k^d \rfloor$. But

$$|F|/k^d \ge x^d|F| \ge \varepsilon^c|G| > 1,$$

and so

$$\lfloor |F|/k^d \rfloor \ge |F|/(2k^d) \ge |F|/k^{d+1} = |F|/k^{d'}.$$

This proves (1).

From 4.1, with d replaced by d', we deduce that there exists $S \subseteq V(G)$ with $|S| \ge x^{d'+1}|G| = \varepsilon^c|G|$ such that one of $G[S], \overline{G}[S]$ has edge-density at most ε. This proves that $\mathcal{H}$ is viral, and so proves 5.1. ∎

Actually, we only need the hypothesis of 5.1 to hold in sparse graphs G. More exactly:

5.2 *Let $\mathcal{H}$ be a set of graphs, closed under taking complements, and let $d > 0$, such that for every $x \in (0, d^{-1})$ and every graph G with $\mu_{\mathcal{H}}(x^d, G) \le 1$ and maximum degree at most $|G|/d$, there is an x-sparse or $(1-x)$-dense blockade in G of length k and width at least $\lfloor |G|/k^d \rfloor$ for some $k \in [2, 1/x]$. Then $\mathcal{H}$ is viral.*

Proof. By Nikiforov's theorem 1.5, there exists $c > 0$ such that for every graph G with $\mu_{\mathcal{H}}(c, G) \le 1$, there exists $S \subseteq V(G)$ with $|S| \ge c|G|$ such that one of $G[S], \overline{G}[S]$ has maximum degree at most $|S|/d$. Let $f := d + \log(c^{-1})$. We claim:

(1) *For every $x \in (0, f^{-1})$ and every graph G with $\mu_{\mathcal{H}}(x^f, G) \le 1$, there exists $k \in [2, 1/x]$ such that G contains an x-sparse or $(1-x)$-dense blockade of length at least k and width at least $\lfloor |G|/k^f \rfloor$.*

Let $0 < x < 1/f$ and let G be such that $\mu_{\mathcal{H}}(x^f, G) \le 1$; then $\mu_{\mathcal{H}}(c, G) \le \mu_{\mathcal{H}}(x^f, G) \le 1$, and so there exists $S \subseteq V(G)$ with $|S| \ge c|G|$ such that one of $G[S], \overline{G}[S]$ has maximum degree at most $|S|/d$. Since $\mathcal{H}$ is closed under taking complements, we can replace G by its complement if necessary; so we may assume that $G[S]$ has maximum degree at most $|S|/d$. Since $|S| \ge c|G|$ and $f = d + \log(1/c)$, it follows that

$$\mu_{\mathcal{H}}(x^d, G[S]) \le \mu_{\mathcal{H}}(x^d c, G) \le \mu_{\mathcal{H}}(x^f, G) \le 1.$$

Consequently, the hypothesis implies that there is an x-sparse or $(1-x)$-dense $(k, \lfloor |S|/k^d \rfloor)$-blockade in $G[S]$ (and hence in G) for some integer $k \in [2, 1/x]$. Since $|S|/k^d \geq |G|/k^{d+\log(1/c)} = |G|/k^f$, it follows that G contains an x-sparse or $(1-x)$-dense $(k, \lfloor |G|/k^f \rfloor)$-blockade. This proves (1).

From (1) and 5.1, it follows that $\mathcal{H}$ is viral. This proves 5.2. ∎

6 Reducing the maximum degree

As we just saw, to show 1.10 it is enough to show that $\mathcal{H} = \{\widehat{C_5}, \overline{\widehat{C_5}}\}$ has the property of 5.2, and we will prove this in the remainder of the paper. Thus our focus now is a graph G, with maximum degree at most some small constant times $|G|$, in which there is no long, wide blockade that is x-sparse or $(1-x)$-dense; and we want to prove that the graph contains $\text{poly}(x)|G|^6$ copies of one of $\widehat{C_5}$, $\overline{\widehat{C_5}}$. We will do so by means of the "comb" lemma of [5]. That will tell us that if $v \in V(G)$ is nice enough, then either it belongs to $\text{poly}(x)|G|^5$ copies of $\widehat{C_5}$ or $\overline{\widehat{C_5}}$, or there are $\text{poly}(x)|G|^5$ copies of $\widehat{C_5}$ or $\overline{\widehat{C_5}}$ within its neighbour set (and if the second ever happens, we are done). But we need $\text{poly}(x)|G|^6$ copies, not just $\text{poly}(x)|G|^5$; so we will need $\text{poly}(x)|G|$ "nice enough" vertices. What that means is, we need there to be many vertices with degree at least half the maximum degree, that have many second neighbours that are at least x-dense to the first neighbours. That is the purpose of the next result.

6.1 *Let $q > 0$; then for all sufficiently large d, the following holds. Let $x \in (0, d^{-1})$, and let G have maximum degree at most $|G|/d$, such that there is no x-sparse or $(1-x)$-dense $(k, \lfloor |G|/k^d \rfloor)$-blockade in G with $k \in [2, 1/x]$. Then there is a number D with $2x^3|G| \leq D \leq |G|/d$, and an induced subgraph G' of G such that, denoting by T the set of vertices in G' that have degree at least $D/2$:*

- *$|G'| \geq |G|/2$, and G' has maximum degree at most D;*
- *$|T| \geq x^2|G|$; and*
- *for every vertex $v \in T$, with neighbour set A in G' say, there are at least $q(D|G|)^{1/2}$ vertices in $V(G') \setminus A$ that have at least $x|A|$ neighbours in A.*

Proof. Let
$$K := \max\left(\frac{2^{1/3}}{2^{1/3}-1}, \frac{2^{1/3}}{1-2^{-2/3}}, \frac{q \cdot 2^{1/3}}{1-2^{-1/6}}\right),$$
and choose $c > 0$ such that $c \leq 1/8$ and $K(2c^{2/3} + c^{1/6}) < 1/4$. We claim that $d := 1/c$ satisfies the theorem. Let $x \in (0, c)$, $\ell := \lfloor 1/x \rfloor$ and $m := \lfloor \log(x^{-2}) \rfloor$. Consequently $\ell \geq 1/(2x)$, and so $x \geq 1/(2\ell) \geq 1/\ell^2$.

Let G be a graph with maximum degree at most $c|G|$, that does not contain an x-sparse or $(1-x)$-dense $(k, \lfloor |G|/k^d \rfloor)$-blockade with $k \in [2, 1/x]$. Then $|G| \geq \ell^d$, since otherwise we could take $k = \ell$ and choose a blockade with k empty blocks to satisfy the theorem.

The idea of the proof is, if we can delete a few vertices to halve the maximum degree, we do so, and repeat until either the graph becomes very sparse (which will contradict our assumption about blockades) or the process stops. At that stage, we try to find the desired blockade directly. We must failx; and at that point we have the properties of the theorem.

For $0 \le s \le m$, define:

$$N_s := c^{-1/3} 2^{(s+1)/3} + 2^{(1-2s)/3} c^{2/3} |G| + 2^{(2-s)/6} c^{1/6} q |G|.$$

We need:

(1) $\sum_{0 \le s \le m-1} (N_s + x^2 |G|) \le |G|/2$.

Observe that, by the definition of m,

$$\begin{aligned}
\sum_{s=0}^{m-1} 2^{(s+1)/3} &= 2^{1/3} \frac{2^{m/3}-1}{2^{1/3}-1} \le K 2^{m/3} \le K x^{-2/3}, \\
\sum_{s=0}^{m-1} 2^{-2s/3} &\le \frac{1}{1-2^{-2/3}} \le K 2^{-1/3}, \\
\sum_{s=0}^{m-1} 2^{-s/6} &\le \frac{1}{1-2^{-1/6}} \le K (2^{1/3} q)^{-1}.
\end{aligned}$$

Therefore, since $m \le \log(1/x^2) \le 2/x$, we have:

$$\begin{aligned}
&\sum_{0 \le i \le m-1} (N_s + x^2 |G|) \\
&\le c^{-1/3} \sum_{s=0}^{m-1} 2^{(s+1)/3} + 2^{1/3} c^{2/3} |G| \sum_{s=0}^{m-1} 2^{-2s/3} + 2^{1/3} c^{1/6} q |G| \sum_{s=0}^{m-1} 2^{-s/6} + m \cdot x^2 |G| \\
&\le K c^{-1/3} x^{-2/3} + K c^{2/3} |G| + K c^{1/6} |G| + 2x|G| \le K(2c^{2/3} + c^{1/6})|G| + \tfrac{1}{4}|G| < \tfrac{1}{2}|G|
\end{aligned}$$

where the last inequality holds by the choice of c and since $x \le 1/8$. This proves (1).

