Pure pairs. V. Excluding some long subdivision

Alex Scott[1]
Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK

Paul Seymour[2]
Princeton University, Princeton, NJ 08544

Sophie Spirkl[3]
Princeton University, Princeton, NJ 08544

December 13, 2019; revised January 24, 2023

[1]Research supported by EPSRC grant EP/V007327/1.
[2]Supported by AFOSR grants A9550-19-1-0187 and FA9550-22-1-0234, and NSF grants DMS-1800053 and DMS-2154169.
[3]This material is based upon work supported by the National Science Foundation under Award No. DMS-1802201. We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) (funding reference number RGPIN-2020-03912). Cette recherche a été finance par le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) (numéro de référence RGPIN-2020-03912).

Abstract

A "pure pair" in a graph G is a pair A, B of disjoint subsets of $V(G)$ such that A is complete or anticomplete to B. Jacob Fox showed that for all $\varepsilon > 0$, there is a comparability graph G with n vertices, where n is large, in which there is no pure pair A, B with $|A|, |B| \geq \varepsilon n$. He also proved that for all $c > 0$ there exists $\varepsilon > 0$ such that for every comparability graph G with $n > 1$ vertices, there is a pure pair A, B with $|A|, |B| \geq \varepsilon n^{1-c}$; and conjectured that the same holds for every perfect graph G. We prove this conjecture and strengthen it in several ways.

In particular, we show that for all $c > 0$, and all $\ell_1, \ell_2 \geq 4/c + 9$, there exists $\varepsilon > 0$ such that, if G is an $(n > 1)$-vertex graph with no hole of length exactly ℓ_1 and no antihole of length exactly ℓ_2, then there is a pure pair A, B in G with $|A| \geq \varepsilon n$ and $|B| \geq \varepsilon n^{1-c}$. This is further strengthened, replacing excluding a hole by excluding some "long" subdivision of a general graph.

1 Introduction

Graphs in this paper are finite, and without loops or parallel edges. Let $A, B \subseteq V(G)$ be disjoint. We say that A is *complete* to B, or A, B are *complete*, if every vertex in A is adjacent to every vertex in B, and similarly A, B are *anticomplete* if no vertex in A has a neighbour in B. A *pure pair* in G is a pair A, B of disjoint subsets of $V(G)$ such that A, B are complete or anticomplete, and $|G|$ denotes the number of vertices of a graph G.

Let $\mathcal{H}$ be a set of graphs: we say G is *$\mathcal{H}$-free* if no induced subgraph of G is isomorphic to a member of $\mathcal{H}$. For some choices of $\mathcal{H}$, every $\mathcal{H}$-free graph admits a pure pair A, B with both $|A|, |B|$ large in terms of $|G|$. Pure pairs with both $|A|, |B|$ linear in $|G|$ are particularly of interest because of connections with the Erdős-Hajnal conjecture [4, 5], and the following was shown in [2]:

1.1 *Let $\mathcal{H}$ be a finite set of graphs.*

- *If $\mathcal{H}$ contains a forest and the complement of a forest, then there exists $\varepsilon > 0$ such that every $\mathcal{H}$-free graph G with $|G| > 1$ admits a pure pair A, B with $|A|, |B| \geq \varepsilon|G|$;*
- *If $\mathcal{H}$ does not contain both a forest and the complement of a forest, then there exists $c > 0$ and arbitrarily large $\mathcal{H}$-free graphs G in which there is no pure pair A, B with $|A|, |B| \geq |G|^{1-c}$.*

But if we allow $\mathcal{H}$ to be infinite, the pretty dichotomy of 1.1 disappears: the first bullet remains true, but the second may be false. For example, it was shown in [3] that:

1.2 *Let H be a graph and let $\mathcal{H}$ be the class of all subdivisions of H and their complements; then there exists $\varepsilon > 0$ such that every $\mathcal{H}$-free graph G with $|G| > 1$ admits a pure pair A, B with $|A|, |B| \geq \varepsilon|G|$.*

And also, there are classes that do not admit linear pure pairs, but for all $c > 0$, do admit pure pairs A, B with $|A|, |B| > |G|^{1-c}$. For instance, Jacob Fox [6] proved:

1.3 *For every sufficiently large positive integer n:*

- *for every n-vertex comparability graph G, there is a pure pair A, B in G with $|A|, |B| > \frac{n}{4 \log_2 n}$;*
- *there is an n-vertex comparability graph G such that there is no pure pair A, B in G with $|A|, |B| \geq \frac{15n}{\log_2 n}$.*

There is also a related asymmetric result, by Fox, Pach and Tóth [8]:

1.4 *There exists $\varepsilon > 0$ such that for every comparability graph G with $|G| > 1$, either there is a complete pair A, B with $|A|, |B| \geq \varepsilon|G|$, or there is an anticomplete pair A, B with $|A|, |B| \geq \varepsilon|G|/\log|G|$.*

Comparability graphs are perfect, and Fox [6] (and see also [7]) conjectured that something like 1.3 holds for all perfect graphs; more exactly:

1.5 Conjecture: *For every sufficiently large positive integer n and every n-vertex perfect graph G, there is a pure pair A, B in G with $|A|, |B| \geq n^{1-o(1)}$.*

We will prove this conjecture, and several strengthenings. To prove 1.5 itself, we will show that

1.6 *For all $c > 0$, and all sufficiently large n, if G is an n-vertex perfect graph, then there is a pure pair A, B in G with $|A|, |B| \geq n^{1-c}$.*

This can be strengthened: we can make one of the two sets of linear size (and replace the "sufficiently large" condition in 1.6 with a multiplicative constant). We will show:

1.7 *For all $c > 0$ there exists $\varepsilon > 0$ such that if G is a perfect graph with $|G| > 1$, then there is a pure pair A, B in G with $|A| \geq \varepsilon|G|$ and $|B| \geq \varepsilon|G|^{1-c}$.*

The complement graph of G is denoted by $\overline{G}$. A *hole* in G is an induced cycle of length at least four, and an *antihole* in G is an induced subgraph whose complement graph is a hole in $\overline{G}$. Perfect graphs are the graphs that have have no holes or antiholes of odd length [1], but we will show that it is not necessary to exclude all odd holes and odd antiholes to have the result 1.7; it is enough to exclude one of each, of sufficient length. The next result is a strengthening of 1.7:

1.8 *Let $c > 0$ with $1/c$ an integer, and let $\ell_1, \ell_2 \geq 4/c + 5$ be integers. Then there exists $\varepsilon > 0$ such that if G is a graph with $|G| > 1$, with no hole of length exactly ℓ_1 and no antihole of length exactly ℓ_2, then there is a pure pair A, B in G with $|A| \geq \varepsilon|G|$ and $|B| \geq \varepsilon|G|^{1-c}$.*

This can be further strengthened, as follows. Let us say G *contains* H if some induced subgraph of G is isomorphic to H, and G is *H-free* otherwise. If $X \subseteq V(G)$, $G[X]$ denotes the subgraph induced on X. We say that a graph H has *branch-length* at least ℓ if every cycle of H has length at least ℓ, and every two vertices of H with degree at least three have distance at least ℓ in H. Since a cycle of length ℓ has branch-length ℓ, the next result strengthens 1.8 and is the main result of the paper:

1.9 *Let $c > 0$ with $1/c$ an integer, and let H_1, H_2 be graphs with branch-length at least $4/c + 5$. Then there exists $\varepsilon > 0$ such that if G is a graph with $|G| > 1$ that is H_1-free and $\overline{H_2}$-free, then there is a pure pair A, B in G with $|A| \geq \varepsilon|G|$ and $|B| \geq \varepsilon|G|^{1-c}$.*

2 Reduction to the sparse case

Let us say a graph G is *ε-sparse* if every vertex has degree less than $\varepsilon|G|$. We say G is *(α, β)-coherent* if there do not exist disjoint subsets A, B of $V(G)$, anticomplete to each other, such that $|A| \geq \alpha$ and $|B| \geq \beta$.

A one-vertex graph does not admit any non-trivial pure pair, but it is ε-sparse for all $\varepsilon > 0$, and (α, β)-coherent for all $\alpha, \beta > 0$; so our standard hypothesis that G is suitably sparse and suitably coherent does not exclude the case $|G| = 1$, and we always need to assume separately that $|G| > 1$. But we observe:

2.1 *Let $0 < \varepsilon \leq 1/2$; if G is ε-sparse and $(\varepsilon|G|, \varepsilon|G|)$-coherent, with $|G| > 1$, then $|G| > 1/\varepsilon$.*

Proof. Suppose that $|G| \leq 1/\varepsilon$. If some distinct $u, v \in V(G)$ are non-adjacent, $\{u\}, \{v\}$ form an anticomplete pair, both of cardinality at least $\varepsilon|G|$, a contradiction. So G is a complete graph; but its maximum degree is less than $\varepsilon|G|$ and $\varepsilon \leq 1/2$, which is impossible since $|G| > 1$. This proves 2.1. ∎

If G is a graph and $v \in V(G)$, a *G-neighbour* of v means a vertex of G adjacent to v in G. A theorem of Rödl [11] implies the following:

2.2 *For every graph H and all $\eta > 0$ there exists $\delta > 0$ with the following property. Let G be an H-free graph. Then there exists $X \subseteq V(G)$ with $|X| \geq \delta|G|$, such that one of $G[X]$, $\overline{G}[X]$ is η-sparse.*

Consequently, in order to prove 1.9, it suffices to prove the following:

2.3 *Let $c > 0$ with $1/c$ an integer, and let H be a graph with branch-length at least $4/c + 5$. Then there exists $\varepsilon > 0$ such that every ε-sparse $(\varepsilon|G|^{1-c}, \varepsilon|G|)$-coherent graph G with $|G| > 1$ contains H.*

Proof of 1.9, assuming 2.3. Let $c > 0$ with $1/c$ an integer, and let H_1, H_2 have branch-length at least $4/c + 5$. For $i = 1, 2$, choose ε_i such that 2.3 holds with $\varepsilon = \varepsilon_i$ and $H = H_i$. Let $\eta = \min(\varepsilon_1, \varepsilon_2, 1/2)$. Choose δ such that 2.2 holds taking $H = H_1$. Let $\varepsilon = \eta\delta$. We claim that ε satisfies 1.9.

Let G be a graph with $|G| > 1$ that is H_1-free and $\overline{H_2}$-free. We must show that there is a pure pair A, B in G with $|A| \geq \varepsilon|G|$ and $|B| \geq \varepsilon|G|^{1-c}$. From the choice of δ, there exists $X \subseteq V(G)$ with $|X| \geq \delta|G|$, such that one of $G[X]$, $\overline{G}[X]$ is η-sparse; and by 2.1 we may assume that $|G| > 1/\varepsilon \geq 1/\delta$, and so $|X| > 1$. In the first case, since $\eta \leq \varepsilon_1$, 2.3 applied to $G[X]$ implies that there is an anticomplete pair A, B in $G[X]$ with $|A| \geq \eta|X|$ and $|B| \geq \eta|X|^{1-c}$. Thus

$$|A| \geq \eta|X| \geq \eta\delta|G| = \varepsilon|G|$$

and

$$|B| \geq \eta|X|^{1-c} \geq \eta\delta^{1-c}|G|^{1-c} \geq \eta\delta|G|^{1-c} = \varepsilon|G|^{1-c},$$

as required. In the second case we argue similarly, working in $\overline{G}[X]$. This proves 1.9. ∎

The remainder of the paper is devoted to proving 2.3.

