Preprint

This is the submitted version of a paper published in *Journal of Graph Theory*.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-144161

Full subgraphs

Victor Falgas-Ravry[*]   Klas Markström[†]   Jacques Verstraëte[‡]

Abstract

Let $G = (V, E)$ be a graph of density p on n vertices. Following Erdős, Łuczak and Spencer, an m-vertex subgraph H of G is called *full* if H has minimum degree at least $p(m-1)$. Let $f(G)$ denote the order of a largest full subgraph of G. If $p\binom{n}{2}$ is a non-negative integer, define

$$f(n, p) = \min\{f(G) : |V(G)| = n,\ |E(G)| = p\tbinom{n}{2}\}.$$

Erdős, Łuczak and Spencer proved that for $n \geq 2$,

$$(2n)^{\frac{1}{2}} - 2 \leq f(n, \tfrac{1}{2}) \leq 4n^{\frac{2}{3}}(\log n)^{\frac{1}{3}}.$$

In this paper, we prove the following lower bound: for $n^{-\frac{2}{3}} < p_n < 1 - n^{-\frac{1}{7}}$,

$$f(n, p) \geq \frac{1}{4}(1 - p)^{\frac{2}{3}} n^{\frac{2}{3}} - 1.$$

Furthermore we show that this is tight up to a multiplicative constant factor for infinitely many p near the elements of $\{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \dots\}$. In contrast, we show that for any n-vertex graph G, either G or G^c contains a full subgraph on $\Omega(\frac{n}{\log n})$ vertices. Finally, we discuss full subgraphs of random and pseudo-random graphs, and several open problems.

1   Introduction

1.1   Full subgraphs

A *full subgraph* of a graph G of density p is an m-vertex subgraph H of minimum degree at least $p(m-1)$. This notion was introduced by Erdős, Łuczak and Spencer [8]. We may think of a full subgraph as a particularly 'rich' subgraph, with 'unusually high' minimum degree: if we select m vertices of G uniformly at random, then the expected *average* degree of the subgraph they induce is exactly $p(m-1)$, which is the minimum degree we require for the subgraph to

be full. One cannot in general expect to find m-vertex subgraphs of higher minimum degree, as may be seen for example by considering complete multipartite graphs with parts of equal sizes. Fixing p and n, one may ask for the largest m such that every n-vertex graph G has a full subgraph with m vertices. For a graph G, let $f(G)$ denote the largest number of vertices in a full subgraph of G. If $p\binom{n}{2}$ is a non-negative integer, define

$$f(n,p) = \min\{f(G) : |V(G)| = n, |E(G)| = p\tbinom{n}{2}\}.$$

Erdős, Łuczak and Spencer [8] raised a problem which, in our terminology, amounts to determining $f(n,p)$ when $p = \frac{1}{2}$, and showed $(2n)^{\frac{1}{2}} - 2 \le f(n, \frac{1}{2}) \le (2 + \frac{2}{\sqrt{3}})n^{\frac{2}{3}}(\log n)^{\frac{1}{3}}$. In this paper, we prove the following theorem for general p, improving the lower bound of [8]:

Theorem 1. *For all $p = p_n$ such that $n^{-\frac{2}{3}} < p_n < 1 - n^{-\frac{1}{7}}$,*

$$f(n,p) \ge \frac{1}{4}(1-p)^{\frac{2}{3}}n^{\frac{2}{3}} - 1.$$

Moreover for each $c \ge 1$, if $p = \frac{r}{r+1} + cn^{-\frac{2}{3}}$ for some $r \in \mathbb{N}$, then $f(n,p) = \Theta(n^{\frac{2}{3}})$.

We also show in Section 3 (after the proof of Theorem 2) that if $p \le n^{-\frac{2}{3}}$ then $|f(n,p) - p^{\frac{1}{2}}n| \le 1$. A case of particular interest is $p = \frac{1}{2}$, where Theorem 1 together with the results of Erdős, Łuczak and Spencer [8] gives

$$\frac{1}{4\sqrt[3]{4}}n^{\frac{2}{3}} - 1 \le f(n, \tfrac{1}{2}) \le \left(2 + \frac{2}{\sqrt{3}}\right)n^{\frac{2}{3}}(\log n)^{\frac{1}{3}}.$$

A similar construction to that of Erdős, Łuczak and Spencer shows that $f(n,p)$ is within a logarithmic factor of $n^{\frac{2}{3}}$ when $p \in \{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \dots\}$. The order of magnitude of $f(n,p)$ is not known in general, and we pose the following problem:

Problem 1. *For each fixed $p \in (0,1)$, determine the order of magnitude of $f(n,p)$.*

It may also be interesting to determine the order of magnitude of the minimum possible value of $f(G)$ when G is from a certain class of n-vertex graph of density p, such as K_r-free graphs or graphs of chromatic number at most r.

1.2   Discrepancy and full subgraphs

Full subgraphs are related to subgraphs of large positive discrepancy. For a graph G of density p and an m-vertex set $X \subseteq V(G)$, let $\delta(X) = e(X) - p\binom{m}{2}$, where $e(X)$ denotes the number of edges of G that lie in X. The *positive* and *negative discrepancy* of G are respectively defined by

$$\mathrm{disc}^+(G) = \max_{X \subseteq V(G)} \delta(X) \qquad \mathrm{disc}^-(G) = \max_{X \subseteq V(G)} (-\delta(X)).$$

The *discrepancy* of G is $\mathrm{disc}(G) = \max\{\mathrm{disc}^+(G), \mathrm{disc}^-(G)\}$. We prove the following simple bound via a greedy algorithm in Section 3, relating the positive discrepancy to the order of a largest full subgraph:

Theorem 2. *Let G be a graph of density p. Then*

$$f(G) \geq (1-p)^{-\frac{1}{2}} (2 \mathit{disc}^{+}(G))^{\frac{1}{2}}.$$

Theorem 2 is best possible, since the graph G consisting of a clique with $\binom{m}{2}$ edges and $n - m$ isolated vertices has $f(G) = m$ and

$$\mathrm{disc}^{+}(G) = \binom{m}{2}\left(1 - \frac{\binom{m}{2}}{\binom{n}{2}}\right),$$

whereas the bound given by Theorem 2 is $f(G) \geq \left\lceil \sqrt{m(m-1)} \right\rceil = m$. On the other hand, if G is any n-vertex graph obtained by adding or removing $o(n^{\frac{4}{3}})$ edges in a complete multipartite graph with a bounded number of parts of equal size, then $\mathrm{disc}^{+}(G) = o(n^{\frac{4}{3}})$ and the lower bound in Theorem 2 is superseded by Theorem 1.

That there should be a relation between full subgraphs (which have unexpectedly high minimum degree) and subgraphs with large positive discrepancy (which have unexpectedly many edges) is not surprising. Indeed, an easy observation is that any subgraph maximising the positive discrepancy must be a full subgraph (see Lemma 9).

1.3   Random and pseudo-random graphs

In a random or pseudo-random setting, we are able to improve our bounds on the size of a largest full subgraph by drawing on previous work on discrepancy and jumbledness. Jumbledness was introduced in a seminal paper of Thomason [23] as a measure of the 'pseudo-randomness' of a graph.

Definition. *A graph G is (p,j)-jumbled if for every $X \subseteq V(G)$, $|\delta_p(X)| \leq j|X|$.*

We prove that graphs which are 'well-jumbled' — meaning that they are (p,j)-jumbled for some small j, and so look 'random-like' — have large full subgraphs.

Theorem 3. *Suppose G is a (p,j)-jumbled graph of density p. Then*

$$f(G) \geq \frac{\mathrm{disc}^{+}(G)}{j}.$$

This result, which for small values of j improves on Theorems 1 and 2, is proved in Section 4.