Let us say a subset $Z \subseteq V(G)$ is *s-good*, where $0 \le s \le m$, if $G[Z]$ has maximum degree at most $2^{-s} c|G|$, and

$$|Z| \ge |G| - \sum_{0 \le i \le s-1} (N_s + x^2 |G|).$$

Thus $V(G)$ is 0-good.

(2) *There is no m-good subset.*

Suppose that Z is m-good. By (1), $|Z| \ge |G|/2$. Let $\mathcal{B}$ be an $(\ell, \lfloor x|Z| \rfloor)$-blockade in $G[Z]$. Since $\ell^2 \ge \frac{1}{4} x^{-2} \ge 1/x$ and $G[Z]$ has maximum degree at most

$$2^{-m} c|G| \le 2x^2 (c|G|) \le 4cx^2 |Z| \le x \lfloor x|Z| \rfloor,$$

it follows that $\mathcal{B}$ is an x-sparse $(\ell, \lfloor |G_m|/\ell^2 \rfloor)$-blockade, a contradiction. This proves (2).

Consequently, from (2), there exists an integer $s < m$ such that there is an s-good subset Z, but there is no $(s+1)$-good subset. Let $C_0 = Z$, let $D = 2^{-s}c|G|$, and choose $j \geq 0$ maximum such that there are vertices $v_1, \dots, v_j$ in Z, and subsets $A_1, \dots, A_j$, $B_1, \dots, B_j$ and $C_1, \dots, C_j$ of Z, such that for $1 \leq i \leq j$:

- $v_i \in C_{i-1}$; A_i is the set of all neighbours of v_i in C_{i-1}; B_i is the set of all vertices in $C_{i-1} \setminus (A_i \cup \{v_i\})$ that have at least $x|A_i|$ neighbours in A_i; and $C_i = C_{i-1} \setminus (A_i \cup B_i \cup \{v_i\})$;
- $|A_i| \geq D/2$, and $|B_i| \leq q\sqrt{|G| \cdot D} = 2^{-s/2}c^{1/2}q|G|$.

Consequently, the sets $\{v_1\}, \dots, \{v_j\}, A_1, \dots, A_j, B_1, \dots, B_j$ are all pairwise disjoint subsets of Z.

(3) *$j < \ell$ and $j < c^{-1/3}2^{(s+1)/3}$.*

Since C_i is x-sparse to A_i for $1 \leq i \leq j$, and $|A_i| \geq D/2$ for $1 \leq i \leq j$, it follows that $(A_1, \dots, A_j)$ is an x-sparse $(j, D/2)$-blockade in G. Moreover,

$$D/2 \geq x^2c|G| \geq x^3|G| \geq |G|/\ell^6.$$

Consequently, $D/2 \geq |G|/\ell^d$, and $2 \leq \ell \leq 1/x$, so if $j \geq \ell$, then $(A_1, \dots, A_\ell)$ would be a blockade with the desired property, a contradiction. If $|G|/j^3 \leq D/2$, then $j \geq 2$, and $(A_1, \dots, A_j)$ would have the desired property, a contradiction. Thus $2^{-s-1}c|G| = D/2 < |G|/j^3$ and so

$$j < (2^{s+1}/c)^{1/3} = c^{-1/3}2^{(s+1)/3}.$$

This proves (3).

(4) *$|A_1 \cup \dots \cup A_j| \leq 2^{(1-2s)/3}c^{2/3}|G|$, and $|B_1 \cup \dots \cup B_j| \leq 2^{(2-s)/6}c^{1/6}q|G|$.*

Since $v_1, \dots, v_j$ all have degree at most D in $G[Z]$, it follows that $|A_1|, \dots, |A_j| \leq D$, and so from (3),

$$|A_1 \cup \dots \cup A_j| \leq c^{-1/3}2^{(s+1)/3}2^{-s}c|G| = 2^{1/3}c^{2/3}|G|2^{-2s/3}.$$

Since $|B_i| \leq 2^{-s/2}c^{1/2}q|G|$ for all $i \in [j]$, we have from (3) that

$$|B_1 \cup \dots \cup B_j| \leq j \cdot 2^{-s/2}c^{1/2}q|G| < 2^{(s+1)/3}c^{-1/3} \cdot 2^{-s/2}c^{1/2}q|G| = 2^{(2-s)/6}c^{1/6}q|G|.$$

This proves (4).

Let T be the set of vertices in C_j that have degree at least $D/2$ in $G[C_j]$. Now $C_j \setminus T$ has maximum degree at most $D/2 = 2^{-s-1}c|G|$; but it is not $(s+1)$-good, and so from (3), (4) it follows that $|T| \geq x^2|G|$. From the maximality of j, if $v \in T$ with neighbour set A in $G[C_j]$ say, there are at least $q(D|G|)^{1/2}$ vertices in $C_j \setminus (A \cup \{v\})$ that have at least $x|A|$ neighbours in A. But then $G' = G[C_j]$ satisfies the theorem. This proves 6.1. ∎

7 Rainbow forests

A *pure pair* in a graph G is a pair P, Q of disjoint subsets of $V(G)$ such that P is either complete or anticomplete to Q. Let $\mathcal{B}$ be a blockade in a graph G. An induced subgraph H of J is *$\mathcal{B}$-rainbow* if every vertex of H is in a block of $\mathcal{B}$ and every two vertices belong to different blocks; and a copy of a graph T in G is *$\mathcal{B}$-rainbow* if it maps T to a $\mathcal{B}$-rainbow induced subgraph of G. The following was proved in [5], extending a theorem of [6]:

7.1 *For every forest T, there exist $d > 0$ and an integer K, such that, for every graph G with a blockade $\mathcal{B}$ of length at least K, either:*

- *there is a pure pair P, Q in G with $|P|, |Q| \geq W/d$, where W is the width of $\mathcal{B}$; or*
- *there is a $\mathcal{B}$-rainbow copy of one of $T, \overline{T}$ in G.*

(The meaning of "blockade" in [6] was different in that all blocks had to be nonempty, but 7.1 is trivially true if $W = 0$.)

To prove 1.10, we need to modify 7.1 in several ways, which we will do below. In fact, we only need these results when T is the four-vertex path P_4, but the proof for P_4 is no simpler than that for general trees, so we have kept it all in full generality. First, we show:

7.2 *For every forest T, there is an integer $L > 0$, such that for every graph G with a blockade $\mathcal{B}$ of length at least L, either:*

- *there is a pure pair P, Q in G with $|P|, |Q| \geq W/L$, where W is the width of $\mathcal{W}$, such that $P \subseteq B$ and $Q \subseteq B'$ for some two distinct blocks B, B' of $\mathcal{B}$; or*
- *there is a $\mathcal{B}$-rainbow copy of one of $T, \overline{T}$ in G.*

Proof. Let d, K be as in 7.1. We may assume that d, K are integers. We claim that $L = 2dK$ satisfies the theorem.

To see this, let $\mathcal{B}$ in a graph G, with length at least L, and width W. We may assume that all the blocks of $\mathcal{B}$ have the same size W. Let $\mathcal{B} = (B_1, \dots, B_L)$. We may assume that there is no $\mathcal{B}$-rainbow copy of one of $T, \overline{T}$ in G. For $1 \leq j \leq K$ let $C_j = \bigcup_{1 \leq i \leq 2d} B_{2d(j-1)+i}$. Then $\mathcal{C} = (C_1, \dots, C_K)$ is a blockade of length K and width $2dW$.

Let G' be the subgraph of G induced on the union of the blocks of $\mathcal{C}$. Since $\mathcal{C}$ has length K, and width $2dW$, we may apply 7.1 to G' and $\mathcal{C}$, and deduce that either:

- there is a pure pair P, Q of G' with $|P|, |Q| \geq (2dW)/d$; or
- there is a $\mathcal{C}$-rainbow copy of one of $T, \overline{T}$ in G'.

