Towards a polynomial form of Rödl's theorem

Jacob Fox[1]
Stanford University,
Stanford, CA 94305, USA

Tung Nguyen[2]
Princeton University,
Princeton, NJ 08544, USA

Alex Scott
Oxford University,
Oxford, UK

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

July 4, 2022; revised January 15, 2023

[1]Supported by a Packard Fellowship and by NSF grants DMS-1953990 and DMS-2154129.
[2]Supported by AFOSR grants A9550-19-1-0187 and FA9550-22-1-0234, and by NSF grants DMS-1800053 and DMS-2154169.
[3]Supported by AFOSR grants A9550-19-1-0187 and FA9550-22-1-0234, and by NSF grants DMS-1800053 and DMS-2154169.

Abstract

A theorem of Rödl says that for every graph H, and every $\varepsilon > 0$, there exists $\delta > 0$ such that if G is a graph that has no induced subgraph isomorphic to H, then there exists $X \subseteq V(G)$ with $|X| \geq \delta|G|$ such that one of $G[X], \overline{G}[X]$ has maximum degree at most $\varepsilon|X|$. But for fixed H, how does δ depends on ε? If the dependence is polynomial for a graph H, then H satisfies the Erdős-Hajnal conjecture; and it has been conjectured that the dependence is *always* polynomial. Here we prove this conjecture in an assortment of special cases, for instance when H can be obtained from subgraphs of the four-vertex path P_4 by repeated vertex-substitution.

We also show that when H is P_4 itself, the dependence is linear, and indeed for every cograph G and every $\varepsilon > 0$, 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 hence at most $\varepsilon|X|$). We sharpen this as much as we can.

We prove a similar "polynomial Rödl" result for the class of comparability graphs and for every hereditary class with the "strong Erdős-Hajnal property". Finally, we show that if H can be obtained by repeated vertex-substitution, starting from forests, then the class of graphs not containing H or its complement has the same "polynomial Rödl" property.

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.

A very useful theorem of V. Rödl [11] 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 at most $\varepsilon|X|(|X|-1)$ edges.*

How does δ depend on ε, for a given graph H? B. Sudakov and the first author [9] studied this, and proposed the conjecture (conjecture 7.1 in their paper) that the dependence is polynomial, or more exactly:

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

Our first result, proved in the next section, shows that 1.2 holds in a particularly nice form when $H = P_4$. We will show:

1.3 *For every $\varepsilon > 0$ 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 $\varepsilon|X|$).*

This is not tight, and with more work we can find sharper bounds, discussed later.

This is really a digression from our main objective, which is to prove 1.2 for as many graphs as we can. So, which other graphs H satisfy 1.2? Let us say a graph H is a *sniffle* if it satisfies 1.2, that is, there exists $c > 0$ such that for every ε with $0 < \varepsilon \leq 1/2$ and every H-free graph G, there exists $X \subseteq V(G)$ with $|X| \geq \varepsilon^c|G|$ such that one of $G[X], \overline{G}[X]$ has at most $\varepsilon|X|(|X|-1)$ edges. It is shown in [9] that every sniffle satisfies the Erdős-Hajnal conjecture [7, 8], and the latter is only known to hold for a rather paltry set of graphs H: the "bull" (which is the graph obtained from a triangle by adding two leaves adjacent to different vertices of the triangle), the cycle C_5 of length five, and graphs that can be obtained from induced subgraphs of these by repeated vertex-substitution. But can we at least show that these few graphs H are sniffles? No, not yet; indeed, we have no proof that either the bull or C_5 is a sniffle. But in this paper we make a start. We will show:

1.4 *If H can be built by repeated vertex-substitution starting from subgraphs of P_4 then H is a sniffle.*

Since we just used it twice, we had better 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 have not been able to show that the class of all sniffles is closed under vertex-substitution; but the class that satisfy a stronger property is closed under vertex-substitution, and we will show that P_4 and its subgraphs have this stronger property, and 1.4 follows from this. The "stronger property" is as follows. Let us say a *copy* of H in G is an isomorphism from H to an induced subgraph of G. We say a graph H is a *virus* if there exist $c, d > 0$ such that for every graph G and every ε with $0 < \varepsilon \leq 1/2$, either

- there exists $X \subseteq V(G)$ with $|X| \geq \varepsilon^c |G|$ such that one of $G[X], \overline{G}[X]$ has at most $\varepsilon|X|(|X|-1)$ edges; or
- there are at least $\varepsilon^d |G|^{|H|}$ copies of H in G.

(So either large areas of the graph are under tight lock-down, or there are a huge number of copies of the virus.) It seems to us that viruses, rather than sniffles, are the "right" concept to investigate. Perhaps every graph is a virus, but we are of course far from proving this. All viruses are sniffles, and we will show the following two results, which together imply 1.4:

1.5 *If H_1, H_2 are viruses and H is obtained by substituting H_2 for a vertex of H_1, then H is a virus.*

1.6 *Every graph on at most four vertices is a virus.*

A *proper hereditary class* is a class of graphs closed under isomorphism and under taking induced subgraphs, and not the class of all graphs. The conjecture 1.2 can be reformulated as:

1.7 Conjecture: *If $\mathcal{F}$ is a proper hereditary class, there exists $c > 0$ such that for every ε with $0 < \varepsilon \leq 1/2$ and every $G \in \mathcal{F}$, there exists $X \subseteq V(G)$ with $|X| \geq \varepsilon^c |G|$ such that one of $G[X], \overline{G}[X]$ has at most $\varepsilon|X|(|X|-1)$ edges.*

So far we have only considered hereditary classes defined by excluding one induced subgraph, but the question also makes sense for other hereditary classes. We will show that 1.7 holds for the class of comparability graphs and for classes of graphs with the "strong Erdős-Hajnal property", which we define later. There is also a "viral" version of the property when more than one graph is excluded, and we will show that for every forest T, the set $\{T, \overline{T}\}$ is viral. This will imply that:

1.8 *Let H be constructed by repeated vertex-substitution starting from forests. Then the class of all graphs not containing H or its complement satisfies 1.7.*

2 Dense and sparse sets in cographs

In this section we prove 1.3. We remark first that the result is almost tight. For instance, let $m \geq 2$ be an integer, take ε slightly smaller than $1/m$ and n a very large integer, and let G be the cograph consisting of the disjoint union of m complete graphs each with n vertices. One can show that if $X \subseteq V(G)$ and one of $G[X], \overline{G}[X]$ has maximum degree at most $\varepsilon|X|$, then $|X| \leq |G|/m$. Indeed, by the pigeonhole principle, any $X \subseteq V(G)$ contains a clique with at least $|X|/m$ vertices and hence $G[X]$ has maximum degree at least $|X|/m - 1$. If $|X| > |G|/m$, then G has a vertex not in this clique and the degree of this vertex in $\overline{G}[X]$ is at least $|X|/m$. One might hope that the result really is tight, but it is not, at least when $\varepsilon \geq 1/2$, as we shall see 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 with $0 \le x, y \le 1$, either:*

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

Proof. If $|G| \le 1$ the result is true, so we assume that $|G| \ge 2$ and the result holds for all cographs with fewer vertices, and for all choices of $x, y \in [0, 1]$. 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 are are given $x, y \in [0, 1]$. If $x = 0$ then the first bullet holds, taking $X = \emptyset$, so we assume that $x > 0$.

