Induced subgraph density. V. All paths approach Erdős-Hajnal

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

Alex Scott[2]
University of Oxford,
Oxford OX2 6GG, UK

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

July 4, 2022; revised October 12, 2024

Abstract

The Erdős-Hajnal conjecture says that, for every graph H, there exists $c > 0$ such that every H-free graph on n vertices has a clique or stable set of size at least n^c. In this paper we are concerned with the case when H is a path. The conjecture has been proved for paths with at most five vertices, but not for longer paths. We prove that the conjecture is "nearly" true for all paths: for every path H, all H-free graphs with n vertices have cliques or stable sets of size at least $2^{(\log n)^{1-o(1)}}$.

1 Introduction

A graph is *H-free* if it has no induced subgraph isomorphic to H. A famous conjecture of Erdős and Hajnal [6, 7] from 1977 says:

1.1 The Erdős-Hajnal Conjecture: *For every graph H there exists $c > 0$ such that every H-free graph G has a stable set or clique of size at least $|G|^c$.*

This remains open, and has been proved only for a very limited set of graphs H (although see [12, 14] for a variety of new graphs H that satisfy 1.1).

If H is a graph, for each $n > 0$ let $f_H(n)$ be the largest integer such that every H-free graph with at least n vertices has a stable set or clique with size at least $f_H(n)$. Thus, the Erdős-Hajnal conjecture says that

1.2 Conjecture: *For every graph H there exists $c > 0$ such that $f_H(n) \geq n^c$ for all $n \geq 0$.*

What can we actually prove about $f_H(n)$? Erdős and Hajnal [7] proved (logarithms are to base two, throughout the paper):

1.3 *For every graph H, there exists $c > 0$ such that $f_H(n) \geq 2^{c\sqrt{\log n}}$ for all $n > 0$.*

In [3], with Bucić, we improved this, showing:

1.4 *For every graph H, there exists $c > 0$ such that $f_H(n) \geq 2^{c\sqrt{\log n \log \log n}}$ for all $n > 0$.*

For general graphs H, this is the best bound known.

The case when H is a path is of particular interest. Paths are very simple graphs, and yet until recently, the Erdős-Hajnal conjecture was open when H is the five-vertex path P_5. This was the smallest graph for which the conjecture was open, and was therefore a focus of attention. In a substantial breakthrough for the case $H = P_5$, Blanco and Bucić [1] improved the power in the exponent from $1/2$ to $2/3$:

1.5 *There exists $c > 0$ such that $f_{P_5}(n) \geq 2^{c(\log n)^{2/3}}$ for all $n > 0$.*

Their argument uses a complex structural analysis of P_5-free graphs, and does not appear to extend to longer paths. However, more recently, using completely different methods, we fully resolved the case $H = P_5$, showing in [14] that P_5 satisfies the Erdős-Hajnal conjecture:

1.6 *There exists $c > 0$ such that $f_{P_5}(n) \geq n^c$ for all $n > 0$.*

But the Erdős-Hajnal conjecture for general paths remains open, and it is natural to ask whether it is possible to obtain an improvement in the exponent such as 1.5 for general paths.

In the present paper, we do better than that: we show that paths of any length have the "near Erdős-Hajnal" property, that is, for every $d < 1$ and every path H, there exists $c > 0$ such that every H-free graph G has a clique or stable set of size at least $2^{(c \log |G|)^d}$ (note that $d = 1$ is the Erdős-Hajnal conjecture itself). In other words:

1.7 *For every path P, $f_P(n) \geq 2^{(\log n)^{1-o(1)}}$.*

We prove a stronger form of this (1.10 below), showing that P-free graphs contain large induced subgraphs which are very sparse or very dense. Before stating this, we need some further discussion about dense and sparse subgraphs.

For $\varepsilon > 0$, a subset $S \subseteq V(G)$ is

- ε-*sparse* if the induced subgraph $G[S]$ has maximum degree at most $\varepsilon |S|$;

- $(1 - \varepsilon)$-*dense* if $\overline{G}[S]$ is ε-sparse, where $\overline{G}$ is the complement graph of G; and

- ε-*restricted* if S is either ε-sparse or $(1 - \varepsilon)$-dense.[1]

Note that if $|S| \leq 2$ then S is trivially ϵ-restricted.

An important result of Rödl [16] shows that, for fixed $\epsilon > 0$, H-free graphs have ϵ-sparse or $(1 - \epsilon)$-dense subsets of linear size:

1.8 *For all $0 < \varepsilon \leq 1/2$, there exists $\delta > 0$ such that for every H-free graph G, there is an ε-restricted subset $S \subseteq V(G)$ with $|S| \geq \delta|G|$.*

This does not imply the Erdős-Hajnal conjecture, as the dependence of δ on ϵ could be very poor. Indeed, Rödl's proof used Szemerédi's Regularity Lemma, which leads to tower-type dependence. However, Fox and Sudakov [8] conjectured that a *polynomial* version of Rödl's theorem should hold, where δ depends polynomially on ε. More exactly:

1.9 Fox-Sudakov 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 $S \subseteq V(G)$ with $|S| \geq \varepsilon^c|G|$ such that S is ε-restricted.*

Let us say that a graph H is *polynomial Rödl* if it satisfies 1.9. It is straightforward to show that if H is polynomial Rödl then H also satisfies the Erdős-Hajnal conjecture. Indeed, the Fox-Sudakov conjecture interpolates between Rödl's theorem and the Erdős-Hajnal conjecture: taking ε to be a constant gives 1.8, and taking ε to be a small negative power of n yields 1.1. Recently, Bucić, Fox and Pham [2] showed the reverse implication: the Erdős-Hajnal conjecture and the Fox-Sudakov conjecture are in fact equivalent.