The second case is impossible, since such a copy would also be $\mathcal{B}$-rainbow. Thus the first case holds. Let P, Q be the corresponding pure pair. Since P is a subset of $B_1 \cup \cdots \cup B_{2K}$, there is a block B of $\mathcal{B}$ with $|P \cap B| \geq |P|/(2K) \geq W/K \geq W/L$. Since $|B| = W$ and $|Q| \geq 2W$, it follows that $|Q \setminus B| \geq W$, and so there is a block $B' \neq B$ with $|Q \cap B'| \geq W/(2K) \geq W/L$. But then the first bullet of the theorem holds. This proves 7.2. ∎

Next, we need a second modification, weakening the "complete/anticomplete" outcome of 7.2 to a "dense/sparse" outcome, and strengthening its "$\mathcal{B}$-rainbow" outcome to say that there are many $\mathcal{B}$-rainbow copies instead of just one. For that, we need the following lemma:

7.3 *Let $a \in (0, \frac{1}{8})$ and $x \in (0, \frac{1}{2})$ such that $x \log \frac{1}{x} \leq \frac{1}{4}a$. Let $m \geq x^{-2}$ and $w \geq 2m$ be integers. Let G be a graph, and let $A, B \subseteq V(G)$ be disjoint, with $|A| = |B| = w$, such that there do not exist $P \subseteq A$ and $Q \subseteq B$ with $|P|, |Q| \geq aw$ for which P is x-sparse to Q. Choose $P \subseteq A$ of cardinality m, uniformly at random, and choose $Q \subseteq B$ similarly. Then the probability that P, Q are anticomplete is at most $2(2a)^{m/2}$.*

Proof. Choose vertices $a_1, \ldots, a_m$ one by one such that for each i, a_i is a uniformly random element of $A \setminus \{a_1, \ldots, a_{i-1}\}$. For $0 \leq i \leq m$, let $X_i := \{a_1, \ldots, a_i\}$. Then for each i, X_i is a uniformly random i-subset of A. For $0 \leq i \leq m$, let Y_i be the set of vertices in B with no neighbour in X_i; thus, $Y_0 \supseteq Y_1 \supseteq \cdots$. For $0 \leq i \leq m$, let A_i be the set of vertices in $A \setminus X_{i-1}$ with at most xaw neighbours in Y_{i-1}. Let $I := \{i : \ 1 \leq i \leq m \text{ and } a_i \notin A_i\}$. For each $i \in I$ there are at least xaw vertices in B adjacent to a_i and to none of $a_1, \ldots, a_{i-1}$, and so $|I| \leq (xa)^{-1}$. Let $r := (xa)^{-1}$.

(1) *For every $J \subseteq \{1, \ldots, m\}$ with $|J| \leq r$, $\mathsf{P}[(|Y_m| \geq aw) \wedge (I \subseteq J)] \leq (2a)^{m-|J|} \leq (2a)^{m-r}$.*

For $1 \leq i \leq m$, let us say that i *behaves* if either $i \in J$, or both $|Y_i| \geq aw$ and $a_i \in A_i$. If $|Y_i| \geq aw$, then since A_i is x-sparse to Y_i, it follows from the hypothesis that $|A_i| \leq aw$. Consequently, for each $i \notin J$, and for every choice of $a_1, \ldots, a_{i-1}$, the probability that i behaves is at most

$$\frac{|A_i|}{w - i + 1} \leq \frac{aw}{w - m} \leq 2a$$

since $m \leq w/2$. Let p_i be the probability that i behaves given that $1, \ldots, i-1$ all behave. The event that $1, \ldots, i-1$ all behave only depends on the choice of $a_1, \ldots, a_{i-1}$; so $p_i \leq 2a$ if $i \notin J$ (and $p_i \leq 1$ if $i \in J$). The probability that $1, \ldots, m$ all behave is the product of $p_1, p_2, \ldots, p_m$, and so at most $(2a)^{m-|J|}$. But $|Y_m| \geq aw$ and $I \subseteq J$ only if $1, \ldots, m$ all behave. This proves (1).

(2) $(2a)^{m-r} m^r \leq (2a)^{m/2}$.

It suffices to show that $(2a)^{m/2-r} m^r \leq 1$. To this end, observe that

$$r = (xa)^{-1} \leq \frac{x^{-2}}{4 \log(x^{-1})} = \frac{x^{-2}}{2 \log(x^{-2})} \leq \frac{m}{2 \log m} \leq \frac{m}{4},$$

since $x \log(x^{-1}) \leq \frac{1}{4}a$ and $4 \leq x^{-2} \leq m$. Thus $m/2 - r \geq m/4$ and $m^r = 2^{r \log m} \leq 2^{m/2}$, and so

$$(2a)^{m/2-r} m^r \leq (2a)^{m/4} 2^{m/2} = (8a)^{m/4} \leq 1.$$

This proves (2).

Since there are at most m^r subsets J of $\{1, \ldots, m\}$ of size at most r (because $r \geq 2$), and since $|I| \leq r$, (1) and (2) imply that the probability that $|Y_m| \geq aw$ is at most

$$\sum_{J \subseteq \{1, \ldots, m\}, |J| \leq r} \mathsf{P}[(|Y_m| \geq aw) \wedge (I \subseteq J)] \leq (2a)^{m-r} m^r \leq (2a)^{m/2}.$$

Let $Y \subseteq B$ be a set of cardinality m chosen uniformly at random; and let E be the event that Y, X_m are anticomplete, that is, $Y \subseteq Y_m$. It suffices to prove $\mathsf{P}[E] \le 2(2a)^{m/2}$. But $Y \subseteq Y_m$ only if either $|Y_m| \ge aw$, or $Y \subseteq Y_m$ and $|Y_m| < aw$, and we will bound the probability of these two events separately. We have seen that the probability that $|Y_m| \ge aw$ is at most $(2a)^{m/2}$; and the probability that both $Y \subseteq Y_m$ and $|Y_m| < aw$ is at most $\binom{aw}{m}/\binom{w}{m} \le a^m$. Consequently

$$\mathsf{P}[E] \le (2a)^{m/2} + a^m \le 2(2a)^{m/2}.$$

This proves 7.3. ∎

We use 7.3 to prove the following variation on 7.2:

7.4 *For every forest T, there is an integer $d > 0$ such that for all $x \in (0, d^{-2})$ and for every graph G with a (d, w)-blockade $\mathcal{B}$, either:*

- *there are distinct blocks B, B' of $\mathcal{B}$, and subsets $P \subseteq B$ and $Q \subseteq B'$, with $|P|, |Q| \ge w/d$, such that Q is either x-sparse or x-dense to P; or*
- *for some subblockade $\mathcal{D}$ of $\mathcal{B}$ with length $|T|$, there are at least $(d^{-2}x^2w)^{|T|}$ $\mathcal{D}$-rainbow copies of one of $T, \overline{T}$.*

Proof. Let $L > 0$ be as in 7.2; we claim that $d := \lceil 8(eL)^8 \rceil$ satisfies the theorem. Thus, $d \ge 1000$. We may assume that $d^{-2}x^2w > 1$, since otherwise we are done by 7.2; and consequently $x^2w \ge 10^6$, and $x^2w \ge 2L$, and $x \log(1/x) \le 1/(4d)$. Let $\mathcal{B} = (B_1, \ldots, B_d)$; we may assume that $|B_i| = w$ for $1 \le i \le d$, and that the first outcome does not hold. Let $m := \lceil x^{-2} \rceil \le 2x^{-2} \le \frac{1}{2}w$. For all distinct $i, j \in \{1, \ldots, d\}$, 7.3 (with $a = 1/d$) implies that a uniformly random m-subset of B_i and a uniformly random m-subset of B_j make a pure pair with probability at most $2(2/d)^{m/2}$. Let $\mathcal{A}_{ij}$ be the collection of pure pairs (A_i, A_j) with $A_i \subseteq B_i$, $A_j \subseteq B_j$ and $|A_i| = |A_j| = m$; then

$$|\mathcal{A}_{ij}| \le 2\left(\frac{d}{2}\right)^{-m/2}\binom{w}{m}^2 \le 2\left(\frac{d}{2}\right)^{-m/2}\left(\frac{ew}{m}\right)^{2m}.$$

