THE ROYAL SWEDISH ACADEMY OF SCIENCES
INSTITUT MITTAG-LEFFLER
Auravägen 17, SE-182 60 Djursholm, Sweden
Tel. +46 8 622 05 60 Fax. +46 8 622 05 89
info@mittag-leffler.se www.mittag-leffler.se

Full subgraphs

V. Falgas-Ravry, K. Markström and J. Verstraete

REPORT No. 32, 2013/2014, spring

ISSN 1103-467X
ISRN IML-R- -32-13/14- -SE+spring

Full subgraphs

Victor Falgas-Ravry* Klas Markström† Jacques Verstraëte‡

Abstract

Let G be a graph of density p on n vertices. Following Erdős, Łuczak and Spencer [10], 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 positive integer, define

$$f_p(n) = \min\{f(G) : v(G) = n, e(G) = p\binom{n}{2}\}.$$

Erdős, Łuczak and Spencer [10] proved that if $p = \frac{1}{2}$, then for $n \geq 2$,

$$\sqrt{2n} - 2 \leq f_p(n) \leq 4n^{\frac{2}{3}}(\log n)^{\frac{1}{3}}.$$

In this paper, we show that for infinitely many p near the elements of $\{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \dots\}$, $f_p(n) = \Theta(n^{\frac{2}{3}})$, and more generally, for all $p:\ n^{-\frac{2}{3}} < p < 1 - n^{-\frac{1}{5}}$,

$$f_p(n) = \Omega((1-p)^{\frac{1}{3}} n^{\frac{2}{3}}).$$

As an ingredient of the proof, we also show that every graph G on n vertices has a subgraph H with $\lfloor \frac{n}{r} \rfloor$, $\lceil fracnr \rceil$ or $\lceil \frac{n}{r} \rceil + 1$ vertices such that for every $v \in V(H)$, $d_H(v) \geq \frac{1}{r} d_G(v)$. Finally, we discuss full subgraphs of random and pseudorandom graphs.

1 Introduction

A *full subgraph* of a graph G with (edge) density p is an m-vertex subgraph H of minimum degree at least $p(m-1)$. This is motivated by the fact that if we select m vertices in the random graph $G_{n,p}$, then $p(m-1)$ is exactly the expected number of selected neighbors of any vertex. In addition, one cannot in general expect subgraphs of higher minimum degree in complete multipartite graphs with parts of equal sizes. 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

$$f_p(n) = \min\{f(G) : v(G) = n, e(G) = p\binom{n}{2}\}.$$

*Department of Mathematics, Vanderbilt University, 1326 Stevenson Center, Nashville, Tennessee 37240, USA. Email: victor.falgas-ravry@vanderbilt.edu. Research supported by postdoctoral fellowships from the Kempe foundation and the Mittag-Leffler Institute.

†Institutionen för matematik och matematisk statistik, Umeå Universitet, 901 87 Umeå, Sweden. E-mail: klas.markstrom@math.umu.se

‡Department of Mathematics, University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92093-0112, USA. E-mail: jverstra@math.ucsd.edu. Research supported by NSF Grant DMS-110 1489.

We shall observe in Section 3 that if $p \leq n^{-\frac{2}{3}}$ then

$$|f_p(n) - p^{\frac{1}{2}}n| \leq 1. \tag{1}$$

In the same section we shall also prove that $f_p(n) = \Omega(\sqrt{\frac{n}{1-p}})$ for $p < 1$. Henceforth we discuss $f_p(n)$ for $n^{-\frac{2}{3}} < p < 1 - n^{-\frac{1}{5}}$. Erdős, Łuczak and Spencer [10] raised the problem of determining $f_p(n)$ when $p = \frac{1}{2}$, and showed

$$\sqrt{2n} - 2 \leq f_{\frac{1}{2}}(n) \leq (2 + \tfrac{2}{\sqrt{3}})n^{\frac{2}{3}}(\log n)^{\frac{1}{3}}.$$

In this paper, we prove the following theorem:

Theorem 1. *For all* $p = p_n : \ n^{-\frac{2}{3}} < p_n < 1 - n^{\frac{1}{5}}$,

$$f_p(n) = \Omega((1-p)^{\frac{1}{3}} n^{\frac{2}{3}}).$$

If $p = \frac{r}{r+1} + cn^{-\frac{2}{3}}$ *for some* $r \in \mathbb{N}$ *and* $c > 1$*, then* $f_p(n) = \Theta(n^{\frac{2}{3}})$.