If for $i = 1, 2$ there exists $X_i \subseteq V(G_i)$ with $|X_i| \ge x|G_i|$ such that $G[X_i]$ has maximum degree at most $xy|G|$, then $|X_1 \cup X_2| \ge x|G|$ and $G[X_1 \cup X_2]$ has maximum degree at most $xy|G|$, and the first bullet of the theorem holds. We may therefore assume that:

(1) *There does not exist $X_1 \subseteq V(G_1)$ with $|X_1| \ge x|G_1|$ such that $G[X_1]$ has maximum degree at most $xy|G|$.*

Let $|G_1| = c|G|$. Since $|G_1| \ge x|G_1|$, there exists $X_1 \subseteq V(G_1)$ with $|X_1| = \lceil x|G_1| \rceil$, and hence $G[X_1]$ has maximum degree at most $|X_1| - 1 \le x|G_1| = xc|G|$; and therefore $xc|G| > xy|G|$ by (1), and so $y < c$ (because $x > 0$). Let $y' = y/c$; then $0 \le y' \le 1$. By (1), it follows from the inductive hypothesis, applied to G_1, x, y', that there exists $Y \subseteq V(G_1)$ with $|Y| \ge y'|G_1|$ such that $\overline{G_1}[Y]$ has maximum degree at most $xy'|G_1|$. Since $y'|G_1| = y|G|$ and $\overline{G_1}[Y] = \overline{G}[Y]$, the second bullet of the theorem holds. This proves 2.2. ∎

This has several useful consequences, and here is one, a strengthening of 1.3:

2.3 *Let G be a cograph, and let $0 \le \varepsilon \le 1$. 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 $|X| \cdot |Y| \ge \varepsilon|G|^2$.*

Proof. Let I be the set of $x \in [0, 1]$ such that for some $X \subseteq V(G)$, $|X| \ge x|G|$ 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 $Y \subseteq V(G)$, $x|Y| \ge \varepsilon|G|$ 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. ∎

Let us say G is *good* if for all x, y with $0 \le x, y \le 1$, either:

- there exists $X \subseteq V(G)$ with $|X| \ge x|G|$ such that $G[X]$ has maximum degree at most $xy|G|$; or
- there exists $Y \subseteq V(G)$ with $|Y| \ge y|G|$ 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 here are some remarks that we state without proof:

- 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 A tighter bound for cographs

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. The result 1.3 is neat, and it seemed plausible that it would be tight, but it is not. For $\varepsilon \in [0,1]$, let δ_ε be the supremum of all δ 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 $\varepsilon\delta|G|$. Thus 1.3 says that $\delta_\varepsilon \geq \varepsilon$, but we will show the following, which implies that $\delta_\varepsilon \geq 1/(2-\varepsilon) > \varepsilon$ when $1/2 \leq \varepsilon < 1$:

3.1 *For every non-null cograph G and every ε with $1/2 \leq \varepsilon < 1$, 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 $\varepsilon\delta|G|$, where $\delta = 1/(2-\varepsilon)$.*

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 $X_1, \ldots, X_k$ as follows. Suppose that $i \geq 1$ and we have defined $X_1, \ldots, X_{i-1}$, such that $V(G) \neq X_1 \cup \cdots \cup X_{i-1}$. Let $Y = V(G) \setminus (X_1 \cup \cdots \cup X_{i-1})$. 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 $V(G) \neq X_1 \cup \cdots \cup X_i$. 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 $Y = V(G) \setminus (X_1 \cup \cdots \cup X_i)$. 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 $A \cap B$ 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 $A \cap B$ has no neighbours in B and has at most $|A| - 1 \leq \delta|G|/2 \leq \varepsilon\delta|G|$ neighbours in A, and the same 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 $|X_{h+1} \cup \cdots \cup X_k| \leq d|G| + 1$. (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 $Y = X_{h+1} \cup \cdots \cup X_k$. Let I be the set of all $i \in \{1, \ldots, h\}$ 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 3.1. ∎

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

3.2 *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; 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 3.2. ∎

What can we say about the values of δ_ε through the remainder of the range of ε? As far as we know, though it seems unlikely, it may be that $\delta_\varepsilon = \varepsilon$ for all $\varepsilon < 1/2$. To try to settle this, we focussed on $\delta_{1/3}$, but all we found is an example that shows that $\delta_{1/3} \leq 3/8$, and more generally $\delta_\varepsilon \leq 1/(4 - 4\varepsilon)$ when $1/3 \leq \varepsilon < 2/5$, the following. Suppose that $1/3 \leq \varepsilon \leq 2/3$ and $\delta > 1/(4 - 4\varepsilon)$; then for sufficiently large integer n, we can choose four disjoint sets A, B, C, D with

$$\begin{aligned}
&|A| < \delta n - 100, \\
&|B| < (1 - \varepsilon)\delta n - 100, \\
&|C| < (1 - 2\varepsilon)\delta n - 100, \text{ and} \\
&|D| < (1 - \varepsilon)\delta n - 100,
\end{aligned}$$

and with union of cardinality n. Make B a clique, and A, C, D stable sets, and make C complete to D, and A complete to $B \cup C \cup D$. One can check (we omit the details) that there is no $X \subseteq V(G)$ with $|X| \geq \delta|G|$ such that one of $G[X], \overline{G}[X]$ has maximum degree at most $\varepsilon\delta|G|$.

4 Comparability graphs and other hereditary classes

A graph is a *comparability graph* if there is a partial order on the element set $V(G)$ such that for all distinct $u, v \in V(G)$, u, v are adjacent in G if and only if one of u, v is less than the other in the partial order. We will show that the class of comparability graphs satisfies 1.7, because of the following:

4.1 *If G is a comparability graph, then for every ε with $0 \le \varepsilon \le 1$, either*

- *there exists $X \subseteq V(G)$ with $|X| \ge \frac{\varepsilon^2}{16(1+\varepsilon)}|G|$ such that $\overline{G}[X]$ has at most $\varepsilon|X|(|X|-1)$ edges; or*
- *there exists $X \subseteq V(G)$ with $|X| \ge \frac{\varepsilon}{4(1+\varepsilon)}|G|$ such that $G[X]$ has at most $\varepsilon|X|^2$ edges.*

Proof. We may assume that $\varepsilon \ne 0$. Let G be a comparability graph and let $<$ be the corresponding partial order on $V(G)$. Let $t = \lceil \varepsilon^3|G|/(16(1+\varepsilon))\rceil$. Define another partial order $<^*$ on $V(G)$ where $a <^* b$ if there are at least t elements c in $V(G)$ such that $a < c < b$. By Mirsky's theorem (the dual of Dilworth's theorem), either there is a sequence

$$a_1 <^* a_2 <^* a_3 <^* \cdots$$

with at least $1/\varepsilon + 1$ terms, or we can partition $V(G)$ into fewer than $1/\varepsilon + 1$ sets that are antichains under $<^*$. In the former case, any t of the intermediate elements between a_i and a_{i+1} form a part of a complete $\lceil 1/\varepsilon \rceil$-partite subgraph of G with parts of size t, and which therefore has at least $t/\varepsilon \ge \varepsilon^2|G|/(16(1+\varepsilon))$ vertices and has edge density at least $1-\varepsilon$, completing this case. So we may assume that there exists $A \subseteq V(G)$ with $|A| \ge |G|/(1+1/\varepsilon)$ that is an antichain under the partial order $<^*$.

Next we obtain an upper bound on the number of triples $a < c < b$ with $a, b, c \in A$. For fixed a, b, there are at most $t-1$ choices of c, and so there are at most $(t-1)|A|^2/2$ such triples. Let $r = ((t-1)|A|)^{1/2}$, and let A_0 be the set of $c \in A$ such that there are at least r elements $a \in A$ with $a < c$ and at least r elements $b \in A$ with $b > c$. So $|A_0| \le r^{-2}(t-1)|A|^2/2 = |A|/2$. We can partition $A \setminus A_0$ into two subsets A_1 and A_2, where for every element in A_1 there are fewer than r elements in A below it, and for every element in A_2 there are fewer than r elements in A above it. Without loss of generality, $|A_1| \ge |A_2|$, so $|A_1| \ge |A|/4 \ge |G|/(4+4/\varepsilon)$.