3 The pathfinder lemma: finding a path of specified length

In this section we will prove the main technical tool that we need, which we call the "pathfinder". If $A, B \subseteq V(G)$ are disjoint, we say A *covers* B if every vertex in B has a neighbour in A. A *levelling* in G is a sequence $\mathcal{L} = (L_0, L_1, \ldots, L_k)$ of disjoint subsets of $V(G)$ with $k \geq 1$ such that

- $|L_0| = 1$;
- L_{i-1} covers L_i for $1 \leq i \leq k$; and
- $L_0 \cup \cdots \cup L_{i-2}$ is anticomplete to L_i for all $i \in \{2, \ldots, k\}$.

We denote $L_0 \cup L_1 \cup \cdots \cup L_k$ by $V(\mathcal{L})$. We call L_k the *base* of the levelling $\mathcal{L} = (L_0, L_1, \ldots, L_k)$, and $V(\mathcal{L}) \setminus L_k$ is called the *heart* of $\mathcal{L}$. We call k the *height* of the levelling, and the unique vertex in L_0 is the *apex*. We call L_{k-1} the *penultimate level* of the levelling (for want of a better name). A path P is *$\mathcal{L}$-vertical* if $V(P) \subseteq V(\mathcal{L})$ and $|V(P) \cap L_i| \leq 1$ for $0 \leq i \leq k$.

The pathfinder says that if a graph G is suitably sparse and suitably coherent, and we are given two levellings with disjoint vertex sets and with bases of size linear in $|G|$, and there are suitable constraints on the edges between the two levellings, then we can find an induced path between the two apexes of any given length greater than the sum of the two heights. (And there is also a version when the two apexes are equal, and in this case we will find a cycle rather than a path.)

Let us explain how the pathfinder will be used to prove 2.3 and hence 1.9. We can assume (by extending H if necessary) that the graph H of 2.3 is obtained from some stable set X by adding paths, each of length at least $4/c + 5$, where each path has both ends in X and no other vertices in X, and all these paths are pairwise vertex-disjoint except for their vertices in X. (For numerical reasons, we will also allow the addition of cycles, but let us skip that for now.) We are given a graph G which is suitably sparse and suitably coherent, and we need to show that it has an induced subgraph isomorphic to H. To obtain a copy of H in G, we will choose an appropriate set $X \subseteq V(G)$, and then try to route paths of G of the right length between the correct pairs of vertices of X. We will find each such path by applying the pathfinder to some pair of levellings with apexes the corresponding pair of vertices of X. But we cannot apply the pathfinder twice to the same levelling, because the paths we want to produce need to be pairwise disjoint and anticomplete except for their ends. Thus, if some vertex in X is supposed to be an end of several paths of H, we will need several levellings all with this apex. So we need a way to find a good supply of levellings, each with base of linear size, and pairwise disjoint except for their apexes, grouped into several sets each with a common apex; and we want the edges between them to be under control. And another thing: the pathfinder can only provide paths between the two apexes of length greater than the sum of the two heights of the corresponding levellings, and we need paths which might be as short as $4/c + 5$, so we need the levellings to have height at most something like $2/c$.

The paper is organized as follows. In this section we prove the pathfinder; and in the next we explain why we can get levellings of height about $1/c$ (later, when we try to get several levellings with a common apex, this height will double). In section 5 we relax the conditions on levellings, and instead just look for subgraphs of radius about $1/c$ that have a linear set of neighbours (we call this a "covering", and the subgraph of bounded radius is its "heart"); we find that we can obtain many coverings, with hearts that are disjoint and pairwise anticomplete. What we really want is something slightly more: we want there also to be a vertex with a neighbour in each of the hearts. To prove this, we prove something stronger, that there is a "multicovering", but this is just a tool to get one neighbour in common.

So we have many coverings, with hearts pairwise anticomplete and with a common neighbour a. Add a to each of the hearts; then we get many coverings, with hearts pairwise anticomplete except for one common vertex, which we call its "apex". We call this group of coverings a "spider", and this is the topic of section 6. By making each of the hearts only just big enough that it has linearly many neighbours, we can find a spider such that most vertices of the graph have no neighbours in any of the hearts of the spider; and so, among them we can do it again, and find another spider. This way we get a "troupe" of spiders, with no edges between their hearts. The next step is to convert the hearts of the coverings in each spider to levellings (so the spiders become "lobsters"); this is also done in section 6. Then we are ready to apply the pathfinder, which is done in section 7, and this completes the proof of 2.3.

Let us begin by proving the pathfinder. First we need the following lemma:

3.1 *Let $\rho \geq 1$ be some real number, let $K, k > 0$ be integers with $K > k$, and let $n_1, \ldots, n_K$ be non-negative integers, all less than $\rho^{K/k-2-1/k}$. Then there exists $i \in \{1, \ldots, K-k\}$ such that $\rho n_i \geq n_j$ for $j = i+1, \ldots, i+k$.*

Proof. Suppose not; then for each $i \in \{1, \ldots, K-k\}$ there exists $f(i)$ such that $i < f(i) \leq i+k$ and $\rho n_i < n_{f(i)}$. Define $x_1 = 1$ and $x_{i+1} = f(x_i)$ provided $x_i \leq K-k$. Let $x_1, \ldots, x_t$ be defined by this process; thus $K-k < x_t \leq K$. Since $x_{i+1} - x_i \leq k$ for each i, it follows that $tk \geq K-1$. Since $n_{x_2} > \rho n_{x_1}$ and n_{x_2} is an integer, it follows that $n_{x_2} \geq 1$. Thus for $2 \leq i \leq t$, $n_{x_i} \geq \rho^{i-2}$, and so $n_{x_t} \geq \rho^{t-2} \geq \rho^{K/k-2-1/k}$, contrary to the hypothesis. This proves 3.1. ∎

Next we need:

3.2 *Let $c > 0$ such that $1/c$ is an integer, and define $r = 2 + 1/c$. Let $\ell \geq 1$ be an integer, and define $K = r^\ell - 1$, and $k = r^{\ell-1} - 1$. Let $\varepsilon > 0$, and let G be an ε-sparse $(\varepsilon|G|^{1-c}, \varepsilon|G|)$-coherent graph. Let $B_0, B_1, \ldots, B_K \subseteq V(G)$ be disjoint, where $B_0 \neq \emptyset$ and $B_1, \ldots, B_K$ each have cardinality at least $r^{2\ell}\varepsilon|G|$. Then either:*

- *there is an induced path of length ℓ, with vertices $p_0, p_1, \ldots, p_\ell$ in order, and*

$$1 \leq t_1 < t_2 < \cdots < t_\ell \leq K,$$

  such that $p_0 \in B_0$, and $p_i \in B_{t_i}$ for $1 \leq i \leq \ell$; or

- *$|B_0| \leq K\varepsilon|G|^{1-c}$, and there are sets $C_1, \ldots, C_{K-k}$ with union B_0, such that for each i with $1 \leq i \leq K-k$, and each j with $i \leq j \leq i+k$, at least $r^{2\ell-2}\varepsilon|G|$ vertices in B_j have no neighbour in C_i.*

Proof. We proceed by induction on ℓ. Suppose first that $\ell = 1$. If there is an edge between B_0 and $B_1 \cup \cdots \cup B_K$, then the first bullet holds; and if B_0 is anticomplete to $B_1 \cup \cdots \cup B_K$, then since H is $(\varepsilon|G|^{1-c}, \varepsilon|G|)$-coherent and $|B_1| \geq r^{2\ell}\varepsilon|G| \geq \varepsilon|G|$, it follows that $|B_0| < \varepsilon|G|^{1-c}$, and the second bullet holds, taking $C_1, \ldots, C_{K-k} = B_0$. Thus we may assume that $\ell \geq 2$ and the result holds for $\ell - 1$. Define $\rho = |G|^c$.

Let $B_0 = \{v_1, \ldots, v_n\}$. For all $i \in \{1, \ldots, K-k\}$ and all $j \in \{i, \ldots, i+k\}$, define $A^0_{i,j} = \emptyset$, and $A^0 = \emptyset$. Inductively for $h = 1, \ldots, n$ we will define

- a set $X^h_i \subseteq B_i$ for each $i \in \{1, \ldots, K\}$
- the *type* of v_h (one of the numbers $1, \ldots, K-k$);
- a set $A^h_{i,j} \subseteq B_j$ for each $i \in \{1, \ldots, K-k\}$ and each $j \in \{i, \ldots, i+k\}$; and
- a set A^h, which is the union of $A^h_{i,j}$ over all $i \in \{1, \ldots, K-k\}$ and all $j \in \{i, \ldots, i+k\}$

as follows. Suppose that $1 \leq h \leq n$, and A^{h-1} and $A^{h-1}_{i,j}$ are defined for all i, j with $1 \leq i \leq K-k$ and $i \leq j \leq i+k$. For $1 \leq i \leq K$ let X^h_i be the set of vertices in $B_i \setminus A^{h-1}$ adjacent to v_h. Since $(K+1)/(k+1) = 2 + 1/c$, it follows that $K > (2+1/c)k + 1$, and so $1/c < K/k - 2 - 1/k$. Hence for $1 \leq i \leq K$, $|X^h_i| \leq |G| < \rho^{K/k-2-1/k}$. By 3.1 applied to the numbers $|X^h_1|, \ldots, |X^h_K|$, there exists t with $1 \leq t \leq K-k$ such that $\rho|X^h_t| \geq |X^h_j|$ for $j = t, \ldots, t+k$. Choose some such t, which we call the type of v_h. For each $j \in \{t, \dots, t+k\}$ define $A^h_{t,j} = A^{h-1}_{i,j} \cup X^h_j$; and For each $i \in \{1, \dots, K-k\} \setminus \{t\}$ and each $j \in \{i, \dots, i+k\}$ define $A^h_{i,j} = A^{h-1}_{i,j}$. This completes the inductive definition.

(1) *$\rho|A^h_{i,j}| \ge |A^h_{i,i}|$ for all $h \in \{1, \dots, n\}$ and all $i \in \{1, \dots, K-k\}$ and all $j \in \{i, \dots, i+k\}$.*

$A^h_{i,j}$ is the disjoint union of the sets X^h_j for all $h \in \{1, \dots, n\}$ such that v_h has type i; and $A^h_{i,i}$ is the disjoint union of X^h_i for the same values of h. Since $\rho|X^h_i| \ge |X^h_j|$ for each such h, this proves (1).

(2) *If v_h has type i, then every vertex of B_j adjacent to v_h belongs to A^h, for all $h \in \{1, \dots, n\}$, all $1 \le i \le K-k$, and all $j \in \{i, \dots, i+k\}$.*

Let $x \in B_j$ be adjacent to v_h. If $x \in A^{h-1}$, then the claim holds since $A^{h-1} \subseteq A^h$. If $x \notin A^{h-1}$ then $x \in X^h_j$ from the definition of X^h_j, and since v_h has type i, it follows that

$$x \in X^h_j \subseteq A^h_{i,j} \subseteq A^h.$$

This proves (2).