A case of particular interest is $p = \frac{1}{2}$, where Theorem 1 together with the results of Erdős, Łuczak and Spencer gives

$$n^{\frac{2}{3}} \leq f_{\frac{1}{2}}(n) \leq (2 + \tfrac{2}{\sqrt{3}})n^{\frac{2}{3}}(\log n)^{\frac{1}{3}}.$$

A similar construction shows that $f_p(n)$ 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 $f_p(n)$ 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* $f_p(n)$*.*

1.1 Discrepancy and full subgraphs

For a graph G of density p and a v-vertex subgraph set $X \subset V(G)$, let $\delta(X) = e(X) - p\binom{v}{2}$. 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 result via a greedy algorithm in Section 3, relating positive discrepancy to full subgraphs:

Theorem 2. *Let* G *be a graph with disc*$^+(G) = \alpha > 0$ *and edge density* p*. Then*

$$f(G) \geq \sqrt{\frac{2\alpha}{1-p}}.$$

It is easy to see that Theorem 2 is best possible, since the graph G consisting of a clique with $\binom{m}{2}$ edges and $n - m$ isolated vertices has $\mathrm{disc}^+(G) = \binom{m}{2}(1 - \frac{\binom{m}{2}}{\binom{n}{2}})$ and $f(G) = m$, which, provided m, n are chosen sufficiently large, is strictly less than $\sqrt{2\mathrm{disc}^+(G)} + 1$. However, it is possible for $\mathrm{disc}^+(G)$ to be linear in n, for example for complete multipartite graphs with parts of equal sizes, in which case Theorem 1 gives much better lower bounds on $f(G)$.

1.2 Random and pseudorandom graphs

Extending earlier results of Erdős and Spencer [8] and Erdős, Goldberg, Pach and Spencer [9], Bollobás and Scott [7] showed that if G is an n-vertex graph with density p satisfying $p(1-p) \geq \frac{1}{n}$, then

$$\mathrm{disc}^{+}(G) \cdot \mathrm{disc}^{-}(G) = \Omega(p(1-p)n^3). \tag{2}$$

By the results of Erdős and Spencer [8], this is tight for $p = \frac{1}{2}$, and it is also tight for $p(1-p) \geq \frac{1}{n}$ by extending the arguments in [8] to show that asymptotically almost surely,

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

and the same holds for $\mathrm{disc}^{-}(G_{n,p})$. While it is not difficult to show that $\mathrm{disc}(G_{n,p})$ is concentrated, the asymptotic expected value of the discrepancy is not known.

For a graph G whose adjacency matrix has real eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, define $\lambda(G) = \max\{|\lambda_i| : 2 \leq i \leq n\}$. Then Alon [1] proved $|\delta(X)| \leq \lambda(G)|X|$ for every set $X \subset V(G)$, and in particular $\mathrm{disc}(G) \leq \lambda(G)n$. This result is sometimes called the Expander Mixing Lemma (see the survey of Krivelevich and Sudakov [17]). We relate full subgraphs to positive discrepancy and eigenvalues as follows:

Theorem 3. *Let G be an n-vertex graph with $\mathrm{disc}^{+}(G) = \alpha$ and $\lambda(G) = \lambda$. Then*

$$f(G) \geq \frac{\alpha}{\lambda}.$$

This theorem is proved in Section 4.

