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

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 July 27, 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**

The Erdős-Hajnal conjecture says that for every graph $H$, there exists $c > 0$ such that every $H$-free graph $G$ has a clique or stable set of size at least $2^{c \log |G|}$ (a graph is "$H$-free" if no induced subgraph is isomorphic to $H$). The conjecture is known when $H$ is a path with at most four vertices, but remains open for longer paths. For the five-vertex path, Blanco and Bucić recently proved a bound of $2^{c(\log |G|)^{2/3}}$; for longer paths, the best existing bound is $2^{c(\log |G| \log\log |G|)^{1/2}}$.

We prove a much stronger result: for any path $P$, every $P$-free graph $G$ has a clique or stable set of size at least $2^{(\log |G|)^{1-o(1)}}$. We strengthen this further, weakening the hypothesis that $G$ is $P$-free by a hypothesis that $G$ does not contain "many" copies of $P$, and strengthening the conclusion, replacing the large clique or stable set outcome with a "near-polynomial" version of Nikiforov's theorem.

# 1 Introduction

Some terminology and notation: if $G$ is a graph, $G[X]$ 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 $H$. A well-known conjecture of Erdős and Hajnal from 1977 [3, 4] 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 [7] for a variety of new graphs $H$ that satisfy 1.1). In particular, it remains open for $H = P_5$, the five-vertex path.

How large a clique or stable set must a $P_5$-free $G$ graph have, in terms of $|G|$? 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 $n > 0$.*

Erdős and Hajnal [4] proved:

**1.3** *For every graph $H$, there exists $c > 0$ such that $f_H(n) \geq 2^{c(\log n)^{1/2}}$ for all $n > 0$.*

In [2], with Bucić, we improved this: we showed:

**1.4** *For every graph $H$, there exists $c > 0$ such that $f_H(n) \geq 2^{c(\log n \log\log n)^{1/2}}$ for all $n > 0$.*

But when $H = P_5$, more can be said. In a substantial breakthrough, in [1], Blanco and Bucić improved 1/2 to 2/3; they proved:

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

In the present paper, we prove a much stronger result: $P_5$ can be replaced by any path, and 2/3 can be replaced by any $d < 1$ ($d = 1$ is the Erdős-Hajnal conjecture itself). More exactly:

**1.6** *For every path $P$, and all $d < 1$, there exists $c > 0$ such that $f_P(n) \geq 2^{c(\log n)^d}$ for all $n > 0$.*

This is equivalent to saying that for every path $P$, $f_P(n) \geq 2^{(\log n)^{1-o(1)}}$.

This will be further strengthened, in two ways both of which need more definitions. If $\varepsilon > 0$, a subset $S \subseteq V(G)$ is

- *$\varepsilon$-sparse* if $G[S]$ has maximum degree at most $\varepsilon|S|$;
- $(1-\varepsilon)$*-dense* if $\overline{G}[S]$ is $\varepsilon$-sparse, where $\overline{G}$ is the complement graph of $G$; and
- *$\varepsilon$-restricted* if $S$ is either $\varepsilon$-sparse or $(1-\varepsilon)$-dense.

A (mildly strengthened) result of Rödl [10] says:

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

Fox and Sudakov [5] proposed the conjecture that the dependence of $\delta$ on $\varepsilon$ is polynomial; or more exactly:

**1.8 Conjecture:** *For every graph $H$ there exists $c > 0$ such that for every $\varepsilon$ 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 $\varepsilon$-restricted.*

Every graph $H$ satisfying this also satisfies the Erdős-Hajnal conjecture; and in the converse direction, we proved in [6, 7] that all the graphs currently known to satisfy the Erdős-Hajnal conjecture also satisfy conjecture 1.8.

As we said, 1.6 will be strengthened in two ways. The first strengthening is, we will prove that every path "nearly" satisfies the Fox-Sudakov conjecture. More exactly:

**1.9** *For every path $P$ and all $d < 1$, there exists $c > 0$ such that for all $0 < \varepsilon \leq 1/2$ and every $P$-free graph $G$, there is an $\varepsilon$-restricted subset $S \subseteq V(G)$ with $|S| \geq 2^{-c(\log \varepsilon^{-1})^{1/d}}|G|$.*

This first strengthening is crucial to the proof of 1.6.