For $1 \le i \le K-k$, let C_i be the set of vertices in B_0 that have type i. Thus $C_1, \dots, C_{K-k}$ are pairwise disjoint and have union B_0. We note that

$$r^{2\ell} - r^{2\ell-2} = (3 + 4/c + 1/c^2)(k+1)^2 \ge 2(k+1)^2.$$

(3) *We may assume that $|A^n_{i,j}| > k\varepsilon|G|$ for some $i \in \{1, \dots, K-k\}$ and some $j \in \{i, \dots, i+k\}$.*

Suppose not. Let $1 \le j \le K$. Since $A^n \cap B_j$ is the union of the sets $A^n_{i,j}$ for all $i \in \{1, \dots, K\}$ with $j - i \in \{0, \dots, k\}$, it follows that $|A^n \cap B_j| \le k(k+1)\varepsilon|G|$. Now let $1 \le i \le K-k$; by (2), C_i is anticomplete to $B_j \setminus A^n$, for all $j \in \{i, \dots, i+k\}$. Since

$$|B_j \setminus A^n| = |B_j| - |B_j \cap A^n| \ge r^{2\ell}\varepsilon|G| - k(k+1)\varepsilon|G| \ge r^{2\ell-2}\varepsilon|G| \ge \varepsilon|G|$$

and G is $(\varepsilon|G|^{1-c}, \varepsilon|G|)$-coherent, it follows that $|C_i| < \varepsilon|G|^{1-c}$. Hence $|B_0| \le K\varepsilon|G|^{1-c}$; and so the second bullet of the theorem holds. This proves (3).

From (3), we may choose $h \in \{1, \dots, n\}$ minimum such that $|A^h_{i,j}| > k\varepsilon|G|$ for some $i \in \{1, \dots, K-k\}$ and some $j \in \{i, \dots, i+k\}$. Define D to be the set of all $v_{h'} \in C_i$ with $1 \le h' \le h$. From the minimality of h, and since G is ε-sparse, it follows that $|A^h_{i,j}| \le (k+1)\varepsilon|G|$ for all $i \in \{1, \dots, K-k\}$ and all $j \in \{i, \dots, i+k\}$. Consequently $|A^h \cap B_i| \le (k+1)^2\varepsilon|G|$ for $1 \le i \le K$. Now choose $i \in \{1, \dots, K-k\}$ such that $|A^h_{i,j}| > k\varepsilon|G|$ for some $j \in \{i, \dots, i+k\}$. By (1), $|A^h_{i,i}| > k\varepsilon|G|/\rho = k\varepsilon|G|^{1-c}$. For $j = i+1, \dots, i+k$, let D_j be the set of all vertices in B_j that have no neighbour in D. Thus $B_j \setminus A^h \subseteq D_j$, and so

$$|D_j| \ge r^{2\ell}\varepsilon|G| - (k+1)^2\varepsilon|G| \ge r^{2\ell-2}\varepsilon|G|.$$

Since $|A^h_{i,i}| > k\varepsilon|G|^{1-c}$, it follows from the inductive hypothesis (with ℓ replaced by $\ell - 1$, and B_0 replaced by $A^h_{i,i}$, and $B_1, \ldots, B_K$ replaced by $D_{i+1}, \ldots, D_{i+k}$) that there is an induced path of length $\ell - 1$, with vertices $p_1, \ldots, p_\ell$ in order, and

$$i + 1 \leq t_2 < \cdots < t_\ell \leq i + k,$$

such that $p_1 \in A^h_{i,i}$, and $p_i \in B_{t_i}$ for $2 \leq i \leq \ell$. Choose $p_0 \in D$ adjacent to p_1, and define $t_1 = i$; then the path with vertices $p_0, p_1, \ldots, p_\ell$ is induced and satisfies the first bullet of the theorem. This proves 3.2. ∎

The "pathfinder", the main result of this section, is the following:

3.3 *Let $c > 0$, such that $1/c$ is an integer. Let $\ell, s, t > 0$ be integers, and let $d > 0$. Let $\varepsilon > 0$ with $(2 + 1/c)^{(t+1)(\ell+t)}\varepsilon < d$. Let G be an ε-sparse $(\varepsilon|G|^{1-c}, \varepsilon|G|)$-coherent graph, and for $i = 1, 2$, let $\mathcal{L}_i$ be a levelling in G, with heart H_i, apex a_i, and base B_i, satisfying:*

- *$V(\mathcal{L}_1) \cap V(\mathcal{L}_2) = \{a_1\} \cap \{a_2\}$;*
- *$V(\mathcal{L}_2) \setminus \{a_2\}$ is anticomplete to H_1, and if $a_1 \neq a_2$ then a_2 is anticomplete to H_1;*
- *$\mathcal{L}_1, \mathcal{L}_2$ have heights s, t respectively; and*
- *$|B_i| \geq d|G|$ for $i = 1, 2$.*

Then there is an induced path (or cycle, if $a_1 = a_2$) of length $\ell + s + t$ between a_1, a_2, with vertex set a subset of $H_1 \cup B_1 \cup H_2 \cup B_2$.

Proof. For each integer $i \geq 0$, let $k_i = (2 + 1/c)^i - 1$. It follows that $k_\ell k_{\ell+1} \ldots k_{\ell+t}\varepsilon < d$. Moreover, since $\varepsilon < d$, it follows that $t \geq 2$ (because $|B_2| > \varepsilon|G|$ and G is ε-sparse), and so $k_{\ell+t}(k_{2\ell+2t} + 2)\varepsilon \leq d$. Define $d_i = (2 + 1/c)^{2i}\varepsilon$ for each integer $i \geq 0$.

Let G, $\mathcal{L}_1, \mathcal{L}_2$ and H_i, a_i, B_i $(i = 1, 2)$ satisfy the hypotheses of the theorem. Let $\mathcal{L}_1 = (L_0, \ldots, L_s)$ and $\mathcal{L}_2 = (M_0, \ldots, M_t)$ say; thus $L_s = B_1$ and $M_t = B_2$. Let $Z_0 = \emptyset$. For $i = 1, \ldots, k_{\ell+t}$, we will inductively define $Z_i \subseteq L_{s-1}$ with $Z_{i-1} \subseteq Z_i$, and $D_i \subseteq L_s$ with $D_1, \ldots, D_i$ pairwise disjoint, satisfying

- $d_{\ell+t}|G| \leq |D_i| \leq (d_{\ell+t} + \varepsilon)|G|$
- D_i is the set of all vertices in L_s that have a neighbour in Z_i and have no neighbour in Z_{i-1} (and so $D_1 \cup \cdots \cup D_i$ is the set of all vertices in L_s that have a neighbour in Z_i).

Thus, suppose that $1 \leq i < k_{\ell+t}$, and $Z_0, \ldots, Z_{i-1}$ and $D_1, \ldots, D_{i-1}$ are defined satisfying the conditions above. It follows that

$$|D_1 \cup \cdots \cup D_{i-1}| \leq (i - 1)(d_{\ell+t} + \varepsilon)|G| \leq k_{\ell+t}(d_{\ell+t} + \varepsilon)|G| - d_{\ell+t}|G|.$$

But $d_{\ell+t} + \varepsilon = (k_{2(\ell+t)} + 2)\varepsilon$, and $k_{\ell+t}(k_{2\ell+2t} + 2)\varepsilon \leq d$, so

$$|D_1 \cup \cdots \cup D_{i-1}| \leq k_{\ell+t}(k_{2(\ell+t)} + 2)\varepsilon|G| - d_{\ell+t}|G| \leq (d - d_{\ell+t})|G|.$$

Hence at least $d_{\ell+t}|G|$ vertices in L_s do not belong to $D_1 \cup \cdots \cup D_{i-1}$. All these vertices have a neighbour in $L_{s-1} \setminus Z_{i-1}$ and have no neighbour in Z_{i-1}; and so there exists Z_i with $Z_{i-1} \subseteq Z_i \subseteq L_{s-1}$, minimal such that at least $d_{\ell+t}|G|$ vertices in L_s have a neighbour in Z_i and have none in Z_{i-1}. Let this set of vertices be D_i. Since G is ε-sparse, the minimality of Z_i implies that $|D_i| \leq (d_{\ell+t} + \varepsilon)|G|$. This completes the inductive definition.

We will try to construct a path (or cycle) satisfying the theorem that starts from a_2, runs down through layers of $\mathcal{L}_2$, jumps to some D_i, runs through some of $D_{i+1}, D_{i+2}, \ldots$, to make it the right length, and then runs up to a_1 through the layers of $\mathcal{L}_1$. The sets Z_i are designed so that when the path has run through enough D_i's to make its length correct, we can exit into the heart of $\mathcal{L}_1$ without picking up unwanted chords. Note that the only edges between $V(\mathcal{L}_2)$ and $V(\mathcal{L}_1)$ have an end in the base of $\mathcal{L}_1$ (or are incident with a_1, if $a_1 = a_2$).

Let $\mathcal{Q} = (Q_0, \ldots, Q_t)$ be a levelling in G. We say it is a *sub-levelling* of $\mathcal{L}_2$ if $Q_i \subseteq M_i$ for $0 \leq i \leq t$. For $0 \leq h \leq t$, we say that such a sub-levelling $\mathcal{Q} = (Q_0, \ldots, Q_t)$ is *h-good* if

- there exists $g \in \{1, \ldots, k_{\ell+t} - k_{\ell+t-h} + 1\}$, and for each $j \in \{g, \ldots, g + k_{\ell+t-h} - 1\}$ there exists $F_j \subseteq D_j$, such that F_j is anticomplete to $Q_0 \cup Q_1 \cup \cdots \cup Q_{h-1}$, and $|F_j| \geq d_{\ell+t-h}|G|$; and
- $|Q_t| > k_\ell k_{\ell+1} \ldots k_{\ell+t-h} \varepsilon |G|^{1-c}$.

Since $d|G| > k_\ell k_{\ell+1} \ldots k_{\ell+t} \varepsilon |G|^{1-c}$ it follows that $\mathcal{L}_2$ is 0-good. Choose $h \leq t$ maximum such that some sub-levelling $\mathcal{Q} = (Q_0, \ldots, Q_t\}$ of $\mathcal{L}_2$ is h-good, and let g and the sets F_j ($j \in \{g, \ldots, g + k_{\ell+t-h} - 1\}$) be as in the definition. Let $K = k_{\ell+t-h}$. Since each $|F_j| \geq d_{\ell+t-h}|G|$, we may apply 3.2, replacing B_0 by Q_h, and replacing ℓ by $\ell + t - h$, and replacing the sequence $B_1, \ldots, B_{k_\ell}$ by $F_g, \ldots, F_{g+K-1}$. There are two possible outcomes of 3.2.