Erdős, Łuczak and Spencer [10] showed that for $p = \frac{1}{2} + o(1)$, $f(G_{n,p}) \geq \beta_1 n - o(n)$ asymptotically almost surely, where $\beta_1 \approx 0.227$. If $G = G_{n,p}$, then asymptotically almost surely $\lambda(G) \sim 2\sqrt{pn}$ (see Füredi and Kómlos [14] and Vu [24]). Combining this with (3), we obtain asymptotically almost surely:

$$f(G_{n,p}) = \Omega\Big(\frac{p^{\frac{1}{2}}(1-p)^{\frac{1}{2}}n^{\frac{3}{2}}}{2(pn)^{\frac{1}{2}}}\Big) = \Omega((1-p)^{\frac{1}{2}}n). \tag{4}$$

In the other direction, results of Riordan and Selby [21] imply $f(G_{n,p}) \leq \beta_2 n + o(n)$ asymptotically almost surely, where $\beta_2 \approx 0.851\ldots$. We believe that $f(G_{n,p})$ is concentrated around $\alpha n + o(n)$ for some function $\alpha = \alpha(p)$ (note that f is not a Lipschitz function, so that this is not immediate from standard martingale arguments). An interesting problem is to prove this concentration and determine the correct value of $\alpha(p)$.

1.3 Full and co-full subgraphs

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, if H has m vertices and G has density p, then H is co-full if $V(H)$ induces a subgraph of G with 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. By using Theorem 3 we obtain the following:

Theorem 4. *For any n-vertex graph G of density p with $p(1-p) \geq \frac{1}{n}$, $g(G) = \Omega(\frac{n}{(\log n)^2})$.*

The proof of Theorem 4, given in Section 4, relies on results of Bollobás and Nikiforov [6] (see also Bilu and Linial [18]). A consequence of their results is that for every n-vertex graph G of density p with $\lambda(G) = \lambda$, $\mathrm{disc}(G) = \Omega(\frac{\lambda n}{\log n)^2})$. Let $g(n) = \min\{g(G) : |V(G)| = n\}$. Our work leaves the following open problem:

Problem 2. *Determine the order of magnitude of $g(n)$.*

By (4) with $p = \frac{1}{2}$, note that $g(G)$ is linear in n for almost all n-vertex graphs G. See also Erdős and Pach [12] and Kang, Pach, Patel and Regts [16] for related work on quasi-Ramsey numbers.

1.4 Half-full subgraphs

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

Theorem 5. *Let G be an n-vertex graph. Then G contains a 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 half-full subgraph of K_n has $\lfloor \frac{n}{2} \rfloor + 1$ vertices and the smallest half-full subgraph of $K_{n,n}$ has $n+1$ vertices when n is odd. For regular graphs, we obtain:

Corollary 1. *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 [2] 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 $\frac{1}{2}d + cd^{\frac{1}{2}}$, exceeding the requirement for a full subgraph and improving Corollary 1 by an additive factor of $cd^{\frac{1}{2}}$. However, as observed by Alon [2], such a result does not hold for large d, as for example complete graphs and complete bipartite graphs show.

1.5 q-Full subgraphs

For $q \in [0,1]$, define a subgraph H to be *q-full* if $d_H(v) \geq q d_G(v)$ for all $v \in V(H)$. We prove:

Theorem 6. *Let G be a graph on n vertices. Then for every $q \in [0,1]$, G contains one of the following:*

(i) a q-full subgraph on $\lceil qn \rceil$ vertices,

(ii) a $(1-q)$-full subgraph on $\lfloor (1-q)n \rfloor$ vertices,

(iii) a q-full subgraph on $\lceil qn \rceil + 1$ vertices and a $(1-q)$-full subgraph on $\lfloor (1-q)n \rfloor + 1$ vertices.