The second strengthening is, we can replace the hypothesis of 1.9 that $G$ is $P_5$-free, with a weaker hypothesis that $G$ does not contain many copies of $H$. 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 [9], strengthening Rödl's theorem:

**1.10** *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 is an $\varepsilon$-restricted subset $S \subseteq V(G)$ with $|S| \geq \delta|G|$.*

We will prove that, when $H$ is a path, this is satisfied taking $\delta$ to be a "near-polynomial" function of $\varepsilon$. More exactly:

**1.11** *For every path $P$ and all $d < 1$, there exists $c > 0$ such that for all $0 < \varepsilon \leq 1/2$, if $\delta$ satisfies*

$$\delta = 2^{-c(\log \varepsilon^{-1})^{1/d}},$$

*then every graph $G$ with $\operatorname{ind}_P(G) \leq (\delta|G|)^{|P|}$, there is an $\varepsilon$-restricted subset $S \subseteq V(G)$ with $|S| \geq \delta|G|$.*

This second strengthening is not crucial for the proof, but it has the advantage that the class of graphs $P$ (not necessarily paths) that satisfy 1.11 is closed under vertex-substitution, while we cannot prove the same for the the class of graphs $P$ that satisfy 1.9.

Thus, 1.11 is our main result. Obviously it implies 1.9, but that it implies 1.6 is not so obvious. Let us see that, via the following.

**1.12** *Let $G, H$ be graphs with $|G| \geq 2$, and assume that $a > 0$ is such that for every $\varepsilon > 0$ with $\varepsilon \leq 1/2$, there is an $\varepsilon$-restricted $S \subseteq V(G)$ with*

$$|S| \geq 2^{-a(\log \frac{1}{\varepsilon})^{1/d}}|G|.$$

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

**Proof.** If $2^{c(\log|G|)^d} \le 2$, then the result holds since $G$ has a clique or stable set of size two; so we assume that $2^{c(\log|G|)^d} > 2$. Let $\varepsilon := 2^{-2c(\log|G|)^d}$; then $\varepsilon \in (0, \frac{1}{4})$. Let $\delta > 0$ be such that

$$\log\frac{1}{\delta} = a\left(\log\frac{1}{\varepsilon}\right)^{1/d} = a(2c)^{1/d}\log|G| \le 2ac\log|G|.$$

Then $\delta \ge |G|^{-2ac}$. From the hypothesis, there is an $\varepsilon$-restricted $S \subseteq V(G)$ with

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

Thus, since $S$ is $\varepsilon$-restricted, $G[S]$ (and so $G$) contains a clique or stable set of size at least

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

This proves 1.12. ∎

It is easier to prove that graphs are near-viral than to prove they are viral, and we can prove that several other types of graph are near-viral. We will return to this in a subsequent paper [8].

We can strengthen the current result by looking at ordered graphs. An *ordered graph* $G$ is a pair $(G^\natural, \le_G)$, where $G^\natural$ is a graph and $\le_G$ 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 $(G^\natural, \le_G)$ where $G^\natural$ is a path and the ordering is as in figure 1. The proof of this paper works for ordered graphs, with minor adjustments; and it shows that every zigzag path is near-viral (defining "near-viral" for ordered graphs in the natural way). We omit the details.

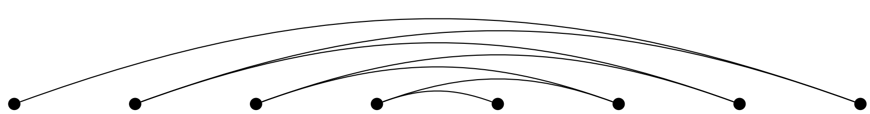

Figure 1: A zigzag path

## 2 Blockades

As in previous papers of this series, we say a graph $H$ is *viral* if there exists $c > 0$ such that for all $0 < \varepsilon \le 1/2$ and every graph $G$ with $\mathrm{ind}_H(G) \le (\varepsilon^c|G|)^{|H|}$, there is an $\varepsilon$-restricted subset $S \subseteq V(G)$ with $|S| \ge \varepsilon^c|G|$. Let us say that a graph $H$ is *near-viral* if for every $d < 1$, there exists $c > 0$ such that for every $\varepsilon \in (0, \frac{1}{2})$, if $\delta$ satisfies

$$\log\frac{1}{\delta} = c\left(\log\frac{1}{\varepsilon}\right)^{1/d},$$

then for every graph $G$ with $\mathrm{ind}_H(G) \le (\delta|G|)^{|H|}$, there is an $\varepsilon$-restricted $S \subseteq V(G)$ with $|S| \ge \delta|G|$. Thus, our main theorem 1.11 says that every path is near-viral.

A *blockade* in a graph $G$ is a finite sequence $(B_1, \dots, B_n)$ of (possibly empty) disjoint subsets of $V(G)$; its *length* is $n$ and its *width* is $\min_{i\in[n]}|B_i|$. For $k, w \geq 0$, $(B_1, \dots, B_n)$ is a $(k,w)$-*blockade* if its length is at least $k$ and its width is at least $w$. For $x \in (0, \frac{1}{2})$, this blockade 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 $i < j$.

Thus, 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 $\varepsilon$-restricted sets. But the reason blockades are useful is that if a graph $G$ and all its large induced subgraphs admit blockades with certain parameters, then $G$ must contain a large $\varepsilon$-restricted set.

There are now several theorems of this type, with a family resemblance, but sufficiently different to be confusing, and perhaps it would be helpful to summarize them here.