For random graphs, Erdős, Łuczak and Spencer [8] showed that for $p = \frac{1}{2}$, $f(G_{n,p}) \geq \beta_1 n - o(n)$ asymptotically almost surely, where $\beta_1 \approx 0.227$. Using Theorem 3, we can extend this linear lower bound to arbitrary, fixed $p \in (0,1)$. Erdős and Spencer [6] proved that for $p = \frac{1}{2}$ we have asymptotically almost surely

$$\mathrm{disc}^{+}(G_{n,p}) = \Theta\big(p^{\frac{1}{2}}(1-p)^{\frac{1}{2}} n^{\frac{3}{2}}\big), \tag{1}$$

and that the same bound holds for $\text{disc}^-(G_{n,p})$. By extending their arguments, it is easily shown that (1) holds for arbitrary fixed $p \in (0,1)$. Further it is well-known that $G = G_{n,p}$ asymptotically almost surely has $|\delta_p(X)| = O(\sqrt{p(1-p)n}|X|)$ for all $X \subseteq V(G_{n,p})$ (see for example [16]), so that $G_{n,p}$ is (p,j)-jumbled for some $j = O(\sqrt{p(1-p)n})$. Combining this with (1) and Theorem 3, we obtain that for fixed $p \in (0,1)$ asymptotically almost surely,

$$f(G_{n,p}) = \Omega(n). \tag{2}$$

In the other direction, results of Riordan and Selby [20] imply that for all fixed $p \in (0,1)$, $f(G_{n,p}) \leq \beta_2 n + o(n)$ asymptotically almost surely, where $\beta_2 \approx 0.851\ldots$. We believe that $f(G_{n,p})$ is concentrated around $\beta n + o(n)$ for some function $\beta = \beta_p$, and pose the following problem.

Problem 2. *For each fixed $p \in (0,1)$, prove the existence and determine the value of a real number $\beta = \beta_p$ such that for all $\delta > 0$, $\mathbb{P}(|f(G_{n,p}) - \beta_p n| > \delta n) \to 0$ as $n \to \infty$.*

1.4 Full and co-full subgraphs

We also a consider a variant of our problem with a Ramsey-theoretic flavour. A subgraph H of a graph G is *co-full* if $V(H)$ induces a full subgraph of G^c, the complement of G. Equivalently, an induced m-vertex subgraph H of a graph G with density p is co-full if it has maximum degree at most $p(m-1)$. Let $g(G)$ be the largest integer m such that G has a full subgraph with at least m vertices or a co-full subgraph with at least m vertices. In other words, $g(G) = \max\{f(G), f(G^c)\}$. Setting $g(n) = \min\{g(G) : |V(G)| = n\}$, we prove the following theorem:

Theorem 4. *There exist constants $c_1, c_2 > 0$ such that*

$$c_1 \frac{n}{\log n} \leq g(n) \leq c_2 \frac{n \log \log n}{\log n}.$$

Bounding $g(G)$ is related to, but distinct from, a problem of Erdős and Pach [9] on *quasi-Ramsey numbers* (see also [14, 15]). Erdős and Pach [9] showed that for every n-vertex graph G, in either G or G^c there exists a subgraph with $m = \Omega(\frac{n}{\log n})$ vertices and minimum degree at least $\frac{1}{2}(m-1)$. In particular, when G has density $\frac{1}{2}$, this shows $g(G) = \Omega(\frac{n}{\log n})$. Erdős and Pach in addition gave an unusual weighted random graph construction G' to show their quasi-Ramsey bound was sharp up to a $\log \log n$ factor. While G' does not have density $\frac{1}{2}$, a simple modification (see Section 5.2) gives a graph $G^\star$ of density $\frac{1}{2}$ such that $g(G^\star) = O(\frac{n \log \log n}{\log n})$, and this gives the upper bound in Theorem 4. This leaves the following problem open:

Problem 3. *Determine the order of magnitude of $g(n)$.*

By (2) with $p = \frac{1}{2}$, note that $g(G)$ is linear in n for almost all n-vertex graphs G. We may also define $g(n, p) = \min\{g(G) : |V(G)| = n, |E(G)| = p\binom{n}{2}\}$, and ask for the order of magnitude of $g(n, p)$. Note Theorem 4 gives $g(n, p) = \Omega(\frac{n}{\log n})$ for all p.

4

1.5 Relatively half-full subgraphs

If G is a graph, then a *relatively half-full* subgraph of G is a subgraph H of G such that $d_H(v) \geq \frac{1}{2} d_G(v)$ for every $v \in V(H)$. A key ingredient in the proof of Theorem 1 is the following theorem on relatively half-full subgraphs:

Theorem 5. *Let G be an n-vertex graph. Then G contains a relatively half-full subgraph with $\lfloor \frac{n}{2} \rfloor$ or $\lfloor \frac{n}{2} \rfloor + 1$ vertices.*

Theorem 5 is best possible, in the sense that the smallest non-empty relatively half-full subgraph of K_n has $\lfloor \frac{n}{2} \rfloor + 1$ vertices and the smallest relatively half-full subgraph of $K_{n,n}$ has $n + 1$ vertices when n is odd. For regular graphs, we obtain:

Corollary 6. *Let G be an n-vertex d-regular graph. Then G contains a full subgraph with $\lfloor \frac{n}{2} \rfloor$ or $\lfloor \frac{n}{2} \rfloor + 1$ vertices.*

(Note that of course G itself is full, as it is regular.) When d is very small relative to n, Alon [1] showed that any d-regular n-vertex graph contains a subgraph on $\lceil \frac{n}{2} \rceil$ vertices in which the minimum degree is at least $\frac{1}{2} d + c d^{\frac{1}{2}}$, exceeding the requirement for a full subgraph by an additive factor of $c d^{\frac{1}{2}}$. However, as observed by Alon [1], such a result does not hold for large d, as for example complete graphs and complete bipartite graphs show.

1.6 Relatively q-full subgraphs

Let $q \in [0, 1]$. A subgraph H of a graph G is *relatively q-full* if $d_H(v) \geq q d_G(v)$ for all $v \in V(H)$. We prove:

Theorem 7. *Let G be a graph on n vertices. Then for every $q \in [0, 1]$, G contains one of the following:*

   (i) *a relatively q-full subgraph on $\lceil qn \rceil$ vertices, or*
  (ii) *a relatively $(1-q)$-full subgraph on $\lfloor (1-q)n \rfloor$ vertices, or*
 (iii) *a relatively q-full subgraph on $\lceil qn \rceil + 1$ vertices and a relatively $(1-q)$-full subgraph on $\lfloor (1-q)n \rfloor + 1$ vertices.*

Using Theorem 7, we prove Theorem 5 and an extension to relatively $\frac{1}{r}$-full subgraphs for $r \geq 3$:

Theorem 8. *Let G be a graph on n vertices, and let $r \in \mathbb{N}$. Then G contains a relatively $\frac{1}{r}$-full subgraph on $\lfloor \frac{n}{r} \rfloor$, $\lceil \frac{n}{r} \rceil$ or $\lceil \frac{n}{r} \rceil + 1$ vertices.*

Theorem 8 is best possible in the following sense: if $r \geq 3$, consider the complete graph K_n for some $n \geq r + 2$ with $n \equiv 2 \mod r$. A smallest non-empty relatively $\frac{1}{r}$-full subgraph of K_n has exactly $\lceil \frac{n-1}{r} \rceil + 1 = \lceil \frac{n}{r} \rceil + 1$ vertices.

It is natural to ask whether Theorem 8 can be extended further to cover other q.