The first outcome is: there is an induced path P of length $\ell + t - h$, with vertices $p_0, p_1, \ldots, p_{\ell+t-h}$ in order, and

$$g \leq t_1 < t_2 < \cdots < t_{\ell+t-h} \leq g + K - 1,$$

such that $p_0 \in Q_h$, and $p_i \in F_{t_i}$ for $1 \leq i \leq \ell + t - h$. In this case, choose a $\mathcal{Q}$-vertical path Q between a_2 and p_0 (therefore of length h); choose a neighbour v of $p_{\ell+t-h}$ in $Z_{t_{\ell+t-h}}$; and choose an $\mathcal{L}_1$-vertical path R between a_1, v (therefore of length $s - 1$). We claim that

$$a_2\text{-}Q\text{-}p_0\text{-}P\text{-}p_{\ell+t-h}\text{-}v\text{-}R\text{-}a_1$$

is an induced path or cycle. To show this, we must check that

- $V(P) \cap V(Q) = \{p_0\}$, and $V(P) \setminus \{p_0\}$ is anticomplete to $V(Q) \setminus \{p_0\}$; this is true since $Q_0, \ldots, Q_{h-1}$ are anticomplete to $F_g, \ldots, F_{g+K-1}$ from the definition of h-good.
- $V(P) \cap V(R) = \emptyset$, and the edge with ends $p_{\ell+t-h}$ and v is the only edge between $V(P)$ and $V(R)$; this is true since $L_0, \ldots, L_{s-2}$ are anticomplete to L_s, and $v \in Z_{t_{\ell+t-h}}$ is anticomplete to $D_{t_1}, \ldots, D_{t_{\ell+t-h-1}}$.
- $V(Q) \cap V(R) = \{a_1\} \cap \{a_2\}$, and every edge between $V(Q)$ and $V(R)$ has an end in $\{a_1\} \cap \{a_2\}$; this is true from the hypothesis.

This proves the path or cycle is indeed induced, and since it has length ℓ, the theorem holds.

The second outcome of 3.2 is: $\ell + t - h > 0$, and $|Q_h| \le K\varepsilon|G|^{1-c}$, and, writing $k = k_{\ell+t-h-1}$, there are sets $C_g, \ldots, C_{g+K-k-1}$ with union Q_h, such that for each i with $g \le i \le g+K-k-1$, and each j with $i \le j \le i+k$, at least $d_{\ell+t-h-1}|G|$ vertices in F_j have no neighbour in C_i. Since

$$|Q_h| \le K\varepsilon|G|^{1-c} < k_\ell k_{\ell+1} \ldots k_{\ell+t-h}\varepsilon|G|^{1-c} < |Q_t|$$

it follows that $h < t$. For $g \le i \le g+K-k-1$, let X_i be the set of vertices in Q_t that are joined to a vertex in C_i by a $\mathcal{Q}$-vertical path. Since $\mathcal{Q}$ is a levelling and $C_g, \ldots, C_{g+K-k-1}$ have union Q_h, it follows that $X_g, \ldots, X_{g+K-k-1}$ have union Q_t; and since $|Q_t| > k_\ell k_{\ell+1} \ldots k_{\ell+t-h}\varepsilon|G|^{1-c}$, there exists i with $g \le i \le g+K-k-1$ such that

$$|X_i| \ge |Q_t|/K > k_\ell k_{\ell+1} \ldots k_{\ell+t-h-1}\varepsilon|G|^{1-c}.$$

For $h \le h' \le t$ let $Q'_{h'}$ be the set of vertices in $Q_{h'}$ that are joined to a vertex in C_i by a $\mathcal{Q}$-vertical path. Thus $Q'_h = C_i$, and

$$(Q_0, \ldots, Q_{h-1}, Q'_h, Q'_{h+1}, \ldots, Q'_t)$$

is an $(h+1)$-good sub-levelling of $\mathcal{L}_2$, a contradiction. This proves 3.3. ∎

The next result is a form of 3.3 with similar hypotheses except that the bases of the two levellings need not be disjoint, and we weaken slightly the condition about edges between the heart of $\mathcal{L}_1$ and $V(\mathcal{L}_2)$.

3.4 *Let $c > 0$, such that $1/c$ is an integer. Let $\ell, s, t > 0$ be integers, and let $d > 0$. Let $\varepsilon > 0$ with $(2+1/c)^{(t+1)(\ell+t)}\varepsilon < d/3$. Let G be an ε-sparse $(\varepsilon|G|^{1-c}, \varepsilon|G|)$-coherent graph. For $i = 1, 2$, let $\mathcal{L}_i$ be a levelling in G, with heart H_i, apex a_i, and base B_i. Suppose that:*

- *for $i = 1, 2$, $|B_i| \ge d|G|$;*
- *$V(\mathcal{L}_1) \cap V(\mathcal{L}_2) = (\{a_1\} \cap \{a_2\}) \cup (B_1 \cap B_2)$; and*
- *every edge between H_1 and $V(\mathcal{L}_2)$ has one end in the penultimate level of $\mathcal{L}_1$ and the other end in B_2.*

Let $\mathcal{L}_1, \mathcal{L}_2$ have heights s, t respectively. Then there is an induced path (or cycle, if $a_1 = a_2$) of length $\ell + s + t$ between a_1, a_2, with vertex set a subset of $H_1 \cup B_1 \cup H_2 \cup B_2$.

Proof. Let $d' = d/3$, and let G, $\mathcal{L}_1$, $\mathcal{L}_2$ and H_i, a_i, B_i $(i = 1, 2)$ satisfy the hypotheses of the theorem. Let $\mathcal{L}_1 = (L_0, \ldots, L_s)$; thus $L_s = B_1$. Choose $L'_{s-1} \subseteq L_{s-1}$ minimal such that at least $d'|G|$ vertices in $B_1 \cup B_2$ have a neighbour in L'_{s-1}. Let L'_s be the set of vertices in $B_1 \cup B_2$ that have a neighbour in L'_{s-1}. Thus

$$d'|G| \le |L'_s| \le (d' + \varepsilon)|G| \le 2d'|G|.$$

Let $\mathcal{L}'_1$ be the levelling $(L_0, \ldots, L_{s-1}, L'_{s-1}, L'_s)$. Let $\mathcal{L}'_2$ be the levelling obtained from $\mathcal{L}_2$ by replacing its base by $B_2 \setminus L'_s$. Then $|L'_s| \ge d'|G|$, and

$$|B_2 \setminus L'_s| \ge d|G| - 2d'|G| \ge d'|G|.$$

Hence $\mathcal{L}'_1$, $\mathcal{L}'_2$ satisfy the hypotheses of 3.3, and the result follows. This proves 3.4. ∎

When we apply 3.4, in the final section, it will be to levellings $\mathcal{L}_1, \mathcal{L}_2$ such that every edge between $V(\mathcal{L}_1), V(\mathcal{L}_2)$ either is incident with the common apex (when there is one) or is between the base of one of the levellings and one of the last two terms of the other levelling; so 3.4 is stronger than we need.

4 Expansion

If $X \subseteq V(G)$, $N(X)$ denotes the set of vertices in $V(G) \setminus X$ with a neighbour in X, and $N[X] = N(X) \cup X$. A graph G is *τ-expanding* if $|N[X]| \geq \min(\tau|X|, |G|/2)$ for every subset $X \subseteq V(G)$.

4.1 *Let $c > 0$, and let G be a $(|G|^{1-c}/4, |G|/4)$-coherent graph. Then there exists $Y \subseteq V(G)$ with $|Y| \leq |G|^{1-c}/4$ such that $G \setminus Y$ is $|G|^c$-expanding.*

Proof. Let $\alpha = |G|^{1-c}/4$ and $\tau = |G|^c$. Choose $Y \subseteq V(G)$ maximal such that $|Y| \leq \alpha$ and $|N[Y]| \leq \tau|Y|$ (possibly $Y = \emptyset$). Let $W = V(G) \setminus Y$. If $G[W]$ is τ-expanding then the theorem holds, so we assume not. Thus there exists $X \subseteq W$ such that $|N[X] \cap W| < \min(\tau|X|, |W|/2)$. Consequently $X \neq \emptyset$. But

$$|N[X \cup Y]| \leq |N[Y]| + |N[X] \cap W| \leq \tau|Y| + \tau|X|,$$

and so from the maximality of Y, it follows that $|X \cup Y| > \alpha$. Now $|N[Y]| \leq \tau|Y| \leq \tau\alpha = |G|/4$, and $|N[X] \cap W| \leq |W|/2 \leq |G|/2$; so

$$|N[X \cup Y]| \leq |N[Y]| + |N[X] \cap W| \leq 3|G|/4.$$

Let $U = V(G) \setminus N[X \cup Y]$; then $|U| \geq |G|/4$. But $X \cup Y$ is anticomplete to U, contradicting that G is $(|G|^{1-c}/4, |G|/4)$-coherent. This proves 4.1. ∎

If u, v are vertices of a graph G, it is sometimes convenient to call the distance between u, v in G the *G-distance* between u, v. We deduce:

4.2 *Let $c > 0$, and let G be a $(|G|^{1-c}/4, |G|/4)$-coherent graph. Then there exist $u \in V(G)$ and an integer $k < 1 + 1/c$, such that:*

- *at most $|G|/2$ vertices have G-distance less than k from u; and*
- *at least $|G|/4$ vertices have G-distance exactly k from u.*

Proof. By 4.1, there exists $Y \subseteq V(G)$ with $|Y| \leq |G|^{1-c}/4$ such that $G \setminus Y$ is τ-expanding, where $\tau = |G|^c$. Choose $u \in V(G) \setminus Y$, and for each integer $i \geq 0$ let M_i be the set of vertices of G that have G-distance at most i from u. Since $G \setminus Y$ is τ-expanding, it follows that for all $i \geq 0$, $|M_{i+1} \setminus Y| \geq \min(\tau|M_i \setminus Y|, |G \setminus Y|/2)$. For each $i \geq 1$, let $L_i = M_i \setminus M_{i-1}$.

(1) *There exists $k \leq 1 + 1/c$ such that $|L_k| \geq |G|/4$.*

Since $G \setminus Y$ is τ-expanding, it is connected, and so there exists ℓ such that $V(G) \setminus Y \subseteq M_\ell$. Since

$|V(G) \setminus Y| \geq 3|G|/4$, we may choose $j \geq 0$ minimum such that $|M_j| \geq |G|^{1-c}/4$. Hence for each $i \in \{1, \ldots, j-1\}$, $|M_i| < |G|^{1-c}/4 < |V(G) \setminus Y|/2$, and so $|M_i \setminus Y| \geq \tau |M_{i-1} \setminus Y|$ since $G \setminus Y$ is τ-expanding. Since $|M_0 \setminus Y| = 1$, it follows that $|M_{j-1} \setminus Y| \geq \tau^{j-1}$. Hence

$$|G|^{(j-1)c} = \tau^{j-1} \leq |M_{j-1} \setminus Y| \leq |M_{j-1}| < |G|^{1-c}/4,$$

and so $(j-1)c < 1-c$, that is, $j < 1/c$.

Since G is $(|G|^{1-c}/4, |G|/4)$-coherent, and M_j is anticomplete to $V(G) \setminus N[M_j]$, it follows that $|V(G) \setminus N[M_j]| < |G|/4$. But also $|M_{j-1}| < |G|^{1-c}/4$ (or $j = 0$), and so $|L_j \cup L_{j+1}| \geq |G| - |G|/4 - |G|^{1-c}/4 \geq |G|/2$. Thus some $k \in \{j, j+1\}$ satisfies the claim. This proves (1).

Choose k as in (1), minimum. Thus $|L_{k-1}| < |G|/4$, and $|M_{k-2}| < |G|^{1-c}/4$ since G is $(|G|^{1-c}/4, |G|/4)$-coherent. Thus $|M_{k-1}| \leq |G|/2$. This proves 4.2. ∎

5 Covering sequences

Let us say a *covering* $\mathcal{L}$ in G is a triple (a, H, B) where H, B are disjoint subsets of $V(G)$, $a \in H$, H covers B, and $G[H]$ is connected. We call a the *apex*, H the *heart*, and B the *base* of the covering, and define $V(\mathcal{L}) = H \cup B$. If for every vertex $v \in H$ there is a path of $G[H]$ between a, v of length at most $r - 1$, we say that (a, H, B) has *height* at most r, and the least such r is the *height* of (a, H, B). For instance, if $(L_0, \ldots, L_k)$ is a levelling with $k > 0$, and $L_0 = \{a\}$, then $(a, L_0 \cup \cdots \cup L_{k-1}, L_k)$ is a covering of height k.