- Erdős and Hajnal [4] proved that 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 $(2, \lfloor x^d|G|\rfloor)$ blockade. From this they deduced their result 1.3.

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

- If we could prove that for a graph $H$, there exists $d > 0$ such that for all $x \in (0, 1/2]$, every graph $G$ with $\operatorname{ind}_H(G) \leq (x^d|G|)^{|H|}$ admits an $x$-sparse or $(1-x)$-dense $(1/x, \lfloor x^d|G|\rfloor)$-blockade, then we could deduce that $H$ is viral. It would be just as good if we could prove that for all $x \in (0, 1/2]$, every graph $G$ with $\operatorname{ind}_H(G) \leq (x^d|G|)^{|H|}$ admits an $x$-sparse or $(1-x)$-dense $(k, \lfloor |G|/k^d\rfloor)$-blockade for some $k \in [2, 1/x]$. This was used in [6] to prove the main results of that paper.

- Suppose that there exist $a, b > 0$ and $d > 2$ such that for all $0 < x < y \leq 1/2$, every $y^a$-restricted graph $G$ with $\operatorname{ind}_H(G) \leq (x^{bd^2}|G|)^{|H|}$ admits either a $y^{ad}$-restricted subset of size at least $y^{bd^2}|G|$, or an $x$-sparse or $(1-x)$-dense $(1/y, \lfloor y^{bd^2}|G|\rfloor)$-blockade. Then $H$ is viral. This was the method used in [7].

- Suppose we could prove that there exists $d > 0$ such that for all $x, y$ with $0 < x < y \leq 1/2$, every $\operatorname{poly}(y)$-restricted graph $G$ with $\operatorname{ind}_H(G) \leq (x^d|G|)^{|H|}$ admits an $x$-sparse or $(1-x)$-dense $(1/y, \lfloor x^d|G|\rfloor)$-blockade. Then we could deduce that $H$ is near-viral. This is the approach in this paper.

The *edge-density* of a graph $G$ is $|E(G)|$ divided by $\binom{|G|}{2}$ (or 1 if $|G| \leq 1$). Let us say a subset $S \subseteq V(G)$ is *weakly $\varepsilon$-restricted* if one of $G[S], \overline{G}[S]$ has edge-density at most $\varepsilon$. A function $\ell\colon (0, \frac{1}{2}) \to \mathbb{R}^+$ is *subreciprocal* if it is nonincreasing and $1 < \ell(x) \leq 1/x$ for all $x \in (0, \frac{1}{2})$. For a subreciprocal function $\ell$, a graph $H$ is *$\ell$-divisive* if there are $c \in (0, \frac{1}{2})$ and $d > 1$ such that for every $x \in (0, c)$ and every graph $G$ with $\operatorname{ind}_H(G) \leq (x^d|G|)^{|H|}$, there is an $x$-sparse or $x$-dense $(\ell(x), \lfloor x^d|G|\rfloor)$-blockade in $G$. Here is a theorem proved in [2].

**2.1** *Let $\ell\colon (0,\frac{1}{2}) \to \mathbb{R}^+$ be subreciprocal, and let $H$ be $\ell$-divisive. Then there exists $C > 0$ such that for every $\varepsilon \in (0,\frac{1}{2})$, if we define $\delta > 0$ by*

$$\log\frac{1}{\delta} = \frac{C(\log\frac{1}{\varepsilon})^2}{\log(\ell(\varepsilon))},$$

*then for every graph $G$ with $\operatorname{ind}_H(G) \le (\delta|G|)^{|H|}$, there is a weakly $\varepsilon$-restricted $S \subseteq V(G)$ with $|S| \ge \delta|G|$.*

This provides us with large subsets that are weakly $\varepsilon$-restricted, but we want $\varepsilon$-restricted subsets. These are easy to find:

**2.2** *For $\varepsilon \in (0,\frac{1}{2})$ and a graph $G$, if $S \subseteq V(G)$ is weakly $\frac{1}{4}\varepsilon$-restricted, then there exists an $\varepsilon$-restricted $T \subseteq S$ with $|T| \ge \frac{1}{2}|S|$.*

**Proof.** We may assume that $G[S]$ has at most $\frac{1}{4}\varepsilon\binom{|S|}{2} < \frac{1}{8}\varepsilon|S|^2$ edges. Let $T$ be the set of vertices in $S$ with degree at most $\frac{1}{2}\varepsilon|S|$ in $G[S]$; then $\frac{1}{2}\varepsilon|S||S \setminus T| < \frac{1}{4}\varepsilon|S|^2$ and so $|S \setminus T| < \frac{1}{2}|S|$. Thus $|T| > \frac{1}{2}|S|$ and $G[T]$ has maximum degree at most $\frac{1}{2}\varepsilon|S| < \varepsilon|T|$. This proves 2.2. ∎