Now $w \ge 2Lx^{-2} \ge Lm$. For $1 \le i \le d$, let S_i be a uniformly random Lm-subset of B_i. We say (S_i, S_j) *includes* $(A_i, A_j) \in \mathcal{A}_{ij}$ if $A_i \subseteq S_i$ and $A_j \subseteq S_j$. Now

$$\frac{\binom{w-m}{Lm-m}}{\binom{w}{Lm}} = \frac{Lm}{w} \cdot \frac{Lm-1}{w-1} \cdots \frac{Lm-m+1}{w-m+1} \le \left(\frac{Lm}{w}\right)^m.$$

Consequently, for all distinct $i, j \in \{1, \ldots, d\}$, the probability that (S_i, S_j) includes some member of $\mathcal{A}_{ij}$ is at most

$$|\mathcal{A}_{ij}|\frac{\binom{w-m}{Lm-m}^2}{\binom{w}{Lm}^2} \le 2\left(\frac{d}{2}\right)^{-m/2}\left(\frac{ew}{m}\right)^{2m}\left(\frac{Lm}{w}\right)^{2m} = 2(2e^4L^4/d)^{m/2}.$$

Hence, by the choice of d, the probability that there are distinct $i, j \in \{1, \ldots, d\}$ for which (S_i, S_j) includes some pair in $\mathcal{A}_{ij}$ is at most

$$\binom{d}{2} \cdot 2(2e^4L^4/d)^{m/2} \le d^2(2e^4L^4/d)^{m/2} \le d^2(2e^4L^4/d)^4 = 16(eL)^{16}d^{-2} \le \frac{1}{2}.$$

Thus, with probability at least $\frac{1}{2}$, there is no pure pair (A, B) with $|A| = |B| = m$ contained in different blocks of the random (d, Lm)-blockade $\mathcal{S} := (S_1, \dots, S_d)$; and so by 7.2, there is an $\mathcal{S}$-rainbow copy of T or $\overline{T}$ in G with probability at least $\frac{1}{2}$. Thus there exists $I \subseteq \{1, \dots, d\}$ with $|I| = |T|$ such that with probability at least $\frac{1}{2}\binom{d}{|T|}^{-1}$, there is an $\mathcal{S}_I$-rainbow copy of T or $\overline{T}$ in G where $\mathcal{S}_I = (S_i : i \in I)$. Therefore, for $\mathcal{D} := (B_i : i \in I)$, the number of $\mathcal{D}$-rainbow induced copies of T or $\overline{T}$ in G is at least

$$\frac{1}{2}\binom{d}{|T|}^{-1}\left(\frac{w}{Lm}\right)^{|T|} \geq \left(\frac{w}{2dLm}\right)^{|T|} \geq (d^{-2}x^2w)^{|T|},$$

since $\binom{d}{|T|} \leq d^{|T|}$, and $2dLm \leq 4dLx^{-2} \leq d^2x^{-2}$. This proves 7.4. ∎

For a graph F with vertex set $\{1, \dots, n\}$, an *x-blowup* of F in a graph G is a blockade $(B_1, \dots, B_n)$ in G such that for all distinct i, j with $1 \leq i < j \leq n$, B_j is x-dense to B_i if $ij \in E(F)$ and x-sparse to B_i if $ij \notin E(F)$. 7.4 gives us a pair of sets where one is x-sparse to the other, and thus an x-blowup of a two-vertex cograph. We can make this into an x-blowup of a larger cograph, as follows.

7.5 *Let T be a forest, and let $d > 0$ be as in 7.4. Then for every $x \in (0, d^{-5})$, every integer $s \geq 1$ such that $2^{s-1}d^{2s-1} \leq x^{-1/5}$, and every graph G with an (ℓ, w)-blockade $\mathcal{B}$ where $\ell = 2^{s-1}d^{2s-1}$, either:*

- *G contains a $(2^s, w/\ell)$-blockade that is an x-blowup of a cograph; or*
- *for some subblockade $\mathcal{D}$ of $\mathcal{B}$ of length $|T|$, there are at least $(x^3w)^{|T|}$ $\mathcal{D}$-rainbow copies of T or $\overline{T}$.*

Proof. We proceed by induction $s \geq 1$. The case $s = 1$ follows from 7.4; so we assume the theorem is true for s, and we shall prove it for $s + 1$. Thus, let $x \in (0, d^{-5})$, and let $\ell = 2^s d^{2s+1} \leq x^{-1/5}$. Let $\mathcal{B} = (B_1, \dots, B_\ell)$ be an (ℓ, w)-blockade in a graph G. We assume that the second outcome does not hold.

Let $n = 2^s d^{2s}$, and let $I_1, \dots, I_d$ be pairwise disjoint subsets of $\{1, \dots, \ell\}$ each of cardinality n. Let $B'_j = \bigcup_{i \in I_j} B_i$ for $1 \leq j \leq d$; then $\mathcal{B}' = (B'_1, \dots, B'_d)$ is a (d, nw)-blockade. Since the second outcome does not hold, and since $nd = \ell \leq x^{-1/5}$, it follows that for every subblockade $\mathcal{D}'$ of $\mathcal{B}'$ of length $|T|$, there are fewer than $n^{|T|}(x^3w)^{|T|} \leq (d^{-2}(x/n^2)^2w)^{|T|}$ $\mathcal{D}'$-rainbow copies of Y or $\overline{T}$ in G. By 7.4, there is a pair P, Q of subsets contained in distinct blocks of $\mathcal{B}'$ with $|P|, |Q| = nw/d$ such that Q is (x/n^2)-dense or (x/n^2)-sparse to P.

Let $\ell' = 2^{s-1}d^{2s-1}$; thus, $\ell' \leq x^{-1/5}$. We may assume that $P \subseteq B'_1$. Let I be the set of $i \in I_1$ such that $|P \cap B_i| \geq w/(2d)$; then $|I| \geq |I_1|/(2d) = n/(2d)$. Since the second outcome is false, it follows that for every subblockade $\mathcal{D}$ of $(B_i : i \in I)$ of length $|T|$, there are fewer than $(x^3w)^{|T|}$ $\mathcal{D}$-rainbow copies of T or $\overline{T}$. Consequently, the same holds for the blockade $(P \cap B_i : i \in I)$. This blockade has length at least $n/(2d) = \ell'$ and width at least $w/(2d)$. Consequently, from the inductive hypothesis, since $(w/(2d))/\ell' \geq w/\ell$, there is a $(2^{s-1}, w/\ell)$-blockade $(C_1, \dots, C_{2^{s-1}})$ in $G[P]$ that is an x-blowup of a cograph. Similarly, there is a $(2^{s-1}, w/\ell)$-blockade $(D_1, \dots, D_{s^{s-1}})$ in $G[Q]$ that is an x-blowup of a cograph. Each set C_i satisfies

$$|C_i| \geq w/\ell = d|P|/(n\ell) = |P|n^{-2};$$

and since Q is (x/n^2)-dense or (x/n^2)-sparse to P, it follows that Q, and hence each set D_j, is x-dense or x-sparse to C_i. Consequently, by combining these two blockades and renumbering, we obtain a $(2^s, w/\ell)$-blockade in $G[P \cup Q]$ that is an x-blowup of a cograph, as desired. This proves 7.5. ∎

8 A "sparse" analogue of components

For $x \in (0, \frac{1}{2})$ and a graph G, a subset $S \subseteq V(G)$ is an *x-core* in G if there is no partition A, B of S with $|A|, |B| \geq x|S|$ such that the edge-density between A, B is less than x^5. (The *edge-density between* A, B is the number of edges between A, B divided by $|A| \cdot |B|$.) Every large graph with no large component contains a blockade of large length and width, the blocks of which are pairwise anticomplete. Here we show a similar statement, with "component" and "anticomplete" replaced by "x-core" and "sparse".

8.1 *Let $x \in (0, \frac{1}{2})$, and let $k \in [4, 1/x]$ be an integer. Let G be a graph that contains no x^2-core in G of size at least $|G|/k$. Then G contains an x-sparse $(k, |G|/k^2)$-blockade.*

Proof. We begin with the following:

(1) *There is a partition $S_1, \dots, S_n$ of $V(G)$, such that $|S_i| \leq |G|/k$ for each i, and the edge-density between S_i, S_j is at most x^4 for all distinct i, j.*

Choose a partition $S_1, \dots, S_n$ of $V(G)$ with n maximum subject to the following conditions:

- $|S_i| \geq x^3|G|$ for $1 \leq i \leq n$;
- the number of edges between S_i, S_j is at most $x^{10}|G|^2$ for all distinct i, j.