A *covering sequence* in G is a sequence $(\mathcal{L}_1, \ldots, \mathcal{L}_n)$ of coverings in G, with hearts $H_1, \ldots, H_n$ say, such that $H_1, \ldots, H_n$ are pairwise disjoint and pairwise anticomplete. We call n its *length*. We say such a sequence has *height* at most r if each term has height at most r. If $\mathcal{M} = (\mathcal{L}_1, \ldots, \mathcal{L}_n)$ is a covering sequence, we define $V(\mathcal{M})$ to be the union of the sets $V(\mathcal{L}_i)$ for $1 \leq i \leq n$.

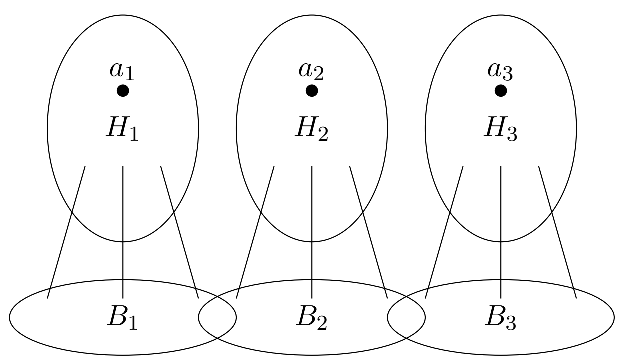

Figure 1: A covering sequence of length three. H_1, H_2, H_3 are disjoint and anticomplete, but B_1, B_2, B_3 need not be; and there may be edges between H_i and $B_j \setminus B_i$.

A covering sequence $(\mathcal{L}_1, \ldots, \mathcal{L}_n)$ is a *multicovering* if $\mathcal{L}_1, \ldots, \mathcal{L}_n$ all have the same base, and then this common base is called the *base* of the multicovering. The main result of this section says that a graph with the usual properties (suitably coherent, suitably sparse) contains a multicovering of length any specified constant, with height at most about $1/c$ and with base of linear cardinality. We prove this in several steps. We begin with:

5.1 *Let $n \geq 0$ be an integer. Let $c > 0$ such that $1/c$ is an integer; let $\varepsilon > 0$ with $\varepsilon \leq 2^{-n-2}$; and let G be an ε-sparse $(\varepsilon|G|^{1-c}, \varepsilon|G|)$-coherent graph. Then there is a covering sequence $(\mathcal{L}_1, \ldots, \mathcal{L}_n)$ in G, where $\mathcal{L}_i = (a_i, H_i, B_i)$ for $1 \leq i \leq n$, such that:*

- *for $1 \leq i < j \leq n$, H_i is anticomplete to B_j;*
- *for $1 \leq i \leq n$, $\mathcal{L}_i$ has height at most $1/c$; and*
- *for $1 \leq i \leq n$, $|B_i| \geq 2^{-i-1}|G|$.*

Proof. We proceed by induction on n. If $n = 0$ the result is trivial, so we assume that $n \geq 1$ and the result holds for $n-1$. By 4.2, there exists $u \in V(G)$ and an integer $k < 1 + 1/c$ (and hence $k \leq 1/c$, since $1/c$ is an integer), such that:

- at most $|G|/2$ vertices of G have distance less than k from u; and
- at least $|G|/4$ vertices of G have distance exactly k from u.

For $0 \leq i \leq k$ let L_i be the set of all vertices of G with distance exactly i from u. Then $(L_0, \ldots, L_k)$ is a levelling, with height at most $1/c$; and $|L_k| \geq |G|/4$, so the theorem holds for $n = 1$. Choose $L'_{k-1} \subseteq L_{k-1}$ minimal such that at least $|G|/4$ vertices in L_k have a neighbour in L'_{k-1}, and let L'_k be the set of vertices in L_k that have a neighbour in L'_{k-1}. Thus $|L'_k| \leq (1/4+\varepsilon)|G|$ since G is ε-sparse. Let $\mathcal{L}_1$ be the levelling $(L_0, \ldots, L_{k-2}, L'_{k-1}, L'_k)$, and let H_1 be its heart. Thus $|V(\mathcal{L}_1)| \leq (3/4+\varepsilon)|G|$. Let W be the set of vertices of G not in $V(\mathcal{L}_1)$; so $|W| \geq (1/4-\varepsilon)|G|$. Since W is anticomplete to H_1, and $1/4 - \varepsilon \geq \varepsilon$ and G is $(\varepsilon|G|^{1-c}, \varepsilon|G|)$-coherent, it follows that $|H_1| \leq \varepsilon|G|^{1-c}$, and so $|W| \geq (3/4-\varepsilon)|G| - \varepsilon|G|^{1-c} \geq |G|/2$. Hence $G[W]$ is (2ε)-sparse and $((2\varepsilon)|W|^{1-c}, (2\varepsilon)|W|)$-coherent. From the inductive hypothesis applied to $G[W]$, there is a covering sequence $(\mathcal{L}_2, \ldots, \mathcal{L}_n)$ in $G[W]$, where $\mathcal{L}_i = (a_i, H_i, B_i)$ for $2 \leq i \leq n$, such that:

- for $2 \leq i < j \leq n$, H_i is anticomplete to B_j;
- for $2 \leq i \leq n$, $\mathcal{L}_i$ has height at most $1/c$; and
- for $2 \leq i \leq n$, $|B_i| \geq 2^{-i}|W| \geq 2^{-i-1}|G|$.

But then $(\mathcal{L}_1, \ldots, \mathcal{L}_n)$ satisfies the theorem. This proves 5.1. ∎

5.2 *Let $n \geq 0$ be an integer, let $m = (n-1)^2 + 1$ and let $\varepsilon = 2^{-m-2} = 2^{-n^2+2n-4}$. Let $c > 0$ such that $1/c$ is an integer, and let G be an ε-sparse $(\varepsilon|G|^{1-c}, \varepsilon|G|)$-coherent graph. Then there is a covering sequence $(\mathcal{L}_1, \ldots, \mathcal{L}_n)$ in G, where $\mathcal{L}_i = (a_i, H_i, B_i)$ for $1 \leq i \leq n$, such that:*

- *for $1 \leq i \leq n$, $\mathcal{L}_i$ has height at most $1/c$;*
- *for $1 \leq i \leq n$, $|B_i| \geq 2^{-m-1}|G|$;*
- *either $B_1, \ldots, B_n$ are pairwise disjoint and H_i is anticomplete to B_j for all distinct $i, j \in \{1, \ldots, n\}$, or $B_1 = B_2 = \cdots = B_n$.*

Proof. Choose $\mathcal{L}_1, \dots, \mathcal{L}_m$ as in 5.1, such that each $\mathcal{L}_i$ has base of cardinality at least $2^{-i-1}|G|$ and height at most $1/c$. Let $\mathcal{L}_i = (a_i, H_i, B_i)$ for $1 \le i \le m$. For $1 \le i \le m$, $H_1, \dots, H_{i-1}$ are anticomplete to B_i, but $H_{i+1}, \dots, H_m$ might have neighbours in B_i. Choose $B'_i \subseteq B_i$ of cardinality at least $|B_i|/2^{m-i} \ge |G|/2^{m+1}$, such that for each $j \in \{i+1, \dots, m\}$, either every vertex in B'_i has a neighbour in H_j, or none do. Let $\mathcal{L}'_i$ be the covering obtained from $\mathcal{L}_i$ by replacing its base by B'_i. Then $(\mathcal{L}'_1, \dots, \mathcal{L}'_m)$ is a covering sequence, and for $1 \le i < j \le m$, H_i is anticomplete to B'_j, and either H_j is anticomplete to B'_i or H_j covers B'_i. If some B'_i is covered by H_j for at least n values of j, the second outcome of the third bullet of the theorem holds, so we assume not. Let $i_1 = 1$, and inductively for $2 \le k \le n$, let $i_k \in \{1, \dots, m\}$ be minimum such that H_{i_k} is anticomplete to each of $B'_{i_1}, \dots, B'_{i_{k-1}}$. This is possible since each of $B'_{i_1}, \dots, B'_{i_{k-1}}$ is covered by H_j for at most $n-1$ values of j, and $m > (n-1)(k-1)$. It follows that $B'_{i_1}, \dots, B'_{i_n}$ are pairwise disjoint. Moreover, $i_1 < \cdots < i_n$, and so the third bullet of the theorem holds. This proves 5.2. ∎

Now we prove the main result of this section. Its proof is closely related to the proof of the main theorem of [10].

5.3 *Let $c > 0$ such that $1/c$ is an integer, and let $n \ge 0$ be an integer. Let $\varepsilon = 2^{-2^{2n}}$. If G is an ε-sparse $(\varepsilon|G|^{1-c}, \varepsilon|G|)$-coherent graph, then G contains a multicovering of length n and height at most $1 + 1/c$, and with base of cardinality at least $3\varepsilon|G|$.*

Proof. Define $q = 2^n$, $p = (q-1)^2 + 1$, and $x = 2^{-p-1}$. It follows that $\varepsilon \le x3^{-n}$ and $\varepsilon \le 2^{-p-2}$. From 5.2 (with m, n replaced by p, q), we may assume that there is a covering sequence $(\mathcal{L}_1, \dots, \mathcal{L}_q)$ in G, such that:

- $V(\mathcal{L}_1), \dots, V(\mathcal{L}_q)$ are pairwise disjoint;
- for $1 \le i \le q$, $\mathcal{L}_i$ has height at most $1/c$;
- for $1 \le i \le q$, the base of $\mathcal{L}_i$ has cardinality at least $x|G|$; and
- for all distinct $i, j \in \{1, \dots, q\}$, every edge between $V(\mathcal{L}_i)$ and $V(\mathcal{L}_j)$ is between the base of $V(\mathcal{L}_i)$ and the base of $V(\mathcal{L}_j)$.

Let $t, d_1, \dots, d_t > 0$ be integers, where $d_1, \dots, d_t \le n$. Let us say a *battery* with *length* t of *type* $(d_1, \dots, d_t)$ is a sequence of t multicoverings $(\mathcal{M}_1, \dots, \mathcal{M}_t)$ in G, such that:

- $V(\mathcal{M}_1), \dots, V(\mathcal{M}_t)$ are pairwise disjoint;
- for $1 \le i \le t$, $\mathcal{M}_i$ has length d_i, and height at most $1+1/c$, and the first term of $\mathcal{M}_i$ has height at most $1/c$;
- for $1 \le i \le t$, the base of $\mathcal{M}_i$ has cardinality at least $x3^{1-d_i}|G|$;
- for all distinct $i, j \in \{1, \dots, q\}$, every edge between $V(\mathcal{M}_i)$ and $V(\mathcal{M}_j)$ is between the base of $V(\mathcal{M}_i)$ and the base of $V(\mathcal{M}_j)$.

