Towards Erdős-Hajnal for graphs with no 5-hole

Maria Chudnovsky[1]
Princeton University, Princeton, NJ 08544

Jacob Fox[2]
Stanford University, Stanford, CA 94305-2125

Alex Scott[3]
Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK

Paul Seymour[4]
Princeton University, Princeton, NJ 08544

Sophie Spirkl
Princeton University, Princeton, NJ 08544

March 3, 2018; revised December 1, 2018

[1]Supported by NSF grant DMS-1550991. This material is based upon work supported in part by the U. S. Army Research Laboratory and the U. S. Army Research Office under grant number W911NF1610404.
[2]Supported by a Packard Fellowship and NSF Career Award DMS-1352121.
[3]Supported by a Leverhulme Trust Research Fellowship.
[4]Supported by ONR grant N00014-14-1-0084 and NSF grant DMS-1265563.

Abstract

The Erdős-Hajnal conjecture says that for every graph H there exists $c > 0$ such that

$$\max(\alpha(G), \omega(G)) \geq n^c$$

for every H-free graph G with n vertices, and this is still open when $H = C_5$. Until now the best bound known on $\max(\alpha(G), \omega(G))$ for C_5-free graphs was the general bound of Erdős and Hajnal, that for all H,

$$\max(\alpha(G), \omega(G)) \geq 2^{\Omega(\sqrt{\log n})}$$

if G is H-free. We improve this when $H = C_5$ to

$$\max(\alpha(G), \omega(G)) \geq 2^{\Omega(\sqrt{\log n \log\log n})}.$$

1 Introduction

All graphs in this paper are finite and have no loops or parallel edges, and the cardinalities of the largest stable sets and cliques in a graph G are denoted by $\alpha(G), \omega(G)$ respectively. If G, H are graphs, we say that G *contains* H if some induced subgraph of G is isomorphic to H, and G is H*-free* otherwise.

The Erdős-Hajnal conjecture [6, 7] asserts:

1.1 Conjecture: *For every graph H, there exists $\epsilon > 0$ such that every H-free graph G satisfies*

$$\max(\alpha(G), \omega(G)) \geq |V(G)|^{\epsilon}.$$

This is true for all H with at most four vertices, but is open when $H = C_5$ (C_5 denotes the cycle of length five). The problem for C_5 has attracted a good deal of unsuccessful attention, for several reasons; not only is C_5 arguably the smallest open case of 1.1, but also it is symmetrical, and more importantly, by excluding C_5 we exclude its complement as well. (Excluding both a graph and its complement is an approach that has been quite fruitful lately, for instance [1, 2].) So we are happy to report some progress at last.

The best general bound for the Erdős-Hajnal conjecture to date was proved by Erdős and Hajnal in [7], namely:

1.2 *For every graph H, there exists $c > 0$ such that*

$$\max(\alpha(G), \omega(G)) \geq 2^{c\sqrt{\log n}}$$

for every H-free graph G with $n > 0$ vertices.

(Logarithms are to base two, throughout the paper.) Until now, this was also the best bound known when $H = C_5$, but in this paper we will improve it to:

1.3 *There exists $c > 0$ such that*

$$\max(\alpha(G), \omega(G)) \geq 2^{c\sqrt{\log n \log\log n}}$$

for every C_5-free graph G with $n > 1$ vertices.

If $A, B \subseteq V(G)$ are disjoint and nonempty, the *edge-density* between them means the number of edges joining A, B, divided by $|A| \cdot |B|$. The proof of 1.3 is via the following conjecture of Conlon, Fox and Sudakov [5]:

1.4 Conjecture: *For every graph H there exist $\epsilon, \sigma > 0$ such that for every H-free graph G on $n > 1$ vertices, and all c with $0 \leq c \leq 1/2$, $V(G)$ contains two disjoint subsets A, B with $|A| \geq \epsilon c^{\sigma} n$ and $|B| \geq \epsilon n$, such that the edge-density between A, B is either at most c or at least $1 - c$.*

This has not been proved so far for any graph H with more than four vertices, but in this paper we prove it for $H = C_5$ (with $\sigma = 1$), and this is the key to proving 1.3. We first prove it for sparse graphs G, and then use a theorem of Rödl to deduce it in general (both in the next section). The proof of 1.3 is completed in section 3.