Since $|G| \le (4+4/\varepsilon)|A_1|$ and $|A| \le 4|A_1|$, it follows that $|G|\cdot|A| \le 16(1+1/\varepsilon)|A_1|^2$. Consequently

$$r^2 = (t-1)|A| \le (\varepsilon^3/(16(1+\varepsilon)))|G|\cdot|A| \le (\varepsilon^3/(16(1+\varepsilon)))(16(1+1/\varepsilon))|A_1|^2 = \varepsilon^2|A_1|^2,$$

and hence $r \le \varepsilon|A_1|$. It follows that the number of edges with both ends in A_1 is at most $|A_1|r \le \varepsilon|A_1|^2$. Since $|A_1| \ge |G|/(4+4/\varepsilon)$, the second bullet of the theorem holds. This proves 4.1. ∎

Thus, we have proved 1.7 for cographs and for comparability graphs; but there are many other hereditary classes for which we could try to prove it. For instance, does it hold for the class of perfect graphs?

Here is another type of hereditary class that we can show satisfies 1.7. A *pure pair* in a graph G is a pair of disjoint subsets A, B of G such that either every vertex in A is adjacent to every vertex in B (that is, A is *complete* to B), or there are no edges between A, B (that is, A is *anticomplete* to

B). Let us say a hereditary class $\mathcal{F}$ has the *strong Erdős-Hajnal property* if there exists $\delta > 0$, such that every graph $G \in \mathcal{F}$ with $|G| \geq 2$ admits a pure pair A, B with $|A|, |B| \geq \delta|G|$. Such classes exist: the following two results were shown in [3, 4].

4.2 *Let $\mathcal{F}$ be a hereditary class of graphs such that some forest is not in $\mathcal{F}$, and the complement of some forest is not in $\mathcal{F}$. Then $\mathcal{F}$ has the strong Erdős-Hajnal property.*

4.3 *Let $\mathcal{F}$ be a hereditary class of graphs such that for some graphs H_1, H_2, there is no subdivision of H_1 that belongs to $\mathcal{F}$, and there is no subdivision of H_2 whose complement belongs to $\mathcal{F}$. Then $\mathcal{F}$ has the strong Erdős-Hajnal property.*

An argument of Sophie Spirkl (private communication) shows the following:

4.4 *Let $\mathcal{F}$ be a hereditary class with the strong Erdős-Hajnal property. Then $\mathcal{F}$ satisfies 1.7.*

Proof. Let δ be as in the definition of the strong Erdős-Hajnal property. Let $c = 5\log(1/\delta)$ (logarithms in this paper are always to base two), and let $0 < \varepsilon \leq 1/2$. We will show that for every $G \in \mathcal{F}$, there exists $X \subseteq V(G)$ with $|X| \geq \varepsilon^c|G|$ such that one of $G[X], \overline{G}[X]$ has at most $\varepsilon|X|(|X|-1)$ edges. Let $t = \lceil 2 + 2\log(1/\varepsilon)\rceil$. Suppose first that $|G| \leq \delta^{-t}$; then

$$\varepsilon^c|G| \leq \varepsilon^c\delta^{-t} \leq \varepsilon^c\delta^{-3-2\log(1/\varepsilon)},$$

and so

$$\log(\varepsilon^c|G|) \leq -(5\log(1/\delta))\log(1/\varepsilon) + (3 + 2\log(1/\varepsilon))\log(1/\delta) \leq 0$$

since $\varepsilon \leq 1/2$. Consequently $\varepsilon^c|G| \leq 1$ and the result holds trivially.

Thus we may assume that $|G| \geq \delta^{-t}$. For each integer s with $0 \leq s \leq t$, we define inductively a set $\mathcal{A}_s$ of pairwise disjoint subsets of $V(G)$, with the following properties:

- $|\mathcal{A}_s| = 2^s$;
- each member of $\mathcal{A}_s$ has cardinality $\lceil\delta^s|G|\rceil$;
- A, B is a pure pair for all distinct $A, B \in \mathcal{A}_s$; and
- the pattern graph of $\mathcal{A}_s$ is a cograph

where the *pattern graph* of $\mathcal{A}_s$ means the graph P_s with vertex set $\mathcal{A}_s$ where distinct $A, B \in \mathcal{A}_s$ are adjacent in P_s if and only if A is complete to B.

The inductive definition is as follows: let $\mathcal{A}_0 = \{V(G)\}$. Now let $1 \leq s \leq t$ and suppose that $\mathcal{A}_{s-1}$ is defined. Since $\delta^t|G| \geq 1$, it follows that $\delta^{s-1}|G| > 1$. Consequently $|A| \geq 2$ for each $A \in \mathcal{A}_{t-1}$, and therefore there is a pure pair $C_A, D_A \subseteq A$ with $|C_A|, |D_A| \geq \delta|A| \geq \delta^s|G|$; and hence we may may choose C_A, D_A with cardinality $\lceil\delta^s|G|\rceil$. Let $\mathcal{A}_s$ be the union of the sets $\{C_A, D_A\}$ over all $A \in \mathcal{A}_{s-1}$. It is easy to see that the four bullets above are satisfied.

Since the pattern graph P_t of $\mathcal{A}_t$ is a cograph, by 1.3 (with ε replaced by $\varepsilon/2$) there exists $|R| \geq (\varepsilon/2)|P_t| = \varepsilon 2^{t-1}$ such that one of $P_t[R], \overline{P_t}[R]$ has maximum degree at most $(\varepsilon/2)^2|P_t| = \varepsilon^2 2^{t-2}$, and from the symmetry we may assume the first. Let $X \subseteq V(G)$ be the union of the sets $A \in R$. Thus

$$|X| \geq \varepsilon 2^{t-1}\lceil\delta^t|G|\rceil \geq (\varepsilon/2)(2\delta)^t|G|.$$

Since

$$\begin{aligned}\log((2/\varepsilon)(2\delta)^{-t}) &= 1+\log(1/\varepsilon)+t(\log(1/\delta)-1)\\ &\le 2\log(1/\varepsilon)+(3+2\log(1/\varepsilon))(\log(1/\delta)-1)\\ &= (3+2\log(1/\varepsilon))\log(1/\delta)-3\\ &\le 5\log(1/\varepsilon)\log(1/\delta)=c\log(1/\varepsilon)\end{aligned}$$

it follows that

$$|X|\ge(\varepsilon/2)(2\delta)^t|G|\ge\varepsilon^c|G|.$$

Moreover, every vertex in X has degree in $G[X]$ at most $(1+2^{t-2}\varepsilon^2)\lceil\delta^t|G|\rceil-1$. Since $1+2^{t-2}\varepsilon^2\le 2^{t-1}\varepsilon^2$ (since $2^{t-2}\varepsilon^2\ge1$) and $\lceil\delta^t|G|\rceil\le 2\delta^t|G|$ (since $\delta^t|G|\ge1$) it follows that $G[X]$ has maximum degree at most $\varepsilon^2(2\delta)^t|G|-1\le 2\varepsilon(|X|-1)$, and so has at most $\varepsilon|X|(|X|-1)$ edges. This proves 4.4. ∎

5 Viruses and vertex-substitution

So far we have been considering classes of graphs in which there is no induced subgraph of a certain type, but it seems better in some respects just to assume that there are not "many" copies of the subgraph, and now we turn to that. First, let us prove 1.5, which we restate:

5.1 *If H_1,H_2 are viruses and H is obtained by substituting H_2 for a vertex of H_1, then H is a virus.*

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 virus, there exist c_i,d_i as in the definition of "virus". Let $c=\max(c_1,c_2+d_1+1)$, and $d=(|H_2|+1)(d_1+1)+d_2$. To show that H is a virus, we will show that:

(1) *For every graph G and all ε with $0<\varepsilon\le1/2$, either*

- *there exists $X\subseteq V(G)$ with $|X|\ge\varepsilon^c|G|$ such that one of $G[X],\overline{G}[X]$ has at most $\varepsilon|X|(|X|-1)$ edges; or*
- *there are at least $\varepsilon^d|G|^{|H|}$ copies of H in G.*