Problem 4. *Determine whether there exists a constant c such that for every $q \in [0, \frac{1}{2}]$, every n-vertex graph G has a relatively q-full subgraph with at least $\lfloor qn \rfloor$ vertices and at most $\lfloor qn \rfloor + c$ vertices.*

For $q > \frac{1}{2}$, a cycle of length n shows that there exist n-vertex graphs with no non-empty relatively q-full subgraphs on fewer than n vertices. We might try to circumvent this example by requiring a weaker degree condition: define a subgraph H of a graph G to be *weakly relatively q-full* if $d_H(v) \geq \lfloor qd_G(v) \rfloor$ for all $v \in V(H)$. However even for this notion of q-fullness a natural generalisation of Theorem 8 fails for rational $q > \frac{1}{2}$: consider the second power of a cycle of length n. If x is a vertex in a weakly relatively $\frac{3}{4}$-full subgraph H, then all but at most one of its neighbours must also belong to H. Thus vertices not in H must lie at distance at least 5 apart in the original cycle, and H must contain at least $\frac{4}{5}n$ vertices, rather than the $\frac{3}{4}n + O(1)$ we might have hoped for. It would be interesting to determine whether powers of paths or cycles provide us with the worst-case scenario for finding weakly relatively q-full subgraphs when $q > \frac{1}{2}$.

Problem 5. *Let $q \in (\frac{1}{2}, 1)$. Determine whether there exist a constant $c_q < 1$ such that every n-vertex graph G has a weakly relatively q-full m-vertex subgraph where $\lfloor qn \rfloor \leq m \leq c_q n$.*

1.7 Notation

We use standard graph theoretic notation. In particular, if X, Y are sets of vertices of a graph $G = (V, E)$, then $e(X)$ denotes the number of edges in the subgraph $G[X]$ of G induced by X, $e(G)$ is the number of edges in G, and $e(X, Y)$ is the number of edges with one end in X and the other end in Y. Denote by $d_X(x)$ the number of neighbours in X of a vertex $x \in V(G)$. The Erdős-Rényi random graph with edge-probability p on n vertices is denoted by $G_{n,p}$. If $(A_n)_{n \in \mathbb{N}}$ is a sequence of events, then we say A_n occurs *asymptotically almost surely* if $\lim_{n \to \infty} \mathbb{P}(A_n) = 1$.

2 Relatively q-full subgraphs : proofs of Theorems $5 - 8$

Proof of Theorem 7. Let G be a graph on n vertices, and let $q \in [0, 1]$ be fixed. Let $X \sqcup Y$ be a bipartition of $V(G)$ with $|X| = \lceil qn \rceil$ and $|Y| = \lfloor (1 - q)n \rfloor$ maximising the value of $(1 - q)e(X) + qe(Y) := M$.

If X is relatively q-full or Y is relatively $(1 - q)$-full, then we are done. Otherwise there exist $x \in X$ and $y \in Y$ with $d_X(x) \leq \lceil qd_G(x) \rceil - 1$ and $d_Y(y) \leq \lceil (1 - q)d_G(y) \rceil - 1$. Let $X' = (X \setminus \{x\}) \cup \{y\}$ and $Y' = (Y \setminus \{y\}) \cup \{x\}$, and let $\mathbb{1}_{xy} = 1$ if $\{x, y\} \in E(G)$ and $\mathbb{1}_{xy} = 0$ otherwise. Write $d_G(x) = d(x)$ for $x \in V(G)$ and set $M' = (1 - q)e(X') + qe(Y')$. Then

$$
\begin{aligned}
M' &= M + (1 - q)d_X(y) - qd_Y(y) + qd_Y(x) - (1 - q)d_X(x) - \mathbb{1}_{xy} \\
&= M + (1 - q)d(y) - d_Y(y) + qd(x) - d_X(x) - \mathbb{1}_{xy} \\
&\geq M + \Big((1 - q)d(y) - \lceil (1 - q)d(y) \rceil\Big) + \Big(qd(x) - \lceil qd(x) \rceil\Big) + 2 - \mathbb{1}_{xy}.
\end{aligned}
$$

6

Since $X \sqcup Y$ maximised $(1-q)e(X) + qe(Y)$ over all bipartitions with $|X| = \lceil qn \rceil$, $|Y| = \lfloor (1-q)n \rfloor$, $M' \leq M$ and we deduce from the inequality above that

    (a) $d_X(x) = \lceil qd(x) \rceil - 1$ and $d_Y(y) = \lceil (1-q)d(y) \rceil - 1$.
    (b) $\{x, y\} \in E(G)$.
    (c) $(1-q)d(y) < \lceil (1-q)d(y) \rceil$ and $qd(x) < \lceil qd(x) \rceil$.

Now let B_X denote the set of $x \in X$ with $d_X(x) \leq \lceil qd(x) \rceil - 1$ and B_Y the set of $y \in Y$ with $d_Y(y) \leq \lceil (1-q)d(y) \rceil - 1$. By our assumption, both sets are non-empty. By (b) above, $B_X \sqcup B_Y$ induces a complete bipartite subgraph of G. Thus by (a) we have that for every $x \in B_X, y \in B_Y$, $X \cup \{y\}$ is a relatively q-full subgraph on $\lceil qn \rceil + 1$ vertices and $Y \cup \{x\}$ is a relatively $(1-q)$-full subgraph on $\lfloor (1-q)n \rfloor + 1$ vertices. $\qquad\square$

Proof of Theorem 5. Apply Theorem 7 with $q = \frac{1}{2}$. $\qquad\square$

Proof of Corollary 6. Suppose at least one of n, d is odd. By Theorem 5, every d-regular graph has an m-vertex subgraph H with $m \in \{\lfloor \frac{1}{2}n \rfloor, \lfloor \frac{1}{2}n \rfloor + 1\}$ such that $d_H(v) \geq \lceil \frac{d}{2} \rceil$ for every $v \in V(H)$. Since for n, d not both even

$$\left\lceil \frac{d}{2} \right\rceil = \left\lceil \frac{d}{n-1} \left\lfloor \frac{n}{2} \right\rfloor \right\rceil \geq \left\lceil \frac{d}{n-1}(m-1) \right\rceil,$$

the subgraph H is a full subgraph.

In the case where both n and d are even, we need to use a slightly stronger form of Theorem 5. In the particular case where G is d-regular with d even and $q = \frac{1}{2}$, condition (c) in the proof of Theorem 7 cannot be satisfied, and in particular one of the alternatives (i) or (ii) must hold in Theorem 7. Thus G must contain a subgraph H on $\frac{n}{2}$ vertices with minimum degree at least $\frac{d}{2}$, which is a full subgraph. $\qquad\square$

Proof of Theorem 8. We use Theorem 7 and induction on r. The base case $r = 1$ is trivial, and Theorem 5 deals with the case $r = 2$. Now apply Theorem 7 with $q = \frac{1}{r}$: given a graph G on n vertices, this gives us a $\frac{1}{r}$-full subgraph on $\lceil \frac{n}{r} \rceil$ or $\lceil \frac{n}{r} \rceil + 1$ vertices (alternatives (i) and (iii)) or an $\frac{r-1}{r}$-full subgraph H on $\lfloor \frac{r-1}{r}n \rfloor$ vertices (alternative (ii)). In the latter case, we use our inductive hypothesis to find a $\frac{1}{r-1}$-full subgraph H' of H on m vertices, for some $m : \lfloor \frac{n}{r} \rfloor \leq m \leq \lceil \frac{n}{r} \rceil + 1$. The subgraph H' is easily seen to be a $\frac{1}{r}$-full subgraph of G, and so we are done. $\qquad\square$

3   A greedy algorithm : proof of Theorem 2

A natural strategy for obtaining a full subgraph in a graph G of density p on n vertices is to repeatedly remove vertices of relatively low degree. When there are i vertices left in the graph, such a greedy algorithm finds a vertex of degree at most $\lceil p(i-1) \rceil - 1$ and deletes that vertex, unless no such vertex exists, in which case the i vertices induce a full subgraph. If G

has positive discrepancy $\text{disc}^+(G) = \alpha$, then we apply this algorithm in a subgraph H on m vertices with $e(H) \geq p\binom{m}{2} + \alpha$ to obtain Theorem 2.