We remark that 1.4 (for all H) is equivalent to the same statement for sparse graphs (for all H), because of the theorem of Rödl discussed in the next section; a graph H satisfies the original version of 1.4 if and only if both H and its complement satisfy the sparse version. We can prove the sparse version of 1.4 for many more graphs H than just C_5 (for instance, for all bipartite H, and all cycles of length at least four); these results will appear in a later paper [3]. But C_5 is still the largest graph H for which we can show that both H and its complement satisfy the sparse version of 1.4, and so the largest for which we can prove the original version of 1.4.

2 Sparse graphs

In this section we prove 1.4 for $H = C_5$, and first we prove it when G is sufficiently sparse. For disjoint $A, B \subseteq V(G)$, we say A is *anticomplete* to B if there are no edges between A and B, and A *covers* B if every vertex in B has a neighbour in A. We will prove:

2.1 *For all c with $0 < c \le 1/2$, and every graph G with $n > 0$ vertices, if G satisfies:*

- *every vertex has degree at most $n/16 - 1$, and*
- *for every two disjoint subsets $A, B \subseteq V(G)$ with $|A| \ge cn/2$ and $|B| \ge n/16$, the edge-density between A, B is at least c,*

then G contains C_5.

Proof. Let $0 < c \le 1/2$, and let G, n be as in the theorem. Since every vertex has degree at most $n/16 - 1$, it follows that $n \ge 16$ and in particular, $\lfloor n/2 \rfloor \ge n/4$. Choose a set $N_0 \subseteq V(G)$ of cardinality $\lfloor n/2 \rfloor$. It follows that $|N_0| \ge n/4 \ge cn/2$, and so the edge-density between N_0 and its complement is at least c. In particular, some vertex in N_0 has at least $cn/2$ neighbours.

Let v_1 be a vertex of degree at least $cn/2$, let N_1 be the set of all neighbours of v_1, and let $Z_2 = V(G) \setminus (N_1 \cup \{v_1\})$. Since $|N_1| + 1 \le n/16$, it follows that $|Z_2| \ge 15n/16$. But $|N_1| \ge cn/2$, and so fewer than $n/16$ vertices in Z_2 have no neighbour in N_1, since $c > 0$. Hence at least $7n/8$ vertices in Z_2 do have such a neighbour. Choose $B_1 \subseteq N_1$ minimal such that B_1 covers at least $5n/16$ vertices in Z_2. Let B_2 be the set of vertices in Z_2 covered by B_1. Thus $5n/16 \le |B_2| \le 3n/8$ from the minimality of B_1, and since every vertex has degree at most $n/16$. Let $A_2 = Z_2 \setminus B_2$. Thus A_2 is anticomplete to B_1, and $|A_2| = |Z_2| - |B_2| \ge (15n/16 - 3n/8) = 9n/16$.

Let $A_1 = N_1 \setminus B_1$. Since $|N_1| \ge cn/2$, the edge-density between N_1, A_2 is at least c. In particular there is a vertex $v_2 \in A_1$ with at least $c|A_2| \ge 9cn/16 \ge cn/2$ neighbours in A_2. (Note that $v_2 \notin B_1$ since B_1 is anticomplete to A_2.) Let N_2 be the set of neighbours of v_2 in A_2. Thus $N_2 \cap B_2 = \emptyset$, but v_2 might also have neighbours in B_2. Let P_1 be the set of vertices in B_1 adjacent to v_2, and let Q be the set of vertices in B_2 that have a neighbour in $B_1 \setminus P_1$.

(1) *If $|Q| \ge n/8$ then G contains C_5.*

Assume that $|Q| \ge n/8$. Since v_2 has degree at most $n/16$, there is a set $Q' \subseteq Q$ of at least $n/16$ vertices that are nonadjacent to v_2. The edge-density between N_2 and Q' is at least c, since

$|N_2| \geq cn/2$, and in particular some vertex $q \in Q'$ has a neighbour $w \in N_2$. Since $q \in Q' \subseteq Q$, it is adjacent to some vertex $b_1 \in B_1$ that is nonadjacent to v_2; but then

$$b_1\text{-}v_1\text{-}v_2\text{-}w\text{-}q\text{-}b_1$$

is an induced cycle of length 5. (Note that b_1 is nonadjacent to w since B_1 is anticomplete to A_2.) This proves (1).

