

# Induced subgraph density. II. Sparse and dense sets in cographs

Jacob Fox<sup>1</sup>  
Stanford University,  
Stanford, CA 94305, USA

Tung Nguyen<sup>2</sup>  
Princeton University,  
Princeton, NJ 08544, USA

Alex Scott<sup>3</sup>  
Oxford University,  
Oxford, UK

Paul Seymour<sup>4</sup>  
Princeton University,  
Princeton, NJ 08544, USA

July 4, 2022; revised July 4, 2023

<sup>1</sup>Supported by a Packard Fellowship and by NSF grants DMS-1953990 and DMS-2154129.

<sup>2</sup>Supported by AFOSR grants A9550-19-1-0187 and FA9550-22-1-0234, and by NSF grants DMS-1800053 and DMS-2154169.

<sup>3</sup>Supported by EPSRC grant EP/X013642/1

<sup>4</sup>Supported by AFOSR grants A9550-19-1-0187 and FA9550-22-1-0234, and by NSF grants DMS-1800053 and DMS-2154169.

## Abstract

A well-known theorem of Rödl says that for every graph  $H$ , and every  $\varepsilon > 0$ , there exists  $\delta > 0$  such that if  $G$  does not contain an induced copy of  $H$ , then 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  $\varepsilon$ . But how does  $\delta$  depend on  $\varepsilon$ ? Fox and Sudakov conjectured that the dependence can be taken to be polynomial: for all  $H$  there exists  $c > 0$  such that for all  $\varepsilon$  with  $0 < \varepsilon \leq 1/2$ , Rödl's theorem holds with  $\delta = \varepsilon^c$ . This conjecture implies the Erdős-Hajnal conjecture, and until now it had not been verified for any non-trivial graphs  $H$ . Our first result shows that it is true when  $H = P_4$ . (Indeed, in that case we can take  $\delta = \varepsilon$ , and insist that one of  $G[X], \overline{G}[X]$  has maximum degree at most  $\varepsilon^2|G|$ .)

We will generalize this, and to do so, we need to work with an even stronger property. Let us say  $H$  is *viral* if there exists  $c > 0$  such that for all  $\varepsilon$  with  $0 < \varepsilon \leq 1/2$ , if  $G$  contains at most  $\varepsilon^c|G|^{|H|}$  copies of  $H$  as induced subgraphs, then there exists  $X \subseteq V(G)$  with  $|X| \geq \varepsilon^c|G|$  such that one of  $G[X], \overline{G}[X]$  has density at most  $\varepsilon$ . We will show that  $P_4$  is viral, using a “polynomial  $P_4$ -removal lemma” of Alon and Fox. We will also show that viral graphs are closed under vertex-substitution, and so all graphs that can be obtained by substitution from copies of  $P_4$  are viral. (In a subsequent paper, it will be shown that all graphs currently known to satisfy the Erdős-Hajnal conjecture are in fact viral.)

Finally, we give a different strengthening of Rödl's theorem: we show that if  $G$  does not contain an induced copy of  $P_4$ , then its vertices can be partitioned into at most  $480\varepsilon^{-4}$  subsets  $X$  such that one of  $G[X], \overline{G}[X]$  has maximum degree at most  $\varepsilon|X|$ .

# 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  $G$ ;  $P_4$  denotes the path with four vertices; a graph is  $H$ -free if it has no induced subgraph isomorphic to  $H$ ; and a *cograph* is a  $P_4$ -free graph. The *edge-density* of a graph  $G$  is its number of edges divided by  $\binom{|G|}{2}$ .

A very useful theorem of Rödl [13] 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  $\varepsilon$ .*

How does  $\delta$  depend on  $\varepsilon$ , for a given graph  $H$ ? Sudakov and the first author [9] proposed the conjecture that the dependence is polynomial:

**1.2 Conjecture ([9], conjecture 7.1):** *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  $X \subseteq V(G)$  with  $|X| \geq \varepsilon^c|G|$  such that one of  $G[X], \overline{G}[X]$  has edge-density at most  $\varepsilon$ .*

This conjecture is very strong, and until now had not been verified for any nontrivial graphs  $H$ . It was motivated by the Erdős-Hajnal conjecture [7, 8], which it implies, but which we do not discuss here.

We first prove that 1.2 holds in a particularly nice form when  $H = P_4$ . We will show:

**1.3** *For every  $\varepsilon \in [0, 1]$  and every cograph  $G$ , there exists  $X \subseteq V(G)$  with  $|X| \geq \varepsilon|G|$  such that one of  $G[X], \overline{G}[X]$  has maximum degree at most  $\varepsilon^2|G|$  (and so at most  $\varepsilon|X|$ ).*

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*.

We would like to prove that more graphs than just  $P_4$  satisfy 1.2, and one natural way is via vertex-substitution (for example, Alon, Pach and Solymosi [2] showed that graphs satisfying the Erdős-Hajnal conjecture are closed under vertex-substitution). We have not been able to show that the graphs that satisfy 1.2 are closed under vertex-substitution. But we *have* been able to show that  $P_4$  itself has an even stronger property than 1.2, and graphs with this stronger property are closed under vertex-substitution. Consequently:

**1.4** *All graphs that can be obtained by vertex-substitution starting from copies of  $P_4$  and its subgraphs satisfy 1.2.*

Let us say a *copy* of  $H$  in  $G$  is an isomorphism from  $H$  to an induced subgraph of  $G$ . Let  $\text{ind}_H G$  be the number of copies of  $H$  in  $G$ . There is a theorem of Nikiforov [12], 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  $\text{ind}_H(G) \leq \delta|G|^{|H|}$ , then 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  $\varepsilon$ .*

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

- $\text{ind}_H(G) \geq \varepsilon^d |G|^{|H|}$ ; or
- 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  $\varepsilon$ .

We will show, using a recent “polynomial removal lemma” for  $P_4$ , proved by Alon and Fox [3], that:

**1.6** *All graphs with at most four vertices are viral.*

We will also show:

**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.*

We deduce:

**1.8** *All graphs that can be obtained by vertex-substitution starting from graphs with at most four vertices are viral.*

In a later paper [11], it is shown that all graphs currently known to satisfy the Erdős-Hajnal conjecture are viral, and therefore satisfy 1.2. That result does not use 1.6, and so gives an alternative proof of 1.6.

In the final section, we will discuss a different strengthening of Rödl’s theorem, and prove:

**1.9** *If  $G$  is a cograph, then for every  $\varepsilon$  with  $0 < \varepsilon \leq 1$ , there is a partition of  $V(G)$  into at most  $480\varepsilon^{-4}$  sets such that for each of them, say  $X$ , one of  $G[X], \overline{G}[X]$  has maximum degree at most  $\varepsilon|X|$ .*

## 2 Cographs have large dense or sparse sets

In this section we prove 1.3. We will discuss how close it is to best possible in the next section.

Cographs are well understood. There is a theorem discovered independently by several authors (see [6]), that:

**2.1** *If  $G$  is a cograph with  $|G| \geq 2$  then one of  $G, \overline{G}$  is disconnected.*

We will use 2.1 to prove 1.3 by induction on  $|G|$ . Applying it directly does not seem to work, and to use induction we will use a strengthening of 1.3, the following (1.3 follows by setting  $x = y = \varepsilon$ ):

**2.2** *If  $G$  is a cograph then, for all  $x, y \geq 0$  with  $\min(x, y) \leq 1$ , either:*

- *there exists  $X \subseteq V(G)$  with  $|X| \geq x|G|$  such that  $G[X]$  has maximum degree at most  $xy|G|$ ; or*
- *there exists  $Y \subseteq V(G)$  with  $|Y| \geq y|G|$  such that  $\overline{G}[Y]$  has maximum degree at most  $xy|G|$ .*

**Proof.** If  $|G| \leq 1$  the result is true, so we assume that  $|G| \geq 2$  and the result holds for all cographs with fewer vertices, and for all choices of  $x, y \geq 0$  with  $\min(x, y) \leq 1$ . If  $x > 1$ , then  $y \leq 1$  and the second bullet holds choosing  $Y \subseteq V(G)$  with  $|Y| = \lceil y|G| \rceil$ ; so we may assume that  $x \leq 1$  and similarly  $y \leq 1$ . By 2.1, taking complements if necessary, we may assume that  $G$  is not connected; let  $G_1, G_2$  be two non-null subgraphs of  $G$ , with union  $G$  and with  $V(G_1) \cap V(G_2) = \emptyset$ . Now we are given  $x, y \geq 0$  with  $\min(x, y) \leq 1$ .