Since $\varepsilon^{c_1}|G|\ge\varepsilon^c|G|$, we may assume that there is no $X\subseteq V(G)$ with $|X|\ge\varepsilon^{c_1}|G|$ such that one of $G[X],\overline{G}[X]$ has at most $\varepsilon|X|(|X|-1)$ edges, since otherwise the first bullet of (1) holds. Consequently, from the choice of c_1,d_1, there are at least $\varepsilon^{d_1}|G|^{|H_1|}$ copies of H_1 in G. For each copy ϕ of $H_1\setminus\{v\}$ in G, let $N(\phi)$ be the set of all vertices $u\in V(G)$ such that extending ϕ by mapping v to u gives a copy of H_1 in G, and let $n(\phi)=|N(\phi)|$. Let Φ be the set of all copies of $H_1\setminus\{v\}$ in G; then

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

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

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

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 c_2, d_2, either there exists $X \subseteq N(\phi)$ with $|X| \geq \varepsilon^{c_2}|N(\phi)|$ such that one of $G[X], \overline{G}[X]$ has at most $\varepsilon|X|(|X|-1)$ edges, or there are $\varepsilon^{d_2}|N(\phi)|^{|H_2|}$ copies of H_2 in $G[N(\phi)]$. In the first case, since $\varepsilon^{c_2}|N(\phi)| \geq \varepsilon^{c_2}\varepsilon^{d_1+1}|G| \geq \varepsilon^{c}|G|$, 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

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

copies of H in G, and hence the second bullet of (1) holds. This proves (1), and hence shows that H is a virus, and proves 5.1. ∎

Let us deduce 1.6, which we restate:

5.2 *Every graph on at most four vertices is a virus.*

Proof. Every graph on at most two vertices is a virus 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 1.5 it suffices to prove that P_4 is a virus. 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 [2, 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 [1] proved a polynomial bound, that is, we can take $\delta = \varepsilon^d$, where d is an absolute constant.

To show that P_4 is a virus, let d be as above, and let $0 < \varepsilon \leq 1/2$. We will show that either

- there exists $X \subseteq V(G)$ with $|X| \geq \varepsilon|G|/2 \geq \varepsilon^2|G|$ such that one of $G[X], \overline{G}[X]$ has at most $\varepsilon|X|(|X|-1)$ edges; or
- there are at least $(\varepsilon/2)^{3d}|G|^4 \geq \varepsilon^{6d}|G|^4$ copies of P_4 in G.

Let us first dispose of a trivial case, when $\varepsilon|G|/2 \leq 3$. 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 $\varepsilon|G|/2 \leq 2$). So we may assume that $\varepsilon|G|/2 > 3$. From the choice of d (with ε replaced by $(\varepsilon/2)^3$), either G contains at least $(\varepsilon/2)^{3d}|G|^4$ 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 $(\varepsilon/2)^3|G|^2$ edges from G. In the latter case, by 1.3, there exists $X \subseteq V(G)$ with $|X| \geq \varepsilon|G|/2$ such that one of $G'[X], \overline{G'}[X]$ has maximum degree at most $(\varepsilon/2)^2|G|$. Then one of $G[X], \overline{G}[X]$ has at most

$$(\varepsilon/2)^3|G|^2 + (\varepsilon/2)^2|G||X|/2 \leq \frac{3}{4}\varepsilon|X|^2 \leq \varepsilon|X|(|X|-1)$$

edges (since $\varepsilon|G|/2 > 3$). This proves 5.2. ∎

6 Viral sets

There is a natural extension of the concept of a virus to finite sets of excluded induced subgraphs. Let us say a set $\mathcal{H}$ of graphs is *viral* if there exist $c, d > 0$ such that for every graph G and every ε with $0 < \varepsilon \leq 1/2$, either

- there exists $X \subseteq V(G)$ with $|X| \geq \varepsilon^c|G|$ such that one of $G[X], \overline{G}[X]$ has at most $\varepsilon|X|(|X|-1)$ edges; or
- for some $H \in \mathcal{H}$, there are at least $\varepsilon^d|G|^{|H|}$ copies of H in G.

(This makes sense if $\mathcal{H}$ is infinite, but we will use it only for finite sets.) We call c, d *viral exponents* for $\mathcal{H}$. We will have more to say about viral sets, but first, what about an analogue of 1.5? We introduced the concept of "viral" to be able to handle vertex-substitution, so we had better check that vertex-substitution still works for viral sets. It is easy to see that if $\mathcal{F}, \mathcal{H}$ are finite viral sets, and for each $F \in \mathcal{F}$ and $H \in \mathcal{H}$, $J(F, H)$ is a graph obtained by substituting H for some vertex of F, then the set of graphs $\{J(F, H) : F \in \mathcal{F}, H \in \mathcal{H}\}$ is viral; but one can be more delicate. We have:

6.1 *Let $\mathcal{H} = \{H_1, \ldots, H_k\}$ be viral, and for $1 \leq i \leq k$ let J_i be obtained by substituting a copy of H_i for some vertex of H_i. Then $\{J_1, \ldots, J_k\}$ is viral.*

This follows by repeated application of the following:

6.2 *Let $\mathcal{H} = \{H_1, \ldots, H_k\}$ be viral, and let J_1 be obtained by substituting a copy of H_1 for some vertex of H_1. Then $\{J_1, H_2, \ldots, H_k\}$ is viral.*

Proof. Let $c, d > 0$ be as in the definition of a viral set, for $\mathcal{H}$. Let $c' = c + d + 1$ and

$$d' = 2d + 1 + (d + 1)\max(|H_1|, \ldots, |H_k|) + \log(k).$$

Let G be a graph. We must show that:

(1) *One of the following holds:*

- *there exists $X \subseteq V(G)$ with $|X| \geq \varepsilon^{c'}|G|$ such that one of $G[X], \overline{G}[X]$ has at most $\varepsilon|X|(|X|-1)$ edges; or*
- *for some $i \in \{2, \ldots, k\}$, there are at least $\varepsilon^{d'}|G|^{|H_i|}$ copies of H_i in G; or*
- *there are at least $\varepsilon^{d'}|G|^{|J_1|}$ copies of J_1 in G.*

From the choice of c, d, and since $c' \geq c$, we may assume that for some $i \in \{2, \ldots, k\}$, there are at least $\varepsilon^d|G|^{|H_i|}$ copies of H_i in G; and since $d' \geq d$, we may assume that $i = 1$. Let J_1 be obtained from H_1 by substituting a copy of H_1 for some vertex v_1 of H_1; and for $2 \leq j \leq k$, let J_j be the graph obtained by substituting H_i for the vertex v_1 of H_1. Let Φ be the set of all copies of $H_1 \setminus \{v_1\}$ in G. For each $\phi \in \Phi$, let $N(\phi)$ be the set of all $v \in V(G)$ such that mapping v_1 to v extends ϕ to a copy of H_1; and let Ψ be the set of all $\phi \in \Phi$ such that $|n(\phi)i| \geq \varepsilon^{d+1}|G|$. As in the proof of 5.1, it follows that $|\Psi| \geq \varepsilon^{d+1}|G|^{|H_1|-1}$.

(2) *We may assume that there exists $j \in \{1, \dots, k\}$ such that there are at least $\varepsilon^{d'}|G|^{|J_j|}$ copies of J_j in G.*

Let $\phi \in \Psi$. From the choice of c, d, either there exists $X \subseteq N(\phi)$ with $|X| \geq \varepsilon^c|N(\phi)|$ such that one of $G[X], \overline{G}[X]$ has at most $\varepsilon|X|(|X|-1)$ edges, or for some $j \in \{1, \dots, k\}$ there are $\varepsilon^d|N(\phi)|^{|H_j|}$ copies of H_j in $G[N(\phi)]$. In the first case, since $\varepsilon^c|N(\phi)| \geq \varepsilon^c\varepsilon^{d+1}|G| \geq \varepsilon^{c'}|G|$, the theorem holds; so we may assume that for each $\phi \in \Psi$, there exists $j \in \{1, \dots, k\}$ such that ϕ extends to at least

$$\varepsilon^d|N(\phi)|^{|H_j|} \geq \varepsilon^d(\varepsilon^{d+1}|G|)^{|H_j|} = \varepsilon^{d+(d+1)|H_j|}|G|^{|H_j|}$$

copies of J_j.