Both conjectures are currently out of reach. But the bound on δ in 1.8 has been strengthened: Fox and Sudakov [8] proved that 1.8 holds with

$$\delta = \varepsilon^{c_H \log(1/\varepsilon)},$$

and this was improved in [3] to

$$\delta = \varepsilon^{c_H \frac{\log(1/\varepsilon)}{\log\log(1/\varepsilon)}}.$$

In order to prove 1.7, we will show that a much better bound than this holds when H is a path: we will prove that every path "nearly" satisfies the Fox-Sudakov conjecture. Here is our main result:

---

[1]Note that we are using 'local' density conditions: for ϵ-sparse sets, we demand that every vertex has degree at most $\varepsilon |S|$, rather than the 'global' condition that there are at most $\varepsilon \binom{|S|}{2}$ edges; and similarly for $(1 - \epsilon)$-dense sets. It is straightforward to move between the two, at the cost of a constant factor in the value of ε and the size of S.

1.10 *For every path P and every $\alpha > 0$, there exists $c > 0$ such that for all $0 < \varepsilon \leq 1/2$, every P-free graph G contains an ε-restricted subset $S \subseteq V(G)$ with $|S| \geq \delta|G|$, where*

$$\delta = \varepsilon^{c(\log \frac{1}{\varepsilon})^{\alpha}}.$$

Let us show that 1.10 implies 1.7.

1.11 *Let G be a graph with $|G| \geq 2$, let $a > 0$ and $c := (2a+2)^{-1}$. Let $\alpha \geq 0$, and assume that, for every $\varepsilon \in (0, 1/2)$, there is an ε-restricted $S \subseteq V(G)$ with*

$$|S| \geq \varepsilon^{a(\log \frac{1}{\varepsilon})^{\alpha}}|G|.$$

Then G contains a clique or stable set of size at least $2^{c(\log |G|)^{\beta}}$, where $\beta = 1/(1 + \alpha)$.

Proof. Since G has a clique or stable set of size two, we may assume that $2^{c(\log |G|)^{\beta}} > 2$. Set

$$\varepsilon := 2^{-2c(\log |G|)^{1/(1+\alpha)}}$$

and let

$$\delta = \varepsilon^{a(\log \frac{1}{\varepsilon})^{\alpha}} = 2^{-a(\log \frac{1}{\varepsilon})^{1+\alpha}} = 2^{-a(2c)^{1+\alpha} \log |G|} \geq |G|^{-2ac}.$$

From the hypothesis, there is an ε-restricted $S \subseteq V(G)$ with

$$|S| \geq \delta|G| \geq |G|^{1-2ac} = |G|^{2c} = 2^{2c \log |G|} \geq 2^{2c(\log |G|)^{\beta}} = \varepsilon^{-1}.$$

Thus, since S is ε-restricted, $G[S]$ or its complement has maximum degree at most $\varepsilon|S|$ and so contains a clique or stable set of size at least

$$\frac{|S|}{\varepsilon|S| + 1} \geq \frac{1}{2\varepsilon} \geq \varepsilon^{-1/2} = 2^{c(\log |G|)^{\beta}}.$$

This proves 1.11. ∎

Note that if the Fox-Sudakov conjecture held, then we could take $\alpha = 0$ in 1.10. We could then choose ϵ to be some small negative power of $|G|$ (so that $\log(1/\varepsilon) \approx \log |G|$), and applying 1.10 would then give an ε-restricted subset of polynomial size. By Turán's theorem or a greedy argument, this would yield a clique or stable set of polynomial size, giving the Erdős-Hajnal theorem for P-free graphs. However we do not quite have this: since we must take $\alpha > 0$ we do not quite get polynomial dependence, as if we take $\log(1/\varepsilon) \approx \log n$ then the set S given by 1.10 is too small. Instead, we need to take $\log(1/\varepsilon) \approx (\log n)^{\beta}$ in the proof of 1.11, where $\beta = 1/(1 + \alpha)$ is slightly less than 1.

The rest of the paper is organized as follows. In section 2, we discuss the relationship between dense or sparse multipartite structures (blockades) and dense or sparse subgraphs. We discuss first the (realtively straightforward) sparse case and then the (harder) dense case in section 3, and we put the proof of 1.10 together in section 4. We conclude with some further discussion in section 5.

We use standard terminology throughout. If G is a graph, $G[X]$ denotes the induced subgraph with vertex set X; $|G|$ denotes the number of vertices of G; and $\overline{G}$ is the complement graph of G.

2 Blockades

Recall that a graph H is *polynomial Rödl* if there exists $c > 0$ such that for all $\varepsilon \in (0, 1/2)$, every H-free graph G contains an ε-restricted subset $S \subseteq V(G)$ with $|S| \geq \delta|G|$, where

$$\delta = \varepsilon^c |G|.$$

We will say that a graph H is *near-polynomial Rödl* if for every $\alpha > 0$, there exists $c > 0$ such that for every $\varepsilon \in (0, \frac{1}{2})$, every H-free graph G contains an ε-restricted $S \subseteq V(G)$ with $|S| \geq \delta|G|$, where

$$\delta = \varepsilon^{c(\log(1/\varepsilon))^\alpha}.$$

Thus, our main theorem 1.10 says that every path is near-polynomial Rödl.

In order to prove 1.10, we will need to find dense or sparse multipartite structures which we refer to as blockades. A *blockade* in a graph G is a finite sequence $(B_1, \ldots, B_n)$ of (possibly empty) disjoint subsets of $V(G)$; its *length* is n and its *width* is $\min_{i \in [n]} |B_i|$. For $k, w \geq 0$, $(B_1, \ldots, B_n)$ is a (k, w)-*blockade* if its length is at least k and its width is at least w.