Let $Y_2 = A_2 \setminus N_2$; it follows that $|Y_2| \geq |A_2| - n/16 \geq n/2$. Since $|N_2| \geq cn/2$, the edge-density between N_2, Y_2 is at least c, and so some vertex $v_3 \in N_2$ has at least $c|Y_2| \geq cn/2$ neighbours in Y_2. Let N_3 be the set of neighbours of v_3 in Y_2. Let P_2 be the set of vertices in B_2 with a neighbour in P_1.

(2) *If $|P_2| \geq 3n/16$ then G contains C_5.*

Assume that $|P_2| \geq 3n/16$. It follows that there is a set $P_2' \subseteq P_2$ of at least $n/16$ vertices that are nonadjacent to both v_2, v_3. The edge-density between N_3 and P_2' is at least c, since $|N_3| \geq cn/2$, and in particular some vertex $p_2 \in P_2'$ has a neighbour $u \in N_3$. Since $p_2 \in P_2' \subseteq P_2$, it is adjacent to some vertex $p_1 \in P_1$; but then

$$p_1\text{-}v_2\text{-}v_3\text{-}u\text{-}p_2\text{-}p_1$$

is an induced cycle of length 5. (Note that p_1 is nonadjacent to v_3, u since B_1 is anticomplete to A_2.) This proves (2).

Since B_1 covers B_2, it follows that $P_2 \cup Q = B_2$, and since $|B_2| \geq 5n/16$, the result follows from (1) and (2). This proves 2.1. ∎

Next we apply a theorem of Rödl [9], the following. ($\overline{G}$ denotes the complement graph of G.)

2.2 *For every graph H and all $d > 0$ there exists $\delta > 0$ such that for every H-free graph G, there exists $X \subseteq V(G)$ with $|X| \geq \delta|V(G)|$ such that in one of $G[X]$, $\overline{G}[X]$, every vertex in X has degree at most $d|X|$.*

We deduce:

2.3 *There exists $\epsilon > 0$ such that for all c with $0 \leq c \leq 1/2$, if G is C_5-free with $n > 1$ vertices, then there exist disjoint $A, B \subseteq V(G)$ with $|A| \geq \epsilon cn$ and $|B| \geq \epsilon n$, such that the edge-density between A, B is either less than c or more than $1 - c$.*

Proof. Let δ satisfy 2.2, taking $d = 1/32$ and $H = C_5$. Now let $\epsilon = \delta/16$, and let G be C_5-free with $n > 1$ vertices. Let v be a vertex; then it has either at least $(n-1)/2$ neighbours or at least $(n-1)/2$ non-neighbours; and since $(n-1)/2 \geq \epsilon n$, we may assume that $1 < \epsilon cn$, for otherwise the theorem holds taking $A = \{v\}$. In particular $n > 2\epsilon^{-1} \geq 32\delta^{-1}$.

By 2.2, there exists $X \subseteq V(G)$ with $|X| \geq \delta n$ such that every vertex of J has degree at most $|V(J)|/32$, where J is one of $G[X]$, $\overline{G}[X]$. Since $|V(J)| \geq \delta n \geq 32$, it follows that $1 + |V(J)|/32 \leq |V(J)|/16$, and so every vertex of J has degree at most $|V(J)|/16 - 1$. Since C_5 is isomorphic to its complement, J is C_5-free, and so from 2.1, there are two disjoint subsets $A, B \subseteq V(J)$ with $|A| \geq c|V(J)|/2$ and $|B| \geq |V(J)|/16$, such that the edge-density between A, B in J is less than c. Thus $|A| \geq c\delta n/2 \geq \epsilon cn$ and $|B| \geq \delta n/16 = \epsilon n$, and the edge-density between A, B in G is either at most c or at least $1 - c$. This proves 2.3. ∎

It is possible to deduce versions of 1.2 from versions of Rödl's theorem 2.2 directly, as follows. If we have d, δ satisfying 2.2, then for any n, if we choose $k \leq \min(\frac{1}{2d}, \frac{\delta n}{2})$ then we can use Turán's theorem to obtain a stable set or clique on k vertices from the set of at least $2k$ vertices with density at most $\frac{1}{2k}$ or at least $1 - \frac{1}{2k}$ that 2.2 gives us. This motivates trying to improve the bound in 2.2.