There are at least $|\Psi|/k \geq \varepsilon^{d+1}|G|^{|H_1|-1}/k$ copies $\phi \in \Psi$ with the same value of j; and so there exists $j \in \{1, \dots, k\}$ such that there are at least

$$\varepsilon^{d+1}|G|^{|H_1|-1}\varepsilon^{d+(d+1)|H_j|}|G|^{|H_j|}/k = \varepsilon^{2d+1+(d+1)|H_j|}|G|^{|J_j|}/k \geq \varepsilon^{d'}|G|^{|J_j|},$$

copies of J_j in G (the last because $\varepsilon^{\log k} \leq (1/2)^{\log k} = 1/k$). This proves (2).

Now there are two cases, $j = 1$ and $j \geq 2$. If $j = 1$, then the third bullet of (1) holds. If $j > 1$, then since each copy of H_j in G only extends to at most $|G|^{|H_1|-1}$ copies of J_j, there are at least

$$\varepsilon^{d'}|G|^{|J_j|}/|G|^{|H_1|-1} = \varepsilon^{d'}|G|^{|H_j|}$$

copies of H_j in G and the second bullet of (1) holds. This proves (1), and hence proves 6.2 and 6.1. ∎

Here is another useful lemma for manipulating viral sets:

6.3 *Let $\mathcal{H}, \mathcal{H}'$ be sets of graphs, and suppose that each member of $\mathcal{H}$ has an induced subgraph isomorphic to a member of $\mathcal{H}'$. If $\mathcal{H}$ is viral then $\mathcal{H}'$ is viral.*

Proof. Let c, d be viral exponents for $\mathcal{H}$. We will show that c, d are also viral exponents for $\mathcal{H}'$. Let G be a graph and let $0 < \varepsilon \leq 1/2$. It follows that either

- there exists $X \subseteq V(G)$ with $|X| \geq \varepsilon^c|G|$ such that one of $G[X], \overline{G}[X]$ has at most $\varepsilon|X|(|X|-1)$ edges; or
- for some $H \in \mathcal{H}$, there are at least $\varepsilon^d|G|^{|H|}$ copies of H in G.

If the first holds, we are done, so we assume that the second holds. Choose $H' \in \mathcal{H}'$ isomorphic to an induced subgraph of H. Each copy of H in G is an extension of a copy of H' in G, and each copy of H' in G extends to at most $|G|^{|H|-|H'|}$ copies of H; so there are at least $\varepsilon^d|G|^{|H'|}$ copies of H' in G, and again we are done. This proves 6.3. ∎

7 Sparse pairs of sets

There is also a viral version of pure pairs and 4.4 that we found useful. Let us say that disjoint subsets A, B of the vertex set of a graph are *c-sparse* to each other if there are at most $c|A| \cdot |B|$ edges of G between A, B; and and *c-dense* to each other if they are c-sparse to each other in $\overline{G}$. Such pairs of sets will provide an analogue of pure pairs that work better with the viral version of 1.2. First, we need a lemma that allows us to tidy up a sparse pair.

7.1 *Let $0 < c, d < 1$, and let A, B be disjoint subsets of $V(G)$, both of cardinality at least $d|G|$, and c-sparse to each other. Then there exist $A' \subseteq A$ and $B' \subseteq B$, both of cardinality at least $d|G|/2$, such that each vertex of A' has at most $4cd|G|$ neighbours in B' and vice versa.*

Proof. If $d|G| \leq 1$, the result is clear (since $c < 1$), so we assume that $d|G| > 1$. By averaging over all subsets A_1 of A with cardinality $\lceil d|G| \rceil$ it follows that there exists $A_1 \subseteq A$ with $|A_1| = \lceil d|G| \rceil$ such that A_1, B are c-sparse to each other; and similarly there exists $B_1 \subseteq B$ with $|B_1| = \lceil d|G| \rceil$ such that A_1, B_1 are c-sparse to each other. Since there are at most $c|A_1| \cdot |B_1|$ edges between A_1, B_1, at most $|A_1|/2$ vertices in A_1 have at least $2c|B_1|$ neighbours in B_1, and so there exists $A_2 \subseteq A_1$ with $|A_2| \geq |A_1|/2$ such that every vertex in A_2 has fewer than $2c|B_1|$ neighbours in B_1. Since $|A_1|/2 \geq d|G|/2$, there exists $A' \subseteq A_2$ with $|A'| = \lceil d|G|/2 \rceil$ such that every vertex in A' has fewer than $2c|B_1|$ neighbours in B_1. Similarly there exists $B' \subseteq B_1$ with $|B'| = \lceil d|G|/2 \rceil$ such that every vertex in B' has fewer than $2c|A_1|$ neighbours in A_1. Now $|B_1| \leq 2d|G|$ since $|B_1| = \lceil d|G| \rceil$ and $d|G| \geq 1$. Hence every vertex in A' has fewer than $4cd|G|$ neighbours in B_1; and similarly every vertex in B' has fewer than $4cd|G|$ neighbours in A_1. But then A', B' satisfy the theorem. This proves 7.1. ∎

One way to prove that a set of graphs is viral is to apply the following, a relative of 4.4:

7.2 *Let $\mathcal{H}$ be a finite set of graphs. Suppose that there exist $a, b > 0$ such that for every graph G with $|G| \geq 2$ and every ε with $0 \leq \varepsilon \leq 1/2$, either:*

- *there exist disjoint $A, B \subseteq V(G)$, both with cardinality at least $a|G|$, that are ε-sparse or ε-dense to each other; or*
- *for some $H \in \mathcal{H}$, there are at least $\varepsilon^b |G|^{|H|}$ copies of H in G.*

Then $\mathcal{H}$ is viral.

Proof. We may assume that $a \leq 1/4$. Choose p such that $a = (1/2)^p$. Let $c = 5p + 2$ and let d be the maximum of $b(5p+6) + 5(p+1)|H|$ over all $H \in \mathcal{H}$. Let G be a graph and let $0 \leq \varepsilon \leq 1/2$. We will show that:

(1) *One of the following holds:*

- *there exists $X \subseteq V(G)$ with $|X| \geq \varepsilon^c |G|$ such that one of $G[X], \overline{G}[X]$ has at most $\varepsilon|X|(|X|-1)$ edges; or*
- *for some $H \in \mathcal{H}$, there are at least $\varepsilon^d |G|^{|H|}$ copies of H in G.*

Choose t minimum such that $(1/2)^t \le \varepsilon^4$. So $(1/2)^t \ge \varepsilon^4/2 \ge \varepsilon^5$. Since $c \ge 5p+2$, and $2^t \le \varepsilon^{-5}$, it follows that

$$\varepsilon^c \le \varepsilon^{5(p+1)} \le (1/2)^{(p+1)t}.$$

Moreover, we may assume that $\varepsilon^c|G| > 2$, since otherwise the first bullet above holds; so $(1/2)^{(p+1)t}|G| > 2$, and therefore $(a/2)^t|G| > 2$.