We need to make a corresponding adjustment to 2.1:

**2.3** *Let $\ell$ be subreciprocal, and let $H$ be an $\ell$-divisive graph. Then there exists $C > 0$ such that for every $\varepsilon \in (0,\frac{1}{2})$, if we define $\delta > 0$ by*

$$\log\frac{1}{\delta} = \frac{C(\log\frac{1}{\varepsilon})^2}{\log(\ell(\varepsilon))},$$

*then for every graph $G$ with $\operatorname{ind}_H(G) \le (\delta|G|)^{|H|}$, there is an $\varepsilon$-restricted $S \subseteq V(G)$ with $|S| \ge \delta|G|$.*

**Proof.** Choose $C'$ such that 2.1 holds with $C$ replaced by $C'$. We claim that $C := 18C'$ satisfies the theorem. To show this, let $\varepsilon \in (0,\frac{1}{2})$, let $\delta$ be as in 2.3, and let $G$ be a graph with $\operatorname{ind}_H(G) \le (\delta|G|)^{|H|}$. We must show that there is an $\varepsilon$-restricted $S \subseteq V(G)$ with $|S| \ge \delta|G|$.

Let $\varepsilon' := \frac{1}{4}\varepsilon \in (\varepsilon^3,\varepsilon)$, and define $\delta'$ by

$$\log\frac{1}{\delta'} := \frac{C'(\log\frac{1}{\varepsilon'})^2}{\log(\ell(\varepsilon'))} < \frac{C'(\log\frac{1}{\varepsilon^3})^2}{\log(\ell(\varepsilon'))} \le \frac{9C'(\log\frac{1}{\varepsilon})^2}{\log(\ell(\varepsilon))} = \frac{1}{2}\log\frac{1}{\delta},$$

and so $\delta' \ge \sqrt{\delta} > 2\delta$. Since $\operatorname{ind}_H(G) \le (\delta|G|)^{|H|} \le (\delta'|G|)^{|H|}$, there is a weakly $\varepsilon'$-restricted $S \subseteq V(G)$ in $G$ with $|S| \ge \delta'|G|$. By 2.2, there is an $\varepsilon$-restricted $T \subseteq S$ with $|T| \ge \frac{1}{2}|S| \ge \frac{1}{2}\delta'|G| > \delta|G|$. This proves 2.3. ∎

For each integer $s \ge 0$, let $\ell_s\colon (0,\frac{1}{2}) \to \mathbb{R}^+$ be the function defined by

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

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

**2.4** *Every path $P$ is $\ell_s$-divisive for all integers $s \ge 0$.*

Let us deduce 1.11, which we restate:

**2.5** *Every path $P$ is near-viral.*

**Proof (assuming 2.4).** We must show that for all $d < 1$, there exists $c > 0$ such that for all $\varepsilon \in (0, 1/2)$, if $\delta$ satisfies

$$\log \frac{1}{\delta} = c\left(\log \frac{1}{\varepsilon}\right)^{1/d},$$

then for every graph $G$ with $\text{ind}_H(G) \le (\delta|G|)^{|H|}$, there is an $\varepsilon$-restricted $S \subseteq V(G)$ with $|S| \ge \delta|G|$.

Choose $s$ with $\frac{s+1}{s+2} \ge d$. Since $P$ is $\ell_s$-divisive, by 2.3 there exists $C > 0$ such that for every $\varepsilon \in (0, \frac{1}{2})$, if we define $\delta' > 0$ by

$$\log \frac{1}{\delta'} = \frac{C\left(\log \frac{1}{\varepsilon}\right)^2}{\log\left(\ell_s(\varepsilon)\right)} = \frac{C\left(\log \frac{1}{\varepsilon}\right)^2}{\left(\log\left(\frac{1}{\varepsilon}\right)\right)^{s/(s+1)}} = C\left(\log \frac{1}{\varepsilon}\right)^{\frac{s+2}{s+1}},$$

then for every graph $G$ with $\text{ind}_H(G) \le (\delta'|G|)^{|H|}$, there is an $\varepsilon$-restricted $S \subseteq V(G)$ with $|S| \ge \delta'|G|$. We claim that we make take $c = C$. To see this, let $\varepsilon \in (0, 1/2)$, let $\delta$ satisfy

$$\log \frac{1}{\delta} = C\left(\log \frac{1}{\varepsilon}\right)^{1/d},$$

and let $G$ be a graph with $\text{ind}_H(G) \le (\delta|G|)^{|H|}$. Then

$$\log \frac{1}{\delta'} = C\left(\log \frac{1}{\varepsilon}\right)^{\frac{s+2}{s+1}} \le C\left(\log \frac{1}{\varepsilon}\right)^{1/d} = \log \frac{1}{\delta}$$

and so $\delta' \ge \delta$. Since $\text{ind}_P(G) \le (\delta'|G|)^k$, there is an $\varepsilon$-restricted $S \subseteq V(G)$ with $|S| \ge \delta'|G| \ge \delta|G|$. This proves 2.5. ▮

# 3 In a sparse graph

The remainder of the paper is devoted to the proof of 2.4. Its proof proceeds by induction on $s$; so we may assume that the path $P$ is $\ell_{s-1}$-divisive, and therefore $V(G)$, and every large subset of $V(G)$, includes a somewhat smaller subset that is appropriately restricted. This subset might be very dense or very sparse, but if ever the subset is very sparse, we can win easily, using the result of this section. As usual with problems about excluding a path, our task is easier if the "host" graph is sparse, and we use a modified version of the well-known "Gyárfás path argument".