For  $i = 1, 2$ , let  $y_i = y|G|/|G_i|$ . If for some  $i \in \{1, 2\}$  there exists  $Y_i \subseteq V(G_i)$  with  $|Y_i| \geq y_i|G_i| = y|G|$  such that  $\overline{G_i}[Y_i]$  has maximum degree at most  $xy_i|G_i| = xy|G|$ , then the second bullet holds. Hence we assume that for  $i = 1, 2$  there is no such  $Y_i$ . But  $\min(x, y_i) \leq 1$ , so from the inductive hypothesis, for  $i = 1, 2$  there exists  $X_i \subseteq V(G_i)$  with  $|X_i| \geq x|G_i|$  such that  $G[X_i]$  has maximum degree at most  $xy_i|G_i| = xy|G|$ . Then  $|X_1 \cup X_2| \geq x|G|$  and  $G[X_1 \cup X_2]$  has maximum degree at most  $xy|G|$ , and the first bullet of the theorem holds. This proves 2.2. ■

Here is a consequence, strengthening 1.3:

**2.3** *Let  $G$  be a cograph, and let  $0 \leq \varepsilon \leq 1$ . Then there exists  $X, Y \subseteq V(G)$ , such that  $G[X], \overline{G}[Y]$  both have maximum degree at most  $\varepsilon|G|$ , and with  $|X| \cdot |Y| \geq \varepsilon|G|^2$ .*

**Proof.** Let  $I$  be the set of  $x \in [0, 1]$  such that for some  $X \subseteq V(G)$ ,  $|X| \geq x|G|$  and  $G[X]$  has maximum degree at most  $\varepsilon|G|$ ; and let  $J$  be the set of  $x \in [0, 1]$  such that for some  $Y \subseteq V(G)$ ,  $x|Y| \geq \varepsilon|G|$  and  $\overline{G}[Y]$  has maximum degree at most  $\varepsilon|G|$ . By 2.2,  $I \cup J = [0, 1]$ . Since  $I, J$  are nonempty closed sets (because  $G$  is finite), it follows that  $I \cap J \neq \emptyset$ . This proves 2.3. ■

The form of 2.2 seems novel, and suggests that we ask which other graphs have the same property. Let us say  $G$  is *good* if for all  $x, y$  with  $0 \leq x, y \leq 1$ , either:

- there exists  $X \subseteq V(G)$  with  $|X| \geq x|G|$  such that  $G[X]$  has maximum degree at most  $xy|G|$ ;  
or
- there exists  $Y \subseteq V(G)$  with  $|Y| \geq y|G|$  such that  $\overline{G}[Y]$  has maximum degree at most  $xy|G|$ .

Thus, complements of good graphs are good; 2.2 says that all cographs are good; and its proof shows that goodness is preserved under taking disjoint unions. Which other graphs are good? This is still open, but we can show (we omit the proofs):

- all forests are good;
- the bull is not good;
- a cycle of length at least five is good if and only if its length is a multiple of six; and
- goodness is *not* preserved under vertex-substitution; indeed, substituting a two-vertex graph for a vertex of a good graph does not always preserve goodness.

### 3 The tightness of 1.3

Let us say two disjoint subsets  $A, B$  are *complete* to each other if every vertex in  $A$  is adjacent to every vertex in  $B$ , and *anticomplete* if there are no edges between  $A, B$ .

For  $\varepsilon \in [0, 1]$ , let  $\delta_\varepsilon$  be the supremum of all  $\delta$  such that for every cograph  $G$ , there exists  $X \subseteq V(G)$  such that  $|X| \geq \delta|G|$  and one of  $G[X], \overline{G}[X]$  has maximum degree at most  $\varepsilon\delta|G|$ . The next result shows that 1.3 is almost tight.

**3.1** *If  $\varepsilon \in [0, 1)$ , then  $\varepsilon \leq \delta_\varepsilon \leq (\lceil \varepsilon^{-1} \rceil - 1)^{-1}$ .*

**Proof.** By 1.3,  $\varepsilon \leq \delta_\varepsilon$ . Let  $m = \lceil \varepsilon^{-1} \rceil - 1$ ; that is, the largest integer strictly less than  $1/\varepsilon$ . Take an integer  $n \geq (1/m - \varepsilon)^{-1}$ , and let  $G$  be the cograph consisting of the disjoint union of  $m$  complete graphs  $C_1, \dots, C_m$ , each with  $n$  vertices. We will show that if  $X \subseteq V(G)$  and one of  $G[X], \overline{G}[X]$  has maximum degree at most  $\varepsilon|X|$ , then  $|X| \leq |G|/m$ , and consequently  $\delta_\varepsilon \leq 1/m$ . Let  $X \subseteq V(G)$  with  $|X| > |G|/m = n$ . By the pigeonhole principle,  $|X \cap C_i| \geq |X|/m$  for some  $i$ , and so  $G[X]$  has maximum degree at least  $|X|/m - 1 > \varepsilon|X|$  (because  $(1/m - \varepsilon)|X| \geq (1/m - \varepsilon)n > 1$ ). But since  $|X| > |G|/m = |C_i|$ , there is a vertex in  $X \setminus C_i$ , and the degree of this vertex in  $\overline{G}[X]$  is at least  $|X|/m > \varepsilon|X|$ . This proves 3.1.  $\blacksquare$

The result 1.3 is neat, and one might think it should be tight, but it is not; and indeed, neither of the bounds of 3.1 is tight when  $1/2 \leq \varepsilon < 1$ . We will show that  $\delta_\varepsilon = 1/(2 - \varepsilon)$  in this range. To do so, we first show the following, which implies that  $\delta_\varepsilon \geq 1/(2 - \varepsilon) > \varepsilon$  when  $1/2 \leq \varepsilon < 1$ :

**3.2** *Let  $1/2 \leq \varepsilon < 1$  and let  $\delta = 1/(2 - \varepsilon)$ . For every non-null cograph  $G$ , there is a set  $X \subseteq V(G)$  with  $|X| > \delta|G|$  such that one of  $G[X], \overline{G}[X]$  has maximum degree at most  $\varepsilon\delta|G|$ .*

**Proof.** Let  $G$  be a non-null cograph, and let  $1/2 \leq \varepsilon < 1$ . Let  $\delta = 1/(2 - \varepsilon)$  and  $d = \varepsilon/(2 - \varepsilon)$ ; we must show that there is a set  $X \subseteq V(G)$  with  $|X| > \delta|G|$  such that one of  $G[X], \overline{G}[X]$  has maximum degree at most  $d|G|$ .

We partition  $V(G)$  into sets  $X_1, \dots, X_k$  as follows. Suppose that  $i \geq 1$  and we have defined  $X_1, \dots, X_{i-1}$ , such that  $V(G) \neq X_1 \cup \dots \cup X_{i-1}$ . Let  $Y = V(G) \setminus (X_1 \cup \dots \cup X_{i-1})$ . If  $|Y| = 1$ , let  $X_i = Y$  and  $k = i$ . Now we assume that  $|Y| > 1$ , and define  $X_i$  as follows. By 2.1, one of  $G[Y], \overline{G}[Y]$  is not connected. Let  $X_i$  be a subset of  $Y$  that is the vertex set of a component of one of  $G[Y], \overline{G}[Y]$ , chosen with  $|X_i|$  minimum. Thus  $|X_i| \leq |Y|/2$ , and in particular  $V(G) \neq X_1 \cup \dots \cup X_i$ . This completes the inductive definition.

(1) *We may assume that  $|X_i| \leq \delta|G|/2$  for  $1 \leq i \leq k - 1$ .*

Suppose that some  $|X_i| > \delta|G|/2$ , and let  $Y = V(G) \setminus (X_1 \cup \dots \cup X_i)$ . Choose  $A \subseteq X_i$  with  $|A| = \lfloor \delta|G|/2 + 1 \rfloor$ . As we saw,  $|Y| \geq |X_i|$ , and so there exists  $B \subseteq Y$  with  $|B| = |A|$ . Now the set  $A \cup B$  has cardinality more than  $\delta|G|$ . Moreover, from the construction,  $X_i$  is either complete or anticomplete to  $Y$ , and by taking complements if necessary, we may assume the former. But then every vertex in  $A$  has no neighbours in  $B$  and has at most  $|A| - 1 \leq \delta|G|/2 \leq \varepsilon\delta|G|$  neighbours in  $A$ , and similarly for  $B$ ; and so setting  $X = A \cup B$  satisfies the theorem. This proves (1).

We may assume that  $|G| \geq 2$  and so  $k \geq 2$ . If  $d|G| \geq |G| - 1$ , then the theorem is satisfied with  $X = V(G)$  (because  $\delta < 1$  and every vertex has at most  $d|G|$  neighbours in  $G$ ). So we may assume that  $d|G| < |G| - 1$ . Choose  $h$  with  $0 \leq h \leq k - 1$ , minimum such that  $|X_{h+1} \cup \dots \cup X_k| \leq d|G| + 1$ . (This is possible since the condition is satisfied when  $h = k - 1$ ). Since  $|G| > d|G| + 1$  it follows that  $h \geq 1$ . By moving to the complement if necessary, we may assume that there  $X_h, Y$  are anticomplete,

where  $Y = X_{h+1} \cup \dots \cup X_k$ . Let  $I$  be the set of all  $i \in \{1, \dots, h\}$  such that  $X_i, Y$  are anticomplete, and let  $J$  be the set of all  $i \in \{1, \dots, h\}$  such that  $X_i, Y$  are complete. Thus  $h \in I$ . Moreover, all the sets  $X_i$  ( $i \in I$ ) are pairwise anticomplete, and the sets  $X_i$  ( $i \in J$ ) are pairwise complete.