Define $\varepsilon_1 = \varepsilon^{5p+6}$. For each integer s with $0 \le s \le t$, we define inductively a set $\mathcal{A}_s$ of 2^s pairwise disjoint subsets of $V(G)$, each of cardinality $\lceil (a/2)^s|G|\rceil$, and a cograph P_s with vertex set $\mathcal{A}_s$, with the following requirement, where f_s is defined by

$$f_s = 4\varepsilon_1 \left(1 + (a/2) + (a/2)^2 + \cdots + (a/2)^{s-1}\right)|G|.$$

Let us say a pair u, v of vertices is *s-contrary* if there exist distinct $A, B \in \mathcal{A}_s$ with $u \in A$ and $v \in B$, and u, v are adjacent in G if and only if A, B are A, B are nonadjacent in P_s. Then for each $A \in \mathcal{A}_s$ and each $u \in A$, we require that there are at most f_s vertices v such that u, v are s-contrary.

The inductive definition is as follows. Define $\mathcal{A}_0 = \{V(G)\}$, and let P_0 be the graph with vertex set $\mathcal{A}_0$ (and therefore with only one vertex). Now let $1 \le s \le t$, and suppose that $\mathcal{A}_{s-1}$ and P_{s-1} are defined. Let $A \in \mathcal{A}_{s-1}$. Thus $|A| \ge (a/2)^t|G| > 2$, and so, from the hypothesis applied to $G[A]$, with ε replaced by ε_1, either

- there exist disjoint $C, D \subseteq A$, both with cardinality at least $a|A|$, that are ε_1-sparse or ε_1-dense to each other; or
- for some $H \in \mathcal{H}$, there are at least $\varepsilon_1^b|A|^{|H|}$ copies of H in $G[A]$.

If the second bullet here holds, then the second bullet of (1) holds, since

$$\varepsilon_1^b|A|^{|H|} \ge \varepsilon_1^b(a/2)^{(s-1)|H|}|G|^{|H|} \ge \varepsilon^d|G|^{|H|},$$

the last because

$$\varepsilon_1^b(a/2)^{t|H|} = \varepsilon^{b(5p+6)}(1/2)^{(p+1)t|H|} \ge \varepsilon^{b(5p+6)}(\varepsilon^5)^{(p+1)|H|} \ge \varepsilon^{b(5p+6)+5(p+1)|H|} \ge \varepsilon^d.$$

Thus we may assume that there exist disjoint $C, D \subseteq A$, both with cardinality at least $a|A|$, that are ε_1-sparse or ε_1-dense to each other. Since $|A| \ge (a/2)^{s-1}|G|$, 7.1 (applied in the complement if C, D are ε_1-dense) implies that there exist $C_A \subseteq C$ and $D_A \subseteq D$, both with cardinality at least $(a/2)^s|G|$, such that either every vertex of C_A has at most $4\varepsilon_1(a/2)^{s-1}|G|$ neighbours in B_A and vice versa, or the same holds in $\overline{G}$. By replacing C_A, D_A by subsets if necessary, we may assume that $|C_A|, |D_A| = \lceil (a/2)^s|G|\rceil$. Let $\mathcal{H}_s$ be the union of the sets $\{C_A, D_A\}$ over all $A \in \mathcal{A}_{s-1}$, and define P_s in the natural way (that is, C_A, D_A are adjacent in P_s if and only if every vertex in C_A has at most $4\varepsilon^c(a/2)^{s-1}|G|$ neighbours in D_A and vice versa, and for distinct $A, A' \in \mathcal{A}_{s-1}$, if A, A' are adjacent in P_{s-1} then C_A, D_A are both adjacent to both of $C_{A'}, D_{A'}$ in P_s, and otherwise the four pairs are nonadjacent in P_s.)

To complete the inductive definition, we must check that for each $A \in \mathcal{A}_{s-1}$, and each $u \in C_A \cap D_A$, there are at most f_s vertices v such that u, v are s-contrary. To show this we may assume from the symmetry that $u \in C_A$ and C_A, D_A are nonadjacent in P_s. Now u, v are s-contrary if and only if either u, v are $(s-1)$-contrary, or $v \in D_A$ and u, v are adjacent in G. There are only f_{s-1} vertices v such that u, v are $(s-1)$-contrary, and there are only $4\varepsilon_1(a/2)^{s-1}|G|$ vertices $v \in D_A$ such that u, v are adjacent in G. Since $f_s = f_{s-1} + 4k\varepsilon_1(a/2)^{s-1}|G|$, this verifies the conditions of the inductive definition and so completes the definition.

From 1.3 applied to P_t (with ε replaced by $\varepsilon/2$), there exists $|R| \geq \varepsilon/2|P_t|$ such that one of $P_t[R], \overline{P_t}[R]$ has maximum degree at most $(\varepsilon/2)^2|P_t| = (\varepsilon/2)^2 2^t$, and from the symmetry we may assume the first. Let $X \subseteq V(G)$ be the union of the sets $A \in R$. Thus

$$|X| \geq (\varepsilon/2)2^t(a/2)^t|G| = (\varepsilon/2)a^t|G|.$$

Since $(1/2)^t \geq \varepsilon^2/8 \geq \varepsilon^5$, it follows that

$$(\varepsilon/2)a^t \geq \varepsilon^2(1/2)^{pt} \geq \varepsilon^2(\varepsilon^5)^p = \varepsilon^{5p+2} = \varepsilon^c,$$

and so $|X| \geq \varepsilon^c|G|$.

Moreover, let $m = \lceil (a/2)^t|G| \rceil$. Since

$$f_t \leq 4\varepsilon_1|G|/(1-a/2) = 4\varepsilon^{5p+6}|G|/(1-a/2) \leq \varepsilon^{5p+3}|G|$$

(because $\varepsilon \leq 1/2$ and $1 - a/2 \geq 1/2$), it follows that every vertex in X has degree in $G[X]$ at most

$$(1 + (\varepsilon/2)^2 2^t)m + f_t - 1 \leq m + (\varepsilon/2)^2 2^t m + \varepsilon^{5p+3}|G| - 1.$$

We will show that the last expression in the above is at most $2\varepsilon(|X| - 1)$, and since $|X| \geq (\varepsilon/2)2^t m$, it suffices to show that

$$m + (\varepsilon/2)^2 2^t m + \varepsilon^{5p+3}|G| - 1 \leq \varepsilon^2 2^t m - 2\varepsilon.$$

To show this, observe that

$$m \leq \varepsilon^4 2^t m \leq \varepsilon^2 2^t m/4$$
$$(\varepsilon/2)^2 2^t m = \varepsilon^2 2^t m/4$$
$$\varepsilon 5p + 3|G| \leq \varepsilon^2(1/2)^{tp}|G|/2 \leq \varepsilon^2 2^t(a/2)^t|G|/2 \leq \varepsilon^2 2^t m/2$$
$$-1 \leq -2\varepsilon$$

and the claim follows by adding. This proves 7.2. ∎

Here is a slight strengthening of the previous result:

7.3 *Let $\mathcal{H}$ be a finite set of graphs. Suppose that there exist $a, b, \varepsilon_0 > 0$ such that for every graph G with $|G| \geq 2$ and every ε with $0 \leq \varepsilon \leq \varepsilon_0$, either:*

- *there exist disjoint $A, B \subseteq V(G)$, both with cardinality at least $a|G|$, that are ε-sparse or ε-dense to each other; or*
- *for some $H \in \mathcal{H}$, there are at least $\varepsilon^b|G|^{|H|}$ copies of H in G.*

Then $\mathcal{H}$ is viral.