Proof of Theorem 2. If G has positive discrepancy $\text{disc}^+(G) = \alpha > 0$, then its density p is strictly less than 1. Let H be a subgraph of G with m vertices such that $e(H) = p\binom{m}{2} + \alpha$. At stage i we delete a vertex of degree at most $\lceil p(m-i) \rceil - 1$ in the remaining graph, or stop if no such vertex exists. The number of edges remaining after stage i is at least

$$p\binom{m}{2} + \alpha - \sum_{j=1}^{i} p(m-j) = p\binom{m}{2} + \alpha - p\binom{m}{2} + p\binom{m-i}{2} = \alpha + p\binom{m-i}{2}.$$

Therefore the greedy algorithm must terminate with a full subgraph on $m-i$ vertices for some i satisfying $(1-p)\binom{m-i}{2} \geq \alpha$. We conclude $f(G) \geq m - i \geq (1-p)^{-\frac{1}{2}}(2\alpha)^{\frac{1}{2}}$. □

An alternate proof may be obtained by appealing to Lemma 9, which states that a subgraph attaining the maximum positive discrepancy must be full. The example of a clique with m vertices and $n - m$ isolated vertices which shows that Theorem 2 is tight is the same example which shows $f(n,p) = O(p^{\frac{1}{2}}n)$ for $p \leq n^{-\frac{2}{3}}$. We now prove that $|f(n,p) - p^{\frac{1}{2}}n| \leq 1$ for $p \leq n^{-\frac{2}{3}}$.

Proof that $|f(n,p) - p^{\frac{1}{2}}n| \leq 1$ for $p \leq n^{-\frac{2}{3}}$. First we show $f(n,p) < p^{\frac{1}{2}}n + 1$ for all $p : 0 < p \leq 1$. If m is defined by $\binom{m-1}{2} < p\binom{n}{2} \leq \binom{m}{2}$, then the n-vertex graph G consisting of a subgraph of a clique of size m with $p\binom{n}{2}$ edges, together with $n - m$ isolated vertices has $f(G) \leq m \leq p^{\frac{1}{2}}n + 1$. Next we show that every n-vertex graph G of density p has a full subgraph with at least $p^{\frac{1}{2}}n - 1$ vertices if $p \leq n^{-\frac{2}{3}}$. Remove all isolated vertices from G. The number of isolated vertices is clearly at most $n - p^{\frac{1}{2}}n$, otherwise the remaining graph has $p\binom{n}{2}$ edges and fewer than $p^{\frac{1}{2}}n$ vertices, which is impossible since this is denser than a complete graph. So we have a subgraph H with at least $p^{\frac{1}{2}}n$ vertices and $p\binom{n}{2}$ edges with no isolated vertices. Clearly H has a subgraph of minimum degree at least 1 with at least $p^{\frac{1}{2}}n - 1$ vertices and at most $p^{\frac{1}{2}}n + 1$ vertices, since the removal of a leaf in a spanning forest creates at most one new isolated vertex. This subgraph is full since

$$\lceil p(p^{\frac{1}{2}}n + 1 - 1) \rceil = \lceil p^{\frac{3}{2}}n \rceil \leq 1$$

when $p \leq n^{-\frac{2}{3}}$, as required. □

Remarks. The analysis of the greedy algorithm in the proof of Theorem 2 above is not optimal; in fact by considering the asymptotic behavior of

$$\phi = \liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} (p(n-i) + 1 - \lceil p(n-i) \rceil),$$

8

and performing our greedy algorithm directly on G rather than on a maximum discrepancy subgraph it follows that

$$f(n,p) \geq \begin{cases} \left(\frac{n(q+1)}{q(1-p)}\right)^{\frac{1}{2}} & \text{if } p \text{ is rational with denominator } q > 1 \\ \left(\frac{n}{1-p}\right)^{\frac{1}{2}} & \text{if } p < 1 - \varepsilon \text{ for some } \varepsilon > 0 \text{ and } p \geq \frac{1}{n}. \end{cases}$$

For instance if say $p_n = \frac{1}{2} - o(1) < \frac{1}{2}$, then $f_{p_n}(n) \geq (1 - o(1))\sqrt{2n}$, as shown by Erdős, Łuczak and Spencer [8]. These lower bounds on $f(n,p)$ will be superseded by the better bounds given in Theorem 1.

We note that there exist examples of n-vertex graphs with density $p = \frac{1}{2} + o(1)$ where greedily removing a vertex of minimal degree could yield a full subgraph of order only $O(\sqrt{n})$. Consider the graph G on $V = \{0, 1, ..., 4n + 1\}$ obtained by taking the nth power of the Hamiltonian cycle through $0, 1, 2, \ldots, 4n + 1$, adding edges between all antipodal pairs $\{i, i + (2n + 1)\}$ (with addition modulo $4n + 2$), and adding a complete bipartite graph $K_{m,m}$ with parts $\{0, 1, ..., m-1\}$ and $\{2n+1, 2n+2, ..., 2n+m\}$, where $m = (3n)^{\frac{1}{2}} + O(1)$. It is an easy exercise to show that by removing antipodal pairs of minimum degree vertices a greedy algorithm could fail to find a full subgraph until it has stripped the graph down to the planted complete bipartite graph $K_{m,m}$.

4 Jumbledness: proof of Theorem 3

As our arguments involve passing to subgraphs with different edge densities, it shall be useful to adapt our notion of full subgraphs, discrepancy and jumbledness as follows.

Let $G = (V, E)$ be a graph. An induced subgraph H of G on m vertices is called p-*full* if its minimum degree is at least $p(m - 1)$, and is called p-*co-full* if its maximum degree is at most $p(m - 1)$. Let $f_p(G)$ be the largest number of vertices in a p-full subgraph of G, and let $g_p(G)$ be the largest size of a p-full or p-co-full subgraph of G. If p happens to be the density of G, then in fact $f_p(G) = f(G)$ and $g_p(G) = g(G)$. Determining the smallest possible values of $f_p(G)$ and $g_p(G)$ given the number of edges and number of vertices in G can be viewed as generalisations of the Turán and of the Ramsey problems, respectively, which comprise the case $p = 1$; see [14] and the references therein.

For $X \subseteq V(G)$, set $\delta_p(X) = e(X) - p\binom{|X|}{2}$. The *positive p-discrepancy* of G is defined to be $\text{disc}_p^+(G) = \max_{X \subseteq V} \delta_p(X)$. The *negative p-discrepancy* of G is $\text{disc}_p^-(G) = \max_{X \subseteq V}(-\delta_p(X))$. The *$p$-discrepancy* of G is $\text{disc}_p(G) = \max\left(\text{disc}_p^+(G), \text{disc}_p^-(G)\right)$. Finally, the *$p$-jumbledness* of G is

$$j_p(G) = \max_{X \subseteq V} \frac{|\delta_p(X)|}{|X|}.$$

We begin with the following simple observation:

Lemma 9. *Let X be a subset of $V(G)$ such that $\delta_p(X) = \text{disc}_p^+(G)$. Then $G[X]$ is p-full.*

Proof. Indeed, otherwise deleting a minimum degree vertex from X would strictly increase the p-discrepancy. $\qquad\square$

We shall prove the following, slightly more general form of Theorem 3.