Using Theorem 6, we prove Theorem 5 and an extension to $\frac{1}{r}$-full subgraphs for $r \geq 3$:

Theorem 7. *Let G be a graph on n vertices, and let $r \in \mathbb{N}$. Then G contains a $\frac{1}{r}$-full subgraph on $\lfloor \frac{n}{r} \rfloor$, $\lceil \frac{n}{r} \rceil$ or $\lceil \frac{n}{r} \rceil + 1$ vertices.*

Theorem 7 is best possible in the following sense: if $r \geq 3$, consider the complete graph K_n for some $n \geq r+1$ with $n \equiv 1 \mod r$. A smallest $\frac{1}{r}$-full subgraph of K_n has exactly $\lceil \frac{n}{r} \rceil + 1$ vertices.

It is natural to ask whether Theorem 7 can be extended further to cover other q.

Question 1. *Does there exist a constant c such that for every $q \in [0, \frac{1}{2}]$, every n-vertex graph G has a q-full subgraph on m vertices for some $m: \lfloor qn \rfloor \leq m \leq \lfloor qn \rfloor + c$?*

For $q > \frac{1}{2}$, a cycle of length n shows that there exist n-vertex graphs with no non-empty 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 q-full* if $d_H(v) \geq \lfloor q d_G(v) \rfloor$ for all $v \in V(H)$. However even for this notion of q-fullness a natural generalization of Theorem 7 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 $\frac{3}{4}$-full subgraph H, then all but at most one of its neighbors 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 q-full subgraphs when $q > \frac{1}{2}$.

Question 2. *Let $q \in (\frac{1}{2}, 1)$ be a fixed rational number. Does there exist a constant $f(q)$: $q < f(q) < 1$ such that every n-vertex graph G has a weakly q-full subgraph on m vertices for some m: $\lfloor qn \rfloor \leq m \leq f(q)n + O(1)$?*

1.6 Notation

We use standard graph theoretic notation. In particular, if X, Y are sets of vertices of a graph G, then $e(X)$ denotes the number of edges in the subgraph of G induced by X and $e(X, Y)$ the number of edges with one end in X and the other end in Y. Denote by $d_X(x)$ the number of neighbors in X of a vertex $x \in V(G)$. The Erdős-Rényi random graph with edge-probability p on n vertices is denoted $G_{n,p}$. 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 q-full subgraphs : Proofs of Theorems 5 – 7

Proof of Theorem 6. 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 $|Y| = \lfloor (1-q)n \rfloor$ maximizing the value of $(1-q)e(X) + qe(Y)$.

If X is q-full or Y is $(1-q)$-full, then we are done. Otherwise there exist $x \in X$ and $y \in Y$ with $d_X(x) \le \lceil q d_G(x) \rceil - 1$ and $d_Y(y) \le \lceil (1-q) d_G(y) \rceil - 1$. Let $X' = (X \setminus \{x\}) \cup \{y\}$ and $Y' = (Y \setminus \{y\}) \cup \{x\}$. Then

$$\begin{aligned}
&(1-q)e(X') + qe(Y') \\
&\quad = (1-q)e(X) + qe(Y) + \quad (1-q)d_X(y) - qd_Y(y) + qd_Y(x) - (1-q)d_X(x) - \mathbf{1}_{xy\in E(G)} \\
&\quad = (1-q)e(X) + qe(Y) + \quad (1-q)d_G(y) - d_Y(y) + qd_G(x) - d_X(x) - \mathbf{1}_{xy\in E(G)} \\
&\quad \ge (1-q)e(X) + qe(Y) + \quad (1-q)d_G(y) - \lceil (1-q)d_G(y) \rceil + qd_G(x) - \lceil qd_G(x) \rceil \\
&\qquad\qquad + 2 - \mathbf{1}_{xy\in E(G)}.
\end{aligned}$$

Since $X \sqcup Y$ maximized $(1-q)e(X) + qe(Y)$ over all bipartitions with $|X| = \lceil qn \rceil$, $|Y| = \lfloor (1-q)n \rfloor$, we deduce from the inequality above that (a) $d_X(x) = \lceil q d_G(x) \rceil - 1$, $d_Y(y) = \lceil (1-q) d_G(y) \rceil - 1$, and, crucially, (b) $\{x, y\} \in E(G)$.