We will need certain density conditions to hold. If $A, B \subseteq V(G)$ are disjoint, we say A is x-*sparse* to B if every vertex in A has at most $x|B|$ neighbours in B, and A is $(1-x)$-*dense* to B if every vertex in A has at least $(1-x)|B|$ neighbours in B. For $x \in (0, \frac{1}{2})$, a blockade $(B_1, \ldots, B_n)$ is x-*sparse* if B_j is x-sparse to B_i for all $i, j \in [n]$ with $i < j$, and $(1-x)$-*dense* if B_j is $(1-x)$-dense to B_i for all $i, j \in [n]$ with $i < j$ (note the asymmetry: vertices in sets of the blockade are dense or sparse to earlier sets in the blockade, but not necessarily to later sets). We say that a blockade is *complete* if, for all $i \neq j$, every vertex of B_i is adjacent to every vertex of B_j; *anticomplete* if (for all $i \neq j$) every vertex of B_i is nonadjacent to every vertex of B_j; and x-*restricted* if it is x-sparse or $(1-x)$-dense.

There are three parameters we care about: the length, width, and sparsity (or density). It is easier to prove that certain graphs contain blockades with some desired combination of the three parameters, than to prove directly that they contain large ε-restricted sets. But the reason blockades are useful for us is that there is a transference theorem (2.1 below), that says that if a graph G and all its large induced subgraphs admit blockades with certain parameters, then G must contain a large ε-restricted set.

A function $h \colon (0, \frac{1}{2}) \to \mathbb{R}^+$ is *subreciprocal* if it is nonincreasing and $1 < h(x) \leq 1/x$ for all $x \in (0, \frac{1}{2})$. For a subreciprocal function h, a graph H is h-*dividing*[2] if there are $c \in (0, \frac{1}{2})$ and $d > 1$ such that for every $x \in (0, c)$ every H-free graph G contains an x-sparse or $(1-x)$-dense $(h(x), \lfloor x^d |G| \rfloor)$-blockade in G. Here is the transference theorem that we will use (see [3], or [12] for an alternative proof of a weaker statement).[3]

2.1 *Let h be subreciprocal, and let H be an h-dividing graph. Then there exists $C > 0$ such that for every $\varepsilon \in (0, \frac{1}{2})$, every H-free graph G contains an ε-restricted $S \subseteq V(G)$ with $|S| \geq \delta|G|$, where*

$$\delta = \varepsilon^{C \log(\frac{1}{\varepsilon})/\log h(\varepsilon)}.$$

---

[2]In previous papers, we use the term h-*divisive* to refer to graphs having this property when a few copies of H are allowed. In this paper, we work with the slightly simpler property of being h-dividing. Bucić, Fox and Pham [2] showed that these two properties are equivalent.

[3]The version in [3] uses edge density rather than ε-restricted subsets. However, it is straightforward to adjust for this by changing the constants.

This result allows us to move between dense or sparse multipartite substructures (blockades) and dense or sparse subsets. This approach has been used a number of times: the challenge is to find blockades that are both long and wide, and satisfy strong restriction properties. For general graphs H, it is possible to find blockades that very restricted but not very long:

- Erdős and Hajnal [7] proved that for every graph H, there exists $d > 0$ such that for all $x \in (0, 1/2]$, every H-free graph admits disjoint sets A, B of size at least $\lfloor x^d |G| \rfloor$ such that B is x-sparse or $(1-x)$-dense to A; in other words, an x-sparse or $(1-x)$-dense $(2, \lfloor x^d |G| \rfloor)$ blockade. From this they deduced their result 1.3.

- In [3], we (with Bucić) proved a strengthening: for every graph H, there exists $d > 0$ such that for all $x \in (0, 1/2]$, every H-free graph admits an x-sparse or $(1-x)$-dense $(\log(1/x), \lfloor x^d |G| \rfloor)$-blockade. This allowed us to deduce 1.4.

To prove results with *polynomial* dependence, blockades with much better parameters are required. In general, it is a very hard challenge to find suitable blockades, but there has been some progress. Let $\mathcal{G}$ be a class of graphs that is closed under taking induced subgraphs.

- If we can find complete or anticomplete pairs of *linear* size, then we can obtain polynomial-size cliques or stable sets. More precisely, we say that $\mathcal{G}$ has the *strong Erdős-Hajnal property* if there is $c > 0$ such that every $G \in \mathcal{G}$ admits a complete or anticomplete $(2, \lfloor c|G| \rfloor)$-blockade. It is not hard to show that this is enough to give cliques or stable sets of polynomial size. Unfortunately, for almost all H, the class of H-free graphs does not have the strong Erdős-Hajnal property, but there are interesting classes where it does hold (for example if we exclude both a forest and the complement of a forest [4]).

- It may be possible to trade off between length and width. Suppose that there is some $d > 0$ such that every $G \in \mathcal{G}$ with at least two vertices admits a $(k, |G|/k^d)$-blockade for some $k \geq 2$ (which may be different for different G) . This is the "quasi-Erdős-Hajnal property" [15, 17], and is again sufficient to give cliques or stable sets of polynomial size. It was shown in [5] that the class of C_5-free graphs has the quasi-Erdős-Hajnal property, proving the Erdős-Hajnal conjecture for C_5.

- We can sometimes also use the density of G. Suppose that every $G \in \mathcal{G}$ contains either a suitable blockade or a large induced subgraph with much better density parameters: say, every y-restricted $G \in \mathcal{G}$ contains either a suitably restricted $(\text{poly}(1/y), \lfloor \text{poly}(y)|G| \rfloor)$-blockade, or a λy-restricted subset of size at least $\text{poly}(\lambda)|G|$, where $\lambda \in (0,1)$ may depend on G. Iterating this, we can pass to a succession of ever more restricted subgraphs until we reach either a polynomial size clique or stable set or a suitable blockade. This is the method of iterative sparsification (introduced in [12], and further developed in [13, 14]).