Theorem 10. *Let G be a graph and let $p \in [0,1]$. Then*

$$f_p(G) \geq \frac{\mathrm{disc}_p^+(G)}{j_p(G)} \quad and \quad g_p(G) \geq \frac{\mathrm{disc}_p(G)}{j_p(G)}.$$

Proof. Let X be a subset of $V(G)$ such that $\delta_p(X) = \mathrm{disc}_p^+(G)$. Then $G[X]$ is p-full by Lemma 9, and in particular, $f_p(G) \geq |X|$. By definition of p-jumbledness we have

$$\mathrm{disc}_p^+(G) = \left|\delta_p(X)\right| \leq j_p(G)|X| \leq j_p(G)f_p(G).$$

Applying the resulting lower bound for $f_p(G)$ to the complement G^c of G (and noting that $\mathrm{disc}_p^-(G) = \mathrm{disc}_{1-p}^+(G^c)$ and $j_p(G) = j_{1-p}(G^c)$), we have

$$g_p(G) \geq \max\left(\frac{\mathrm{disc}_p^+(G)}{j_p(G)}, \frac{\mathrm{disc}_{1-p}^+(G^c)}{j_{1-p}(G^c)}\right) = \frac{\mathrm{disc}_p(G)}{j_p(G)}.$$

$\qquad\square$

5 Proof of Theorem 4

Let $G = (V,E)$ be a graph on n vertices, and let $k : 1 \leq k \leq n$ be an integer. The *p-jumbledness of G on k-sets* is defined to be

$$j_{k,p}(G) = \max_{X \subseteq V: \, |X|=k} \frac{|\delta_p(X)|}{|X|}.$$

Similarly, the *positive p-discrepancy of G on k-sets* is defined to be

$$\mathrm{disc}_{k,p}^+(G) = \max_{X \subseteq V: \, |X|=k} \delta_p(X),$$

with the *negative p-discrepancy on k-sets* $\mathrm{disc}_{k,p}^-(G)$ and the *p-discrepancy on k-sets* $\mathrm{disc}_{k,p}(G)$ defined mutatis mutandis. Note that by definition we have $j_p(G) = \max\{j_{k,p}(G) : 1 \leq k \leq n\}$ and $\mathrm{disc}_p(G) = \max\{\mathrm{disc}_{k,p}(G) : 1 \leq k \leq n\}$. In addition for each $k: 1 \leq k \leq n$ we have $\mathrm{disc}_{k,p}(G) = k \cdot j_{k,p}(G)$.

In general, it is not true that the 'local' jumbledness on k-sets for linear-sized k is of the same order as the global jumbledness of G. Indeed consider an Erdős–Rényi random graph G with edge probability $p = \frac{1}{2}$ within which we plant a clique and a disjoint independent set, each of order $m = n^{\frac{3}{4}}$. It is straightforward to show with $k = \lceil\frac{n}{2}\rceil$ that asymptotically almost surely $j_{k,\frac{1}{2}}(G) = \Theta(n^{\frac{1}{2}})$ whilst $j_{\frac{1}{2}}(G) = \Theta(m)$. To prove Theorem 4, we use the following theorem of Thomason [23] to show that every n-vertex graph has an induced subgraph $G^\star$ on $\Omega(n)$ vertices such that the p-jumbledness of $G^\star$ on $\lceil\frac{n^\star}{e}\rceil$-sets and the global p-jumbledness of $G^\star$ differ by a factor of only $O(\log n)$.

Proposition 11 (Thomason [23])**.** *Let G be a graph of order n, let ηn be an integer between 2 and $n-2$, and let $M > 1$ and $p \in [0,1]$. Suppose that every set X of ηn vertices of G satisfies*

$$|\delta_p(X)| \leq \eta n \alpha.$$

Then G contains an induced subgraph G' of order

$$|V(G')| \geq \left(1 - \frac{880}{\eta(1-\eta)^2 M}\right) n$$

such that G' is $(p, M\alpha)$-jumbled.

Lemma 12. *There exists an absolute constant $c > 0$ such that for every $p \in [0,1]$ and n-vertex graph G, G contains an induced subgraph $G^\star$ on $n^\star \geq \frac{n}{e}$ vertices for which*

$$j_p(G^\star) \leq (c \log n)\, j_{\lceil \frac{n^\star}{e} \rceil, p}(G^\star).$$

Proof. Let $G = G_0$ be a graph on $n = n_0$ vertices. Let $\eta = \frac{1}{e}$ and $M = \frac{880}{\eta(1-\eta)^2} \log n$. For $i \geq 0$, if

$$j_{\lceil \frac{n_i}{e} \rceil, p}(G_i) < \lfloor \tfrac{\eta}{M} j_p(G_i) \rfloor, \tag{3}$$

then apply Proposition 11 to find an induced subgraph G_{i+1} of G_i on $n_{i+1} \geq (1 - \frac{880}{\eta(1-\eta)^2 M})n_i$ vertices with

$$j_p(G_{i+1}) \leq M j_{\lceil \frac{n_i}{e} \rceil, p}(G_i). \tag{4}$$

Combining (3) and (4) and iterating, we see that

$$j_p(G_{i+1}) \leq M j_{\lceil \frac{n_i}{e} \rceil, p}(G_i) < \lfloor \eta j_p(G_i) \rfloor \leq \lfloor \eta^{i+1} j_p(G_0) \rfloor \leq \lfloor e^{-i-1}(n/2) \rfloor. \tag{5}$$

Since $j_p(G_{i+1}) \geq 0$, we deduce from (5) that this procedure must terminate for some integer $i < \log n - 1$ with a graph $G^\star = G_i$ on $n^\star = n_i$ vertices, where

$$n_i \;\geq\; \left(1 - \frac{880}{\eta(1-\eta)^2 M}\right)^i n \geq \left(1 - \frac{1}{\log n}\right)^{\log n - 1} n \geq \frac{n}{e}$$

and where $j_{\lceil \frac{n^\star}{e} \rceil, p}(G^\star) \geq \lfloor \frac{\eta}{M} j_p(G^\star) \rfloor = \Omega\left(j(G^\star)/\log n\right)$. $\qquad\square$

5.1 Proof of $g(n) = \Omega(\frac{n}{\log n})$.

Let G be a graph on n vertices. By Lemma 12, it has an induced subgraph $G^\star$ on $n^\star \geq \frac{n}{e}$ vertices such that the p-jumbledness of $G^\star$ on $\lceil \frac{n^\star}{e} \rceil$-sets and the global p-jumbledness of $G^\star$ differ by a multiplicative factor of at most $c \log n$.

Applying Theorem 10 to $G^\star$ we have:

$$g(G) \geq g_p(G^\star) \geq \frac{\operatorname{disc}_p(G^\star)}{j_p(G^\star)} \geq \frac{\operatorname{disc}_{\lceil \frac{n^\star}{e} \rceil, p}(G^\star)}{(c \log n)\, j_{\lceil \frac{n^\star}{e} \rceil, p}(G^\star)} \geq \frac{n^\star}{ec \log n} \geq \frac{n}{e^2 c \log n}. \qquad\square$$

11

5.2 Proof of $g(n) = O\left(\frac{n \log \log n}{\log n}\right)$.

Our bound is based on an unusual weighted random graph construction due to Erdős and Pach [9], and a subsequent careful analysis of its properties by Kang, Pach, Patel and Regts [14], who established the following.

Proposition 13 (Theorem 1.4 in [14]). *For every $v > 0$ there exists $C_v > 0, n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ there exists an n-vertex graph G_n^{EP} in which every m-set of vertices with $m \geq C_v n \log \log n / \log n$ induces a subgraph of G_n^{EP} with minimum degree strictly less than $\frac{1}{2}(m-1) - m^{1-v}$ and maximum degree strictly greater than $\frac{1}{2}(m-1) + m^{1-v}$.*

Proposition 13 almost gives us what we want, namely a graph with no large full or co-full subgraph, with one caveat: it does not have density $\frac{1}{2}$. We circumvent this problem by taking disjoint copies of the Erdős–Pach construction and its complement and adding a carefully chosen random bipartite subgraph with density $\frac{1}{2}$ between them.

Let $\varepsilon > 0$. Fix $v = \frac{1}{3} - \varepsilon$, and let $2n \in \mathbb{N}$ be sufficiently large to ensure the existence of a graph G_{2n}^{EP} on $2n$ vertices satisfying the conclusion of Proposition 13.