Choose  $Z \subseteq X_h$  such that  $|Y \cup Z| = \lfloor d|G| + 1 \rfloor$  (this is possible since  $|X_h \cup Y| > d|G| + 1$  from the minimality of  $h$ ). Let  $A$  be the union of  $Y$  and the sets  $X_i$  ( $i \in I$ ). Since each of the sets  $X_i$  ( $i \in I$ ) and  $Y$  have cardinality at most  $d|G| + 1$  by (1), and there are no edges between them, it follows that  $G[A]$  has maximum degree at most  $d|G|$ . Similarly, let  $B$  be the union of  $Y \cup Z$  and the sets  $X_i$  ( $i \in J$ ); then since these sets all have cardinality at most  $d|G| + 1$ , and there are no edges of  $\overline{G}$  between any two of them, it follows that  $\overline{G}[B]$  has maximum degree at most  $d|G|$ . But  $|A| + |B| = |G| + |Y| + |Z|$ , and so one of  $|A|, |B|$  has cardinality at least  $(|G| + |Y| + |Z|)/2$ . To complete the proof it suffices to show that  $(|G| + |Y| + |Z|)/2 \geq \delta|G|$ . Certainly  $|Y \cup Z| > d|G|$ ; and hence

$$(|G| + |Y| + |Z|)/2 > (1 + d)|G|/2 = (1 + \varepsilon/(2 - \varepsilon))|G|/2 = \delta|G|.$$

This proves 3.2. ■

3.2 says that  $|X| > \delta|G|$ , and hence  $|X| \geq \lfloor \delta|G| + 1 \rfloor$ . Next we show that this is tight.

**3.3** *Let  $1/2 \leq \varepsilon \leq 1$  and  $\delta = 1/(2 - \varepsilon)$ . For each even integer  $2n$ , there is a cograph  $G$  with  $2n$  vertices such that if  $X \subseteq V(G)$  and one of  $G[X], \overline{G}[X]$  has maximum degree at most  $\varepsilon\delta|G|$ , then  $|X| \leq \delta|G| + 1$  (and hence  $|X| \leq \lfloor \delta|G| + 1 \rfloor$ ).*

**Proof.** Let  $G$  be the ‘‘half-graph’’ with vertex set  $\{a_1, \dots, a_n, b_1, \dots, b_n\}$ , in which  $\{a_1, \dots, a_n\}$  is a stable set,  $\{b_1, \dots, b_n\}$  is a clique, and  $a_i, b_j$  are adjacent if and only if  $i \leq j$ . This graph is a cograph. Now choose  $X \subseteq V(G)$  such that  $G[X]$  has maximum degree at most  $\varepsilon\delta|G|$ , with  $|X|$  maximum, and subject to that with  $|X \cap A|$  maximum. Since  $|X| > |G|/2$ , it contains a vertex  $b \in B$ ; and so for each  $a \in A$ , since  $b$  dominates  $a$ , and we cannot trade  $b$  for  $a$ , it follows that  $a \in X$  and so  $A \subseteq X$ . Let  $|X \cap B| = i$  say; then there is a vertex in  $X \cap B$  with  $i$  neighbours in  $A$  and adjacent to all other vertices in  $X \cap B$ , and since its degree in  $G[X]$  is at most  $\varepsilon\delta|G|$ , we deduce that  $2i - 1 \leq \varepsilon\delta|G|$ . So  $|X \cap B| \leq (\varepsilon\delta|G| + 1)/2$ , and hence  $|X| \leq |G|/2 + (\varepsilon\delta|G| + 1)/2 = \delta|G| + 1/2$ . Similarly (the graph is not quite self-complementary), if  $X \subseteq V(G)$  and  $\overline{G}[X]$  has maximum degree at most  $\varepsilon\delta|G|$ , it follows that  $|X| \leq \delta|G| + 1$ . This proves 3.3. ■

We deduce:

**3.4** *If  $1/2 \leq \varepsilon \leq 1$ , then  $\delta_\varepsilon = 1/(2 - \varepsilon)$ .*

**Proof.** By 3.2,  $\delta_\varepsilon \geq 1/(2 - \varepsilon)$ , and by 3.3,  $\delta_\varepsilon \leq 1/(2 - \varepsilon)$ . This proves 3.4. ■

## 4 Viral graphs and vertex-substitution

Let us prove 1.7, which we restate:

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

**Proof.** Let  $H$  be obtained by substituting  $H_2$  for a vertex  $v$  say of  $H_1$ . For  $i = 1, 2$ , since  $H_i$  is a viral, there exists  $d_i$  as in the definition of ‘‘viral’’. Let  $d = (|H_2| + 1)(d_1 + 1) + d_2$ . To show that  $H$  is a viral, we will show that:

(1) For every graph  $G$  and all  $\varepsilon$  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  $\varepsilon$ ; or
- there are at least  $\varepsilon^d |G|^{|H|}$  copies of  $H$  in  $G$ .

Since  $\varepsilon^{d_1} |G| \geq \varepsilon^d |G|$ , we may assume that there is no  $X \subseteq V(G)$  with  $|X| \geq \varepsilon^{d_1} |G|$  such that one of  $G[X], \overline{G}[X]$  has edge-density at most  $\varepsilon$ , since otherwise the first bullet of (1) holds. Consequently, from the choice of  $d_1$ , there are at least  $\varepsilon^{d_1} |G|^{|H_1|}$  copies of  $H_1$  in  $G$ . For each copy  $\phi$  of  $H_1 \setminus \{v\}$  in  $G$ , let  $N(\phi)$  be the set of all vertices  $u \in V(G)$  such that extending  $\phi$  by mapping  $v$  to  $u$  gives a copy of  $H_1$  in  $G$ , and let  $n(\phi) = |N(\phi)|$ . Let  $\Phi$  be the set of all copies of  $H_1 \setminus \{v\}$  in  $G$ ; then

$$\sum_{\phi \in \Phi} n(\phi) \geq \varepsilon^{d_1} |G|^{|H_1|}.$$

Let  $\Psi$  be the set of all  $\phi \in \Phi$  such that  $n(\phi) \geq \varepsilon^{d_1+1} |G|$ . Since

$$\sum_{\phi \in \Phi \setminus \Psi} n(\phi) \leq \sum_{\phi \in \Phi \setminus \Psi} \varepsilon^{d_1+1} |G| \leq |G|^{|H_1|-1} \varepsilon^{d_1+1} |G|,$$

it follows that

$$\sum_{\phi \in \Psi} n(\phi) \geq \varepsilon^{d_1} |G|^{|H_1|} (1 - \varepsilon) \geq \varepsilon^{d_1+1} |G|^{|H_1|}.$$

Since  $n(\phi) \leq |G|$ , we deduce that  $|\Psi| \geq \varepsilon^{d_1+1} |G|^{|H_1|-1}$ .

Let  $\phi \in \Psi$ . Thus  $|N(\phi)| = n(\phi) \geq \varepsilon^{d_1+1} |G|$ . From the choice of  $d_2$ , either there exists  $X \subseteq N(\phi)$  with  $|X| \geq \varepsilon^{d_2} |N(\phi)|$  such that one of  $G[X], \overline{G}[X]$  has edge-density at most  $\varepsilon$ , or there are  $\varepsilon^{d_2} |N(\phi)|^{|H_2|}$  copies of  $H_2$  in  $G[N(\phi)]$ . In the first case, since  $\varepsilon^{d_2} |N(\phi)| \geq \varepsilon^{d_2} \varepsilon^{d_1+1} |G| \geq \varepsilon^d |G|$ , the first bullet of (1) holds; so we may assume that there are at least

$$\varepsilon^{d_2} |N(\phi)|^{|H_2|} \geq \varepsilon^{d_2} \varepsilon^{(d_1+1)|H_2|} |G|^{|H_2|}$$

copies of  $H_2$  in  $G[N(\phi)]$ , and hence each  $\phi \in \Psi$  can be extended to at least  $\varepsilon^{d_2} \varepsilon^{(d_1+1)|H_2|} |G|^{|H_2|}$  copies of  $H$ . Since  $|\Psi| \geq \varepsilon^{d_1+1} |G|^{|H_1|-1}$ , there are at least

$$\varepsilon^{d_1+1} |G|^{|H_1|-1} \varepsilon^{d_2} \varepsilon^{(d_1+1)|H_2|} |G|^{|H_2|} = \varepsilon^{d_1+1+d_2+(d_1+1)|H_2|} |G|^{|H|} = \varepsilon^d |G|^{|H|}$$

copies of  $H$  in  $G$ , and hence the second bullet of (1) holds. This proves (1), and hence shows that  $H$  is viral, and proves 4.1. ■