In all these results, the main obstacle is proving the existence of blockades that are sufficiently long, wide and restricted.

In this paper we are interested in P-free graphs, where P is a path. If we could show that P is h^*-dividing, where

$$h^*(x) = (1/x)^c = 2^{c \log \frac{1}{x}},$$

then 2.1 would yield the polynomial Rödl property. However, the blockades that we obtain are not sufficiently long to show this. Instead, we will will work with a sequence of non-polynomial approximations to h^*. For each integer $s \geq 0$, let $h_s \colon (0, \frac{1}{2}) \to \mathbb{R}^+$ be the function defined by

$$h_s(x) := 2^{(\log \frac{1}{x})^{\frac{s}{s+1}}}$$

for all $x \in (0, \frac{1}{2})$. (Thus h_s is subreciprocal.) We will show that:

2.2 *Every path P is h_s-dividing for all integers $s \geq 0$.*

Let us first deduce our main result 1.10.

Proof of 1.10 (assuming 2.2). We must show that for all $\alpha > 0$, there exists $c > 0$ such that for all $\varepsilon \in (0, 1/2)$, every H-free graph G contains an ε-restricted $S \subseteq V(G)$ with $|S| \geq \delta|G|$, where

$$\delta = \epsilon^{c(\log \frac{1}{\varepsilon})^\alpha}.$$

Choose an integer s such that $\frac{1}{s+1} \leq \alpha$. Since P is h_s-dividing, by 2.1 there exists $C > 0$ such that for every $\varepsilon \in (0, \frac{1}{2})$, every H-free graph G contains an ε-restricted $S \subseteq V(G)$ with $|S| \geq \eta|G|$, where

$$\eta = \varepsilon^{C \log(\frac{1}{\varepsilon})/\log(h_s(\varepsilon))} = \varepsilon^{C \log(\frac{1}{\varepsilon})/(\log \frac{1}{\varepsilon})^{\frac{s}{s+1}}} = \varepsilon^{C(\log(\frac{1}{\varepsilon}))^{\frac{1}{s+1}}} \geq \varepsilon^{C(\log(\frac{1}{\varepsilon}))^\alpha}.$$

Taking $c = C$, we have $\eta \geq \delta$, and the result follows. This proves 1.10. ∎

As noted in the previous section, 1.10 implies that every path is near-polynomial Rödl.

The proof of 2.2 will run by induction on s: to show that P is h_s-dividing, we first use the fact that P is h_{s-1}-dividing to find a large restricted subset, and then work inside this to obtain the required blockade. We will need to deal with two cases: an easier sparse case and a harder dense case. We address these separately in the next section. The main argument is then given in section 4.

3 Sparse and dense graphs

The proof of 2.2 proceeds by induction on s. Every graph is h_0-dividing, so we may assume that $s > 0$ and that the path P is h_{s-1}-dividing. We are given a P-free graph G, and to show that P is h_s-dividing we need to find a blockade in G with certain parameters. By 2.1, we know that every large subset of $V(G)$ includes a somewhat smaller subset S that is either very dense or very sparse.

There is no symmetry between the dense and sparse cases, because we are excluding a path but not its complement. As usual with problems about excluding a path, our task is easier if the host graph is sparse (see, for example, [4]). We first prove a result that will allow us to win if ever the subset S provided by 2.1 turns out to be sparse.

3.1 *Let P be a path with $k \geq 2$ vertices, and let $0 < y \leq 1/(60k)$. Let G be a y^2-sparse graph. Then either:*

- *G contains a copy of P, or*

- *G contains an anticomplete $(1/y, \lfloor y^2|G| \rfloor)$-blockade in G.*

Proof. We may assume that $|G| \geq 1/y^2$ or the result is trivial. Let $\mu = 1/(12y)$.

Let us say that a set A of vertices in a graph G is *poorly expanding* (in G) if $|N_G(A)| \leq \mu|A|$, where $N_G(A)$ is the set of vertices $u \notin A$ that have a neighbour in A. We repeatedly remove poorly expanding sets together with their neighbourhoods, as follows: let $G_0 = G$ and, for $i \geq 0$, if G_i contains a set A_{i+1} of vertices of size at most $y^2|G|$ that is poorly expanding in G_i, then let $B_{i+1} = N_{G_i}(A_{i+1})$ and set $G_{i+1} = G_i \setminus (A_{i+1} \cup B_{i+1})$. The process terminates with some induced subgraph G_t that has no poorly expanding set of size at most $y^2|G|$.

Suppose first that $|G_t| \leq |G|/2$. Let $A = \bigcup_{i=1}^{t} A_i$ and $B = \bigcup_{i=1}^{t} B_i$, so $G_t = G \setminus (A \cup B)$ and $|B| \leq \mu|A|$. Since $|A \cup B| \geq |G|/2$, it follows that $|A| \geq |G|/(2 + 2\mu) > 3y|G|$. The sets A_i are pairwise anticomplete, so each component of $G[A]$ has size at most $y^2|G|$. We group the components together into as many sets as possible such that each contains between $y^2|G|$ and $2y^2|G|$ vertices. If we get at least $1/y$ sets then we have found our blockade. If not, then there are at most $y^2|G|$ unused vertices and so $|A| \leq (1/y)2y^2|G| + y^2|G| < 3y|G|$, giving a contradiction.