Let A and B be disjoint sets of $2n$ vertices, each split into n pairs. Given a pair in A and a pair in B, place one of the two possible matchings between them selected uniformly at random, and do this independently for each of the n^2 pairs (pair from A, pair from B). This gives a random bipartite graph H between A and B with density precisely $\frac{1}{2}$. Add a copy of G_{2n}^{EP} to A and a copy of its complement $\left(G_{2n}^{EP}\right)^c$ to B to obtain a graph $G^\star$ on $4n$ vertices with density exactly $\frac{1}{2}$.

Let $m = m(n) = \lceil C_v(2n) \log \log(2n) / \log(2n) \rceil$, and let $2\lambda = m^{1-v}$. Let Y be a set of at least $4m$ vertices in $G^\star$. Without loss of generality, assume that $\ell = |Y \cap B| \leq |Y \cap A|$ and thus $|Y \cap A| \geq 2m \geq \ell$. Set X' to be the collection of vertices in $Y \cap A$ that have degree at most $\frac{1}{2}(|Y \cap A| - 1) - \lambda$ in $Y \cap A$. There are at least λ such vertices, for otherwise $X = (Y \cap A) \setminus X'$ is a set of at least $2m - \lambda > m$ vertices inducing a subgraph of G_{2n}^{EP} with minimum degree at least $\frac{1}{2}(|Y \cap A| - 1) - 2\lambda > \frac{1}{2}(|X| - 1) - |X|^{1-v}$, a contradiction. For each pair from A discard if necessary one of its two vertices from X' to obtain a set $X'' \subseteq X'$ of at least $\lambda/2$ vertices, each coming from a distinct pair. Note that by construction this means the degrees of the vertices from X'' into $Y \cap B$ are independent random variables with mean $\frac{1}{2}\ell$.

For Y to induce a full subgraph of $G^\star$, each vertex in X' would need to have at least $\frac{1}{2}\ell + \lambda$ neighbours in $Y \cap B$. By standard concentration inequalities (e.g. the Chernoff bound), the probability that a given vertex in X'' has that many neighbours in the ℓ-set $Y \cap B$ is at most $\exp\left(-\lambda^2/2\ell\right)$. By the independence noted above, the probability that all vertices in X'' have the right degree in $Y \cap B$ is thus at most

$$\exp\left(-\frac{\lambda^2}{2\ell}|X''|\right) \leq \exp\left(-\frac{\lambda^3}{4\ell}\right) \leq \exp\left(-\frac{1}{64}m^{3(1-v)-1}\right) = o(2^{-4n}).$$

It follows that asymptotically almost surely as $n \to \infty$ there is no pair (X', Z) where $X' \subseteq A$ is a collection of at least λ vertices and $Z \subseteq B$ is such that every vertex in X' has degree

at least $\frac{1}{2}|Z| + \lambda$ in Z. By symmetry, asymptotically almost surely no such pair exists either when $X' \subseteq B$ and $Z \subseteq A$, and in particular $G^\star$ contains no full subgraph on $4m$ vertices. Still by symmetry, asymptotically almost surely the complement $(G^\star)^c$ also fails to contain a full subgraph on $4m$ vertices. Thus asymptotically almost surely we have $g(G^\star) < 4m$.

In particular, for all $n \cong 0 \mod 4$ sufficiently large, there exist n-vertex graphs containing no full or co-full subgraph on $4m(n)$ vertices or more, and $g(4n) = O\left(\frac{n \log \log n}{\log n}\right)$. It is straightforward to adapt our construction of $G^\star$ to the case of sufficiently large n with $n \not\cong 0 \mod 4$ to show that more generally $g(n) = O\left(n \log \log n / \log n\right)$. $\square$

6   Proof of Theorem 1

Proof of $f(n, p) = O(n^{\frac{2}{3}})$ *for* $p = \frac{r}{r+1} + cn^{-\frac{2}{3}}$ *and* $c \geq 1$ *fixed.* Let $n \in \mathbb{N}$, and let $p = \frac{r}{r+1} + cn^{-\frac{2}{3}}$ for some $c \geq 1$ be such that $p\binom{(r+1)n}{2} \in \mathbb{N}$. Set $\delta = cn^{-\frac{2}{3}}$. Take a complete $(r+1)$-partite graph with parts $S_1, S_2, \ldots, S_{r+1}$, and n vertices in each part, and for $i = 1, 2, \ldots, r+1$, add a clique T_i of size k in S_i, such that

$$\binom{r+1}{2}n^2 + (r+1)\binom{k-1}{2} < p\binom{(r+1)n}{2} \leq \binom{r+1}{2}n^2 + (r+1)\binom{k}{2}.$$

A quick calculation shows $k \geq \sqrt{\delta r} n$ for n sufficiently large. Delete edges from the T_i in an equitable manner as necessary to obtain a graph G_n on $(r+1)n$ vertices with precisely $p\binom{(r+1)n}{2}$ edges. Suppose H is a full subgraph of G_n on $m > (r+1)k$ vertices, induced by sets $X_i \subseteq S_i$ where X_i has size s_i for $i = 1, 2, \ldots, r+1$. Without loss of generality, we may assume $s_1 = \max_i s_i > k$. For a vertex $v \in X_1$, let $d_j(v)$ denote the number of neighbours of v in X_j. Then for $x \in X_1 \setminus V(T_1)$, which is non-empty since $s_1 > k$,

$$d_H(x) = \sum_{j=2}^{r+1} d_j(x) \leq m - s_1 \leq \frac{r}{r+1}m.$$

On the other hand, since H is full,

$$d_H(x) \geq p(m-1) = \frac{r}{r+1}m + \delta m - p.$$

It follows that $\delta m \leq p$ and thus $m \leq \delta^{-1}p \leq c^{-1}n^{\frac{2}{3}}$. However we had assumed that $m > (r+1)k > r^{\frac{3}{2}}\sqrt{c}n^{\frac{2}{3}}$. Taken together, our bounds for m imply $c^{\frac{3}{2}} < r^{-\frac{3}{2}}$, and in particular $c < 1$, a contradiction. Thus

$$f(G_n) \leq (r+1)k = O\left(((r+1)n)^{\frac{2}{3}}\right).$$

This proves the second part of Theorem 1. $\square$

Proof of $f(n, p) \geq \frac{1}{4}(1-p)^{\frac{2}{3}}n^{\frac{2}{3}} - 1$ *for* $p = p_n :\ n^{-\frac{2}{3}} < p_n < 1 - n^{-\frac{1}{7}}$. Let G be an n-vertex graph of density p. We shall repeatedly delete vertices of minimum degree to obtain a sequence of subgraphs $G = G_1, G_2, G_3, \ldots$, with G_i having $n - i + 1$ vertices.

13

Let $m = \lceil \frac{n}{2} \rceil$ and $d_i = \lceil p(n-i) \rceil$. Note that d_i is the minimum degree required for G_i to be full. Let t be a positive integer so that $(1-p)^{-\frac{2}{3}} n^{\frac{1}{3}} \le 2^t < 2(1-p)^{-\frac{2}{3}} n^{\frac{1}{3}}$, and let r_i be the remainder when d_i is divided by 2^t. For at least $\frac{(1-p)}{2} m$ of the values $i: 1 \le i \le m$, we have $r_i \le (1-p)2^t$. At stage $i \le m$ of the algorithm, we delete a vertex of minimum degree from G_i. If for some $i \le m$ such that $r_i \le (1-p)2^t$, all $n-i+1$ vertices in the graph G_i have degree at least $d_i - r_i + 1$, then, by Theorem 8 (or Theorem 5 applied t times), G_i has a $\frac{1}{2^t}$-full subgraph H on N vertices, where

$$\left\lfloor \frac{n-i+1}{2^t} \right\rfloor \le N \le \left\lceil \frac{n-i+1}{2^t} \right\rceil + 1 \le \frac{n-i}{2^t} + 2 - \frac{1}{2^t}.$$