Suppose that $|S_i| \geq |G|/k$ for some i. Since S_i is not an x^2-core, there is a partition A, B of S_i with $|A|, |B| \geq x^2|S_i|$ such that the edge-density between A, B is less than x^{10}, and hence the number of edges between A, B is at most $x^{10}|A| \cdot |B| \leq x^{10}|G|^2$. But $|A|, |B| \geq x^2|S_i| \geq x^2|G|/k \geq x^3|G|$. Thus, the partition obtained from $S_1, \dots, S_n$ by removing S_i and adding A and B satisfies the two bullets above, contrary to the maximality of n. This proves that $|S_i| < |G|/k$ for each i. For all distinct i, j, there are at most $x^{10}|G|^2$ edges between S_i, S_j, and since $|S_i|, |S_j| \geq x^3|G|$, the edge-density between S_i, S_j is at most x^4. This proves (1).

(2) *There is a partition $S_1, \dots, S_n$ of $V(G)$, such that $|S_i| \leq |G|/k$ for $1 \leq i \leq n$, and $|S_i| \geq |G|/(2k)$ for $1 \leq i \leq n-1$, and the edge-density between S_i, S_j is at most x^4 for all distinct i, j.*

By (1), we may choose a partition $S_1, \dots, S_n$ of $V(G)$ with n minimum subject to the following conditions:

- $|S_i| \leq |G|/k$ for $1 \leq i \leq n$;
- the edge-density between S_i, S_j is at most x^4 for all distinct i, j.

If two of $S_1, \dots, S_n$ both have cardinality at most $|G|/(2k)$, we can replace them by their union, contrary to the minimality of k. Thus we may renumber such that $|S_i| \ge |G|/(2k)$ for $1 \le i \le n-1$. This proves (2).

Since $|S_i| \le |G|/k$ for each i, it follows that $n \ge k$, and if equality holds then $|S_n| = |G|/k$. So $(S_1, \dots, S_k)$ is a blockade of width at least $|G|/k$. For all i, j with $1 \le i < j \le k$, let D_{ij} be the set of vertices in S_j that have more than $x^2|S_i|$ neighbours in S_i; and let $B_j := S_j \setminus (\bigcup_{1 \le i < j} D_{ij})$. Since the edge-density between S_i, S_j is at most x^4, we have $|D_{ij}| < x^2|S_j|$ for all i, j with $1 \le i < j \le k$; and so $|B_j| > (1-(k-1)x^2)|S_j| \ge (1-x)|S_j|$. Since B_j is x^2-sparse to S_i and $x^2 < x(1-x)$, we see that B_j is x-sparse to B_i. Therefore $(B_1, \dots, B_k)$ is an x-sparse blockade in G of length k and width at least $(1-x)|G|/(2k) \ge |G|/k^2$. This proves 8.1. ∎

9 Three covering lemmas

In this section we give three lemmas about covering in bipartite graphs. If $A \subseteq V(G)$ and $v \in V(G)$, we say A *covers* v if v has a neighbour in A; and if $A, B \subseteq V(G)$, we say A *covers* B if A covers every member of B. First, we need the following:

9.1 *Let $x \in (0, \frac{1}{2})$. Let (A, B) be a bipartition of a graph G, such that every vertex in B has at least $x|A|$ neighbours in A. Let $\mathcal{Z}$ be a set of nonempty subsets of A, such that*

$$\sum_{Z \in \mathcal{Z}} (x|A|)^{-|Z|} \le 1/4.$$

Then there exists $A' \subseteq A$ with $|A'| \le 1/x$ that covers at least $\frac{1}{4}|B|$ vertices in B and includes no member of $\mathcal{Z}$.

Proof. Let S be a subset of A with size $k := \lfloor 1/x \rfloor$, chosen uniformly at random. For each $v \in B$, since v has at least $x|A|$ neighbours in A, the probability that v has no neighbours in S is at most

$$\frac{\binom{|A|-x|A|}{k}}{\binom{|A|}{k}} \le \left(\frac{|A|-x|A|}{|A|}\right)^k = (1-x)^{\lfloor 1/x \rfloor} \le e^{-x\lfloor 1/x \rfloor} \le e^{-2/3}$$

where the last inequality holds since $x\lfloor 1/x \rfloor \ge 2/3$ for $x \in (0, \frac{1}{2})$. Thus, Markov's inequality implies that the expected number of vertices in B with no neighbour in S is at most $e^{-2/3}|B|$; and so S covers fewer than $\frac{1}{4}|B|$ vertices in B with probability at most $\frac{4}{3}e^{-2/3}$.

For every $Z \in \mathcal{Z}$, the probability that $Z \subseteq S$ is

$$\frac{\binom{|A|-|Z|}{k-|Z|}}{\binom{|A|}{k}} \le \left(\frac{k}{|A|}\right)^{|Z|} \le (x|A|)^{-|Z|},$$

since $k \le 1/x$. Thus, the probability that S includes some member of $\mathcal{Z}$ is at most

$$\sum_{Z \in \mathcal{Z}} (x|A|)^{-|Z|} \le 1/4.$$

Consequently, the probability that S either covers fewer than $\frac{1}{4}|B|$ vertices in B or includes some member of $\mathcal{Z}$ is at most $\frac{4}{3}e^{-2/3} + \frac{1}{4} < 1$. Hence there is a choice of $A' \subseteq A$ with the desired property. This proves 9.1. ∎

Our second lemma will help us to handle pairs of subsets with middling edge-density.

9.2 *Let $x \in (0, \frac{1}{2})$, let G be a graph, and let $A, B \subseteq V(G)$ be nonempty and disjoint, such that the edge-density between A and B is at least x and at most $1 - x$. Then either*

- *there exists $A' \subseteq A$ with $|A'| \geq (1-x)|A|$ such that every vertex in A' has at least $\frac{1}{4}x|B|$ neighbours and at least $\frac{1}{4}x|B|$ nonneighbours in B, or*
- *there exists $B' \subseteq B$ with $|B'| \geq \frac{1}{4}x|B|$ such that every vertex in B' has at least $\frac{1}{4}x|A|$ neighbours and at least $\frac{1}{4}x|A|$ nonneighbours in A.*

Proof. Let B_1 be the set of vertices in B with at most $\frac{1}{4}x|A|$ neighbours in A, and let B_2 be the set of vertices in B with at most $\frac{1}{4}x|A|$ nonneighbours in A. Let $B' = B \setminus (B_1 \cup B_2)$. If $|B'| \geq \frac{1}{4}x|B|$ then we are done, so we may assume that $|B'| < \frac{1}{4}x|B|$. Let A_1 be the set of vertices in A with at most $\frac{1}{2}|B_1|$ neighbours in B_1; then $|A_1| \geq (1 - \frac{x}{2})|A|$. Similarly, let A_2 be the set of vertices in A with at most $\frac{1}{2}|B_2|$ nonneighbours in B_2; then $|A_2| \geq (1 - \frac{x}{2})|A|$. Let $A' := A_1 \cap A_2$; then $|A'| \geq (1-x)|A|$, and every vertex in A' has at least $\frac{1}{2}|B_1|$ nonneighbours in B_1 and at least $\frac{1}{2}|B_2|$ neighbours in B_2. Thus, if $|B_1|, |B_2| \geq \frac{1}{2}x|B|$ then we are done, and so by the symmetry we may assume that $|B_1| < \frac{1}{2}x|B|$. Therefore the number of nonedges between A and B is at most

$$|A|(|B'| + |B_1|) + \tfrac{1}{4}x|A||B_2| < (\tfrac{1}{4}x + \tfrac{1}{2}x)|A||B| + \tfrac{1}{4}x|A||B| = x|A||B|,$$

a contradiction. This proves 9.2. ∎

The third lemma is a key theorem from [5], a result about "combs". Let G be a graph, and let $t, w \geq 0$ where t is an integer. We say $((a_i, B_i) : 1 \leq i \leq t)$ is a (t, w)-*comb* in G if:

- $a_1, \dots, a_t \in V(G)$ are distinct, and $B_1, \dots, B_t$ are pairwise disjoint subsets of $V(G) \setminus \{a_1, \dots, a_t\}$;
- for $1 \leq i \leq t$, a_i is adjacent to every vertex in B_i;
- for $i, j \in \{1, \dots, t\}$ with $i \neq j$, a_i has no neighbour in B_j; and
- $B_1, \dots, B_t$ all have cardinality at least w.