**3.1** *Let $P$ be a path, let $0 < x \le y \le 1/(2|P|)$, and let $G$ be a $y^2$-sparse graph. Then either:*

- $\text{ind}_P(G) \ge (x^4|G|)^{|P|}$*, or*
- *there is an $x$-sparse $(1/y, \lfloor x^5|G|\rfloor)$-blockade in $G$.*

**Proof.** Let $|P| = k \geq 1$. If $k = 1$ the first bullet holds, and if $x^5|G| < 1$ then the second bullet holds, so we assume that $k \geq 2$ and $|G| \geq x^{-5}$. Choose an $x$-sparse blockade $(B_1, \ldots, B_{n-1}, C)$ in $G$ with $n$ maximum such that $|B_1|, \ldots, |B_{n-1}|, |C| \geq x^5|G|$ and $|C| \geq (1 - k(n-1)y^2)|G|$. We may assume that $n < 1/y$, and so

$$|C| \geq (1 - k(n-1)y^2)|G| = (1 + ky^2 - kny^2)|G| \geq (1/2 + ky^2)|G|.$$

We claim:

(1) *For every $X \subseteq C$ with $|X| \geq x^4|G|$, and $Y \subseteq C \setminus X$ with $|Y| \geq (1 + 4x^3 - kny^2)|G|$, some vertex in $X$ has at least $2x^4|G|$ neighbours in $Y$.*

Suppose not. Then $|Y| \geq (1 + 4x^3 - kny^2)|G| \geq |G|/2$, since $kny^2 \leq 1/2$. There are most $2x^4|G| \cdot |X| \leq 4x^4|X| \cdot |Y|$ edges between $X$ and $Y$, and so at most $4x^3|G|$ vertices in $Y$ have at least $x|X|$ neighbours in $X$. Thus there is a subset $Y'$ of $Y$ with cardinality at least

$$|Y| - 4x^3|G| \geq (1 - kny^2)|G| \geq |G|/2 \geq x^5|G|$$

that is $x$-sparse to $X$. But then $(B_1, \ldots, B_{n-1}, X, Y')$ contradicts the maximality of $n$. This proves (1).

For $t \geq 1$ an integer, let us say a *$t$-brush* is an induced path $v_1\text{-}\cdots\text{-}v_t$ of $G[C]$, such that $v_t$ has at least $2x^4|G|$ neighbours in $C$ that are different from and nonadjacent to each of $v_1, \ldots, v_{t-1}$.

(2) *For $1 \leq t \leq k-2$, if $v_1\text{-}\cdots\text{-}v_t$ is a $t$-brush of $G[C]$, there are at least $x^4|G|$ vertices $v$ such that $v_1\text{-}\cdots\text{-}v_t\text{-}v$ is a $(t+1)$-brush.*

Let $X$ be the set of neighbours of $v_t$ in $C$ that are different from and nonadjacent to each of $v_1, \ldots, v_{t-1}$; and let $Y$ be the set of all vertices in $C$ that are different from and nonadjacent to each of $v_1, \ldots, v_t$. Thus $|X| \geq 2x^4|G|$ since $v_1\text{-}\cdots\text{-}v_t$ is a $t$-brush. Moreover, $k \geq 3$ since $1 \leq t \leq k-2$, and so, as $G$ is $y^2$-sparse,

$$|Y| \geq |C| - (k-2)y^2|G| \geq (1 - k(n-1)y^2 - (k-2)y^2)|G| = (1 + 2y^2 - kny^2)|G| \geq (1 + 4x^3 - kny^2)|G|$$

(because $4x^3 \leq 2y^2$). By (1), fewer than $x^4|G|$ vertices in $X$ have fewer than $2x^4|G|$ neighbours in $Y$. All the others give $(t+1)$-brushes extending $v_1\text{-}\cdots\text{-}v_t$; and since $|X| - x^4|G| \geq x^4|G|$, this proves (2).