Thus G contains a battery of type $(1, \dots, 1)$, and of length q. Choose a battery $\mathcal{B}$ of type $(d_1, \dots, d_t)$ with t minimum such that $2^{d_1} + \cdots + 2^{d_t} \ge q$. Let $\mathcal{B} = (\mathcal{M}_1, \dots, \mathcal{M}_t)$. For $1 \le i \le t$, let the base of $\mathcal{M}_i$ be B_i. For each i, $|B_i| \ge x3^{1-n}|G| \ge 3\varepsilon|G|$. If some $d_i = n$, then the ith term of $\mathcal{B}$ is

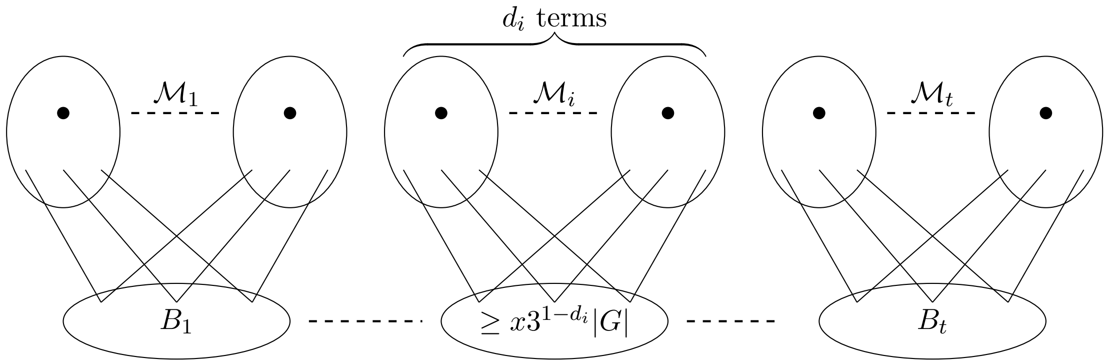

Figure 2: A battery of type $(d_1, \ldots, d_t)$

a multicovering satisfying the theorem; so we assume that $d_1, \ldots, d_t < n$. In particular, $2^{d_1} < 2^n = q$, and so $t \geq 2$. By reordering the terms of the battery, we may assume that $d_t \leq d_1, \ldots, d_{t-1}$. Since G is $(\varepsilon|G|^{1-c}, \varepsilon|G|)$-coherent, and $|B_t| \geq \varepsilon|G|$, for $1 \leq i < t$ there are fewer than $\varepsilon|G|^{1-c} \leq 2|B_i|/3$ vertices in B_i that have no neighbour in B_t. Hence we may choose $X \subseteq B_t$ minimal such that for some $i \in \{1, \ldots, t-1\}$, at least $|B_i|/3$ vertices in B_i have a neighbour in X. For $1 \leq i < t$, let Y_i be the set of vertices in B_i that have a neighbour in X, and $Z_i = B_i \setminus Y_i$. By reordering, we may assume that $|Y_1| \geq |B_1|/3$. From the minimality of X, $|Y_i| \leq |B_i|/3 + \varepsilon|G|$ for $2 \leq i \leq t-1$, and so $|Z_i| \geq 2|B_i|/3 - \varepsilon|G| \geq |B_i|/3$. Let $\mathcal{M}_1 = (\mathcal{L}_1, \ldots, \mathcal{L}_{d_1})$, and let the first term of $\mathcal{M}_t$ be $\mathcal{L} = (a, H, B_t)$. Let $\mathcal{L}'_1$ be the covering $(a, H \cup X, Y_1)$, which therefore has height at most $1 + 1/c$. Let $\mathcal{M}'_1$ be obtained from $\mathcal{M}_1$ by replacing its base by Y_1 and adding a new final term $\mathcal{L}'_1$; so $\mathcal{M}'_1$ has length $d_1 + 1$. For $2 \leq i \leq t-1$, let $\mathcal{M}'_i$ be obtained from $\mathcal{M}_i$ by replacing its base by Z_i. Then $(\mathcal{M}'_1, \ldots, \mathcal{M}'_{t-1})$ is a battery of type $(d_1 + 1, d_2, \ldots, d_{t-1})$. Since $d_1 \geq d_t$, it follows that

$$2^{d_1+1} + \cdots + 2^{d_{t-1}} \geq 2^{d_1} + \cdots + 2^{d_t} \geq 2^q,$$

a contradiction to the choice of $\mathcal{B}$. This proves 5.3. ∎

6 Making spiders

Let $\mathcal{L}_1, \ldots, \mathcal{L}_n$ be coverings in G, such that

- $\mathcal{L}_1, \ldots, \mathcal{L}_n$ all have the same apex a;
- for $1 \leq i \leq n$ let $\mathcal{L}_i$ have heart H_i; then for $1 \leq i < j \leq n$, $H_i \setminus \{a\}$ is disjoint from and anticomplete to $H_j \setminus \{a\}$.

We call $(a, \mathcal{L}_1, \ldots, \mathcal{L}_n)$ a *spider* in G, and a is its *apex*. Its *height* is the maximum of the heights of $\mathcal{L}_1, \ldots, \mathcal{L}_n$, and its *length* is n. It has *mass* b where b is the minimum cardinality of the bases of $\mathcal{L}_1, \ldots, \mathcal{L}_n$. The union of the hearts of $\mathcal{L}_1, \ldots, \mathcal{L}_n$ is called the *heart* of the spider. We call $\mathcal{L}_1, \ldots, \mathcal{L}_n$ the *members* of the spider.

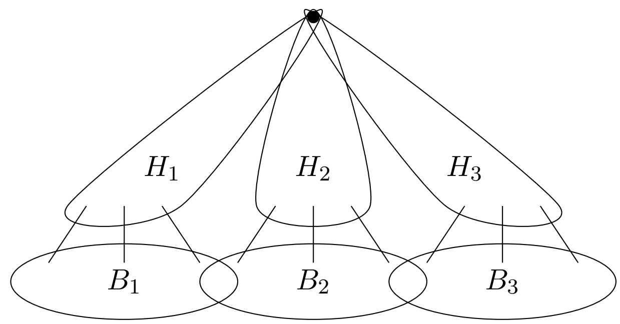

Figure 3: A spider of length three.

6.1 *Let $c > 0$ such that $1/c$ is an integer, let $n \geq 1$ be an integer, and let $\varepsilon = 2^{-2^{2n}}$. If G is an ε-sparse $(\varepsilon|G|^{1-c}, \varepsilon|G|)$-coherent graph with $|G| \geq 2$, then G contains a spider of length n and height at most $2 + 2/c$, and with mass at least $\varepsilon|G|$.*

Proof. 5.3 implies that G contains a multicovering $(\mathcal{L}_1, \ldots, \mathcal{L}_n)$ of length n and height at most $1 + 1/c$, and with base B of cardinality at least $3\varepsilon|G|$. Choose $a \in B$. Let $1 \leq i \leq n$, and let H_i be the heart of $\mathcal{L}_i$. Then every vertex of $H_i \cup \{a\}$ can be joined to a by a path of $G[H_i \cup \{a\}]$ with length at most $1 + 2/c$. Hence $(a, H_i \cup \{a\}, B \setminus \{a\})$ is a covering of height at most $2 + 2/c$, say $\mathcal{L}'_i$. Consequently $(\{a\}, \mathcal{L}'_1, \ldots, \mathcal{L}'_n)$ is a spider of length n and height at most $2 + 2/c$, and mass $|B| - 1 \geq 3\varepsilon|G| - 1 \geq \varepsilon|G|$, since $|G| \geq 1/\varepsilon$ by 2.1. This proves 6.1. ∎

A *troupe* of spiders is a set of spiders such that their hearts are pairwise disjoint and anticomplete.

6.2 *Let $c > 0$ such that $1/c$ is an integer, and let $m, n \geq 1$ be integers. Let let $\varepsilon^{-1} = 2^{2^{2n}} + 3(m-1)n$. If G is an ε-sparse $(\varepsilon|G|^{1-c}, \varepsilon|G|)$-coherent graph, then G contains a troupe of m spiders, each of length n and height at most $2 + 2/c$, and with mass at least $\varepsilon|G|$.*

Proof. We proceed by induction on m. The result is true if $m = 1$, from 6.1; so we assume that $m \geq 2$, and the result holds for $m - 1$. From 6.1 it follows that G contains a spider of length n and height at most $2 + 2/c$, and with mass at least $\varepsilon|G|$; say $\mathcal{S}_1 = (a_1, \mathcal{L}_1, \ldots, \mathcal{L}_n)$. For $1 \leq i \leq n$, let H_i be the heart of $\mathcal{L}_i$. Thus every vertex of H_i has $G[H_i]$-distance from a_1 at most $1 + 2/c$, and there are at least $\varepsilon|G|$ vertices in $V(G) \setminus H_i$ with a neighbour in H_i. Let us choose $\mathcal{S}_1$ such that each H_i is minimal with these two properties (that is, there are at least $\varepsilon|G|$ vertices in $V(G) \setminus H_i$ with a neighbour in H_i, and every vertex of H_i can be joined to a_1 by a path of $G[H_i]$ with length at most $1 + 2/c$.) Let B_i be the set of vertices in $V(G) \setminus H_i$ with a neighbour in H_i.

(1) *For $1 \leq i \leq n$, $|H_i| < \varepsilon|G|^{1-c}$, and $|B_i| < 2\varepsilon|G|$.*

Let $H_i = \{v_1, \ldots, v_t\}$, ordered with increasing $G[H_i]$-distance from a_1 (and hence $v_1 = a_1$). Every vertex in B_i either has a neighbour in $H_i \setminus \{v_t\}$ or is adjacent to v_t; there are fewer than $\varepsilon|G|$ vertices in B_i with a neighbour in $H_i \setminus \{v_t\}$, from the minimality of H_i, and there are fewer than $\varepsilon|G|$ vertices in B_i that are adjacent to v_t, since G is ε-sparse. Thus $|B_i| < 2\varepsilon|G|$. Let $j = \lceil \varepsilon|G|^{1-c} \rceil$, and suppose that $t \geq j$. Let $J = \{v_1, \ldots, v_j\}$. Thus $|J| \leq \varepsilon|G| + 1 \leq 2\varepsilon|G|$ by 2.1, and so $|V(G) \setminus J| \geq (1 - 2\varepsilon)|G|$. Since G is $(\varepsilon|G|^{1-c}, \varepsilon|G|)$-coherent, there are fewer than $\varepsilon|G|$ vertices in $V(G) \setminus J$ that have no neighbour in J, and so there are at least $(1 - 3\varepsilon)|G| \geq 2\varepsilon|G|$ vertices in $V(G) \setminus J$ that have a neighbour in J. Thus $t = j$ and $H_i = J$, from the minimality of H_i; but this is impossible since $|B_i| < 2\varepsilon|G|$. This proves that $t < j$, and so proves (1).