The following was proved in [5], strengthening a result of Pach and Tomon [16]:

9.3 *Let $\Gamma, \theta > 0$ with $\theta < 1$. Let (A, B) be a bipartition of a graph G, such that every vertex in B has a neighbour in A and each vertex in A has at most $\Delta > 0$ neighbours in B. Then either*

- $|B| \leq \frac{3^{\theta+1}}{3/2 - (3/2)^{\theta}} \Gamma^{\theta} \Delta^{1-\theta}$*; or*
- *for some integer $k \geq 1$, G contains a $(k, \Gamma k^{-1/\theta})$-comb $((a_i, B_i) : i \in [k])$ where $a_i \in A$ and $B_i \subseteq B$ for all $i \in [k]$.*

10 Finishing the proof

So far, our results have not been focussed on excluding any particular graphs, but now we turn to excluding $\widehat{C_5}$ and its complement, to complete the proof of 1.10. First, we will need the following, proved in [5]:

10.1 *There exists $a > 0$ such that every $\{\widehat{C_5}, \overline{\widehat{C_5}}\}$-free graph G has a clique or stable set of size at least $|G|^a$.*

We use this and the results of the preceding sections to prove the following, which is the last step in the proof of 1.10:

10.2 *For all sufficiently large d, the following holds. Let $x \in (0, d^{-1})$, and let G be a graph with maximum degree at most $\Delta \in (x^3|G|, |G|/d)$. Let $v \in V(G)$ be a vertex of degree at least $x^3|G|$, let A be the set of neighbours of v, and let B be the set of vertices in $V(G) \setminus (A \cup \{v\})$ with at least $x|A|$ neighbours in A. Then either*

- *there are at least $(x^8|G|)^5$ induced copies of $\widehat{C_5}$ in G whose images contain v; or*
- *there are at least $(x^3|A|)^6$ induced copies of $\widehat{C_5}$ or $\overline{\widehat{C_5}}$ in $G[A]$; or*
- *$|B| \leq 80\sqrt{|G|\Delta}$; or*
- *there is an x-sparse or x-dense $(k, |G|/k^d)$-blockade in G for some integer $k \in [2, 1/x]$.*

Proof. Choose a as in 10.1. Choose $d' > 0$ such that 7.4 holds with d, T replaced by d', P_4, and let $b := 1 + 3\log(d')$; then $(2^s)^b \geq 2^s(d')^{2s+1}$ for all $s \geq 1$. We claim that if $d \geq \max((d')^{10/a}, 30b/a, 8/a)$ then d satisfies the theorem.

Let x, G, Δ, v, A, B be as in the theorem. Let $\mathcal{Z}_6$ be the collection of 6-subsets of A that induce subgraphs isomorphic to $\widehat{C_5}$ or $\overline{\widehat{C_5}}$ in G. Let $\mathcal{Z}_2$ be the collection of nonadjacent pairs $\{a_1, a_2\}$ in A such that there is a $(2, x^3|G|)$-comb $((a_1, B_1), (a_2, B_2))$ where B_1, B_2 are x^2-cores in $G[B]$ and the edge-density between them is at least $4x^2$ and at most $1 - 4x^2$. Let $\mathcal{Z}_3$ be the collection of 3-subsets $\{a_1, a_2, a_3\}$ in A such that either

- $\{a_1, a_2, a_3\}$ is a clique and there is a $(3, x^2|G|)$-comb $((a_i, B_i) : i \in \{1, 2, 3\})$ in G with $B_1, B_2, B_3 \subseteq B$, such that there are at least $x^{19}|G|^3$ induced three-vertex paths u_1-u_2-u_3 with $u_i \in B_i$ for all $i \in \{1, 2, 3\}$; or
- $\{a_1, a_2, a_3\}$ is stable and there is a $(3, x^3|G|)$-comb $((a_i, B_i) : i \in \{1, 2, 3\})$ in G with $B_1, B_2, B_3 \subseteq B$, such that the edge-density between every pair among them is at least $1 - 4x^2$.

(1) *There are at least $2|\mathcal{Z}_2|(x^{24}|G|^3)$ copies of $\widehat{C_5}$ in G whose images contain v.*

Let $\{a_1, a_2\} \in \mathcal{Z}_2$; then there is a comb $((a_1, B_1), (a_2, B_2))$ where B_1, B_2 are x^2-cores in $G[B]$ and the edge-density between them is at least $4x^2$ and at most $1 - 4x^2$. By 9.2, there exists $i \in \{1, 2\}$ such that there are at least $x^2|B_i|$ vertices in B_i each having at least $x^2|B_{3-i}|$ neighbours and at least $x^2|B_{3-i}|$ nonneighbours in B_{3-i}; and we may assume $i = 1$. For each such vertex $u \in B_1$ with set of neighbours N_u in B_2, since B_2 is an x^2-core in $G[B]$, the edge-density between N_u and $B_2 \setminus N_u$ is at least x^{10}; and each such edge together with v, u, a_1, a_2 gives a copy of $\widehat{C_5}$ in G. Actually, it gives two copies, since a "copy" is an isomorphism, not just an induced subgraph. Moreover,

$$|N_u|(|B_2| - |N_u|) \geq x^2(1 - x^2)|B_2|^2 \geq x^3|B_2|^2.$$

Thus there are at least

$$2x^2|B_1| \cdot x^{10} \cdot x^3|B_2|^2 = 2x^{15}|B_1||B_2|^2 \geq 2x^{24}|G|^3$$

copies of $\widehat{C_5}$ in G whose images contain v, a_1, a_2. It follows that the number of copies of $\widehat{C_5}$ in G whose images contain v is at least $2|\mathcal{Z}_2|(x^{24}|G|^3)$. This proves (1).

(2) *There are at least* $2|\mathcal{Z}_3|(x^{19}|G|^2)$ *copies of* $\widehat{C_5}$ *in* G *whose images contain* v.

Let $\{a_1, a_2, a_3\} \in \mathcal{Z}_3$; we consider two cases.

- In the first case, $\{a_1, a_2, a_3\}$ is a clique and there is a $(3, x^2|G|)$-comb $((a_i, B_i) : i \in \{1, 2, 3\})$ in G with $B_1, B_2, B_3 \subseteq B$, such that there are at least $x^{19}|G|^3$ induced three-vertex paths u_1-u_2-u_3 with $u_i \in B_i$ for all $i \in \{1, 2, 3\}$. Then each such induced path together with v, a_1, a_3 gives two copies of $\widehat{C_5}$ in G.
- In the second case, $\{a_1, a_2, a_3\}$ is stable and there is a $(3, x^3|G|)$-comb $((a_i, B_i) : i \in \{1, 2, 3\})$ in G with $B_1, B_2, B_3 \subseteq B$, such that the edge-density between every pair among them is at least $1 - 4x^2$. Then there are at least $(1 - 12x^2)|B_1||B_2||B_3| \geq x(x^3|G|)^3 \geq x^{10}|G|^3$ triangles which are (B_1, B_2, B_3)-rainbow, and each together with v, a_1, a_3 gives two copies of $\widehat{C_5}$ in G.

Since the copies of $\widehat{C_5}$ just produced contain a_1 and a_3 but not a_2, we must beware of double-counting them. But there are at least $|\mathcal{Z}_3|/|G|$ pairs $\{a_1, a_3\}$ in A for which there exists $a_2 \in A \setminus \{a_1, a_3\}$ with $\{a_1, a_2, a_3\} \in \mathcal{Z}_3$, and each such pair gives us $2x^{19}|G|^3$ copies of $\widehat{C_5}$, with no double-counting. It follows that there are at least $2(|\mathcal{Z}_3|/|G|)(x^{19}|G|^3) \geq 2|\mathcal{Z}_3|(x^{19}|G|^2)$ copies of $\widehat{C_5}$ in G whose images contain v. This proves (2).