Next we will deduce 1.6. The proof uses both 1.3 and a polynomial bound in the induced graph removal lemma for  $P_4$ . The induced graph removal lemma (see [1, 5, 10]) says that for each graph  $H$  and  $0 < \varepsilon \leq 1/2$  there exists  $\delta > 0$  such that every graph  $G$  with at most  $\delta |G|^{|H|}$  copies of  $H$  can be made  $H$ -free by adding or deleting at most  $\varepsilon |G|^2$  edges. For  $H = P_4$ , Alon and the first author [3] proved a polynomial bound:



**4.2** *There exists  $d > 0$  such that, if  $0 < \varepsilon \leq 1/2$ , then every graph  $G$  containing at most  $\varepsilon^d |G|^{|H|}$  copies of  $P_4$  can be made  $P_4$ -free by adding or deleting at most  $\varepsilon |G|^2$  edges.*

We deduce 1.6, which we restate:

**4.3** *Every graph on at most four vertices is viral.*

**Proof.** Every graph on at most two vertices is viral from the definition. Every graph on three or four vertices, apart from  $P_4$ , can be obtained through vertex-substitution from smaller graphs. So from 1.7 it suffices to prove that  $P_4$  is a viral. Let  $d$  be as in 4.2, and let  $0 < \varepsilon \leq 1/2$ . We will show that either

- there exists  $X \subseteq V(G)$  with  $|X| \geq \varepsilon |G|/4 \geq \varepsilon^3 |G|$  such that one of  $G[X], \overline{G}[X]$  has at most  $\varepsilon \binom{|X|}{2}$  edges; or
- there are at least  $(\varepsilon/4)^{3d} |G|^4 \geq \varepsilon^{12d} |G|^4$  copies of  $P_4$  in  $G$ .

We first dispose of a trivial case, when  $\varepsilon |G|/4 \leq 3$ . Then the first bullet holds if  $|G| \geq 6$  (because then  $G$  or  $\overline{G}$  has a triangle), and also if  $|G| \leq 5$  (because then  $\varepsilon |G|/4 \leq 5/8$  and we can take  $|X| = 1$ ). So we may assume that  $\varepsilon |G|/4 > 3$ . From the choice of  $d$  (with  $\varepsilon$  replaced by  $(\varepsilon/4)^3$ ), either  $G$  contains at least  $(\varepsilon/4)^{3d} |G|^4$  copies of  $P_4$  (in which case we are done), or we can obtain a  $P_4$ -free graph  $G'$  with the same vertex set as  $G$  by adding or deleting at most  $(\varepsilon/4)^3 |G|^2$  edges from  $G$ . In the latter case, by 1.3, there exists  $X \subseteq V(G)$  with  $|X| \geq \varepsilon |G|/4 > 3$  such that one of  $G'[X], \overline{G'}[X]$  has maximum degree at most  $(\varepsilon/4)^2 |G|$ . Then one of  $G[X], \overline{G}[X]$  has at most

$$(\varepsilon/4)^3 |G|^2 + (\varepsilon/4)^2 |G| |X|/2 \leq \frac{3}{8} \varepsilon |X|^2 \leq \varepsilon \binom{|X|}{2}$$

edges (since  $|X| \geq 4$ ). This proves 4.3. ■

## 5 Partitioning into Rödl sets

For  $\varepsilon > 0$ , let us say  $X \subseteq V(G)$  is  $\varepsilon$ -restricted if one of  $G[X], \overline{G}[X]$  has maximum degree at most  $\varepsilon |X|$ . There is a strengthening of Rödl's theorem proved in [4]:

**5.1** *For every graph  $H$  and every  $\varepsilon > 0$ , there exists  $N > 0$  such that if  $G$  is  $H$ -free, there is a partition of  $V(G)$  into at most  $N$   $\varepsilon$ -restricted subsets.*

(Note that being  $\varepsilon$ -restricted involves maximum degree rather than edge-density; the edge-density version is an easy consequence of 1.1.)

There is a corresponding strengthening of the Fox-Sudakov conjecture: perhaps in 5.1,  $N$  can always be taken to be a polynomial in  $\varepsilon^{-1}$  (depending on  $H$ ). This seems very intractible, and we have not been able to show it even when  $H$  is a triangle. But it works when  $H = P_4$ , as we will show below.

It would be even nicer to get a version of 5.1 that is strong enough to imply 1.3 when  $H = P_4$ , but that eludes us. The obvious attempt is false:

**5.2** For all  $\varepsilon$  with  $0 < \varepsilon < 1/2$  such that  $\varepsilon^{-1}$  is not an integer, there is a cograph  $G$  such that there is no partition of  $V(G)$  into at most  $1/\varepsilon$   $\varepsilon$ -restricted sets.

**Proof.** Let  $k = \lfloor \varepsilon^{-1} \rfloor$ , let  $m$  be an integer with  $(1 - k\varepsilon)m \geq 1$ , and let  $n$  be some large integer. Let  $G$  be the graph consisting of  $k + 1$  disjoint cliques  $C_0, \dots, C_k$ , where  $|C_0| = n$  and  $|C_1|, \dots, |C_k| = mn$ . Suppose that there is a partition of  $V(G)$  into at most  $1/\varepsilon$  (and hence at most  $k$ )  $\varepsilon$ -restricted sets, and so there is an  $\varepsilon$ -restricted set  $X$  with  $|X| \geq mn + n/k$ . Let  $x_i := |X \cap C_i|$  for  $0 \leq i \leq k$ . Suppose first that  $G[X]$  has maximum degree at most  $\varepsilon|X|$ . It follows that  $x_i - 1 \leq \varepsilon|X|$  for  $1 \leq i \leq k$ , and  $x_0 \leq n$ , and summing,

$$|X| = x_0 + x_1 + \dots + x_k \leq n + k\varepsilon|X| + k;$$

so

$$n/(1 - \varepsilon k) + n/k \leq mn + n/k \leq |X| \leq (k + n)/(1 - k\varepsilon),$$

a contradiction when  $n$  is large. Thus,  $\overline{G}[X]$  has maximum degree at most  $\varepsilon|X|$ . Since  $|X| > |C_i|$  for  $0 < i \leq k$ , there exists  $i \in \{0, \dots, k\}$  such that  $0 < x_i \leq |X|/2$ . Choose  $v \in X \cap C_i$ ; then  $v$  has at least  $|X|/2 > \varepsilon|X|$  non-neighbours in  $X$ , contradicting that  $\overline{G}[X]$  has maximum degree at most  $\varepsilon|X|$ . This proves 5.2. ■

What happens in 5.2 when  $\varepsilon^{-1}$  is an integer? Is it true that for every integer  $k \geq 1$ , every cograph  $G$  can be vertex-partitioned into at most  $k$  parts, each  $1/k$ -restricted? For  $k = 2$  this is true, and for  $k = 3$  it is false. Let us see both those things now.

To show it is true for  $k = 2$ , let us say  $X \subseteq V(G)$  is *thin* if every component of  $G[X]$  has at most  $(|X| + 1)/2$  vertices, and *thick* if every component of  $\overline{G}[X]$  has at most  $(|X| + 1)/2$  vertices. So thick and thin sets are both  $1/2$ -restricted. We will prove:

**5.3** If  $G$  is a cograph, there is a partition of  $V(G)$  into a thin set and a thick set.

**Proof.** Let  $G$  be a cograph. Choose a partition  $A, B, C$  of  $V(G)$  with  $C$  minimal such that  $|C| > |G|/2$ ,  $A$  is anticomplete to  $C$ , and  $B$  is complete to  $C$ . (This is possible since we may take  $A = B = \emptyset$ .) We may assume that  $|C| \geq 2$ , and so one of  $G[C], \overline{G}[C]$  is not connected, by 2.1. By taking complements if necessary, we may assume that  $\overline{G}[C]$  is not connected. Partition  $C$  into two nonempty sets  $P, Q$  complete to each other. From the minimality of  $C$ , it follows that  $|P|, |Q| \leq |G|/2$ . In summary, we have a partition of  $V(G)$  into four sets  $A, B, P, Q$ , where  $A$  or  $B$  may be empty, but  $P, Q \neq \emptyset$ ;  $|A| + |B| < n/2$  and  $|P|, |Q| \leq n/2$ ;  $B, P, Q$  are mutually complete, and  $A$  is anticomplete to  $P \cup Q$ . (The edges between  $A, B$  are unrestricted.)

We may assume that  $|Q| \geq |P|$ . Define  $m := \max(0, |Q| - |P| - |B|)$ .

$$(1) |A| + m \leq 2|Q|, \text{ and } |A| - m \leq 2|P|.$$

Suppose first that  $m = 0$ . Then  $|Q| \leq |P| + |B|$ , and we must show that  $|A| \leq 2|P|$  ( $\leq 2|Q|$ ). But  $|P| \geq |Q| - |B|$ , so  $2|P| \geq |P \cup Q| - |B| \geq |A|$  as required. Now suppose that  $m > 0$ , and so  $|Q| = |P| + |B| + m$ ; and we must show that  $|A| + |Q| - |P| - |B| \leq 2|Q|$  and  $|A| - (|Q| - |P| - |B|) \geq 2|P|$ . The first says  $|A| - |B| \leq |P| + |Q|$ , and the second that  $|A| + |B| \geq |P| + |Q|$ , and both of these are true. This proves (1).