Write $d_i = q2^t + r_i$. The minimum degree in H is

$$D \ge \left\lceil \frac{d_i - r_i + 1}{2^t} \right\rceil = q + 1. \tag{6}$$

For H to be a full subgraph of G we require $D \ge p(N-1)$. Now

$$p(N-1) \le p\left(\frac{n-i}{2^t} + 1 - \frac{1}{2^t}\right)$$
$$< \frac{d_i}{2^t} + p = q + \frac{r_i}{2^t} + p, \tag{7}$$

which is at most $q+1$ since $r_i \le (1-p)2^t$. As this is at most our lower bound on D, H is a full subgraph of G. Our choice of t ensures

$$|V(H)| \ge \left\lfloor \frac{m}{2^t} \right\rfloor \ge \frac{(1-p)^{\frac{2}{3}} n^{\frac{2}{3}}}{4} - 1.$$

On the other hand suppose that at every stage $i \le m$ of the greedy algorithm where $r_i \le (1-p)2^t$, we could remove a vertex of degree at most $\lceil p(n-i) \rceil - r_i$ and that at every other stage $i \le m$ we could remove a vertex of degree at most $\lceil p(n-i) \rceil - 1$ (for otherwise we would have found a full subgraph on at least m vertices). Set $I = \{i \le m: r_i \le (1-p)2^t\}$. We know that $|I| \ge \frac{(1-p)m}{2}$. What is more, I can be divided into intervals of consecutive indices i of length at most $(1-p)2^t \cdot (\frac{1}{p})$, and over each of these intervals r_i takes each of the values $1, 2, \dots \lfloor (1-p)2^t \rfloor$ at least $\frac{1-p}{p}$ times. Indeed, suppose $r_{i-1} = j+1$ and $r_i = j$ for some $j \ge 1$. Then there is a $k: \frac{1-p}{p} \le k \le \frac{1}{p}$ such that $r_{i'} = j$ for $i' \in \{i, i+1, \dots i+k-1\}$ and $r_{i+k} = j-1$.

By considering $\sum r_i$ on these intervals and using $m = \lceil \frac{n}{2} \rceil$, $(1-p)^{-\frac{2}{3}} n^{\frac{1}{3}} \le 2^t < 2(1-p)^{-\frac{2}{3}} n^{\frac{1}{3}}$, we get that:

$$\begin{aligned}
\alpha := \operatorname{disc}^+(G) &\ge \left(p\binom{n}{2} - \sum_{i=1}^m (\lceil p(n-i) \rceil - 1) + \sum_{i \in I}(r_i - 1)\right) - p\binom{m}{2} \\
&\ge \left\lfloor \left(\frac{(1-p)m}{2}\right) \Big/ \left(\frac{(1-p)2^t}{p}\right) \right\rfloor \cdot \left(\sum_{j=1}^{\lfloor (1-p)2^t \rfloor} \frac{1-p}{p} j\right) \\
&\ge \left\lfloor \frac{p(1-p)^{\frac{2}{3}} n^{\frac{2}{3}}}{8} \right\rfloor \left(\frac{1-p}{2p}\right) \left(\lfloor (1-p)^{\frac{1}{3}} n^{\frac{1}{3}} \rfloor\right) \left(\lfloor (1-p)^{\frac{1}{3}} n^{\frac{1}{3}} \rfloor + 1\right) \\
&\ge \frac{(1-p)^{\frac{7}{3}}}{32} n^{\frac{4}{3}}.
\end{aligned}$$

14

Then, by Theorem 2, we have

$$f(G) \geq \sqrt{\frac{2\alpha}{1-p}} \rfloor \geq \frac{(1-p)^{\frac{2}{3}}}{4} n^{\frac{2}{3}}.$$

This completes the proof of Theorem 1. □

Remark. We did not optimise the constants in the proof of $f(n,p) = \Omega((1-p)^{\frac{2}{3}} n^{\frac{2}{3}})$, since it is unlikely that this argument gives an asymptotically tight lower bound on $f(n,p)$. Also note that for $p = 1 - o(n^{-\frac{1}{7}})$, this lower bound on $f(n,p)$ is superseded by that given in Theorem 2.

7 Concluding remarks

• We showed that $f(n,p) = \Omega(n^{\frac{2}{3}})$ and that this bound is tight up to constants for many values of p. It remains an open problem to determine the order of magnitude of $f(n,p)$ for each $p = p(n)$. Similarly, we leave it as an open problem to determine the order of magnitude of $g(n)$ and $g(n,p)$, having proved $g(n) = \Omega(n/\log n)$ and $g(n) = O(n \log \log n / \log n)$.

• One motivation for studying relatively half-full subgraphs, apart from their use in the proof of Theorem 1, is a random process on graphs known as majority bootstrap percolation: vertices of a graph are infected at time zero with probability p, and at any later time a vertex becomes infected if more than half of its neighbours are infected. A key quantity of interest in bootstrap percolation is the function $\theta_p(G)$, which is the probability that the process on the graph G infects all the vertices in finite time. The quantity $\theta_p(G)$ is precisely the probability that at time zero there is no relatively half-full subgraph of uninfected vertices. One may ask whether there exist $p > 0$ and $c < 1$ such that for every graph G, $\theta_p(G) \leq c$. In other words, is it the case that if the infection probability is too small (but still positive), then there is an absolute positive probability that on any graph G we fail to infect any vertex from some relatively half-full subgraph (and hence that the infection does not spread to all vertices of G)? Recently, this was answered in the negative by Mitsche, Pérez-Giménez and Prałat [17], who given any arbitrarily small $p > 0$ constructed a sequence of regular graphs $G_1, G_2, \ldots$ such that $\lim_{n \to \infty} \theta_p(G_n) = 1$. Another question, which remains open, is whether the number of relatively half-full subgraphs of an n-vertex graph grows at rate $\exp(\Theta(n))$ as $n \to \infty$.

• **Hypergraphs.** If G is an n-vertex r-uniform hypergraph of density p, then an m-vertex subgraph H of G is *full* if for every $v \in V(H)$, $d_H(v) \geq p\binom{m-1}{r-1}$. Write $f(G)$ for the order of a largest full subgraph of G and $f^r(n,p)$ for the minimum of $f(G)$ over all r-uniform hypergraphs G on n vertices with density p. The problems studied in this paper can be generalised to uniform hypergraphs, and in particular, one may ask for the order of $f^r(n,p)$. It is straightforward to imitate the proof of the upper bound in Theorem 1 to show that for certain values of $p \in (0,1)$ that $f^r(n,p) = O(n^{\frac{r}{r+1}})$. We leave open the problem of determining the tightness of this upper bound, as well as of giving bounds on $f^r(n,p)$ when p decays as n grows:

Problem 6. *Determine the order of magnitude of $f^r(n,p)$ for $p = p(n)$ and $r \geq 3$.*

The greedy algorithm from Section 3 generalises to r-uniform hypergraphs, in which it yields a full subgraph of order only $n^{\frac{1}{r}}$. It is an open question as to how the results on relatively q-full subgraphs in Section 1.6 may be extended to r-uniform hypergraphs. On the other hand, Proposition 11 was generalised to hypergraphs by Haviland and Thomason [12], and using their result, the proof of the lower bound in Theorem 4 extends to the setting of r-uniform hypergraphs exactly as before. In particular, every n-vertex r-uniform hypergraph contains a full or a co-full subgraph with $\Omega(n/\log n)$ vertices. In a recent preprint, Kang, Patel and Regts [15] prove a stronger version of this statement, which applies to multicoloured hypergraphs and gives subgraphs exceeding the degree requirement for fullness or co-fullness.