- Rödl's original proof of 2.2 uses Szemerédi's regularity lemma and gives a tower-type bound for $1/\delta$ in terms of $1/d$, which yields something worse than 1.2.
- In [8], a better bound of $\delta = 2^{-15|V(H)|(\log(1/d))^2}$ in 2.2 is proved, which implies the bound of 1.2.
- It is conjectured that a polynomial dependence of δ on d holds, and this would imply the Erdős-Hajnal conjecture itself.
- For $H = C_5$ we can get mid-way between, and that provides a different route to proving 1.3, as follows. One can prove that for $H = C_5$ we may take

  $$\delta = 2^{-O(\log(1/d)^2/\log\log(1/d))}$$

  in 2.2 by appropriately adapting the proof of 2.2 in [8] using that we now know 1.4 for $H = C_5$. This would imply 1.3. But the details of the proof of this improved bound for 2.2 for C_5 are involved and similar to that of the proof of 1.3 given in the next section, and we omit them for the sake of brevity.

3 The proof of 1.3.

Now we use 2.3 to prove 1.3. Since the argument to come is rather heavy, and works just as well for any graph H satisfying 1.4 instead of C_5, it might be wise to present it in full generality. Thus, let us say a class of graphs $\mathcal{I}$ is *hereditary* if every graph isomorphic to an induced subgraph of a member of the class also belongs to the class. Let ϵ be as in 2.3, and let $\sigma > 1 + \log(\epsilon^{-1})$. Then for $c \leq 1/2$, $c^\sigma \leq \epsilon$, and so by 2.3, if G is C_5-free with $n \geq 2$ vertices, and $0 \leq c \leq 1/2$, then there exist disjoint $A, B \subseteq V(G)$ with $|A| \geq c^\sigma n$ and $|B| \geq \epsilon n$, such that the edge-density between them is either at most c or at least $1 - c$. Then 1.3 follows from 2.3 and the following, applied to the hereditary class of all C_5-free graphs:

3.1 *Let $\mathcal{I}$ be a hereditary class of graphs, and let $\sigma \geq 0$ and $0 \leq \epsilon \leq 1$ with the following property: for every graph $G \in \mathcal{I}$ with at least two vertices, and all c with $0 \leq c \leq 1/2$, there are disjoint subsets $A, B \subseteq V(G)$ with $|A| \geq c^\sigma n$ and $|B| \geq \epsilon n$, such that the edge-density between A, B is either at most c or at least $1 - c$, where $n = |V(G)|$. Then there exists $\kappa > 0$ such that*

$$\max(\alpha(G), \omega(G)) \geq 2^{\kappa\sqrt{\log n \log\log n}}$$

for every $G \in \mathcal{I}$, where $n = |V(G)| \geq 2$.

Proof. Let us define $r(n) = \sqrt{\log n \log\log n}$ for $n \geq 2$, for typographical convenience.

A *cograph* is a graph not containing a 4-vertex path. Thus the disjoint union of two cographs is a cograph, and so is the complement of a cograph. We prove 3.1 by showing that G contains a cograph with at least $2^{2\kappa r(n)}$ vertices. As cographs are perfect, there is a clique or independent set with $2^{\kappa r(n)}$ vertices (and so of the desired cardinality).

For a graph G, let $\phi(G)$ denote the maximum of $|V(H)|$ over all cographs H contained in G. For each real number $x \geq 0$, let $f(x)$ be the minimum of $\phi(G)$, over all graphs $G \in \mathcal{I}$ with $|V(G)| = \lceil x \rceil$ (we may assume there is some such graph G, or else the result is trivially true). Since $\mathcal{I}$ is hereditary, $f(x)$ is non-decreasing with x.