Consequently, we may choose subsets  $P' \subseteq P$  and  $Q' \subseteq Q$  with  $|P \setminus P'| = \lceil (|A| - m)/2 \rceil$  and  $|Q \setminus Q'| = \lfloor (|A| + m)/2 \rfloor$ . We claim that  $X := A \cup (P \setminus P') \cup (Q \setminus Q')$  is thin and  $Y := B \cup P' \cup Q'$

is thick, and so the theorem holds. Since  $|(P \setminus P') \cup (Q \setminus Q')| = |A|$ , and so  $|X| = 2|A|$ , and each of its components has vertex set a subset of one of  $A, P \setminus P') \cup (Q \setminus Q')$ , and therefore has at most  $|A|$  vertices, it follows that  $X$  is thin. To show that  $Y$  is thick, we need:

(2) *Each of  $B, P', Q'$  has cardinality at most  $(|Y| + 1)/2$ .*

Certainly  $|B| \leq n/2 - |A| = |Y|/2$  since  $|A| + |B| < n/2$ . For the other two inequalities, we have  $|Y| = |P'| + |Q'| + |B|$ , and  $|P'| = |P| - \lceil (|A| - m)/2 \rceil$ , and  $|Q'| = |Q| - \lfloor (|A| + m)/2 \rfloor$ , so we must show that

$$|P| - \lceil (|A| - m)/2 \rceil \leq |Q| - \lfloor (|A| + m)/2 \rfloor + |B| + 1$$

and

$$|Q| - \lfloor (|A| + m)/2 \rfloor \leq |P| - \lceil (|A| - m)/2 \rceil + |B| + 1.$$

These simplify to showing that  $|P| + m \leq |Q| + |B| + 1$ , and  $|Q| \leq |P| + |B| + m$ , which both follow from the choice of  $m$ , and since  $|Q| \geq |P|$ . This proves (2).

From (2), we deduce that  $Y$  is thin. This proves 5.3. ■

Now a counterexample for  $k = 3$ . Let  $n$  be a large integer, and take four disjoint sets  $A, B, C, D$ , of sizes  $2n, 3n, 4n, 5n$  respectively. Make  $A, B, C$  stable sets complete to each other, and make  $D$  a clique anticomplete to  $A \cup B \cup C$ , forming a graph  $G$ . We leave the reader to check that  $G$  admits no vertex-partition into three  $1/3$ -restricted sets. (For the convenience of the interested reader, a proof is included in an appendix.)

But at least a “polynomial” version of 5.1 is true when  $H = P_4$ , because of 1.9, which we will now prove. The proof breaks into several steps, that follow.

A pair  $(P, Q)$  of disjoint subsets of  $V(G)$  is *pure* if  $Q$  is either complete or anticomplete to  $P$ .

**5.4** *Let  $G$  be a cograph with  $|G| \geq 2$ , and let  $0 < \varepsilon \leq 1$ . Then there is a partition of  $V(G)$  into four (possibly empty) sets  $A_0, A_1, A_2, A_3$ , with the following properties:*

- $A_0$  is  $\varepsilon$ -restricted;
- every two of  $A_1, A_2, A_3$  form a pure pair;
- for  $1 \leq i \leq 3$ , if  $A_i \neq \emptyset$ , then there exists  $B \subseteq V(G) \setminus A_i$  with  $|B| \geq \frac{\varepsilon^2}{4}|G|$  such that  $(A_i, B)$  is a pure pair.

**Proof.** For convenience, we say a subset of  $V(G)$  is *big* if its cardinality is more than  $\frac{\varepsilon^2}{4}|G|$ , and *small* otherwise. Choose a maximal sequence  $S_1, \dots, S_k$  of nonempty, pairwise disjoint, small subsets of  $V(G)$ , such that

- for  $1 \leq i \leq k$ ,  $S_i$  is complete or anticomplete to  $V(G) \setminus (S_1 \cup \dots \cup S_i)$ , and
- $|S_1 \cup \dots \cup S_k| < |G|/2$ .

Let  $A = V(G) \setminus (S_1 \cup \dots \cup S_k)$ . Let  $I$  be the set of  $i \in \{1, \dots, k\}$  such that  $S_i$  is complete to  $V(G) \setminus (S_1 \cup \dots \cup S_i)$ , and  $J = \{1, \dots, k\} \setminus I$ . Let  $P = \bigcup_{i \in I} S_i$  and  $Q = \bigcup_{j \in J} S_j$ . Thus the sets  $S_i$  ( $i \in I$ ) and  $A$  are pairwise complete, and the sets  $S_j$  ( $j \in J$ ) and  $A$  are pairwise anticomplete.

We may assume that  $|A| \geq 2$ , since otherwise  $|G| \leq 2$  and the theorem is true. So one of  $G[A], \overline{G}[A]$  is not connected, by 2.1. Choose a partition  $B, C$  of  $A$  with  $B, C$  both nonempty, such that  $B$  is complete or anticomplete to  $C$ . We may assume that  $|B| \geq |C|$ , and so  $|B| \geq |G|/4$ .

From the maximality of the sequence, either:

- $|B| \leq |G|/2$  and  $C$  is small; or
- $C$  is big.

In the first case, by taking complements if necessary, we may assume that  $B, C$  are complete. Let  $P' = P \cup C$ . Thus every component of  $\overline{G}[P']$  has at most  $\frac{\varepsilon^2}{4}|G|$  vertices; so if  $|P'| \geq \varepsilon|G|/4$ , then  $P'$  is  $\varepsilon$ -restricted, and the theorem is satisfied taking  $A_0 = P', A_1 = B, A_2 = Q$  and  $A_4 = \emptyset$ , since  $B, Q$  are anticomplete. Note that the third condition of the theorem is satisfied, since  $A_1$  is complete to the big set  $A_0$  and  $A_2$  is complete to the big set  $A_1$ . So we may assume that  $|P'| < \varepsilon|G|/4$ . Similarly, if  $|Q| \geq \varepsilon|G|/4$ , then  $Q$  is  $\varepsilon$ -restricted, and the theorem is satisfied by setting  $A_0 = Q, A_1 = B$  and  $A_2 = P'$ , since  $A_1$  is anticomplete to the big set  $A_0$ , and  $A_2$  is complete to the big set  $A_1$ . But  $P' \cup Q = V(G) \setminus B$ , and  $|B| \leq |G|/2$ , and so one of  $P', Q$  has cardinality at least  $|G|/4 \geq \varepsilon|G|/4$ , and the theorem holds.

In the second case, since  $|B| \geq |G|/4$ , 1.3 implies that there exists  $X \subseteq B$  with  $|X| \geq \varepsilon|B|$  such that one of  $G[X], \overline{G}[X]$  has maximum degree at most  $\varepsilon^2|B|$ ; and by replacing  $X$  by a subset, we may assume that  $|X| = \lceil \varepsilon|B| \rceil$ . Thus  $X$  is  $\varepsilon$ -restricted. By taking complements if necessary, we may assume that  $G[X]$  has maximum degree at most  $\varepsilon|X|$ . Let  $Q' = Q \cup X$ ; then  $|Q'| \geq \varepsilon|G|/4$ . If  $v \in Q$  then its degree in  $G[Q']$  is at most  $\frac{\varepsilon^2}{4}|G| \leq \varepsilon|X| \leq \varepsilon|Q'|$ ; so  $Q'$  is  $\varepsilon$ -restricted, and the theorem is satisfied by setting  $A_0 = Q', A_1 = B \setminus X, A_2 = C$ , and  $A_3 = P$ . To see the last, note that  $(A_1, A_2)$  is a pure pair;  $A_3$  is complete to both  $A_1, A_2$ ;  $A_1$  is complete or anticomplete to the big set  $A_2$ ;  $A_2$  is complete or anticomplete to the big set  $A_1$ ; and  $A_3$  is complete to the big set  $A_1$ . This proves 5.4. ■

Let  $X \subseteq V(G)$  with  $X \neq \emptyset$ . A *ribbon attached to  $X$*  is a sequence  $\mathcal{B} = (B_1, \dots, B_k)$  of pairwise disjoint subsets of  $V(G) \setminus X$ , where  $k \geq 0$ , such that  $B_i$  is complete or anticomplete to  $X \cup B_1 \cup \dots \cup B_{i-1}$  for  $1 \leq i \leq k$ . Its *length* is  $k$ , and its *breadth* is the minimum of  $|B_i|/|X|$  for  $1 \leq i \leq k$  (or 1 if  $k = 0$ ). We say  $X$  is the *attachment* of the ribbon.