Now denote by B_X the set of $x \in X$ with $d_X(x) \le \lceil q d_G(x) \rceil - 1$ and B_Y the set of $y \in Y$ with $d_Y(y) \le \lceil (1-q) d_G(y) \rceil - 1$. By our assumption, both sets are non-empty. Our argument above establishes that (a) $d_X(x) = \lceil q d_G(x) \rceil - 1$ for all $x \in B_X$, $d_Y(y) = \lceil (1-q) d_G(y) \rceil - 1$ for all $y \in B_Y$, and that (b) $B_X \sqcup B_Y$ induces a complete bipartite subgraph of G. Thus for every $x \in B_X, y \in B_Y$, $X \cup \{y\}$ is a q-full subgraph on $\lceil qn \rceil + 1$ vertices and $Y \cup \{x\}$ is a $(1-q)$-full subgraph on $\lfloor (1-q)n \rfloor + 1$ vertices. □

Proof of Theorem 5. Apply Theorem 6 with $q = \frac{1}{2}$. □

Proof of Corollary 1. 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

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

the subgraph H is a full subgraph. □

Proof of Theorem 7 We use Theorem 6 and induction on r. The base case $r = 1$ is trivial, and Theorem 5 deals with the case $r = 2$. Now apply Theorem 6 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. □

3 A greedy algorithm : Proof of Theorem 2

An obvious strategy for obtaining a full subgraph is to repeatedly remove vertices of relatively low degree in a graph G of density p on n vertices. When there are i vertices left in the graph, the 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 α, 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 $\alpha > 0$, then its edge density p is strictly less than 1. Let H be a subgraph of G with m vertices such that $e(H) \geq p\binom{m}{2} + \alpha$. 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

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

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 $f(G) \geq m - i \geq \sqrt{\frac{2\alpha}{1-p}}$. □

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 $f_p(n) = O(p^{\frac{1}{2}} n)$ for $p \leq n^{-\frac{2}{3}}$. We now prove (1), which states that $|f_p(n) - p^{\frac{1}{2}} n| \leq 1$ for $p \leq n^{-\frac{2}{3}}$.

Proof of (1). First we show $f_p(n) \leq p^{\frac{1}{2}} n + 1$ 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 $f(G) \le m \le p^{\frac{1}{2}}n+1$. 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 $p \le n^{-\frac{2}{3}}$. 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 less 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 addition of edges adds at most two vertices at a time. This subgraph is full since

$$p(p^{\frac{1}{2}}n) \le \lceil p^{\frac{3}{2}}n \rceil \le 1$$

when $p \le n^{-\frac{2}{3}}$, 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

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

it follows that

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

For instance if say $p_n = \frac{1}{2} - o(1) < \frac{1}{2}$, then $f_{p_n}(n) \ge (1 - o(1))\sqrt{2n}$, as shown by Erdős, Łuczak and Spencer [10]. These lower bounds on $f_p(n)$ will be superseded by the better bounds given in Theorem 1.

We note that there exist examples of n-vertex graphs with edge 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 n^{th} 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 $K_{m,m}$.

A final remark is that every n-vertex graph G of density p with $\frac{1}{n} \le p \le \frac{1}{2}$ has an m-vertex subgraph with minimum degree at least $p(m-1) + \Omega(\sqrt{n})$. This is a consequence of the fact that every n-vertex graph of density p has positive discrepancy of order $(1-p)n$, due to (2). The random graph $G_{n,p}$ has no m-vertex subgraph of larger minimum degree, by standard concentration inequalities.