We may assume that $\sigma \geq 1$ (by increasing σ if necessary). Let $\mu = (32\sigma)^{-1/2}$. Choose n_0 such that

$$\left\lfloor \frac{\sigma 2\mu r(n) - 1}{\log(2/\epsilon)} \right\rfloor \geq \sqrt{\log n}$$

for all $n \geq n_0$, and also such that $\mu r(n_0) \geq 2$, and $\log n_0 \geq 4\sigma\mu r(n_0)$. Choose $\kappa > 0$ such that $\kappa \leq \mu/2$ and $2\kappa r(n_0) \leq 1$. We will show that κ satisfies the theorem.

(1) *For all $n \geq 2$ and all c with $0 \leq c \leq 1/2$, either $f(n) \geq 1/(4c)$ or $f(n) \geq f(c^\sigma n/2) + f(\epsilon n/2)$.*

Let $G \in \mathcal{I}$ with $n \geq 2$ vertices, such that $\phi(G) = f(n)$. Since $G \in \mathcal{I}$, the hypothesis implies that there are disjoint sets $A, B \subseteq V(G)$ with $|A| \geq c^\sigma n$ and $|B| \geq \epsilon n$ such that the edge-density between A and B is either at most c or at least $1-c$. We suppose without loss of generality that this density is at most c (in the other case, we apply the same argument to $\overline{G}$).

Let A'' be the set of vertices in A with at least $2c|B|$ neighbours in B. As the number of edges between A, B is at least $2c|B||A''|$ and at most $c|A||B|$, it follows that $|A''| \leq |A|/2$. Let $A' = A \setminus A''$; so $|A'| = |A| - |A''| \geq |A|/2$ and every vertex in A' has at most $2c|B|$ neighbours in B. Since $G[A'] \in \mathcal{I}$, it follows from the definition of f that $\phi(G[A']) \geq f(|A'|)$. Let $A_0 \subseteq A'$ induce a cograph, with $|A_0| = f(|A'|)$.

If $|A_0| \geq 1/(4c)$, then $f(n) = \phi(G) \geq |A_0| \geq 1/(4c)$ as required, so we may assume that $|A_0| \leq 1/(4c)$. Let B' be those vertices in B with no neighbours in A_0; so $|B'| \geq |B| - 2c|B||A_0| \geq |B|/2$. Again from the definition of f, $\phi(G[B']) \geq f(|B'|) \geq f(\epsilon n/2)$. Since A_0 is anticomplete to B', it follows that

$$f(n) = \phi(G) \geq |A_0| + \phi(G[B']) \geq f(c^\sigma n/2) + f(\epsilon n/2).$$

This proves (1).

(2) *For all $n \geq 2$ and all c with $0 \leq c \leq 1/2$, if $\log n \geq \sigma \log(1/c)$ then either $f(n) \geq 1/(4c)$ or $f(n) \geq k f(c^{2\sigma} n)$, where*

$$k = \left\lfloor \frac{\sigma \log(1/c) - 1}{\log(2/\epsilon)} \right\rfloor.$$

We may assume that $f(n) < 1/(4c)$, and hence $f(n') < 1/(4c)$ for all $n' \leq n$. From the definition of k, $k \log(2/\epsilon) \leq \sigma \log(1/c) - 1 \leq \log n - 1$, and so $n(\epsilon/2)^k \geq 2$. Hence we may recursively apply (1) k times without violating the condition "$n \geq 2$" in (1); and we obtain

$$f(n) \geq f(c^\sigma n/2) + f(c^\sigma(\epsilon/2)n/2) + f(c^\sigma(\epsilon/2)^2 n/2) + \cdots + f(c^\sigma(\epsilon/2)^k n/2).$$

Each of the $k+1$ terms on the right side is at least $f(c^{2\sigma} n)$, from the definition of k, and so $f(n) \geq k f(c^{2\sigma} n)$. This proves (2).