(3) *There are at least $|G|/2$ 1-brushes.*

Suppose there is a set $X$ of $\lceil x^4|G| \rceil$ vertices in $C$ each with degree less than $2x^4|G|$ in $G[C]$. Let $Y = C \setminus X$; then

$$|Y| \geq |C| - x^4|G| - 1 \geq |C| - x^4|G| - x^5|G| \geq (1 - k(n-1)y^2 - x^4 - x^5)|G| \geq (1 + 4x^3 - kny^2)|G|$$

(since $ky^2 \geq 2x^2 \geq 4x^3 + x^4 + x^5$), contrary to (1). Thus there are fewer than $x^4|G|$ vertices that have degree less than $2x^4|G|$ in $G[C]$. All the others give 1-brushes, and since $|C| - x^4|G| \geq |G|/2$, this proves (3).

From (2) and (3), it follows inductively that for $1 \le t \le k-1$ there are at least $x^{4(t-1)}|G|^t/2$ $t$-brushes, and in particular, there are at least $x^{4(k-1)}|G|^{k-1}/2$ $(k-1)$-brushes. Each extends to at least $2x^4|G|$ induced $k$-vertex paths; and so $\mathrm{ind}_P(G) \ge (2x^4|G|)x^{4(k-1)}|G|^{k-1}/2 = x^{4k}|G|^k$. This proves 3.1. ∎

# 4 The dense case

As we discussed at the start of the previous section, for the inductive proof of 2.4, we will now be able to assume that every large subset of $V(G)$ includes a somewhat smaller subset that is very dense. That motivates the following:

**4.1** *Let $P$ be a path with $|P| \ge 1$, and let $0 < x \le y \le 1/100$. Let $G$ be a graph such that for every $S \subseteq V(G)$ with $|S| \ge x^{3|P|}|G|$, there is a $(1-y^3)$-dense subset $S' \subseteq S$ with $|S'| \ge x|S|$. Then either:*

- $\mathrm{ind}_P(G) \ge (x^{3|P|}|G|)^{|P|}$*; or*
- *there is a $(1-x)$-dense $(1/y, \lfloor x^{3|P|}|G|\rfloor)$-blockade in $G$.*

**Proof.** In the proof of 3.1, we counted "$t$-brushes", induced $t$-vertex paths in which the last vertex had many neighbours that all had no neighbours in the earlier part of the path. The issue there was to prove that, given a $t$-brush, there were many ways to extend it to a $(t+1)$-brush. We will do something similar here, but we need to redefine a $t$-brush. We will be working inside a graph that is very dense, so 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. A *non-neighbour* of $v$ means a vertex different from and nonadjacent to $v$, and the *antidegree* of $v$ is the number of its non-neighbours.

Let $k := |P|$ and $a := 3k$. We may assume that $x^a|G| \ge 1$, since otherwise the second bullet holds. Define $a_1 := x/2$, and $b_1 := x^2y/8$; and for $2 \le t \le k$, define $a_t := (x/2)b_{t-1}$ and $b_t := (x^2/2)b_{t-1}$. For $1 \le t \le k$ 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 $A, B \subseteq V(G)$ 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 $\{v_1, \ldots, v_t\}$;
- $|A| \ge a_t|G|$ and $|B| \ge b_t|G|$;
- for every $Y \subseteq B$ with $|Y| \ge x^a|G|$, there are at least $y|A|/4$ vertices in $A$ that have at least $x|Y|$ non-neighbours in $Y$; and
- every vertex in $B$ has at most $3y^3|A|$ non-neighbours in $A$.

We claim first:

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

Since $|S| \geq x^a|G|$, there exists $S' \subseteq S$ with $|S'| \geq x|S|$ such that $S'$ is $(1-y^3)$-dense. Choose a $(1-x)$-dense blockade $(B_1, \dots, B_{n-1}, C)$ in $G[S']$ with $n$ maximum such that $|B_1|, \dots, |B_{n-1}|, |C| \geq x^a|G|$ 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.

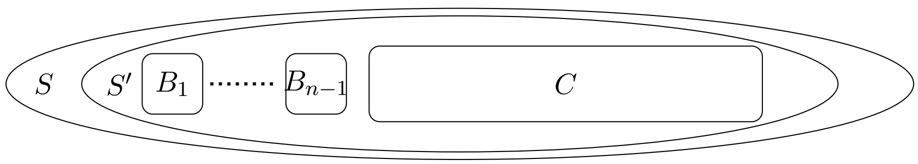


Figure 2: For step (1) of the proof of 4.1.

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).

(2) *There are at least $x|G|/2$ 1-brushes.*

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 that there is a set $Y$ of $\lceil x^a|G| \rceil$ vertices in $C$ each with antidegree less than $(x^2y/8)|G|$ in $G[C]$. Let $X = C \setminus Y$. Then

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

There are at most $(x^2y/8)|G| \cdot |Y|$ nonedges between $X, Y$; and yet from the choice of $C$, since $|X| \geq (1-y/4)|C|$, there are at least $(y|X|/4)(x|Y|) = (xy/4)|X| \cdot |Y|$ such nonedges. So

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

and so $2|X| \leq x|G|$. But

$$|X| \geq |C| - x^a|G| - 1 \geq |C| - 2x^a|G| \geq x(1+y)|G|/2 - 2x^a|G| > x|G|/2,$$

a contradiction.