Again, there may be double-counting between (1) and (2); but (1) and (2) imply that there are at least

$$|\mathcal{Z}_2|(x^{24}|G|^3) + |\mathcal{Z}_3|(x^{19}|G|^2)$$

copies of $\widehat{C_5}$ in G whose images contain v. We may assume that

$$|\mathcal{Z}_2|(x^{24}|G|^3) + |\mathcal{Z}_3|(x^{19}|G|^2) \leq (x^8|G|)^5$$

since otherwise the first outcome holds; and so, since $|A| \geq x^3|G|$, we obtain

$$\begin{aligned}
|\mathcal{Z}_2|(x|A|)^{-2} + |\mathcal{Z}_3|(x|A|)^{-3} &\leq |\mathcal{Z}_2|(x^4|G|)^{-2} + |\mathcal{Z}_3|(x^4|G|)^{-3} \\
&\leq (x^7|G|)^{-5} \cdot (|\mathcal{Z}_2|(x^{24}|G|^3) + |\mathcal{Z}_3|(x^{19}|G|^2)) \\
&\leq (x^7|G|)^{-5} \cdot (x^8|G|)^5 = x^5.
\end{aligned}$$

We may assume that $|\mathcal{Z}_6| \leq (x^2|A|)^6$, since otherwise the second outcome holds; and we deduce that

$$|\mathcal{Z}_2|(x|A|)^{-2} + |\mathcal{Z}_3|(x|A|)^{-3} + |\mathcal{Z}_6|(x|A|)^{-6} \leq x^5 + x^6 < 1/4.$$

By 9.1, there exist $A' \subseteq A$ and $B' \subseteq B$ with $|A'| \le 1/x$ and $|B'| \ge \frac{1}{4}|B|$ such that every vertex in B' has a neighbour in A', and A' includes no member of $\mathcal{Z}_2 \cup \mathcal{Z}_3 \cup \mathcal{Z}_6$. We may assume that $|B'| \ge \frac{1}{4}|B| \ge 20\sqrt{|G|\Delta}$, since otherwise the third outcome holds; and now we will prove that the fourth outcome holds. By 9.3, there is a $(k, |G|/k^2)$-comb $((a_i, B_i) : 1 \le i \le k)$ for some integer $k \in [1, 1/x]$, where $a_i \in A'$ and $B_i \subseteq B'$ for $1 \le i \le k$. Since the degree of a_1 is at most Δ, it follows that $|B_1| \le \Delta$, and so $|G|/k^2 \le \Delta \le |G|/d$. Hence $k^2 \ge d$ and so $k \ge d^{1/2} \ge 4$. The set $\{a_1, \dots, a_k\}$ includes no member of $\mathcal{Z}_6$ and thus induces a $\{\widehat{C_5}, \widehat{\overline{C_5}}\}$-free graph in G. So by the definition of a and the choice of d, there exists $I \subseteq \{1, \dots, k\}$ with $|I| \ge k^a \ge d^{a/2} \ge (d')^5$ such that $S := \{a_i : i \in I\}$ is a clique or a stable set. Let $\mathcal{B} := (B_i : i \in I)$; then $\mathcal{B}$ is a $(|I|, w)$-blockade where $w = \lceil |G|/k^2 \rceil \ge x^2|G|$. We may assume that all the sets B_i ($i \in I$) have cardinality exactly w.

(3) *If S is a clique then the fourth outcome holds.*

Since $|I|^{1/5} \ge d'$, there exists an integer $s \ge 1$ maximum such that $|I|^{1/5} \ge 2^{s-1}(d')^{2s-1}$; and so $(2^s)^b \ge 2^s(d')^{2s+1} \ge |I|^{1/5}$ by the choice of b. Since S contains no member of $\mathcal{Z}_3$, for all distinct $p, q, r \in I$ there are fewer than $x^{19}|G|^3$ induced three-vertex paths u_p-u_q-u_r where $u_p \in B_p$, $u_q \in B_q$ and $u_r \in B_r$. Thus, for every subblockade $\mathcal{D}$ of $\mathcal{B}$ with length four, there are fewer than

$$12x^{19}|G|^3 \cdot w \le x^{18}|G|^3 \cdot w = (x^6|G|)^3 w \le (x^4 w)^3 w = (x^3 w)^4$$

$\mathcal{D}$-rainbow copies of P_4 in G (since all the sets B_i have cardinality w). Hence, since $|I| \le k \le 1/x$, and therefore

$$2^{s-1}(d')^{2s-1} \le |I|^{1/5} \le x^{-1/5},$$

7.5 implies that there is a $(2^s, w/|I|)$-blockade in $G[B]$ which is an x-blowup of a cograph; and thus $G[B]$ contains an x-sparse or x-dense $(\lceil 2^{s/2} \rceil, w/|I|)$-blockade. But then the fourth outcome holds, since

$$2^{s/2} \ge |I|^{1/(10b)} \ge k^{a/(10b)} \ge k^{3/d}$$

by the choice of d, and $w/|I| \ge w/k \ge |G|/k^3$. This proves (3).

(4) *If S is stable then the fourth outcome holds.*

Suppose that there exists $i \in I$ such that B_i includes no x^2-core of size at least $|B_i|/k$. Since $k \ge 4$, 8.1 implies that there is an x-sparse $(k, |B_i|/k^2)$-blockade in $G[B_i]$; but then the fourth outcome holds since

$$|B_i|/k^2 \ge |G|/k^4 \ge |G|/k^d.$$

Thus, we may assume that for every $i \in I$, there is an x^2-core $D_i \subseteq B_i$ with

$$|D_i| \ge |B_i|/k \ge |G|/k^3 \ge x^3|G|.$$

Let $\mathcal{D} := (D_i : i \in I)$. Because S includes no member of $\mathcal{Z}_2$, the edge-density between every pair of blocks of $\mathcal{D}$ is at most $4x^2$ or at least $1 - 4x^2$. Let F be the graph with vertex set I such that for all distinct $i, j \in I$, if D_i, D_j have edge-density at least $1 - 4x^2$ then $ij \in E(F)$, and if D_i, D_j have edge-density at most $4x^2$ then $ij \notin E(F)$. Because S contains no member of $\mathcal{Z}_3$, F is triangle-free, and so contains a stable set with size at least $|I|^{1/2} \ge k^{a/2} \ge k^{4/d}$ by the choice of d; and so we may choose a stable set J in F with $|J| = \lceil k^{4/d} \rceil$, and therefore $2 \le |J| \le \lceil x^{-1/2} \rceil$. Thus $(D_i : i \in J)$ is a $(|J|, |G|/k^3)$-blockade in $G[B]$ where the edge-density between every pair of blocks is at most $4x^2$.

Now, for every $i, j \in J$ with $i < j$, let D_{ij} be the set of vertices in D_j with at least $\frac{1}{2}x|D_i|$ neighbours in D_i; then $|D_{ij}| \le 8x|D_j|$. For each $j \in J$, let $B'_j := D_j \setminus (\bigcup_{i \in J, i<j} D_{ij})$; then

$$|B'_j| \ge (1 - 8x|J|)|D_j| \ge \frac{1}{k}|D_j|$$

since $8x|J| \le 8x\lceil x^{-1/2} \rceil \le \frac{1}{2}$. Then for every $i, j \in J$ with $i < j$, every vertex in B'_j has at most $\frac{1}{2}x|D_i| \le x|B'_i|$ neighbours in B'_i; and so $(B'_i : i \in J)$ is an x-sparse $(|J|, |G|/k^4)$-blockade. Since $k^4 \le |J|^d$, this shows that the fourth outcome holds, and so proves (4).