We will be concerned with partitions of  $V(G)$  into parts, such that for each part  $X$ , either  $X$  is  $\varepsilon$ -restricted, or there is a ribbon attached to  $X$ ; and moreover, that for every two parts  $X, Y$  that are not  $\varepsilon$ -restricted,  $X$  is either complete or anticomplete to  $Y$ . We must take care that the total number of sets in the partition is not too large, that the number of beribboned sets is not too large, and that the ribbons are long enough, and have breadth not too small. For  $\varepsilon > 0$ , let us say an  $(\varepsilon, k)$ -*beribboning* of a graph  $G$  is a partition  $\mathcal{P}$  of  $V(G)$  (together with, implicitly, a choice of ribbons), such that

- for each  $X \in \mathcal{P}$ , either  $X$  is  $\varepsilon$ -restricted or there is a ribbon of length  $k$  attached to  $X$ ; and
- if  $X, Y \in \mathcal{P}$  are different and not  $\varepsilon$ -restricted, then  $(X, Y)$  is a pure pair.

The *dimensions* of the beribboning are  $(m, n)$ , where  $n = |\mathcal{P}|$  and  $m$  is the number of members of  $\mathcal{P}$  that are not  $\varepsilon$ -restricted; and its *breadth* is the minimum of the breadth of its ribbons. From 5.4, we see that every cograph admits an  $(\varepsilon, 1)$ -beribboning with dimensions at most  $(3, 4)$  and breadth at least  $\varepsilon^2/4$ .

**5.5** *If  $0 < \varepsilon \leq 1$ , and  $G$  is a cograph that admits an  $(\varepsilon, k)$ -beribboning with dimensions at most  $(m, n)$  and breadth at least  $\beta$ , then it also admits an  $(\varepsilon, k)$ -beribboning with dimensions at most  $(\varepsilon^{-2}, n)$  and breadth at least  $\beta$ .*

**Proof.** We will prove that if  $G$  admits an  $(\varepsilon, k)$ -beribboning  $\mathcal{P}$  with dimensions  $(m, n)$  and breadth  $\beta$ , where  $m \geq \varepsilon^{-2}$ , then it also admits one with dimensions at most  $(m - 1, n)$  and breadth at least  $\beta$ . Let  $k = \lceil \varepsilon^{-1} \rceil$ . Since  $m \geq \varepsilon^{-2} > (k - 1)^2$ , and every cograph with more than  $(k - 1)^2$  vertices has a clique or stable set of size  $k$ , we may choose distinct  $X_1, \dots, X_k \in \mathcal{P}$ , not  $\varepsilon$ -restricted, such that either  $X_1, \dots, X_k$  are pairwise anticomplete or  $X_1, \dots, X_k$  are pairwise complete. We may assume that  $|X_k| \leq |X_1|, \dots, |X_{k-1}|$ . Choose  $Y_i \subseteq X_i$  with  $|Y_i| = |X_k|$  for  $1 \leq i \leq k - 1$ . Then  $Y_1 \cup \dots \cup Y_{k-1} \cup X_k$  is  $\varepsilon$ -restricted (because  $k \geq \varepsilon^{-1}$ ). Let  $\mathcal{P}'$  be the partition obtained from  $\mathcal{P}$  by replacing the sets  $X_1, \dots, X_k$  by the sets

$$X_1 \setminus Y_1, \dots, X_{k-1} \setminus Y_{k-1}, Y_1 \cup \dots \cup Y_{k-1} \cup X_k.$$

Thus  $\mathcal{P}'$  has the same number of members as  $\mathcal{P}$ , but at least one more of them is  $\varepsilon$ -restricted. Moreover, each of the sets  $X_i \setminus Y_i$  has a ribbon attached of length  $k$  and breadth at least  $\beta$ . So  $\mathcal{P}'$  is an  $(\varepsilon, k)$ -beribboning with dimensions at most  $(m - 1, n)$  and breadth  $\geq \beta$ . By repeating, this proves 5.5. ■

**5.6** *If  $G$  is a cograph, and  $0 < \varepsilon \leq 1$ , and  $k \geq 0$  is an integer, then  $G$  admits an  $(\varepsilon, k)$ -beribboning with dimensions at most  $(\varepsilon^{-2}, 1 + 3k\varepsilon^{-2})$  and breadth at least  $\varepsilon^2/4$ .*

**Proof.** We proceed by induction on  $k$ . For  $k = 0$ , the result is trivial. Inductively, we assume that  $k \geq 1$ , and  $G$  admits an  $(\varepsilon, k - 1)$ -beribboning  $\mathcal{P}$  with dimensions at most  $(\varepsilon^{-2}, 1 + 3(k - 1)\varepsilon^{-2})$  and breadth at least  $\varepsilon^2/4$ . Let  $X_1, \dots, X_m$  be the members of  $\mathcal{P}$  that are not  $\varepsilon$ -restricted. For  $1 \leq i \leq m$ , since  $|X_i| \geq 2$ , there is, by 5.4, a partition of  $X_i$  into four (possibly empty) sets  $A_{i0}, A_{i1}, A_{i2}, A_{i3}$ , with the following properties:

- $A_{i0}$  is  $\varepsilon$ -restricted;
- every two of  $A_{i1}, A_{i2}, A_{i3}$  form a pure pair;
- for  $1 \leq j \leq 3$ , if  $A_{ij} \neq \emptyset$ , then there exists  $B \subseteq X_i \setminus A_{ij}$  such that  $|B| \geq \frac{\varepsilon^2}{4}|X_i|$  and  $(A_{ij}, B)$  is a pure pair.

Let  $\mathcal{P}'$  be obtained from  $\mathcal{P}$  by replacing  $X_i$  by  $A_{i0}, A_{i1}, A_{i2}, A_{i3}$  for  $1 \leq i \leq m$ . We claim that this is a  $(\varepsilon, k)$ -beribboning with dimensions at most  $(3m, 1 + 3k\varepsilon^{-2})$  and breadth at least  $\varepsilon^2/4$ . Certainly

$$|\mathcal{P}'| \leq |\mathcal{P}| + 3m \leq 1 + 3k\varepsilon^{-2}$$

and the number of members of  $\mathcal{P}'$  that are not  $\varepsilon$ -restricted is at most  $3m$ . We need to check the ribbons. Let  $1 \leq i \leq m$ . There is a  $(\varepsilon, k - 1)$ -ribbon  $(B_1, \dots, B_{k-1})$  attached to  $X$  with breadth at least  $\varepsilon^2/4$ . Let  $1 \leq j \leq 3$  with  $A_{ij} \neq \emptyset$ . From the third bullet above, there exists a subset  $B \subseteq X_i \setminus A_{ij}$  such that  $|B| \geq \frac{\varepsilon^2}{4}|X_i|$  and  $(A_{ij}, B)$  is a pure pair. But then  $(B, B_1, \dots, B_{k-1})$  is a ribbon attached to  $A_{ij}$  of breadth at least  $\varepsilon^2/4$  and length  $k$ . This proves our claim that  $\mathcal{P}'$  is a  $(\varepsilon, k)$ -beribboning with dimensions at most  $(3m, 1 + 3k\varepsilon^{-2})$  and breadth at least  $\varepsilon^2/4$ , and then an application of 5.5 gives the result. This proves 5.6. ■

Let us say a ribbon  $(B_1, \dots, B_k)$  attached at  $X$  is *pure* if either all the sets  $X, B_1, \dots, B_k$  are pairwise complete, or all the sets  $X, B_1, \dots, B_k$  are pairwise anticomplete. An  $(\varepsilon, k)$ -beribboning is *pure* if all the ribbons it uses are pure.

**5.7** *If  $G$  is a cograph, and  $0 < \varepsilon \leq 1/2$ , then  $G$  admits a pure  $(\varepsilon, \lceil \varepsilon^{-1} \rceil)$ -beribboning with dimensions at most  $(\varepsilon^{-2}, 10\varepsilon^{-3})$  and breadth at least  $\varepsilon^2/4$ .*

**Proof.** Taking  $k = 2\lceil \varepsilon^{-1} \rceil \leq 3\varepsilon^{-1}$ , we deduce from 5.6 that  $G$  admits an  $(\varepsilon, k)$ -beribboning with dimensions at most  $(\varepsilon^{-2}, 1 + 3k\varepsilon^{-2}) \leq (\varepsilon^{-2}, 10\varepsilon^{-3})$  and breadth at least  $\varepsilon^2/4$ . Let  $(B_1, \dots, B_k)$  be a ribbon attached to some  $X \subseteq V(G)$ . Let  $I$  be the set of  $i \in \{1, \dots, k\}$  such that  $B_i$  is complete to  $X \cup B_1 \cup \dots \cup B_{i-1}$ , and  $J = \{1, \dots, k\} \setminus I$ . Then both of  $(B_i : i \in I)$ ,  $(B_j : j \in J)$  are pure ribbons attached to  $X$ , and one of them has length at least  $k/2 = \lceil \varepsilon^{-1} \rceil$ . Hence, for each  $X \in \mathcal{P}$  that is not  $\varepsilon$ -restricted, there is a pure ribbon of length at least  $\varepsilon^{-1}$  and breadth at least  $\varepsilon^2/4$  attached to  $X$ . This proves 5.7. ■

The purpose of the ribbons is: suppose we have a pure ribbon  $(B_1, \dots, B_k)$  attached to a set  $X$ , and its length is at least  $\varepsilon^{-1}$ , and  $B_1, \dots, B_k$  all have the same size. Then we can partition  $X \cup B_1 \cup \dots \cup B_k$  into  $\varepsilon$ -restricted sets, as follows. We can partition almost all of  $X$  into a few  $\varepsilon$ -restricted subsets, greedily, in such a way that the remainder,  $Y$  say, has size at most  $|B_1|$ ; and then  $Y \cup B_1 \cup \dots \cup B_k$  is also  $\varepsilon$ -restricted. But to use this method, we first need to tidy up the ribbons. We want to arrange that:

- the ribbons are vertex-disjoint from one another;
- for each set of the partition  $\mathcal{P}$ , at most half its vertices belong to ribbons; and
- for each ribbon, all its members have the same size.