Otherwise, $|G_t| > |G|/2$ and G_t contains no poorly expanding set of size at most $y^2|G|$. We work inside G_t for the rest of the argument. We define sets of vertices $R_1, \ldots, R_{k-1}$ and $S_0, \ldots, S_{k-1}$ as follows: let $S_0 \subseteq V(G_t)$ be any set of $\lfloor y^2|G|\rfloor$ vertices of G_t; for $i \geq 1$, let R_i be a minimal subset of S_{i-1} such that $|N_{G_t}(R_i) \setminus \bigcup_{j<i} S_j| \geq y^2|G|$; and let $S_i = N_{G_t}(R_i) \setminus \bigcup_{j<i} S_j$. If we can do this successfully, then we can find an induced copy $v_1\text{-}\cdots\text{-}v_k$ of P: let v_k be any vertex of S_{k-1} and, for $i = k-1, \ldots, 1$, let v_i be any neighbour of v_{i+1} in R_i.

It is therefore enough to show that we can find sets R_i and S_i at each stage. Suppose that $1 \leq i \leq k-1$ and we have defined $S_0, \ldots, S_{i-1}$. Let $T \subseteq S_{i-1}$ be any subset of size $\lfloor y^2|G|\rfloor$. Then, by our assumption on poorly expanding sets, $|N_{G_t}(T)| \geq \mu\lfloor y^2|G|\rfloor \geq y|G|/24$. By the minimality of R_j and since G is y^2-sparse, each set S_j has size at most $2y^2|G|$ and so $|N(T) \setminus \bigcup_{0 \leq j < i} S_j| \geq y|G|/24 - 2ky^2|G| \geq y^2|G|$. It follows that we can pick R_i and S_i. This proves 3.1. ∎

The dense case is much more difficult. This is handled by the next lemma, which will allow us either to extract a dense blockade, or to move to a reasonably large induced subgraph with no large, dense parts.

3.2 *Let P be a path with $|P| \geq 1$, let $a = 3|P|$, and let $0 < x \leq y \leq 1/100$. For every P-free graph G, either:*

(a) there is $S \subseteq V(G)$ with $|S| \geq x^a|G|$ such that every $S' \subseteq S$ with $|S'| \geq x|S|$ has density at most $(1 - y^3)$; or

(b) there is a $(1 - x)$-dense $(1/y, \lfloor x^a|G|\rfloor)$-blockade in G.

Proof. Let us assume that neither (a) nor (b) holds. It follows that $x^a|G| \geq 1$, since otherwise (b) holds. A *non-neighbour* of a vertex v means a vertex different from and nonadjacent to v, and the *antidegree* of v is the number of its non-neighbours. We claim first:

(1) For every $S \subseteq V(G)$ with $|S| \geq 2x^{a-1}|G|$, there exists $C \subseteq S$ with $|C| \geq x(1+y)|S|/2$, such that

- *C is $(1 - 2y^3)$-dense; and*

- *for all disjoint $X, Y \subseteq C$ with $|X| \geq (1 - y/4)|C|$ and $|Y| \geq x^a|G|$, at least $y|X|/4$ vertices in X have at least $x|Y|$ non-neighbours in Y.*

7

Since $|S| \geq x^a|G|$, and (a) does not hold, there exists $S' \subseteq S$ with $|S'| \geq x|S|$ such that S' is $(1-y^3)$-dense. Choose a $(1-x)$-dense blockade $(B_1, \ldots, B_{n-1}, C)$ in $G[S']$ with n maximum such that $|B_1|, \ldots, |B_{n-1}|, |C| \geq x^a|G|$ and $|C| \geq (1-(n-1)y/2)|S'|$. (This is possible because $|S'| \geq x^a|G|$, and so we can take $n=1$ and $C = S'$.) We may assume that $n < 1/y$, and so

$$|C| \geq (1-(n-1)y/2)|S'| = (1+y/2-ny/2)|S'| \geq (1+y)|S'|/2 \geq x(1+y)|S|/2.$$

In particular, $|C| \geq |S'|/2$, and consequently C is $(1-2y^3)$-dense.
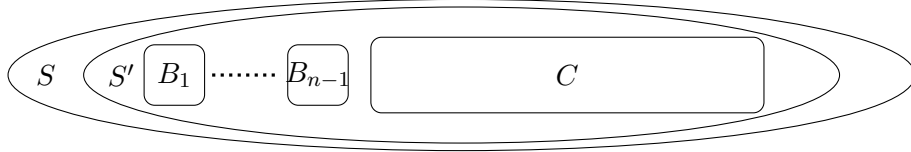

Figure 1: For step (1) of the proof of 3.2.

Suppose that $X, Y \subseteq C$ are disjoint, with $|X| \geq (1-y/4)|C|$ and $|Y| \geq x^a|G|$. It follows that

$$|X| \geq (1-y/4)(1+y)|S'|/2 \geq |S'|/2.$$

Since $|Y| \geq x^a|G|$, and $(1-ny/2)|S'| \geq |S'|/2 \geq x^a|G|$, fewer than $(1-ny/2)|S'|$ vertices in X are $(1-x)$-dense to Y, from the maximality of n. Since $|C| \geq (1-(n-1)y/2)|S'|$, it follows that at least $y|S'|/2 - |Y|$ vertices in X have at least $x|Y|$ non-neighbours in Y. But $|Y| \leq y|C|/4$, since $X \cap Y = \emptyset$, and so $y|S'|/2 - |Y| \geq y|S'|/2 - y|C|/4 \geq y|C|/4 \geq y|X|/4$. This proves (1).

Next we use a new version of the "Gyárfás path" argument, originally used to find induced paths in graphs with large chromatic number (see [9, 10]). The usual form of the argument grows a path $v_1\text{-}\cdots\text{-}v_k$ one vertex at a time, walking towards some part of the graph with large chromatic number: at each stage, v_i has neighbours in some connected part of the graph with large chromatic number that is nonadjacent to $\{v_1, \ldots, v_{i-1}\}$, and we choose v_{i+1} from this part. If the neighbourhood of every vertex has small chromatic number, this can be made to work. However, the current setup is rather different: we are dealing with cardinality rather than chromatic number, and we are working inside a graph that is very dense. There is no problem arranging that the last vertex of the path has many neighbours; the issue is to arrange that there are many vertices with no neighbours in the path, and to maintain this as we grow the path.