Thus there are fewer than $x^a|G|$ vertices that have antidegree less than $(x^2y/8)|G|$ in $G[C]$; and so there are at least $|C| - x^a|G| \geq x|G|/2$ vertices in $C$ with antidegree at least $(x^2y/8)|G|$ in $G[C]$.

We claim that each such vertex forms a 1-brush. Let $v$ be such a vertex, and let $A, B$ be its sets of neighbours and non-neighbours in $G[C]$. Then $b_1|G| = (x^2y/8)|G| \le |B| \le 2y^3|C|$, and so (since $1 \le x^a|G| \le 2x^{a+1}|C| \le y^3|C|$)

$$|A| \ge |C| - 2y^3|C| - 1 \ge (1-3y^3)|C| \ge (1-3y^3)x(1+y)|G|/2 \ge x|G|/2 = a_1|G|.$$

Moreover, since $C$ is $(1-2y^3)$-dense, every vertex in $B$ has at most $2y^3|C| \le 3y^3|A|$ non-neighbours in $A$. Finally, let $Y \subseteq B$ with $|Y| \ge x^a|G|$. Since $|A| \ge (1-3y^3)|C| \ge (1-y/4)|C|$, the choice of $C$ implies that at least $y|A|/4$ vertices in $A$ have at least $x|Y|$ non-neighbours in $Y$. Hence $v$ forms a 1-brush. This proves (2).

(3) *Let $1 \le t \le k-1$, and let $v_1\text{-}\cdots\text{-}v_t$ be a $t$-brush. Then there are at least $ya_t|G|/8$ vertices $v$ 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 $t \le k-1$, and $a \ge 3k$, it follows that

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

By (1), there exists $C \subseteq B$ with $|C| \ge x(1+y)|B|/2$, such that $C$ is $(1-2y^3)$-dense, and for all disjoint $X, Y \subseteq C$ with $|X| \ge (1-y/4)|C|$ and $|Y| \ge x^a|G|$, 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 $|D| \ge y|A|/8$, such that for each $v \in D$, the number of its non-neighbours in $C$ is between $x|C|$ and $24y^2|C|$.

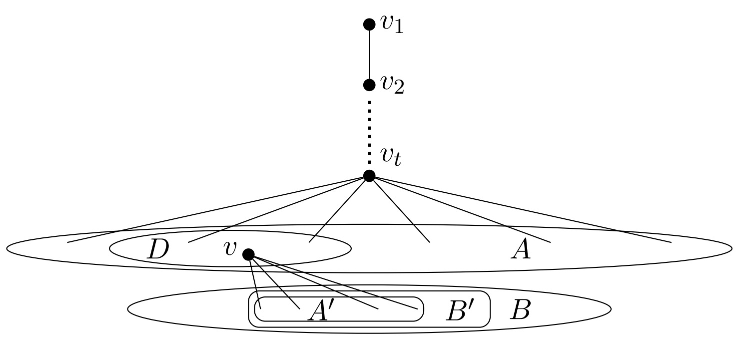


Figure 3: For step (3). $C = A' \cup B'$.

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,

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

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).

From (2) and (3), and some arithmetic which we omit, it follows that there are at least $x^{3k^2}|G|^k$ $k$-brushes, and so $\mathrm{ind}_{P_k}(G) \geq (x^a|G|)^k$. This proves 4.1. ∎

# 5 Decreasing density

We remind the reader that for each integer $s \geq 0$, $\ell_s\colon (0, \frac{1}{2}) \to \mathbb{R}^+$ is the function defined by

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

for all $x \in (0, \frac{1}{2})$. Now we can complete the proof of 2.4, which we restate.

**5.1** *Every path $P$ is $\ell_s$-divisive for all integers $s \geq 0$.*

**Proof.** The proof is by induction on $s$. For $s = 0$, the result is due to Fox and Sudakov [5], extending a theorem of Erdős and Hajnal [4]; indeed, they proved that every graph is $\ell_0$-divisive. So, we assume that $s \geq 1$, and $P$ is $\ell_{s-1}$-divisive. By 2.3, with $\ell = \ell_{s-1}$, we deduce that there exists $C > 0$ such that for every $\varepsilon \in (0, \frac{1}{2})$, if we define $\delta > 0$ by

$$\log \frac{1}{\delta} = \frac{C(\log \frac{1}{\varepsilon})^2}{\log(\ell_{s-1}(\varepsilon))},$$

then for every graph $G$ with $\mathrm{ind}_P(G) \leq (\delta|G|)^{|P|}$, there is an $\varepsilon$-restricted $S \subseteq V(G)$ with $|S| \geq \delta|G|$. But $\log(\ell_{s-1}(\varepsilon)) = (\log \frac{1}{\varepsilon})^{\frac{s-1}{s}}$ and so

$$\log \frac{1}{\delta} = C\left(\log \frac{1}{\varepsilon}\right)^{\frac{s+1}{s}}.$$