(3) *For all* $n \geq 2$ *and all* c *with* $0 \leq c \leq 1/2$, *if* $\log n \geq 2\sigma \log(1/c)$ *and with* k *as in* (2), *either* $f(n) \geq 1/(4c)$ *or* $f(n) \geq k^j$, *where*

$$j = \left\lfloor \frac{\log n}{4\sigma \log(1/c)} \right\rfloor.$$

Again, we may assume that $f(n) < 1/(4c)$, and hence $f(n') < 1/(4c)$ for all $n' \leq n$. From the definition of j, $c^{2\sigma j} n \geq n^{1/2}$, and so $\log(c^{2\sigma j} n) \geq \frac{1}{2} \log n \geq \sigma \log(1/c)$. Moreover, $c^{2\sigma(j-1)} n \geq n^{1/2} c^{-2\sigma} \geq 2$ since $\sigma \geq 1$. Hence we may apply (2) recursively j times, and deduce that $f(n) \geq k^j f(c^{2\sigma j} n) \geq k^j$. This proves (3).

(4) *Let* $n \geq n_0$, *and* $c = 2^{-2\mu r(n)}$. *Then*

- $c \leq 1/2$;
- $\log n \geq 4\sigma\mu r(n) = 2\sigma \log(1/c)$;
- $k \geq \sqrt{\log n}$, *where* k *is as defined in* (2); *and*
- $1/(4c) \geq 2^{\mu r(n)}$.

We observe first that $c \leq 1/2$ if $n \geq n_0$, since $\mu r(n_0) \geq 1$. Also, $\log n_0 \geq 4\sigma\mu r(n_0)$ from the choice of n_0, and since $\frac{\log n}{r(n)}$ increases with n, it follows that $\log n \geq 4\sigma\mu r(n)$ for $n \geq n_0$. But $4\sigma\mu r(n) = 2\sigma \log(1/c)$, and so this proves the second statement. The third statement follows from the choice of n_0. For the final statement, we must check that $\log(1/c) - 2 \geq \mu r(n)$, that is, $\mu r(n) \geq 2$; but this holds from the definition of n_0. This proves (4).

(5) *If* $n \geq n_0$ *then* $f(n) \geq 2^{\mu r(n)}$.

Let c be as in (4) and let $n \geq n_0$. By the first two statements of (4), we may apply (3), and so either $f(n) \geq 1/(4c)$ or $f(n) \geq (\log n)^{j/2}$, by the third statement of (4). In the first case, the claim follows from the final statement of (4), so we may assume that

$$f(n) \geq (\log n)^{j/2} \geq (\log n)^{(\log n)/(16\sigma \log(1/c))} = 2^{(16\sigma \cdot 2\mu)^{-1} r(n)}.$$

As $\mu = (16\sigma \cdot 2\mu)^{-1}$ from the definition of μ, this proves (5).

We recall that $\kappa \leq \mu/2$ and $2\kappa r(n_0) \leq 1$. We claim that $f(n) \geq 2^{2\kappa r(n)}$ for all $n \geq 2$. This is true if $n \leq n_0$, because then $f(n) \geq 2 \geq 2^{2\kappa r(n)}$; and if $n > n_0$ then it follows from (5). This proves 3.1. ∎

References

[1] M. Bonamy, N. Bousquet and S. Thomassé, "The Erdős-Hajnal conjecture for long holes and antiholes", *SIAM J. Discrete Math.* **30** (2015), 1159–1164.

[2] N. Bousquet, A. Lagoutte, and S. Thomassé, "The Erdős-Hajnal conjecture for paths and antipaths", *J. Combinatorial Theory, Ser. B*, **113** (2015), 261–264.

[3] M. Chudnovsky, J. Fox, A. Scott, P. Seymour and S. Spirkl, "Sparse graphs with no polynomial-sized anticomplete pairs", submitted for publication, `arXiv:1810.00058`.

[4] M. Chudnovsky, A. Scott, P. Seymour and S. Spirkl, "Trees and linear anticomplete pairs", submitted for publication, `arXiv:1809.00919`.

[5] D. Conlon, J. Fox and B. Sudakov, "Recent developments in graph Ramsey theory", *Surveys in Combinatorics 2015*, London Math. Soc. Lecture Note Ser., **424** (2015), 49–118, Cambridge Univ. Press, Cambridge, problem 3.13.

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

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

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

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