Let $k := |P|$, so $a = 3k$. Define $a_1 := x/2$, and $b_1 := x^2y/8$; and for $2 \leq t \leq k$, define $a_t := x^{2t-1}y/2^{t+2} = (x/2)b_{t-1}$ and $b_t := x^{2t}y/2^{t+2} = (x^2/2)b_{t-1}$. For $1 \leq t \leq k$ let us say a t-brush is an induced path of G with vertices $v_1\text{-}\cdots\text{-}v_t$ in order, such that there exist subsets $A, B \subseteq V(G) \setminus \{v_1, \ldots, v_t\}$ with the following properties:

- every vertex in A is adjacent to v_t and is nonadjacent to $v_1, \ldots, v_{t-1}$;

- every vertex in B has no neighbours in $\{v_1, \ldots, v_t\}$;

- $|A| \geq a_t|G|$ and $|B| \geq b_t|G|$;

- for every $Y \subseteq B$ with $|Y| \geq x^a|G|$, there are at least $y|A|/4$ vertices in A that have at least $x|Y|$ non-neighbours in Y; and

8

- every vertex in B has at most $3y^3|A|$ non-neighbours in A.

(2) *There is a 1-brush.*

Since $|G| \geq 2x^{a-1}|G|$, (1) implies that there exists $C \subseteq V(G)$ with $|C| \geq x(1+y)|G|/2$, such that C is $(1-2y^3)$-dense, and and for all disjoint $X, Y \subseteq C$ with $|X| \geq (1-y/4)|C|$ and $|Y| \geq x^a|G|$, at least $y|X|/4$ vertices in X have at least $x|Y|$ non-neighbours in Y.

Suppose first that no vertex of C has at least $b_1|G|$ nonneighbours in C. Let $Y \subseteq C$ be a set of $\lceil x^a|G| \rceil$ vertices and let $X = C \setminus Y$. Then the total number of nonedges between X and Y is at most the sum of antidegrees of vertices in Y, which is at most $b_1|G||Y|/2 = x^2y|G||Y|/8$. On the other hand,

$$|Y| \leq x^a|G| + 1 \leq 2x^a|G| \leq 4x^{a-1}|C| \leq (x/4)|C| \leq (y/4)|C|,$$

so by our choice of C the number of nonedges is at least $(y|X|/4)x|Y|$. It follows that

$$(y|X|/4)x|Y| \leq x^2y|G||Y|/8,$$

and so $|X| \leq x|G|/2$, which gives a contradiction as $|X| \geq (1-y/4)|C| > x|G|/2$.

So let v_1 be a vertex of C with at least $b_1|G|$ nonneighbours in C. Let A be the set of its nonneighbours in C, and let $B = C \setminus A$. As C is $(1-2y^3)$-dense, we have

$$|A| \geq (1-2y^3)|C| \geq x|G|/2 = a_1|G|;$$

and we clearly have $|B| \geq b_1|G|$. Since A and B are subsets of C, the last two bullets in the definition of a 1-brush are also satisfied. This proves (2).

(3) *Let $1 \leq t \leq k-1$, and let $v_1\text{-}\cdots\text{-}v_t$ be a t-brush. Then there is a vertex v such that $v_1\text{-}\cdots\text{-}v_t\text{-}v$ is a $(t+1)$-brush.*

Choose A, B satisfying the five bullets in the definition of "t-brush". Since $b_t = (x^2/2)^t y/4$, and $t \leq k-1$, and $a \geq 3k$, it follows that

$$|B| \geq b_t|G| = (x^2/2)^t y|G|/4 \geq x^{3k-4}|G| \geq 2x^{a-1}|G|.$$

By (1), there exists $C \subseteq B$ with $|C| \geq x(1+y)|B|/2$, such that C is $(1-2y^3)$-dense, and for all disjoint $X, Y \subseteq C$ with $|X| \geq (1-y/4)|C|$ and $|Y| \geq x^a|G|$, at least $y|X|/4$ vertices in X have at least $x|Y|$ non-neighbours in Y.

Since $v_1\text{-}\cdots\text{-}v_t$ is a t-brush, each vertex in C has at most $3y^3|A|$ non-neighbours in A, and so at most $y|A|/8$ vertices in A have at least $24y^2|C|$ non-neighbours in C. On the other hand, there are at least $y|A|/4$ vertices in A that have at least $x|C|$ non-neighbours in C; and so there is a set $D \subseteq A$ with $|D| \geq y|A|/8$, such that for each $v \in D$, the number of its non-neighbours in C is between $x|C|$ and $24y^2|C|$.
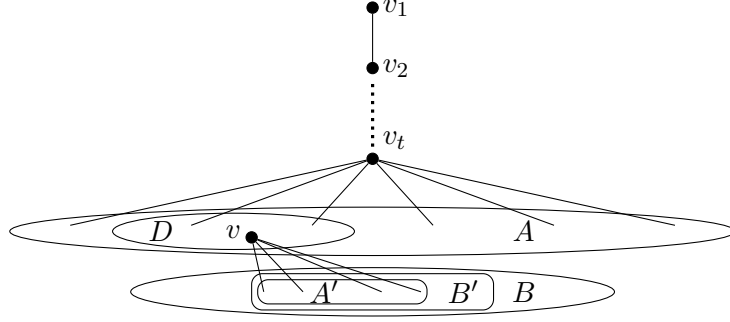
9

Figure 2: For step (3). $C = A' \cup B'$.