From (1),

$$|H_1 \cup \cdots \cup H_n \cup B_1 \cup \cdots \cup B_n| \leq 3\varepsilon n|G|.$$

Let X be the complement in $V(G)$ of this set; thus $|X| \geq (1 - 3\varepsilon n)|G|$. Let

$$(\varepsilon')^{-1} = 2^{2^{2n}} + 3(m-2)n = \varepsilon^{-1} - 3n.$$

Thus $\varepsilon'(1 - 3n\varepsilon) = \varepsilon$, and so $\varepsilon'|X| \geq \varepsilon|G|$. It follows that $G[X]$ is ε'-sparse and $(\varepsilon'|X|^{1-c}, \varepsilon'|X|)$-coherent. From the inductive hypothesis applied to $G[X]$, we deduce that there is a troupe of $m-1$ spiders in $G[X]$, each of length n and height at most $2 + 2/c$, and with mass at least $\varepsilon'|X| \geq \varepsilon|G|$. But then adding $\mathcal{S}_1$ to this troupe gives a troupe of m spiders satisfying the theorem. This proves 6.2. ∎

So, our graph contains a troupe of spiders, of arbitrarily large cardinality, and each with arbitrarily large length, all of height at most $2+2/c$, and with bases of linear cardinality. The next result converts the members of these spiders to levellings, but we need to be careful exactly what we mean. In a levelling, all edges from heart to base start from the penultimate level of the levelling. We need more than this: we need that for every two levellings that are members of spiders in the troupe, every edge from the heart of one to the base of the other starts from the penultimate level of the first, and this is more tricky to arrange.

Let us first state the definition formally. Let $n \geq 1$ and let $\mathcal{L}_1, \ldots, \mathcal{L}_n$ be levellings in a graph G, all with the same apex a, such that

- for $1 \leq i \leq n$, let H_i be the heart of $\mathcal{L}_i$; then $H_1 \setminus \{a\}, \ldots, H_n \setminus \{a\}$ are pairwise disjoint (the bases of $\mathcal{L}_1, \ldots, \mathcal{L}_n$ may intersect);
- for all distinct $i, j \in \{1, \ldots, n\}$, every edge of G between $H_i \setminus \{a\}$ and $V(\mathcal{L}_j) \setminus \{a\}$ has one end in the penultimate level of $\mathcal{L}_i$ and the other in the base of $\mathcal{L}_j$.

We call $(a, \mathcal{L}_1, \ldots, \mathcal{L}_n)$ a *lobster* in G, and a is its *apex*. Its *height* is the maximum height of $\mathcal{L}_1, \ldots, \mathcal{L}_n$, and its *length* is n. It has *mass* b where b is the minimum cardinality of the bases of $\mathcal{L}_1, \ldots, \mathcal{L}_n$. Its *heart* is the union of the hearts of $\mathcal{L}_1, \ldots, \mathcal{L}_n$. We call $\mathcal{L}_1, \ldots, \mathcal{L}_n$ the *members* of the lobster.

A *troupe* of lobsters is a set $\{\mathcal{T}_1, \ldots, \mathcal{T}_m\}$ of lobsters, such that for all $i, j \in \{1, \ldots, m\}$:

- for $1 \leq i < j \leq m$, the heart of $\mathcal{T}_i$ is disjoint from and anticomplete to the heart of $\mathcal{T}_j$;
- let $\mathcal{L}, \mathcal{M}$ each be a member of one of $\mathcal{T}_1, \ldots, \mathcal{T}_m$, with $\mathcal{L} \neq \mathcal{M}$, and let $\mathcal{L} = (L_0, \ldots, L_k)$; then there is no edge between $L_0 \cup \cdots \cup L_{k-2}$ and the base of $\mathcal{M}$.

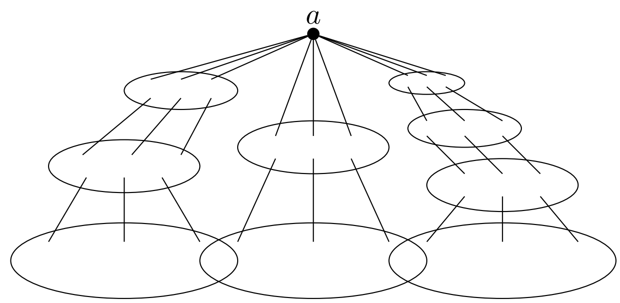

Figure 4: A lobster of length three.

6.3 *Let $c > 0$ such that $1/c$ is an integer, and let $m, n \geq 1$ be integers. Let*

$$\varepsilon^{-1} = (2^{2^{2n}} + 3(m-1)n)(2+2/c)^{mn}.$$

If G is an ε-sparse $(\varepsilon|G|^{1-c}, \varepsilon|G|)$-coherent graph, then G contains a troupe of m lobsters, each of length n and height at most $2+2/c$, and with mass at least $\varepsilon|G|$.

Proof. Let G be an ε-sparse $(\varepsilon|G|^{1-c}, \varepsilon|G|)$-coherent graph. By 6.2 with ε replaced by $(2+2/c)^{mn}\varepsilon$, there is a troupe of spiders $\{\mathcal{S}_1, \ldots, \mathcal{S}_m\}$ in G, each of length n and height at most $2+2/c$, and with mass at least $(2+2/c)^{mn}\varepsilon|G|$. Let the members of these spiders (in some order) be $\mathcal{L}_1, \ldots, \mathcal{L}_{mn}$, and for $1 \leq i \leq mn$ let $\mathcal{L}_i = (a_i, H_i, B_i)$. (Thus, some of $a_1, \ldots, a_{mn}$ may be equal.) We shall convert these members one by one to levellings, at each step shrinking all the bases.

Let $X^0 = B_1 \cup \cdots \cup B_{mn}$, and for $1 \leq i \leq mn$ let X_i^0 be the set of all vertices in X^0 with a neighbour in H_i (thus $B_i \subseteq X_i^0$). Inductively, let $1 \leq h \leq mn$, and suppose that we have defined X^{h-1} and $\mathcal{L}'_1, \ldots, \mathcal{L}'_{h-1}$, and X_i^{h-1} for $1 \leq i \leq mn$, satisfying:

- for $1 \leq i \leq h-1$, $\mathcal{L}'_i$ is a levelling; its heart is a subset of H_i, and a_i is its apex; its height is at most $2+2/c$;
- for $1 \leq i \leq h-1$, X_i^{h-1} is the set of all vertices in X^{h-1} with a neighbour in the heart of $\mathcal{L}'_i$, and for $h \leq i \leq mn$, X_i^{h-1} is the set of all vertices in X^{h-1} with a neighbour in the heart of $\mathcal{L}_i$;
- for $1 \leq i \leq h-1$, every edge between the heart of $\mathcal{L}'_i$ and X^{h-1} has an end in the penultimate level of $\mathcal{L}'_i$; and
- for $1 \leq i \leq mn$, $|X_i^{h-1}| \geq (2+2/c)^{mn+1-h}\varepsilon|G|$.

For $0 \leq j \leq 1+2/c$, let L_j be the set of vertices in H_h with $G[H_h]$-distance to a_h exactly j. Thus every vertex $v \in X_h^{h-1}$ has a neighbour in some L_j where $j \in \{0, \ldots, 1+2/c\}$, and the smallest such j is called the *type* of v. There are only $2+2/c$ possible types, and so there exists $k \in \{0, \ldots, 1+2/c\}$ such that at least $|X_h^{h-1}|/(2+2/c)$ vertices in X_h^{h-1} have type k. Consequently, since

$$|X_h^{h-1}|/(2+2/c) \geq (2+2/c)^{mn+1-h}\varepsilon|G|/(2+2/c) = (2+2/c)^{mn-h}\varepsilon|G|,$$

there exists $k \in \{0, \dots, 1+2/c\}$ minimum such that at least $(2+2/c)^{mn-h}\varepsilon|G|$ vertices in X_h^{h-1} have type k. Let X_h^h be the set of all vertices in X_h^{h-1} that have type k, and let $\mathcal{L}'_h = (L_0, \dots, L_k, X_h^h)$. Thus $\mathcal{L}'_h$ is a levelling with height $k+1 \le 2+2/c$.

Let Z^h be the set of vertices in X_h^{h-1} with type less than k. Thus

$$|Z^h| \le (1+2/c)(2+2/c)^{mn-h}\varepsilon|G|.$$

For $1 \le i \le mn$ with $i \ne h$, define $X_i^h = X_i^{h-1} \setminus Z^h$. Thus $|X_i^h| \ge |X_i^{h-1}| - |Z^h|$, and so

$$|X_i^h| \ge (2+2/c)^{mn+1-h}\varepsilon|G| - (1+2/c)(2+2/c)^{mn-h}\varepsilon|G| = (2+2/c)^{mn-h}\varepsilon|G|.$$

Let X^h be the union of the sets X_i^h $(1 \le i \le n)$. This completes the inductive definition.

For $1 \le i \le m$, let $\mathcal{T}_i$ be the lobster obtained from $\mathcal{S}_i$ by replacing each of its members $\mathcal{L}_j$ by $\mathcal{L}'_j$. This makes a troupe of lobsters satisfying the theorem, and so proves 6.3. ∎

7 Part assembly

Now we put these several pieces together to prove 1.9, which we restate:

7.1 *Let $c > 0$ with $1/c$ an integer, and let H_1, H_2 be graphs with branch-length at least $4/c+5$. Then there exists $\varepsilon > 0$ such that if G is a graph with $|G| > 1$ that is H_1-free and $\overline{H_2}$-free, then there is a pure pair A, B in G with $|A| \ge \varepsilon|G|$ and $|B| \ge \varepsilon|G|^{1-c}$.*

As we saw in section 2, to prove 7.1, it suffices to prove the following:

7.2 *Let $c > 0$ with $1/c$ an integer, and let H be a graph with branch-length at least $4/c+5$. Then there exists $\varepsilon > 0$ such that if G is an ε-sparse $(\varepsilon|G|^{1-c}, \varepsilon|G|)$-coherent graph, then G contains H.*

Proof. By adding more vertices to H, we may assume that if X denotes the set of vertices of H that have degree different from two, then every cycle of H contains at least one vertex in X, and every path in H with both ends in X has length at least $4/c+5$, and every cycle of H has length at least $4/c+5$. Let $X = \{x_1, \dots, x_m\}$. Consequently H can be obtained from the set X of m isolated vertices by adding

- paths with ends in X and each of length at least $4/c+5$, and
- cycles with exactly one vertex in X, of length at least $4/c+5$

where every vertex of $V(H) \setminus X$ belongs to exactly one of these paths and cycles, and has degree exactly two in H. Let the paths be R_i $(i \in I_1)$, and let the cycles be R_i $(i \in I_2)$, where $I_1 \cap I_2 = \emptyset$. For $i \in I_1$, let R_i have ends (u_i, v_i) (ordered arbitrarily) and have length ℓ_i, and for $i \in I_2$, let $u_i = v_i$ be the unique vertex of R_i in X, and let R_i have length ℓ_i. Thus H is determined up to isomorphism by a knowledge of X, the pairs (u_i, v_i) $(i \in I_1 \cup I_2)$, and the numbers ℓ_i $(i \in I_1 \cup I_2)$. Let $I = I_1 \cup I_2$, and for each $i \in I$ let $\alpha_i, \beta_i \in \{1, \dots, m\}$ such that $x_{\alpha_i} = u_i$ and $x_{\beta_i} = v_i$. Let $I = \{1, \dots, p\}$.

Let n be the maximum degree of H, and let $d^{-1} = (2^{2^{2n}} + 3(m-1)n)(2+2/c)^{mn}$. Choose $\varepsilon > 0$ with $3(2+1/c)^{|H|^2}(4p)^p\varepsilon < d$. We claim that ε satisfies the theorem. Let G be an ε-sparse

$(\varepsilon|G|^{1-c}, \varepsilon|G|)$-coherent graph. We must show that G contains H. By 6.3, G contains a troupe $\{\mathcal{S}_1, \ldots, \mathcal{S}_m\}$ of m lobsters, each of length n and height at most $2 + 2/c$, and with mass at least $d|G|$. For $1 \le i \le p$, choose a member $\mathcal{L}_{2i-1}$ of $\mathcal{S}_{\alpha_i}$ and a member $\mathcal{L}_{2i}$ of $\mathcal{S}_{\beta_i}$, in such a way that the levellings $\mathcal{L}_i, \mathcal{M}_i$ $(i \in I)$ are all different. (This is possible from the definition of n).