We deduce:

(1) *Let $0 < x \leq 1/2$ and let $y := 1/\ell_s(x)$. Then for every graph $G$ with $\mathrm{ind}_P(G) \leq (x^{9C}|G|)^{|P|}$, there is a $y^3$-restricted subset $S \subseteq V(G)$ with $|S| \geq x^{9C}|G|$.*

Since $x \leq 1/2$, it follows that $\ell_s(x) \geq 2$ and so $y^3 \leq y \leq 1/2$. By setting $\varepsilon = y^3$ and

$$\log \frac{1}{\delta} = C\left(\log \frac{1}{y^3}\right)^{\frac{s+1}{s}} = C \log \frac{1}{x}$$

(that is, $\delta = x^{9C}$) we deduce that for every graph $G$ with $\text{ind}_P(G) \le (x^{9C}|G|)^{|P|}$, there is a $y^3$-restricted $S \subseteq V(G)$ with $|S| \ge x^{9C}|G|$. This proves (1).

Now, let $d = 27C|P| + 9C + 4$, and choose $c > 0$ with $c \le 1/2$, and sufficiently small that

$$c^{9C} \le \frac{1}{\ell_s(c)} \le \min\left(\frac{1}{2|P|}, \frac{1}{100}\right).$$

Let $x \in (0, c)$ and let $G$ be a graph with $\text{ind}_P(G) \le (x^d|G|)^{|P|}$. We will show that there is an $x$-sparse or $x$-dense $(\ell_s(x), \lfloor x^d|G|\rfloor)$-blockade in $G$, and therefore that $P$ is $\ell_s$-divisive. Suppose (for a contradiction) that there is no such blockade. Let $y := 1/\ell_s(x)$.

(2) *For every $S \subseteq V(G)$ with $|S| \ge x^{d-9C-4}|G|$, there exists a $(1-y^3)$-dense subset $S' \subseteq S$ with $|S'| \ge x^{9C}|S|$.*

Suppose not. By (1) applied to $G[S]$, either $\text{ind}_P(G[S]) > (x^{9C}|S|)^{|P|}$, or there is an $y^3$-sparse subset $S' \subseteq S$ with $|S|' \ge x^{9C}|S|$. In the first case,

$$\text{ind}_P(G) > (x^{9C}|S|)^{|P|} \ge (x^d|G|)^{|P|}$$

(since $x^{9C}|S| \ge x^d|G|$), a contradiction. In the second case, $|S'| \ge x^{9C}|S|$, and by 3.1 applied to $G[S']$, either

- $\text{ind}_P(G[S']) \ge (x^4|S'|)^{|P|}$, or
- there is an $x$-sparse $(1/y, \lfloor x^5|S'|\rfloor)$-blockade in $G[S']$.

The first is impossible since $x^4|S'| \ge x^{9C+4}|S| \ge x^d|G|$. If the second holds, then $G$ admits an $x$-sparse $(1/y, \lfloor x^{9C+1}|G|\rfloor)$-blockade and hence an $x$-sparse $(1/y, \lfloor x^d|G|\rfloor)$-blockade since $d \ge 9C+1$, again a contradiction. This proves (2).

In particular, (1) implies that for every $S \subseteq V(G)$ with $|S| \ge x^{27C|P|}|G|$, there is a $(1-y^3)$-dense subset $S' \subseteq S$ with $|S'| \ge x^{9C}|S|$, since $x^{27C|P|} = x^{d-9C-4}$. By 4.1, with $x$ replaced by $x^{9C}$ (note that $x^{9C} \le y \le 1/100$ from the choice of $c$), we deduce that either:

- $\text{ind}_P(G) \ge (x^{27C|P|}|G|)^{|P|}$; or
- there is a $(1-x^{9C})$-dense $(1/y, \lfloor x^{27C|P|}|G|\rfloor)$-blockade in $G$.

The first is impossible since $\left(x^{27C|P|}|G|\right)^{|P|} \ge \left(x^d|G|\right)^{|P|}$ (because $d < 27C|P|$). Thus there is a $(1-x^{9C})$-dense, and hence $(1-x)$-dense, $(1/y, \lfloor x^d|G|\rfloor)$-blockade in $G$. This proves 5.1. ∎

## References

[1] P. Blanco and M. Bucić, “Towards the Erdős-Hajnal conjecture for $P_5$-free graphs”, `arXiv:2210.10755`.

[2] 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.

[3] 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

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

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

[6] T. Nguyen, A. Scott and P. Seymour, "Induced subgraph density. III. The pentagon and the bull", in preparation, arXiv:2307.06455.

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

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

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

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