Let $v \in D$. We claim that $v_1\text{-}\cdots\text{-}v_t\text{-}v$ is a $(t+1)$-brush. Let A' be the set of all neighbours of v in C, and let $B' = C \setminus A'$. We will show that A', B' satisfy the five conditions in the definition of a $(t+1)$-brush. The first two are immediate. For the third,

$$|A'| \geq (1 - 24y^2)|C| \geq (1 - 24y^2)x(1+y)|B|/2 \geq (x/2)b_t|G| = a_{t+1}|G|,$$

and

$$|B'| \geq x|S| \geq x(x/2)|B| \geq (x^2/2)b_t|G| = b_{t+1}|G|.$$

For the fourth condition, suppose that $Y \subseteq B'$ with $|Y| \geq x^a|G|$. From the choice of C, since $|A'| \geq (1 - 24y^2)|C| \geq (1 - y/4)|C|$, there are at least $y|A'|/4$ vertices in A' that have at least $x|Y|$ non-neighbours in Y. Finally, for the fifth condition, since C is $(1 - 2y^3)$-dense, each vertex in B' has at most $2y^3|C| \leq 3y^3|A'|$ non-neighbours in A'. This proves (3).

The result now follows by induction: (2) shows that there is a 1-brush, and (3) provides the inductive step. Thus there is a k-brush, as required. This proves 3.2. ∎

4 Decreasing density

We remind the reader that if h is subreciprocal, a graph H is h-*dividing* if there are $c \in (0, \frac{1}{2})$ and $d > 1$ such that for every $x \in (0, c)$ and every H-free graph G admits an x-sparse or $(1 - x)$-dense $(h(x), \lfloor x^d|G|\rfloor)$-blockade in G. Also, for each integer $s \geq 0$, $h_s \colon (0, \frac{1}{2}) \to \mathbb{R}^+$ is the function defined by

$$h_s(x) := 2^{(\log \frac{1}{x})^{\frac{s}{s+1}}}$$

for all $x \in (0, \frac{1}{2})$. In this section we use the results of the previous two sections, together with 2.1, to prove 2.2, which we restate:

4.1 *Every path P is h_s-dividing for all integers $s \geq 0$.*

Proof. The proof is by induction on s. The key part of the argument is the inductive step, where we bootstrap from h_{s-1}-dividing to h_s-dividing. Our goal is to find a large blockade that is suitably dense or sparse. As P is h_{s-1}-dividing, we can use the transference lemma 2.1 to obtain a large

10

subset S that induces subgraph that is either dense or sparse. If the induced subgraph is sparse, then we can use 3.1 to obtain the required blockade; otherwise, S induces a large dense subgraph. In fact, we can apply this argument to any large induced subgraph: thus either we win, or we find that every large induced subgraph of G contains a large, dense part. This will allow us to use 3.2 to obtain a dense blockade.

So let G be a P-free graph. For $s = 0$, we need to find disjoint sets B_1 and B_2 of size at least $\lfloor x^d |G| \rfloor$ such that B_2 is x-sparse or $(1-x)$-dense to B_1. This follows from a strengthened version of a result of Erdős and Hajnal [7] given by Fox and Sudakov [8] (see their Lemma 2.1). So, we may assume that $s \geq 1$, and P is h_{s-1}-dividing.

By 2.1, with $h = h_{s-1}$, we deduce that there exists $C > 0$ such that for every $\varepsilon \in (0, \frac{1}{2})$, every P-free graph G' contains an ε-restricted $S \subseteq V(G')$ with $|S| \geq \delta |G|'$, where

$$\delta = \varepsilon^{\frac{C \log \frac{1}{\varepsilon}}{\log(h_{s-1}(\varepsilon))}} = \varepsilon^{C(\log \frac{1}{\varepsilon})^{1/s}}.$$

Let $b = 3^{1+1/s} C$. We deduce:

(1) *Let $0 < x \leq 1/2$ and let $y := 1/h_s(x)$. Then every P-free graph G' contains a y^3-restricted subset $S \subseteq V(G')$ with $|S| \geq x^b |G'|$.*

Since $x \leq 1/2$, it follows that $h_s(x) \geq 2$ and so $y^3 \leq y \leq 1/2$. Moreover,

$$\log \frac{1}{y} = \log(h_s(x)) = \left(\log \frac{1}{x}\right)^{s/(s+1)}.$$

Setting $\varepsilon = y^3$, we deduce that every P-free graph G' contains a y^3-restricted $S \subseteq V(G')$ with $|S| \geq \delta |G'|$, where

$$\delta = \varepsilon^{C(\log \frac{1}{\varepsilon})^{1/s}} = 2^{-C(\log \frac{1}{\varepsilon})^{1+1/s}} = 2^{-3^{1+1/s} C(\log \frac{1}{y})^{1+1/s}} = 2^{-b(\log \frac{1}{y})^{1+1/s}} = x^b.$$

This proves (1).

Now, let $d = 3b|P| + b + 5$, and choose $c > 0$ with $c \leq 1/2$, and sufficiently small that

$$c^b \leq \frac{1}{h_s(c)} \leq \min\left(\frac{1}{60|P|}, \frac{1}{100}\right).$$

Let $x \in (0, c)$ and let G be a P-free graph. We will show that there is an x-sparse or x-dense $(h_s(x), \lfloor x^d |G| \rfloor)$-blockade in G, and therefore that P is h_s-dividing. Suppose (for a contradiction) that there is no such blockade. As before, let $y := 1/h_s(x)$. We note that $y \leq x \leq c \leq 1/100$.

(2) *For every $S \subseteq V(G)$ with $|S| \geq x^{d-b-5}|G|$, there exists a $(1 - y^3)$-dense subset $S' \subseteq S$ with $|S'| \geq x^b|S|$.*

Suppose not. By (1) applied to $G[S]$, there is an y^3-sparse subset $S' \subseteq S$ with $|S'| \geq x^b|S|$. By 3.1 applied to $G[S']$, there is an x-sparse $(1/y, \lfloor x^5|S'| \rfloor)$-blockade in $G[S']$. But then G admits

11

an anticomplete $(1/y, \lfloor x^{b+2}y^2|G|\rfloor)$-blockade: this is an x-sparse $(1/y, \lfloor x^d|G|\rfloor)$-blockade, giving a contradiction. This proves (2).