Proof. Let a, b, ε_0 be as in the theorem. If $\varepsilon_0 \geq 1/2$ then the result follows from 7.2, so we assume that $\varepsilon_0 < 1/2$. Choose c such that $(1/2)^c = \varepsilon_0$ (thus, $c \geq 1$), and let $b' = bc$. We claim that a, b' satisfy the conditions of 7.2 (with b replaced by b'). Let G be a graph with $|G| \geq 2$ and let $0 \leq \varepsilon' \leq 1/2$. Let $\varepsilon = (\varepsilon')^c$. Thus $\varepsilon \leq \min(\varepsilon', \varepsilon_0)$. By the hypothesis, either

- there exist disjoint $A, B \subseteq V(G)$, both with cardinality at least $a|G|$, that are ε-sparse or ε-dense to each other; or
- for some $H \in \mathcal{H}$, there are at least $\varepsilon^b|G|^{|H|}$ copies of H in G

In the first case, since $\varepsilon \le \varepsilon'$, the first bullet of 7.2 is (with ε replaced by ε'); and in the second case, since

$$\varepsilon^b = (\varepsilon')^{bc} = (\varepsilon')^{b'},$$

the second bullet of 7.2 is satisfied. This proves 7.3. ∎

8 Forests and forest complements

If $\mathcal{H}$ is a set of graphs, we say that G is *$\mathcal{H}$-free* if G is H-free for each $H \in \mathcal{H}$. Because of the result of 4.2, it follows from 4.4 that for every two forests T_1, T_2, the hereditary class of graphs that are $\{T_1, \overline{T_2}\}$-free satisfies 1.7. But in this section we will use 7.2 to show the stronger result that $\{T_1, \overline{T_2}\}$ is viral.

We need the following (we recall that logarithms are to base two):

8.1 *Let $0 < \alpha \le 1/8$ and $0 \le \varepsilon \le 1/2$, with $\varepsilon \log(1/\varepsilon) \le \frac{1}{2}\alpha$. Let $m \ge \varepsilon^{-2}$ be an integer. Let G be a graph with $|G| \ge 4m$, and suppose that there do not exist disjoint $P, Q \subseteq V(G)$ with $|P|, |Q| \ge \alpha|G|$ such that every vertex in P has at most $\alpha\varepsilon|G|$ neighbours in Q. Let $A, B \subseteq V(G)$ be disjoint sets both of cardinality m, chosen uniformly at random. Then the probability that A, B are anticomplete is at most $(4\alpha)^{m/2}$.*

Proof. Let $\tau = (\alpha\varepsilon)^{-1}$. If $X \subseteq V(G)$, let $N[X]$ denote the union of X and the set of vertices of G that have a neighbour in X. Let $|G| = n$.

Let us choose $a_1, \dots, a_m \in V(G)$ one by one, distinct and otherwise uniformly at random, and set $A = \{a_1, \dots, a_m\}$. For $1 \le i \le m$ define

$$D_i = N[\{a_1, \dots, a_i\}] \setminus N[\{a_1, \dots, a_{i-1}\}],$$

and $Y_i = V(G) \setminus N[\{a_1, \dots, a_i\}]$. Let I be the set of all $i \in \{1, \dots, m\}$ such that $|D_i| \le n/\tau$ and $|Y_{i-1}| \ge 2\lceil \alpha n \rceil$.

(1) *For $1 \le i \le m$, and for each choice of $a_1, \dots, a_{i-1}$, the probablity that $i \in I$ is at most α.*

If $|Y_{i-1}| < 2\lceil \alpha n \rceil$, then the probability that $i \in I$ is zero, so we assume that $|Y_{i-1}| \ge 2\lceil \alpha n \rceil$. Let R be the set of all vertices in $V(G)$ with at most $\alpha\varepsilon n$ neighbours in Y_{i-1}. If $|R| \ge \alpha n$, we may choose $P \subseteq R$ with $|P| = \lceil \alpha n \rceil$ and $Q \subseteq Y_{i-1} \setminus P$ with $|Q| = \lceil \alpha n \rceil$, and then every vertex in P has at most $\alpha\varepsilon n$ neighbours in Q, contrary to the hypothesis. Thus $|R| < \alpha n$, and so the probability that $a_i \in R$ is at most α. But $i \in I$ only if $a_i \in R$, and so for $1 \le i \le m$, the probability that $i \in I$ is at most α. This proves (1).

(2) *If $m - |I| > \tau$ then $|V(G) \setminus N[A]| < 2\lceil \alpha n \rceil$.*

Since the sets D_i ($i \in \{1, \dots, m\} \setminus I$) are pairwise disjoint, not all of them have cardinality at least n/τ. Thus there exists $i \in \{1, \dots, m\} \setminus I$ such that $|D_i| < n/\tau$, and consequently $|Y_{i-1}| < 2\lceil \alpha n \rceil$ since $i \notin I$. Since $Y_{i-1} \supseteq Y_m = V(G) \setminus N[A]$, it follows that $|V(G) \setminus N[A]| < 2\lceil \alpha n \rceil$. This proves (2).

(3) *The probability that* $|V(G) \setminus N[A]| \geq 2\lceil \alpha n \rceil$ *is at most* $m^\tau \alpha^{m-\tau}$.

By (2), the probability that $|V(G) \setminus N[A]| \geq 2\lceil \alpha n \rceil$ is at most the probability that $m - |I| \leq \tau$. Let J be a subset of $\{1, \dots, m\}$. The probability that $J \subseteq I$ is the product, over all $j \in J$, of the conditional probability that $j \in I$ given that I contains all members of J less than j; and by (1), this conditional probability is at most α, for each $j \in J$. Hence the probability that $J \subseteq I$ is at most $\alpha^{|J|}$. There are at most m^τ subsets J of $\{1, \dots, m\}$ with $m - |J| \leq \tau$ (since $m \geq 2$ and $\tau \geq 2$), and for each such J, the probability that $J \subseteq I$ is at most $\alpha^{|J|} \leq \alpha^{m-\tau}$; and so the probability that I includes some J with $m - |J| \leq \tau$ (that is, the probability that $m - |I| \leq \tau$) is at most $m^\tau \alpha^{m-\tau}$. This proves (3).

(4) *Let* A, B *be disjoint subsets of* $V(G)$, *both of cardinality* m, *chosen uniformly at random. Then the probability that* A, B *are anticomplete is at most*

$$m^\tau \alpha^{m-\tau} + (2\lceil \alpha n \rceil / n)^m.$$

Let us first choose A, and then choose B disjoint from A. By (3), the probability that A, B are anticomplete and $|V(G) \setminus N[A]| \geq 2\lceil \alpha n \rceil$ is at most $m^\tau \alpha^{m-\tau}$. But the probability that A, B are anticomplete and $|V(G) \setminus N[A]| < 2\lceil \alpha n \rceil$ is at most $(2\lceil \alpha n \rceil/(n - m))^m$, since for this to happen, every vertex of B must belong to $V(G) \setminus N[A]$. Hence the probability that A, B are anticomplete is at most the sum of these, that is,

$$m^\tau \alpha^{m-\tau} + (2\lceil \alpha n \rceil/(n - m))^m.$$

This proves (4).

(5) $m^\tau \alpha^{m-\tau} \leq \frac{1}{2}(4\alpha)^{m/2}$.

From the hypothesis, $\varepsilon \log(1/\varepsilon)) \leq \frac{1}{2}\alpha$, and $\alpha \leq 1/8$, and $\varepsilon \leq 1/2$. Consequently, $\varepsilon \leq \varepsilon \log(1/\varepsilon)) \leq 1/16$, so $\log(1/\varepsilon) \geq 4$. Hence $8\varepsilon \leq \alpha$. It follows that

$$m - \tau = m - (\alpha\varepsilon)^{-1} \geq m - (8\varepsilon^2)^{-1} \geq 7m/8 \geq m/2 + 1,$$

since $m \geq \varepsilon^{-2} \geq 3$. Consequently

$$m^\tau \alpha^{m-\tau} \leq m^\tau \alpha^{m/2+1} \leq \frac{1}{2} m^\tau \alpha^{m/2}.$$

Thus, to prove the claim, it suffices to show that

$$\frac{1}{2} m^\tau \alpha^{m/2} \leq \frac{1}{2}(4\alpha)^{m/2},$$

that is, $m^\tau \le 4^{m/2}$, or equivalently $\tau \log m \le m$. But by hypothesis, $m \ge \varepsilon^{-2}$ and $\varepsilon \log(1/\varepsilon)) \le \frac{1}{2}\alpha$, so

$$\tau = (\alpha\varepsilon)^{-1} \le \varepsilon^{-2}/\log(\varepsilon^{-2}) \le m/\log m.$$

This proves (5).