4 Proof of Theorems 3 and 4

Proof of Theorem 3. Let G be a graph on n vertices with $\lambda(G) = \lambda$ and density p and positive discrepancy α. We show $f(G) \geq \frac{\alpha}{\lambda}$. Let X be a set of m vertices with $e(X) \geq p\binom{m}{2} + \alpha$. At stage $t \geq 1$, we remove a vertex of X of degree strictly less than $p(m-t)$ in the subgraph induced by X, if it exists, and otherwise we stop. Alon [1] proved that if $\lambda(G) = \lambda$, then for every set $Y \subset V(G)$,

$$\delta(Y) \leq \lambda |Y|. \tag{5}$$

By (5), at each stage of the process we must have

$$\begin{aligned} p\binom{m-t}{2} + \lambda(m-t) &\geq e(X) - \sum_{i=1}^{t} p(m-i) \\ &\geq p\binom{m}{2} + \alpha - p\binom{m}{2} + p\binom{m-t}{2}, \end{aligned}$$

from which we deduce that the algorithm must terminate by stage t for some t satisfying $\lambda(m-t) \geq \alpha$. This gives us a full subgraph on $m - t \geq \frac{\alpha}{\lambda}$ vertices. $\square$

Proof of Theorem 4. Let G be an n-vertex graph. We use Theorem 3 together with

$$\text{disc}(G) = \Omega\left(\frac{n\lambda(G)}{(\log n)^2}\right) \quad \text{and} \quad \text{disc}(\overline{G}) = \Omega\left(\frac{n\lambda(\overline{G})}{(\log n)^2}\right)$$

as proved by Bollobás and Nikiforov [6]. By Theorem 3,

$$g(G) \geq \max\left\{\frac{\text{disc}^+(G)}{\lambda(G)}, \frac{\text{disc}^+(\overline{G})}{\lambda(\overline{G})}\right\}.$$

Since $\text{disc}^+(\overline{G}) = \text{disc}^-(G)$ and $\text{disc}(G) = \text{disc}(\overline{G})$, we have $\text{disc}^+(G) \geq \text{disc}(G)$ or $\text{disc}^+(\overline{G}) \geq \text{disc}(\overline{G})$. Therefore

$$g(G) \geq \max\left\{\frac{\text{disc}(G)}{\lambda(G)}, \frac{\text{disc}(\overline{G})}{\lambda(\overline{G})}\right\}.$$

Using the lower bounds on $\text{disc}(G)$ and $\text{disc}(\overline{G})$ above, we obtain Theorem 4. $\square$

5 Proof of Theorem 1

Proof of $f_p(n) = O(n^{\frac{2}{3}})$ for $p = \frac{r}{r+1} + cn^{-\frac{2}{3}}$ and $c \geq 1$. It is convenient to let $\delta = cn^{-\frac{2}{3}}$. 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

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

Let this graph be G_n. In particular, a calculation shows $k \geq \sqrt{\delta r} n$. Suppose H is a full subgraph of G_n with more than $(r+1)k$ vertices, induced by sets $X_i \subset S_i$ where X_i has size

s_i for $i = 1, 2, \ldots, r+1$. By symmetry, we may assume $s_1 = s_2 = \cdots = s_{r+1} = t > k$. For a vertex $v \in S_i$, let $d_j(v)$ denote the number of neighbors of v in X_j. Then for $x \in X_1 \backslash V(T_1)$, which is non-empty since $t > k$,

$$d_H(x) = \sum_{j=2}^{r+1} d_j(x) = rt.$$

On the other hand, since H is full,

$$d_H(x) \geq p((r+1)t - 1) = rt + \delta(r+1)t - p.$$

It follows that $\delta rt \leq p$, and since $t > k$, this implies $\delta^{\frac{3}{2}} n \leq (\delta r)^{\frac{3}{2}} n < 1$. Since $\delta \geq n^{-\frac{2}{3}}$, this is a contradiction and shows

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

This proves the last statement of Theorem 1. □

Proof of $f_p(n) = \Omega(n^{\frac{2}{3}})$ for $p = p_n > n^{-\frac{2}{3}}$. Let G be an n-vertex graph with density p. We 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.