We call these the *prettification conditions*. Let us say a  $(\varepsilon, k)$ -beribboning is *prettified* if it is pure and satisfies the three conditions above. All these things will be accomplished by replacing the sets of the ribbons by subsets of themselves. This will reduce the breadth, so we must be careful that the breadth does not get too small. In particular, if the sets of a ribbon are already very small, we may not be able to shrink them by the required factors, and we must treat such ribbons differently. But in this case, the corresponding attachment is also very small, a polynomial in  $\varepsilon^{-1}$ , and we can easily partition it into a few  $\varepsilon$ -restricted sets, and need not use the ribbon at all.

Let us see the last statement above. By repeatedly applying 1.3, we evidently have:

**5.8** *If  $G$  is a cograph, and  $X \subseteq V(G)$ , and  $0 \leq \varepsilon \leq 1$ , and  $t \geq 0$  is an integer, then there are  $t$  pairwise disjoint  $\varepsilon$ -restricted subsets of  $X$  with union  $X \setminus Y$  say, such that  $|Y| \leq (1-\varepsilon)^t |G| \leq e^{-\varepsilon t} |X|$ .*

So in particular, there is a partition of  $X$  into  $t$   $\varepsilon$ -restricted sets, if we take  $t$  so large that  $e^{\varepsilon t} > |X|$ .

**5.9** *For  $0 < \varepsilon \leq 1/2$ , if  $G$  is a cograph, then  $G$  admits a prettified  $(\varepsilon, \lceil \varepsilon^{-1} \rceil)$ -beribboning  $\mathcal{P}$  with dimensions at most  $(\varepsilon^{-2}, 21\varepsilon^{-4})$  and breadth at least  $\varepsilon^4/32$ .*

**Proof.** Let  $k = \lceil \varepsilon^{-1} \rceil$ . By 5.7,  $G$  admits a pure  $(\varepsilon, k)$ -beribboning  $\mathcal{P}_0$  with dimensions at most  $(\varepsilon^{-2}, 10\varepsilon^{-3})$  and breadth at least  $\varepsilon^2/4$ . Let  $q = 16$ . If  $X \in \mathcal{P}_0$  and  $|X| < \varepsilon^{-q}$ , then by 5.8 there is a partition of  $X_i$  into at most  $q\varepsilon^{-1} \log(\varepsilon^{-1}) \leq q\varepsilon^{-2}$   $\varepsilon$ -restricted sets. For each  $X \in \mathcal{P}_0$

that is not  $\varepsilon$ -restricted and has cardinality less than  $\varepsilon^{-q}$ , let us replace  $X$  in  $\mathcal{P}_0$  by the sets of the corresponding partition. We obtain a pure  $(\varepsilon, k)$ -beribboning  $\mathcal{P}$  of  $G$  with dimensions at most  $(\varepsilon^{-2}, 10\varepsilon^{-3} + q\varepsilon^{-4}) \leq (\varepsilon^{-2}, (q+5)\varepsilon^{-4})$  and breadth at least  $\varepsilon^2/4$ , such that each  $X \in \mathcal{P}$  that is not  $\varepsilon$ -restricted has cardinality at least  $\varepsilon^{-q}$ . Now we will prettify the ribbons.

Let  $X_1, \dots, X_m$  be the members of  $\mathcal{P}$  that are not  $\varepsilon$ -restricted, and for  $1 \leq i \leq m$  let  $(B_{i1}, \dots, B_{ik})$  be a pure ribbon attached to  $X_i$  with breadth at least  $\varepsilon^2/4$ .

(1) For  $1 \leq i \leq m$  and  $1 \leq j \leq n$  there exists  $C_{ij} \subseteq B_{ij}$ , with  $|C_{ij}| \geq \lfloor |B_{ij}|/m \rfloor$  such that all the sets  $C_{ij}$  ( $1 \leq i \leq m, 1 \leq j \leq k$ ) are pairwise disjoint.

For each  $i, j$ , let  $\mathcal{Q}_{ij}$  be a partition of  $B_{ij}$  into  $\lfloor |B_{ij}|/m \rfloor$  sets, of which  $\lfloor |B_{ij}|/m \rfloor$  have size  $m$ , and possibly one has size less than  $m$ . For each  $Q \in \mathcal{Q}_{ij}$  take a new vertex  $u_{ij}^Q$ , and let  $U$  be the set of all these new vertices; that is,

$$U = \{u_{ij}^Q : 1 \leq i \leq m, 1 \leq j \leq n, Q \in \mathcal{Q}_{ij}\}.$$

Let  $H$  be the graph with bipartition  $(U, V(G))$  in which  $u_{ij}^Q$  is adjacent to  $v \in V(G)$  if  $v \in Q$ . Then  $H$  has maximum degree at most  $m$ , and so its edge-set can be partitioned into  $m$  matchings. Let  $M$  be one of these matchings. For each vertex  $h \in V(H)$  with degree  $m$  in  $H$ , there is an edge of  $M$  incident with  $h$ . Consequently, for  $1 \leq i \leq m$  and  $1 \leq j \leq k$ , there are at least  $\lfloor |B_{ij}|/m \rfloor$  edges in  $M$  with ends in  $\mathcal{Q}_{ij}$ . Let  $C_{ij}$  be the set of vertices of  $G$  that are joined by an edge of  $M$  to a vertex in  $\mathcal{Q}_{ij}$ . This proves (1).

For  $1 \leq i \leq m$ , since  $|X_i| \geq \varepsilon^{-q}$ , it follows that each  $|B_{ij}| \geq \varepsilon^{2-q}/4 \geq \varepsilon^{-2} \geq m$ , and hence

$$|C_{ij}| \geq \lfloor |B_{ij}|/m \rfloor \geq |B_{ij}|/(2m) \geq (\varepsilon^4/8)|X_i|.$$

Consequently  $(C_{i1}, \dots, C_{ik})$  is a ribbon attached to  $X_i$  with breadth at least  $\varepsilon^4/8$ . These ribbons are pairwise vertex-disjoint, so we have satisfied the first prettification condition.

For the second condition, for each  $i, j$ , choose  $D_{ij} \subseteq C_{ij}$  such that for each  $Y \in \mathcal{P}$ ,  $|D_{ij} \cap Y| = \lfloor |C_{ij} \cap Y|/2 \rfloor$ . Then

$$|D_{ij}| \geq |C_{ij}|/2 - |\mathcal{P}|/2 \geq |C_{ij}|/2 - (q+5)\varepsilon^{-4}/2 \geq |C_{ij}|/4 \geq (\varepsilon^4/32)|X_i|.$$

since

$$|C_{ij}| \geq (\varepsilon^4/8)|X_i| \geq \varepsilon^{4-q}/8 \geq 2(q+5)\varepsilon^{-4}.$$

Then for each  $Y \in \mathcal{P}$ , at most half the vertices of  $Y$  belong to the union of the sets  $D_{ij}$ ; so the ribbons  $(D_{i1}, \dots, D_{ik})$  have breadth at least  $\varepsilon^4/32$  and satisfy the first and second prettification conditions. Finally, for the third, for each  $i, j$  choose  $E_{ij} \subseteq D_{ij}$  of cardinality  $\lfloor (\varepsilon^4/32)|X_i| \rfloor$ ; then the ribbons  $(E_{i1}, \dots, E_{ik})$  for  $1 \leq i \leq m$  satisfy all the prettification conditions, and have breadth at least  $\varepsilon^4/32$ . This proves 5.9. ■

Finally, let us deduce 1.9, which we restate, with  $\varepsilon$  replaced by  $\varepsilon/2$  for convenience:

**5.10** *If  $G$  is a cograph, then for every  $\varepsilon$  with  $0 < \varepsilon \leq 1/2$ , there is a partition of  $V(G)$  into at most  $30\varepsilon^{-4}$   $2\varepsilon$ -restricted sets.*

**Proof.** From 5.9,  $G$  admits a prettified  $(\varepsilon, \lceil \varepsilon^{-1} \rceil)$ -beribboning  $\mathcal{P}$  with dimensions at most  $(\varepsilon^{-2}, 21\varepsilon^{-4})$  and breadth at least  $\varepsilon^4/32$ . Let  $X_1, \dots, X_m$  be the members of  $\mathcal{P}$  that are not  $\varepsilon$ -restricted, and for  $1 \leq i \leq m$  let  $(B_{i1}, \dots, B_{ik})$  be a pure ribbon attached to  $X_i$  with breadth at least  $\varepsilon^4/32$ , satisfying the prettification conditions. Let  $F$  be the union of the sets  $B_{ij}$  for  $1 \leq i \leq m$  and  $1 \leq j \leq k$ . Thus  $|F \cap Y| \leq |Y|/2$  for each  $Y \in \mathcal{P}$ . Let  $Z = F \cup X_1 \cup \dots \cup X_m$ . We will partition  $Z$  and  $V(G) \setminus Z$  into  $2\varepsilon$ -restricted sets.