(3) and (4) together yield the fourth outcome. This proves 10.2. ∎

We deduce:

10.3 *Let $\mathcal{H} = \{\widehat{C_5}, \overline{\widehat{C_5}}\}$; then there exists $d > 0$ with the following property. Let $x \in (0, d^{-1})$, and let G be a graph with $\mu_{\mathcal{H}}(x^d, G) < 1$ and maximum degree at most Δ where $\Delta \in (2x^3|G|, |G|/d)$. Let T be the set of vertices of G with degree at least $\frac{1}{2}\Delta$, and for each $v \in T$ let A_v be the set of neighbours of v in G. Then either*

- *$|T| \le x^{42}|G|$; or*
- *for some $v \in T$, there are at most $80\sqrt{|G|\Delta}$ vertices in $V(G) \setminus (A_v \cup \{v\})$ that have at least $x|A_v|$ neighbours in A_v; or*
- *there is an x-sparse or x-dense $(k, \lfloor |G|/k^d \rfloor)$-blockade in G for some integer $k \in [2, 1/x]$.*

Proof. Choose d as in 10.2. Since all larger numbers also satisfy 10.2, we may assume that $d \ge 100$. We claim that d satisfies the theorem. To see this, let $x \in (0, d^{-1})$, and let G be such that $\mu_{\mathcal{H}}(x^d, G) \le 1$, with maximum degree at most Δ where $\Delta \in (2x^3|G|, |G|/d)$. We assume that none of the three outcomes hold; and in particular, $|T| \ge x^{42}|G| \ge 1$. Let $n := \lceil x^{42}|G| \rceil$, and let $v_1, \dots, v_n \in T$ be distinct. For notational convenience, for each $i \in [n]$, let $A_i := A_{v_i}$, and let B_i be the set of vertices in $V(G) \setminus (A_i \cup \{v_i\})$ that have at least $x|A_i|$ neighbours in A_i; then $|B_i| > 80\sqrt{|G|\Delta}$. By 10.2, since its fourth outcome does not hold, it follows that for $1 \le i \le n$, either

- there are at least $(x^8|G|)^5$ copies of $\widehat{C_5}$ in G whose images contain v_i; or
- there are at least $(x^3|A_i|)^6$ copies of $\widehat{C_5}$ or $\overline{\widehat{C_5}}$ in $G[A_i]$.

Since $\mu_{\mathcal{H}}(x^d, G) < 1$, there are fewer than $x^{6d}|G|^6$ copies of one of each of $\widehat{C_5}$, $\overline{\widehat{C_5}}$ in G. Consequently the second bullet above does not hold, since

$$(x^3|A_i|)^6 \ge (x^6|G|)^6 \ge 2x^{6d}|G|^6.$$

Thus the first holds for all $i \in \{1, \dots, n\}$, and so for $1 \le i \le n$, there are at least $(x^8|G|)^5$ copies of $\widehat{C_5}$ in G whose images contain v_i. Each such copy contains at most six of $v_1, \dots, v_n$, and so is counted at most six times; and hence altogether there are at least

$$(x^8|G|)^5 n/6 \ge (x^8|G|)^5 x^{42}|G|/6 = x^{82}|G|^6/6 \ge x^{83}|G|^6 \ge x^{6d}|G|^6$$

copies of $\widehat{C_5}$ in G, again contradicting that contradicting that $\mu_{\mathcal{H}}(x^d, G) \le 1$. This proves 10.3. ∎

Now we can deduce 1.10, and 1.8:

10.4 *The set $\{\widehat{C_5}, \overline{\widehat{C_5}}\}$ is viral. Consequently C_5 and the bull are viral.*

Proof. Let $\mathcal{J} = \{\widehat{C_5}, \overline{\widehat{C_5}}\}$. Choose b such that setting $d = b$ satisfies 10.3; and choose $d \geq b+1$ satisfying 6.1 taking $q = 80$. We claim that d satisfies the theorem. By 5.2, it suffices to show that if G is a graph with $\mu_{\mathcal{H}}(x^d, G) \leq 1$ and maximum degree at most $|G|/d$, there is an x-sparse or $(1-x)$-dense blockade in G of length k and width at least $\lfloor |G|/k^d \rfloor$ for some $k \in [2, 1/x]$. Suppose not. Then, from 6.1 with $q = 80$, there is a number D with $2x^3|G| \leq D \leq |G|/d$, and an induced subgraph G' of G such that, denoting by T the set of vertices in G' that have degree at least $D/2$:

- $|G'| \geq |G|/2$, and G' has maximum degree at most $D/2$;
- $|T| \geq x^2|G|$; and
- for every vertex $v \in T$, with neighbour set A in G' say, there are at least $q(D|G|)^{1/2}$ vertices in $V(G') \setminus A$ that have at least $x|A|$ neighbours in A.

Now $\mu_{\mathcal{H}}(x^b, G') \leq \mu_{\mathcal{H}}(x^{b+1}, G) \leq \mu_{\mathcal{H}}(x^d, G) \leq 1$ by the choice of d. Thus, since G' has no x-sparse or x-dense $(k, |G'|/k^b)$-blockade for any integer $k \in [2, 1/x]$ (note that $|G'|/k^b \geq |G|/k^{b+1} \geq |G|/k^d$), 10.3 implies that $|T| \leq x^{42}|G|'$, a contradiction, since $|T| \geq x^2|G|$. This proves that there is an x-sparse or $(1-x)$-dense blockade in G of length k and width at least $\lfloor |G|/k^d \rfloor$ for some $k \in [2, 1/x]$, and hence $\mathcal{J}$ is viral from 5.2. Let $\mathcal{H} = \{C_5\}$. Since every graph in $\mathcal{J}$ includes the member of $\mathcal{H}$ as an induced subgraph, it follows from 2.1 that C_5 is viral, and similarly the bull is viral. This proves 10.4. ∎

It was proved in [5] that $\{C_6, \overline{C_6}\}$ has the Erdős-Hajnal property, and a proof similar to that of 10.4 shows that $\{C_6, \overline{C_6}\}$ is viral, as we sketch now. In the proof of 10.2, with the same $S = \{a_i : i \in I\}$ and $\mathcal{B} = (B_i : i \in I)$, every $\mathcal{B}$-rainbow induced copy of P_4 would give an induced copy of C_6 whenever S is a clique or stable set in G. Consequently, up to minor numerical adjustments in the exponents of x, one can change $\widehat{C_5}, \overline{\widehat{C_5}}$ to $C_6, \overline{C_6}$, respectively, in the first two outcomes of 10.2, and 10.3 and 10.4 can be modified accordingly. We omit further details.

Acknowledgement

Thanks to Jacob Fox for his encouragement with this paper.

References

[1] N. Alon and J. Fox, "Easily testable graph properties", *Combinatorics, Probabability and Computing*, **24** (2015), 646–657.

[2] N. Alon, J. Pach and J. Solymosi, "Ramsey-type theorems with forbidden subgraphs", *Combinatorica* **21** (2001), 155–170.

[3] M. Bucić, T. Nguyen, A. Scott, and P. Seymour, "Induced subgraph density. I. A loglog step towards Erdős-Hajnal", submitted for publication, `arXiv:2301.10147`.

[4] M. Chudnovsky and M. Safra, "The Erdős-Hajnal conjecture for bull-free graphs", *J. Combinatorial Theory, Ser. B*, **98** (2008), 1301–1310.

[5] M. Chudnovsky, A. Scott, P. Seymour, and S. Spirkl, "Erdős-Hajnal for graphs with no five-hole", *Proceedings of the London Math. Soc.*, **126** (2023), 997–1014, `arXiv:2102.04994`.

[6] M. Chudnovsky, A. Scott, P. Seymour and S. Spirkl, "Pure pairs. I. Trees and linear anticomplete pairs", *Advances in Math.*, **375** (2020), 107396, `arXiv:1809.00919`.

[7] M. Chudnovsky, A. Scott, P. Seymour and S. Spirkl, "Pure pairs. II. Excluding all subdivisions of a graph", *Combinatorica* **41** (2021), 279–405, `arXiv:1804.01060`.

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

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

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

[11] J. Fox, T. Nguyen, A. Scott and P. Seymour, "Induced subgraph density. II. Sparse and dense sets in cographs", submitted for publication, `arXiv:2307.00801`.

[12] T. Nguyen, A. Scott and P. Seymour, "Induced subgraph density. IV. New graphs with the Erdős-Hajnal property", submitted for publication, `arXiv:2307.06455`.

[13] T. Nguyen, A. Scott and P. Seymour, "Induced subgraph density. VI. Graphs that approach Erdős-Hajnal", in preparation.

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

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

[16] J. Pach and I. Tomon, "Erdős-Hajnal-type results for ordered paths", *J. Combinatorial Theory, Ser. B* **151** (2021), 21–37, `arXiv:2004.04594`.