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 $(1-p)^{-\frac{1}{3}} n^{\frac{1}{3}} \leq 2^t < 2(1-p)^{-\frac{1}{3}} n^{\frac{1}{3}}$, 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 $i: \ 1 \leq i \leq m$, we have $r_i \leq (1-p)2^t$. At stage $i \leq m$ of the algorithm, we delete a vertex of minimum degree from G_i. If for some $i \leq m$ such that $r_i \leq (1-p)2^t$, all $n - i + 1$ vertices in the graph G_i have degree at least $d_i - r_i + 1$, then, by Theorem 7 (or Theorem 5 applied t times), G_i has a $\frac{1}{2^t}$-full subgraph H on N vertices, where

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

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

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

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

$$\begin{aligned} p(N-1) &\leq p\left(\frac{n-i+1}{2^t} + 1 - \frac{1}{2^t}\right) \\ &\leq \frac{d_i}{2^t} + p = q + \frac{r_i}{2^t} + p, \end{aligned}$$

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

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

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 $\lceil p(n-i) \rceil - r_i$. 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 $(1-p)2^t(\frac{2}{p})$, and over each of these intervals r_i takes each of the value $1, 2, \ldots \lfloor (1-p)2^t \rfloor$ at least $\frac{1}{2p}$ times. By considering $\sum r_i$ on these intervals and using $m = \lceil \frac{n}{2} \rceil$, $2^t > n^{\frac{1}{3}}$, we get that:

$$\begin{aligned}
\alpha := \text{disc}^+(G) &\ge p\binom{n}{2} - \sum_{i=1}^{m}(\lceil p(n-i) \rceil - 1) + \sum_{i \in I}(r_i - 1) - p\binom{m}{2} \\
&\ge \left\lfloor \left(\frac{(1-p)m}{2}\right) \Big/ \left(\frac{(1-p)2^t 2}{p}\right) \right\rfloor \sum_{j=1}^{\lfloor (1-p)2^t \rfloor} \frac{1}{2p} j \\
&\ge \frac{(1-p)^{\frac{5}{3}}}{64} n^{\frac{4}{3}}.
\end{aligned}$$

Then by Theorem 2 we have

$$f(G) \ge \sqrt{\frac{2\alpha}{1-p}} \ge \frac{(1-p)^{\frac{1}{3}}}{8} n^{\frac{2}{3}}.$$

This completes the proof of Theorem 1. □

Remark. We did not optimize the constants in the proof of $f_p(n) = \Omega((1-p)^{\frac{1}{3}} n^{\frac{2}{3}})$, since it is unlikely that this argument gives an asymptotically tight lower bound on $f_p(n)$.

6 Concluding remarks

• A motivation for the introduction of half-full subgraphs, apart from their use in the proof of Theorem 1, is a random process on a graph known as bootstrap percolation: vertices of the graph are initially infected at time zero with probability p, and at any later time, a vertex becomes infected if more than half of its neighbors 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 half-full subgraph of uninfected vertices. One may ask the following question:

Question 3. *Do there exist $p > 0$ and $c < 1$ such that for every graph G, $\theta_p(G) \le c$?*

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 the infection does not spread to all vertices of G? It should be remarked that if we only require infection of *at least* *half* of the neighbors of a vertex to infect it, then for any $p > 0$, there are many sequences of graphs G_n, such as $n \times n$ grids [15], for which $\theta_p(G_n) \to 1$ as $n \to \infty$.