Let  $t = \lceil 8\varepsilon^{-2} \rceil \leq 9\varepsilon^{-2}$ . By 5.8, for  $1 \leq i \leq m$ , since

$$|X_i \setminus F| \leq |X_i| \leq (32\varepsilon^{-4})|B_{i1}|,$$

we may choose  $t$  pairwise disjoint  $\varepsilon$ -restricted subsets  $A_{i1}, \dots, A_{it}$  of  $X_i \setminus F$ , with union  $(X_i \setminus F) \setminus Y_i$  say, such that  $|Y_i| \leq |B_{i1}|$ . But then the sets

$$\begin{aligned} A_{ij} \quad (1 \leq i \leq m, 1 \leq j \leq t) \\ Y_i \cup B_{i1} \cup \dots \cup B_{ik} \quad (1 \leq i \leq m) \end{aligned}$$

are all  $\varepsilon$ -restricted, and pairwise disjoint, and have union  $Z$ .

For each  $\varepsilon$ -restricted set  $Y \in \mathcal{P}$ , the set  $Y \setminus F$  is  $2\varepsilon$ -restricted, since  $|Y \setminus F| \geq |Y|/2$ ; and the union of all these sets  $Y \setminus F$  (where  $Y \in \mathcal{P}$  is not  $\varepsilon$ -restricted) is  $V(G) \setminus Z$ . So altogether we have found a partition of  $V(G)$  into at most

$$|\mathcal{P}| + mt \leq 21\varepsilon^{-4} + \varepsilon^{-2}(9\varepsilon^{-2}) \leq 30\varepsilon^{-4}$$

$2\varepsilon$ -restricted sets. This proves 5.10. ■

## Acknowledgement

Thanks to Maria Chudnovsky and Sophie Spirkl for stimulating discussions on the topic of this paper.

## References

- [1] N. Alon, E. Fischer, M. Krivelevich and M. Szegedy, “Efficient testing of large graphs”, *Combinatorica*, **20** (2000), 451–476.
- [2] N. Alon, J. Pach and J. Solymosi, “Ramsey-type theorems with forbidden subgraphs”, *Combinatorica* **21** (2001), 155–170.
- [3] N. Alon and J. Fox, “Easily testable graph properties”, *Combinatorics, Probability and Computing*, **24** (2015), 646–657.
- [4] M. Chudnovsky, A. Scott, P. Seymour and S. Spirkl, “Strengthening Rödl’s theorem”, submitted for publication, [arXiv:2105.07370](https://arxiv.org/abs/2105.07370).
- [5] D. Conlon and J. Fox, “Graph removal lemmas”, *Surveys in Combinatorics 2013*, 1–49, London Math. Soc. Lecture Note Ser., **409**, Cambridge Univ. Press, Cambridge, 2013.



- [6] D. G. Corneil, H. Lerchs, and L. Stewart Burlingham, “Complement reducible graphs”, *Discrete Applied Mathematics*, **3** (1981), 163–174.
- [7] P. Erdős and A. Hajnal, “On spanned subgraphs of graphs”, *Graphentheorie und Ihre Anwendungen* (Oberhof, 1977).
- [8] P. Erdős and A. Hajnal, “Ramsey-type theorems”, *Discrete Applied Mathematics* **25** (1989), 37–52.
- [9] J. Fox and B. Sudakov, “Induced Ramsey-type theorems”, *Advances in Mathematics* **219** (2008), 1771–1800.
- [10] Z. Füredi, “Extremal hypergraphs and combinatorial geometry”, *Proceedings of the International Congress of Mathematicians, Vol. 1*, **2** (Zürich, 1994), 1343–1352, Birkhäuser, Basel, 1995.
- [11] T. Nguyen, A. Scott and P. Seymour, “Induced subgraph density. III. The pentagon and the bull”, in preparation.
- [12] V. Nikiforov, “Edge distribution of graphs with few copies of a given graph”, *Combin. Probab. Comput.* **15** (2006), 895–902.
- [13] V. Rödl, “On universality of graphs with uniformly distributed edges”, *Discrete Mathematics* **59** (1986), 125–134.

## Appendix

For the convenience of the interested reader, here is a proof of a claim in the last section.

**5.11** *There is a cograph that admits no vertex-partition into three  $1/3$ -restricted sets.*

**Proof.** Take four disjoint sets  $A, B, C, D$  (we will specify their sizes later). Make  $A, B, C$  stable sets complete to each other, and make  $D$  a clique anticomplete to  $A \cup B \cup C$ , forming a graph  $G$ .

(1) *Every  $1/3$ -restricted subset  $X$  of  $V(G)$  satisfies either*

- $X \subseteq D$ ; or
- $X$  is disjoint from two of  $A, B, C$ , and its intersection with the third has cardinality at least  $2|X \cap D| - 3$ ; or
- $X$  is disjoint from one of  $A, B, C$ , and its intersections with the other two and with  $D$  have cardinalities that differ by at most one; or
- $X \cap D = \emptyset$ , and  $|X \cap A|, |X \cap B|, |X \cap C|$  differ by at most two; or
- $|X| \leq 6$ .

Suppose first that  $\overline{G}$  has maximum degree at most  $|X|/3$ . Consequently not both  $X \cap (A \cup B \cup C)$  and  $X \cap D$  are nonempty, and so we may assume that  $X \subseteq A \cup B \cup C$ , since otherwise the first bullet holds. We may assume that  $|X \cap A| \geq |X|/3$ ; but each vertex in  $X \cap A$  has at most  $|X|/3$  non-neighbours in  $X$ , and so  $|X \cap A| \leq |X|/3 + 1$ ; and so the fourth bullet holds.

Now we assume that  $G$  has maximum degree at most  $|X|/3$ . Hence  $|X \cap D| \leq |X|/3 + 1$ , and so  $|X \cap (A \cup B \cup C)| \geq 2|X|/3 - 1$ . We may assume that  $X \cap A \neq \emptyset$ ; and so  $|X \cap (B \cup C)| \leq |X|/3$ , and consequently  $|X \cap A| \geq |X|/3 - 1$ . If also  $X \cap B, X \cap C$  are nonempty, then all three have cardinality at least  $|X|/3 - 1$  by the same argument, and so vertices in  $X \cap A$  have at least  $2|X|/3 - 2 > |X|/3$  neighbours in  $X$ ; so  $|X| \leq 6$  and the fifth bullet holds. So we may assume that  $X \cap C = \emptyset$ . If also  $X \cap B = \emptyset$  then  $|X \cap A| \geq 2|X|/3 - 1 \geq 2|X \cap D| - 3$  and the second bullet holds, so we assume that  $X \cap B \neq \emptyset$ . Consequently  $|X \cap A|, |X \cap B| \leq |X|/3$ , and so  $|X \cap D| \geq |X|/3$ , and the third bullet holds. This proves (1).

We say that a  $1/3$ -restricted set  $X$  has *type* 1–5 depending which bullet of (1) it satisfies. Now let  $n$  be a large integer, and let  $|A| = 2n$ ,  $|B| = 3n$ ,  $|C| = 4n$  and  $|D| = 5n$ . Suppose that  $V(G)$  can be partitioned into three  $1/3$ -restricted sets  $X, Y, Z$ . If  $X$  has type 2, 3, 4 or 5, then  $2|X \cap D| \leq |X \cap (A \cup B \cup C)| + 12$ ; so if none of  $X, Y, Z$  has type 1, then, summing these three inequalities, we deduce that  $2|D| \leq |A| + |B| + |C| + 36$ , a contradiction since  $n$  is large. So we may assume that  $Z$  has type 1. If say  $Y$  has type 1, 4 or 5, then  $|X \cap A|, |X \cap B|, |X \cap C|$  pairwise differ by at least  $n - 6$ , contrary to (1). So  $X, Y$  both have types 2 or 3. They cannot both have type 2 since their union includes  $A, B, C$ , so we assume that  $Y$  has type 3. If  $X$  has type 2, then the intersections of two of  $A, B, C$  with  $X \cup Y$  differ by at most one, a contradiction; so both  $X, Y$  have type 3. But then the sum of two of  $|A|, |B|, |C|$  should be equal to the third ( $\pm 2$ ), a contradiction. This proves 5.11. ■