(6) $(2\lceil \alpha n\rceil/(n-m))^m \le \frac{1}{2}(4\alpha)^{m/2}$.

Since by hypothesis $\alpha n \ge 2$ (and indeed $\alpha n \ge 512$, because $\alpha \ge 8\varepsilon$ and $n \ge 4m \ge 4\varepsilon^{-2}$, and $\varepsilon \le 1/16$), it follows that $\lceil \alpha n\rceil \le 3\alpha n/2$; and so $2\lceil \alpha n\rceil/(n-m) \le 4\alpha$ since $m \le n/4$. Thus it suffices to show that $(4\alpha)^m \le \frac{1}{2}(4\alpha)^{m/2}$. But this is true since $m \ge m/2+1$ and $4\alpha \le 1/2$. This proves (6).

From (3)–(6), this proves 8.1. ∎

We deduce the main result of this section:

8.2 *Let H_1, H_2 be forests; then $\{H_1, \overline{H_2}\}$ is viral.*

Proof. By replacing both H_1, H_2 by their disjoint union, we may assume by 6.3 that $H_1 = H_2 = H$ say. By 4.2, there exists $c > 0$ such that every $\{H, \overline{H}\}$-free graph G with $|G| \ge 2$ has a pure pair (A, B) with $|A|, |B| \ge c|G|$. Choose $c \le 1/2$. Let $a = c^4/2^{12}$; choose ε_0 with $0 < \varepsilon_0 \le 1/2$ such that $\varepsilon\log(1/\varepsilon) \le a/2$ for all ε with $0 < \varepsilon \le \varepsilon_0$; and choose b such that $\varepsilon_0^{2-b/|H|} \ge (4/c)(32/c)^{1/|H|}$. Let $m = \lceil \varepsilon^{-2}\rceil$. Thus $m \le 2\varepsilon^{-2}$. To show that $\{H, \overline{H}\}$ is viral, it suffices by 7.2 to show that:

(1) *For every graph G with $|G| \ge 2$ and every ε with $0 < \varepsilon \le \varepsilon_0$, either:*

- *there exist disjoint $P, Q \subseteq V(G)$, both with cardinality at least $a|G|$, that are ε-sparse or ε-dense to each other; or*
- *for some $J \in \{H, \overline{H}\}$, there are at least $\varepsilon^b|G|^{|H|}$ copies of J in G.*

Thus, let G be a graph with $|G| \ge 2$, and let $0 < \varepsilon \le \varepsilon_0$. We may assume that the first bullet above is false. Consequently, there do not exist disjoint $P, Q \subseteq V(G)$ with $|P|, |Q| \ge a|G|$ such that every vertex in P has at most $a\varepsilon|G|$ neighbours in Q. Since $a \le c$, the choice of c implies that G is not $\{H, \overline{H}\}$-free; and so we may assume that $\varepsilon^b|G|^{|H|} > 1$, for otherwise the second bullet of (1) holds. It follows that

$$|G| > \varepsilon^{-b/|H|} \ge (4/c)\varepsilon^{-2} \ge 2m/c \ge 4m.$$

Let $A, B \subseteq V(G)$ be disjoint sets both of cardinality m, chosen uniformly at random. Since $\varepsilon\log(1/\varepsilon) \le \varepsilon_0\log(1/\varepsilon_0) \le a/2$, 8.1 implies that the probability that A is anticomplete to B is at most $(4a)^{m/2}$. Thus, writing $n = |G|$, the number of anticomplete pairs (A, B) of subsets of $V(G)$ of cardinality m is at most

$$(4a)^{m/2}\binom{n}{m}\binom{n-m}{m}.$$

Let $M = \lceil 2m/c\rceil$. Thus $M \le 4m/c$ since $2m/c \ge 1$, and $n \ge M$ since we already saw that $|G| \ge 2m/c$.

It follows that the number of triples (A, B, C) of subsets of $V(G)$, such that A, B are disjoint and anticomplete, and both have size m, and C includes $A \cup B$, and $|C| = M$, is at most

$$(4a)^{m/2}\binom{n}{m}\binom{n-m}{m}\binom{n-2m}{M-2m}.$$

Let $\mathcal{C}$ be the set of all $C \subseteq V(G)$ with cardinality M, such that there exist $A', B' \subseteq C$ that are disjoint and anticomplete and both have cardinality $2m$. If $C \in \mathcal{C}$ then there are at least $\binom{2m}{m}^2$ anticomplete pairs (A, B) both of cardinality m and included in C; because the corresponding A' and B' both have $\binom{2m}{m}$ subsets of cardinality m. Thus

$$\begin{aligned}\frac{|\mathcal{C}|}{\binom{n}{M}} &\le (4a)^{m/2}\frac{\binom{n}{m}\binom{n-m}{m}\binom{n-2m}{M-2m}}{\binom{2m}{m}^2\binom{n}{M}} = (4a)^{m/2}\left(\frac{m!}{(2m)!}\right)^2\frac{M!}{(M-2m)!}\\ &\le (4a)^{m/2}\left(\frac{M}{m}\right)^{2m} \le (4a)^{m/2}\left(\frac{4}{c}\right)^{2m} = (2^{10}a/c^4)^{m/2} \le \frac{1}{4},\end{aligned}$$

since $\frac{m!}{(2m)!} \le \frac{1}{m^m}$, and $\frac{M!}{(M-2m)!} \le M^{2m}$, and $\frac{M}{m} \le \frac{4}{c}$, and $2^{10}a/c^4 \le \frac{1}{4}$ from the choice of a.

Let $C \subseteq V(G)$ with $|C| = M$. From 4.2, either

- there exist $A', B' \subseteq C$, disjoint and anticomplete, and both with cardinality $2m$; or
- there exist $A', B' \subseteq C$, disjoint and complete, and both with cardinality $2m$; or
- $G[C]$ contains H; or
- $G[C]$ contains $\overline{H}$.

We have just seen that, if C is chosen uniformly at random over all subsets of cardinality M, then the probability that $C \in \mathcal{C}$, that is, the first bullet holds, is at most $1/4$; and similarly the probability that the second holds is at most $1/4$. Hence either the third or fourth bullet holds with probability at least $1/4$, and from the symmetry we may assume the third holds with probability at least $1/4$. Since each copy of H in G appears in only $\binom{n-|H|}{M-|H|}$ subsets of cardinality M, it follows that there are at least

$$\frac{\frac{1}{4}\binom{n}{M}}{\binom{n-|H|}{M-|H|}} = \frac{1}{4}\frac{n}{M}\frac{n-1}{M-1}\cdots\frac{n-|H|+1}{m-|H|+1} \ge \frac{1}{4}\left(\frac{n}{M}\right)^{|H|}$$

copies of H in G. Since $M \le 4m/c$ and $m \le 2/\varepsilon^2$, it follows that this number is at least

$$\frac{1}{4}\left(\frac{c\varepsilon^2}{8}\right)^{|H|} n^{|H|} \ge \varepsilon^b|G|^{|H|},$$

and hence (1) holds. This proves 8.2. ∎

Finally, we observe that 1.8 follows from 6.1 and 8.2. Indeed, let H be constructed by repeated vertex-substitution starting from forests. From 6.1 and 8.2 it follows that $\{H, \overline{H}\}$ is viral, which implies 1.8.

Acknowledgement

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

References

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

[2] N. Alon, E. Fischer, M. Krivelevich and M. Szegedy, "Efficient testing of large graphs", *Combinatorica*, **20** (2000), 451–476.

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

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

[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 Csongress of Mathematicians, Vol. 1*, **2** (Zürich, 1994), 1343–1352, Birkhäuser, Basel, 1995.

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