• **Digraphs.** We could also ask about directed graphs. Since every subgraph of a transitive tournament has a vertex of in-degree zero and a vertex of out-degree zero, it is more fruitful to ask about extensions of Theorem 5 than of Theorem 1. Let $d_H^+(v)$ denote the out-degree of a vertex $v \in H$. A subgraph H of a directed graph D is *relatively q-out-full* if for every $v \in H$ we have $d_H^+(v) \geq qd_D^+(v)$. Then the problem is to determine the smallest function $h(n, q)$ such that every digraph on n vertices has a relatively q-out-full subgraph with at most $h(n, q)$ vertices.

• **Weighted graphs.** A *weighted graph* is a pair $W = (V, w)$, where $w : V^{(2)} \to [0, 1]$ is a weighting of pairs of vertices from V. The *density* p of W is then the average pair-weight under w, and the *degree* $d_Y(x)$ of x in a subset $Y \subseteq V$ is the sum over all $y \in Y$ of $w(\{x, y\})$. Our definition of full subgraphs carries over to the weighted graph setting in the natural way, and we can ask:

Problem 7. *Let W be a weighted graph. Determine tight lower bounds for the order of a largest full subgraph of W.*

We may similarly ask about relatively q-full subgraphs for weighted graphs. The proof of Theorem 7 extends naturally to this setting: running through the same argument as before, we obtain a new equality

$$M' = M + (1 - q)d(Y) - d_Y(y) + qd(x) - d_X(x) - w(\{x, y\}).$$

From the maximality of $M = (1 - q)e(X) + qe(Y)$ over all weighted bipartitions, we deduce, in replacement of conditions (a)–(c), that

$$qd(x) - d_X(x) \leq w(\{x, y\}) \qquad \text{and} \qquad (1 - q)d(y) - d_Y(y) \leq w(\{x, y\}).$$

This implies as before that $X \cup \{y\}$ is relatively q-full on $\lceil qn \rceil + 1$ vertices and $Y \cup \{x\}$ is relatively $(1-q)$-full on $\lfloor (1-q)n \rfloor + 1$ vertices. Weighted analogues of Theorems 5 and 8 follow as immediate corollaries. However, this does not quite allow us to prove a weighted version of Theorem 1: in the weighted setting, degrees need not be integers and so the minimum degree lower bound (6) becomes $D > q$, while the upper bound on the degree requirement for being a full subgraph in (7) remains $q + 1 - p2^{-t}$. Thus our argument in this case only yields a 'weakly' full subgraph on N vertices (i.e. with minimum degree greater than $p(N-1) - 1$, rather than $p(N-1)$). Is this the best that can be done? We leave this as an open problem.

16

We note that the proofs of Theorems 2 and 3 carry over to the weighted setting without any changes. To recover a weighted analogue of our result on full and co-full subgraphs, Theorem 4, a weighted version of Thomason's theorem (Proposition 11) would be needed. This problem too is left open.

• **Computational complexity.** It appears likely that the following computational problems, which we have not investigated, are of similar complexity to Max Cut:

(i) find a largest full subgraph of G;

(ii) given an integer k, determine whether or not G contains a full subgraph on k vertices;

(iii) given an integer k, find a k-vertex subgraph with largest minimum degree.

The problem of finding (an approximation to) the densest subgraph of order k has received a significant amount of attention from the computer science community (see e.g. [3, 10]), including in some variants involving degree constraints [2]. From an algorithmic perspective, we thus expect that problems (i)–(iii) above will be hard: densest subgraph of order k is known to be NP-Hard. Further, examples due to Schäffer and Yannakakis [21] and Monien and Tscheuschner [18] for weighted versions of the Max-Cut problem suggest that local search (i.e. algorithms based on flipping vertices between a k-set X whose minimum degree we are trying to maximise and its complement) could take exponential time to converge to a local optimum for problem (iii) (see also the work of Poljak [19]). We note that the proofs of Theorems 1, 2, 5 and 8 yield polynomial time algorithms in each case.

Acknowledgements

References

[1] N. Alon, On the edge-expansion of graphs, Comb. Probab. Comp. 6 (1997), 145–152.

[2] O. Amini, D. Peleg, S. Pérennes, I. Sau, S. Saurabh, "Degree-constrained subgraph problems: Hardness and approximation results", Approximation and Online Algorithms, Springer (2009), pp. 29–42.

[3] A. Bhaskara, M. Charikar, E. Chlamtac, U. Feige and A. Vijayaraghavan, "Detecting High Log-densities: An $O(n^{\frac{1}{4}})$ Approximation for Densest K-subgraph", Proceedings of the Forty-second ACM Symposium on Theory of Computing (STOC), ACM (2010), pp. 201–220.

[4] B. Bollobás, Random Graphs, London: Academic Press, 1985.

[5] B. Bollobás, and A. Scott, "Discrepancy in graphs and hypergraphs", More Sets, Graphs and Numbers, Bolyai Society Mathematical Studies 15, Springer (2006), pp. 33–56.

[6] P. Erdős, J. Spencer, Imbalances in k-colorations, Networks 1 (1972), 379–385.

[7] P. Erdős, M. Goldberg, J. Pach and J. Spencer, Cutting a graph into two dissimilar halves, J. Graph Theory 12 (1988), 33–56.

[8] P. Erdős, T. Łuczak, J. Spencer, "Subgraphs of large minimal degree", Random graphs, Volume 2. Chichester: Wiley, Wiley-Interscience Publication (1992), pp. 59–66.

[9] P. Erdős and J. Pach, On a quasi-Ramsey problem, J. Graph Theory 7 (1983), 137–147.

[10] U. Feige, D. Peleg, and G. Kortsarz, The dense k-subgraph problem, Algorithmica 29 (2001), 410–421.

[11] Z. Füredi and J. Kómlos, The eigenvalues of random symmetric matrices, Combinatorica, 1 (1981), 233–241.

[12] J. Haviland and A. Thomason, Pseudo-random hypergraphs, Discrete Math. 75 (1989), 255–278.

[13] A. Holroyd, Sharp metastability threshold for two-dimensional bootstrap percolation, Probab. Theory Related Fields, 125(2) (2003), 195–224.

[14] R.J. Kang, J. Pach, V. Patel and G. Regts, A precise threshold for quasi-Ramsey numbers, SIAM J. Discrete Math. 29(3) (2015), 1670–1682.

[15] R.J. Kang, V. Patel and G. Regts, Discrepancy and large dense monochromatic subsets, preprint (2016), arXiv ref: 1610.06359.

[16] M. Krivelevich and B. Sudakov, ' Pseudo-random graphs", More Sets, Graphs and Numbers, Bolyai Society Mathematical Studies 15, Springer (2006), pp. 199–262.

[17] D. Mitsche, X. Pérez-Giménez, P. Prałat, Strong-majority bootstrap percolation on regular graphs with low dissemination threshold, Stochastic Processes and their Applications 17 (2017), pp. 3110–3134.

[18] B. Monien and T. Tscheuschner, "On the Power of Nodes of Degree Four in the Local Max-Cut Problem", Algorithms and Complexity, Springer (2010), pp. 264–275.

[19] S. Poljak, Integer Linear Programs and Local Search for Max-Cut, SIAM J. Comput. 24 (1995), 822–839.

[20] O. Riordan, A. Selby, The maximum degree of a random graph, Comb. Probab. Comp. 9 (2000), 549–572.

[21] A. Schäffer and M. Yannakakis, Simple local search problems that are hard to solve, SIAM J. Computing 20 (1991), 56–87.

[22] A. Scott, "Judicious partitions and related problems", Surveys in Combinatorics 2005, London Math. Soc. Lecture Note Ser. 327, Cambridge University Press, Cambridge (2005), pp. 95–117.

[23] A. Thomason, "Pseudo-random graphs", Proceedings of Random Graphs, Poznán 1985,

M. Karoński (ed.), Annals of Discrete Math. 33 (North Holland 1987), pp. 307–331.

[24] P. Turán, On an extremal problem in graph theory, Mat. Fiz. Lapok 48 (1941), 436–452.