We will prove that for all $h \in I$, there is a path P_h (or cycle, if the two apexes are equal) between the apex of $\mathcal{L}_{2h-1}$ and the apex of $\mathcal{L}_{2h}$ of length ℓ_h, such that the union of $P_1, \ldots, P_p$ makes an induced subgraph of G isomorphic to H. We will choose these paths and cycles in order. Also for $1 \le h \le p$ we need to choose a subset X_k^h of the base of each $\mathcal{L}_k$ for $2h < k \le 2p$, and a subset Y_k^h of the penultimate level of $\mathcal{L}_k$, with properties that we will describe below. We denote by P_h^* the set of vertices of P_h different from its ends (if it is a path) or different from the apex of A_h (if it is a cycle). In either case $|P_h^*| = \ell_h - 1$.

For $0 \le h \le p$ let $w_h = (4p)^{-h}d$. Let B be the union of the bases of $\mathcal{L}_1, \ldots, \mathcal{L}_{2\rho}$. For $1 \le i \le 2p$, let Y_i^0 be the penultimate level of $\mathcal{L}_i$, let X_i^0 be the set of vertices in B with a neighbour in Y_i^0, and let a_i be the apex of $\mathcal{L}_i$. Thus, $|X_i^0| \ge w_0|G|$. Now inductively, suppose we have chosen the first $h-1$ paths or cycles, say $P_1, \ldots, P_{h-1}$, where $1 \le h \le p$, satisfying:

- for $1 \le g \le h-1$, if $a_{2g-1} \ne a_{2g}$, then P_g is an induced path joining these apexes, of length ℓ_g; and if the apexes are equal then P_g is a cycle of length ℓ_g containing this apex;
- for $1 \le g \le h-1$, and $2h+1 \le i \le 2p$, every vertex of the heart of $\mathcal{L}_i$ with a neighbour in P_g^* belongs to the penultimate level of $\mathcal{L}_i$.

Suppose moreover that for $2h-1 \le i \le 2p$ we have chosen $X_i^{h-1} \subseteq X_i^0$ and $Y_i^{h-1} \subseteq Y_i^0$, such that for all $i \in \{2h-1, \ldots, 2p\}$:

- X_i^{h-1} is the set of all vertices in B with a neighbour in Y_i^{h-1};
- $X_i^{h-1} \cup Y_i^{h-1}$ is anticomplete to $P_1^*, \ldots, P_{h-1}^*$; and
- $|X_i^{h-1}| \ge w_{h-1}|G|$.

We choose P_h as follows. For $2h+1 \le i \le 2p$, choose $Y_i^h \subseteq Y_i^{h-1}$ minimal such that at least $(w_h + \varepsilon(|H|-1))|G|$ vertices in B (necessarily all in X_i^{h-1}) have a neighbour in Y_i, and let X_i be the set of vertices in B with a neighbour in Y_i. From the minimality of Y_i,

$$(w_h + \varepsilon(|H|-1))|G| \le |X_i| \le (w_h + \varepsilon|H|)|G|.$$

Now $\varepsilon|H| \le w_h$, so $|X_i| \le 2w_h$. Let $Z = X_{2h+1} \cup \cdots \cup X_{2p}$; thus $|Z| \le 2(2p-2)w_h|G|$. For $i = 2h-1, 2h$ let $X_i = X_i^{h-1} \setminus Z$. Thus

$$|X_i| \ge |X_i^{h-1}| - |Z| \ge (w_{h-1} - 4(p-1)w_h)|G| \ge w_h|G| \ge (4p)^{-p}d|G| = 3(2+1/c)^{|H|^2}\varepsilon|G|$$

for $i = 2h-1, 2h$.

For $i = 2h-1, 2h$ let $\mathcal{L}_i'$ be the levelling obtained from $\mathcal{L}_i$ by replacing its base by X_i. Now $\mathcal{L}_{2h-1}'$, $\mathcal{L}_{2h}'$ both have height at most $2+2/c$, and $\ell_h \ge 5 + 4/c$. By 3.4 applied to the levellings $\mathcal{L}_{2h-1}'$, $\mathcal{L}_{2h}'$, there is an induced path P_h of length ℓ_h between a_{2h-1}, a_{2h} (or a cycle, if $a_{2h-1} = a_{2h}$), with vertex set included in $V(\mathcal{L}_{2h-1}') \cup V(\mathcal{L}_{2h}')$. Consequently P_h^* is anticomplete to Y_i for $2h+1 \le i \le 2p$, and to $P_1^*, \ldots, P_{h-1}^*$. It might have neighbours in X_i for $2h+1 \le i \le 2p$, but since $|P_h^*| \le |H| - 1$, there are at most $\varepsilon(|H|-1)|G|$ such vertices. For $2h+1 \le i \le 2p$, let X_i^h be the set of vertices in X_i with no neighbour in P_h^*. Thus $|X_i^h| \ge |X_i| - \varepsilon(|H|-1)|G| \ge w_h|G|$. This completes the inductive definition.

But then the union of $P_1, \dots, P_p$ forms an induced subgraph isomorphic to H. This proves 7.2, and hence completes the proof of 1.9. ∎

One might wonder how ε in 1.9 depends on c, H_1, H_2. For simplicity let us assume that if X denotes the set of vertices of H_1 that have degree different from two, then every cycle of H_1 contains at least one vertex in X, and every path in H_1 with both ends in X has length at least $4/c+5$, and every cycle of H_1 has length at least $4/c+5$; and the same for H_2. Let $r = \max(|H_1|, |H_2|)$. Then one can check that (with $H = H_1, H_2$) defining d, ε as in the proof of 7.2 yields a value of ε that satisfies 7.2, with $\log\log(1/\varepsilon) = O(r)$. Next we need a version of 2.2 with an explicit dependence of δ on η, and for this we can use the proof of 2.2 due to Fox and Sudakov [9]; this can be used to show that 2.2 holds where $\log 1/\delta = O(|H|(\log 1/\eta)^2)$. The argument given in section 2 that 2.3 implies 1.9 shows that if ε satisfies 7.2, and δ satisfies 2.2, then $\varepsilon' = \varepsilon\delta$ satisfies 1.9. Putting these pieces together, we deduce that there exists ε with $\log\log(1/\varepsilon) = O(r)$ that satisfies 1.9.

8 Further extension

We have found a kind of strengthening of 1.9, that we state without proof. For $\ell \ge 2$, let us say a graph H is *ℓ-handled* if there are induced subgraphs $P_0, \dots, P_k$ of H, for some $k \ge 1$, such that:

- P_0 is a forest;
- every path of P_0 has length at most ℓ;
- $P_1, \dots, P_k$ are pairwise vertex-disjoint paths, each of length at least ℓ;
- for $1 \le i \le k$, $V(P_i \cap P_0)$ consists exactly of the two ends of P_i; and
- $H = P_0 \cup P_1 \cup \dots \cup P_k$.

Then:

8.1 *There exists $\gamma > 0$ with the following property. Let $c > 0$ with $1/c$ an integer, and let H_1, H_2 be γ/c-handled graphs. Then there exists $\varepsilon > 0$ such that if G is a graph with $|G| > 1$ that is H_1-free and $\overline{H_2}$-free, then there is a pure pair A, B in G with $|A| \ge \varepsilon|G|$ and $|B| \ge \varepsilon|G|^{1-c}$.*

Then the essentials of 1.9 follow from 8.1 by taking P_0 to be the subgraph of H induced on the set of all vertices of degree at least three and their neighbours. But we feel that 8.1 is not very satisfactory, because if the forest P_0 has long paths, the hypothesis requires the paths $P_1, \dots, P_k$ to be long too. We would prefer a version of 8.1 where we omit the second bullet from the definition of ℓ-handled, but so far we cannot prove it.

A weaker form of 1.9 will be proved for a wider class of graphs in [12]. Let H be a graph. If $E(H) \neq \emptyset$, we define the *congestion* of H to be the maximum of $1 - (|J|-1)/|E(J)|$, taken over all subgraphs J of H with at least one edge; and if $E(H) = \emptyset$, we define the congestion of H to be zero. Thus the congestion of H is always non-negative, and equals zero if and only if H is a forest; and, for instance, long cycles have smaller congestion than short cycles.

In [12] we will prove:

8.2 *Let $c > 0$, and let H_1, H_2 be graphs with congestion at most $c/(9+15c)$. Then there exists $\varepsilon > 0$ such that if G is a graph with $|G| > 1$ that is H_1-free and $\overline{H_2}$-free, then there is a pure pair A, B in G with $|A|, |B| \geq \varepsilon|G|^{1-c}$.*

This is pleasing because of the following weak converse (easily proved with a random graph argument that we omit):

8.3 *Let $c > 0$, and let H_1, H_2 be graphs both with congestion more than c. There is no $\varepsilon > 0$ such that for every graph G with $|G| > 1$ that is H_1-free and $\overline{H_2}$-free, there is a pure pair A, B in G with $|A|, |B| \geq \varepsilon|G|^{1-c}$.*

The result 8.2 does not contain 1.9, because in 8.2 neither of A, B have to have linear cardinality. What if we ask for a strengthened version of 8.2 that would contain 1.9 (by requiring one of $|A|, |B|$ to be linear)? We pose that as a conjecture:

8.4 Conjecture: *For all $c > 0$, there exists $\sigma > 0$ with the following property. Let H_1, H_2 be graphs with congestion at most σ. There exists $\varepsilon > 0$ such that if G is a graph with $|G| > 1$ that is H_1-free and $\overline{H_2}$-free, then there is a pure pair A, B in G with $|A| \geq \varepsilon|G|$ and $|B| \geq \varepsilon|G|^{1-c}$.*

Acknowledgement

The referee reports were very thorough and very hepful, and we would like to express our thanks.

References

[1] M. Chudnovsky, N. Robertson, P. Seymour and R. Thomas, "The strong perfect graph theorem", *Annals of Math.* **164** (2006), 51–229.

[2] 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`.

[3] 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`.

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

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

[6] J. Fox, "A bipartite analogue of Dilworth's theorem", *Order* **23** (2006), 197–209.

[7] J. Fox and J. Pach, "Erdős-Hajnal-type results on intersection patterns of geometric objects", in *Horizons of Combinatorics*, *Bolyai Soc. Math. Stud.* **17**, Springer, Berlin (2008), 79–103.

[8] J. Fox, J. Pach and C. Tóth, "Turán-type results for partial orders and intersection graphs of convex sets", *Israel J. Math.* **178** (2010), 29–50.

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

[10] A. Liebenau, M. Pilipczuk, P. Seymour and S. Spirkl, "Caterpillars in Erdős-Hajnal", *J. Combinatorial Theory, Ser. B*, **136** (2019), 33–43, `arXiv:1810.00811`.

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

[12] A. Scott, P. Seymour and S. Spirkl, "Pure pairs. VIII. Excluding a sparse graph", submitted for publication, `arXiv:2201.04062`.