In particular, (1) implies that for every $S \subseteq V(G)$ with $|S| \geq x^{3b|P|}|G|$, there is a $(1 - y^3)$-dense subset $S' \subseteq S$ with $|S'| \geq x^b|S|$, since $x^{3b|P|} = x^{d-b-5}$. But now, by 3.2 with x replaced by x^b, we deduce that there is a $(1 - x^b)$-dense, and hence $(1 - x)$-dense, $(1/y, \lfloor x^{3b|P|}|G|\rfloor)$-blockade in G. As $d \geq 3b|P|$, this proves 4.1. ∎

5 Concluding remarks

We have proved that all paths are near-polynomial Rödl; can we do the same for any other graphs? A *caterpillar* is a tree in which there is a path that contains at least one end of every edge of the tree; and the proof of this paper for paths can be adapted without difficulty to show that all caterpillars are near-polynomial Rödl. This leads naturally to asking whether all trees are near-polynomial Rödl, but we have not been able to prove that; we cannot handle the tree obtained from the 3-claw $K_{1,3}$ by subdividing each edge once (this is the unique minimal tree that is not a caterpillar). But there are four further graphs (two complementary pairs) that we can prove are near-polynomial Rödl, all with six vertices and shown in the figure 3. The method of proof is similar. We omit all these proofs, which will appear in Tung Nguyen's thesis [11].
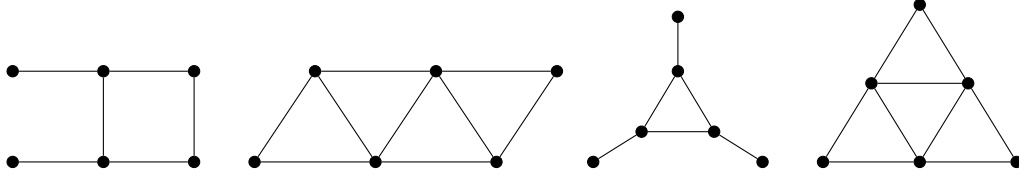

Figure 3: Four near-polynomial Rödl six-vertex graphs.

Another way to strengthen the current result is by looking at ordered graphs. An *ordered graph* G is a pair $(G^\sharp, \leq_G)$, where $G^\sharp$ is a graph and $\leq_G$ is a linear order of its vertex set. Induced subgraph containment for ordered graphs is defined in the natural way, respecting the orders of both graphs. A *zigzag path* is an ordered graph $(G^\sharp, \leq_G)$ where $G^\sharp$ is a path and the ordering is as in figure 4. The proof of this paper works for ordered graphs, with minor adjustments, showing that every zigzag path is near-polynomial Rödl (defining "near-polynomial Rödl" for ordered graphs in the natural way). We omit the details.
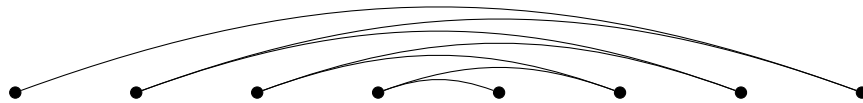

Figure 4: A zigzag path

Note. For the purpose of Open Access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission.

References

[1] P. Blanco and M. Bucić, "Towards the Erdős-Hajnal conjecture for P_5-free graphs", *Res. Math. Sci.*, **11** (2024), `arXiv:2210.10755`.

[2] M. Bucić, J. Fox, and Huy Tuan Pham, "Equivalence between Erdős-Hajnal and polynomial Rödl and Nikiforov conjectures", `arXiv:2403.08303`

[3] M. Bucić, T. Nguyen, A. Scott, and P. Seymour, "Induced subgraph density. I. A loglog step towards Erdős-Hajnal", *International Mathematics Research Notices* **2024**, (2024), 9991-10004.

[4] M. Chudnovsky, A. Scott, P. Seymour, and S. Spirkl, "Pure pairs. I. Trees and linear anticomplete pairs", *Advances in Mathematics* **375** (2020), 107396.

[5] M. Chudnovsky, A. Scott, P. Seymour, and S. Spirkl, "Erdős-Hajnal for graphs with no 5-hole", *Proceedings of the London Mathematical Society* **126** (2023), 997–1014

[6] P. Erdős and A. Hajnal, "On spanned subgraphs of graphs", *Contributions to Graph Theory and its Applications* (Internat. Colloq., Oberhof, 1977), 80–96, Tech. Hochschule Ilmenau, Ilmenau, 1977, https://old.renyi.hu/~p_erdos/1977-19.pdf

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

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

[9] A. Gyárfás, "On Ramsey covering-numbers", *Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday)*, Vol. II, pp. 801-816, *Colloq. Math. Soc. Janos Bolyai* **10**, North-Holland, Amsterdam, 1975.

[10] A. Gyárfás, "Problems from the world surrounding perfect graphs", *Proceedings of the International Conference on Combinatorial Analysis and its Applications* (Pokrzywna, 1985), *Zastos. Mat.* **19** (1987), 413-441.

[11] T. Nguyen, *Induced Subgraph Density*, PhD thesis, Princeton University, May 2025, in preparation.

[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. Bounded VC-dimension", submitted for publication, `arXiv:2312.15572`.

[14] T. Nguyen, A. Scott and P. Seymour, "Induced subgraph density. VII. The five-vertex path", in preparation, `arXiv:2312.15333`.

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

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

[17] István Tomon, "String graphs have the Erdős–Hajnal property", *J. Eur. Math. Soc.* **26** (2024), 275—287.