• **Hypergraphs.** If G is an n-vertex r-uniform hypergraph of density p, then an m-vertex subhypergraph H of G is *full* if for every $v \in V(H)$, $d_H(v) \geq p^{r-1}\binom{m-1}{r-1}$ and *half-full* if for every $v \in V(H)$, $d_H(v) \geq 2^{1-r} d_G(v)$. We may define $f(G)$ to be the largest number of vertices in a full subgraph of an r-uniform hypergraph G of density p, and let

$$f_p^{(r)}(n) = \min\{f(G) : e(G) = p\binom{n}{r}\}.$$

It is straightforward to imitate the proof of Theorem 1 to show for certain values of $p \in (0,1)$ that $f_p^{(r)}(n) = O(n^{\frac{r}{r+1}})$. We leave the problem of determining the tightness of this upper bound, as well as bounds on $f_p^{(r)}(n)$ when p decays as n grows:

Problem 3. *Determine the order of magnitude of $f_p^{(r)}(n)$ for $p \in (0,1)$ and $r \geq 3$.*

The greedy algorithm from Section 3 generalizes to r-uniform hypergraphs, in which it yields a full subgraph of order only $n^{\frac{1}{r}}$. The results on half-full subgraphs (Section 2) do not seem to generalize in an easy way to half-full subgraphs of hypergraphs.

• A subgraph F of an r-uniform hypergraph H is q-full if $d_F(v) \geq q^{r-1} d_H(v)$ for every $v \in V(F)$. Perhaps it is true that every n-vertex r-uniform hypergraph has a q-full subgraph with $\lfloor qn \rfloor + O(1)$ vertices, but we do not have a proof even for $q = \frac{1}{2}$ and $r = 3$. Complete hypergraphs show if this is true, then it is best possible. We leave this as an open problem.

• **Digraphs.** We could also ask about directed graphs. Since every subgraph of a transitive tournament has a vertex of indegree zero and a vertex of outdegree zero, it is more fruitful to ask about extensions of Theorem 5 than of Theorem 1. Let $d_H^+(v)$ denote the outdegree of a vertex $v \in H$. A subgraph H of a directed graph D is *q-out-full* if for every $v \in H$ we have $d_H^+(v) \geq q d_D^+(v)$. Then the problem is to determine the smallest function $h(n,q)$ such that every digraph on n vertices has a q-out-full subgraph with at most $h(n,q)$ vertices. If the outdegrees in the digraph are sufficiently large relative to $\log n$, then a random subset of roughly $qn + (n \log n)^{\frac{1}{2}}$ vertices induces a q-out-full subgraph. Therefore the more precise question is how large $h(n,q) - qn$ can be.

• **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 *edge-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 definitions of full subgraphs carry over to the weighted graph setting in the natural way, and we can ask:

Problem 4. *Let W be a weighted graph. Determine the largest number of vertices in a full subgraph of W.*

A similar question may be asked for finding q-full subgraphs for $q \in [0, \frac{1}{2}]$: what is the smallest c such that every n-vertex weighted graph has a q-full subgraph with at most $\lfloor qn \rfloor + c$ vertices. We note here that the proof of Theorem 1 carries over to the weighted setting modulo a weighted version of Theorem 5. The proof of Theorem 5 itself, however, does not transfer in an obvious way to the weighted setting.

• **Computational complexity.** It appears likely that the following computational problems, which we have not investigated in this paper, 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. [4, 13]), including in some variants involving degree constraints [3]. 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 [22] and Monien and Tscheuschner [19] 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 maximize and its complement) could take exponential time to converge to a local optimum for problem (iii) (see also the work Poljak [20]). The proofs of the theorems in this paper lead to polynomial time algorithms in each case.

Acknowledgement

This research was conducted while the authors were attending the research semester on Graphs, hypergraphs and computing at the Institut Mittag-Leffler (Djursholm, Sweden) , whose hospitality the authors gratefully acknowledge. We also thank L. Lovász who pointed out the work of S. Poljak, and A. Scott for a helpful discussion of counterexamples for partition problems.

References

[1] N. Alon, Eigenvalues and expanders, Combinatorica **6** (1986), 83–96.

[2] N. Alon, On the edge-expansion of graphs, Combinatorics, Probability and Computing **6** (1997), 145–152.

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

[4] 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, pp 201–220 in: Proceedings of the Forty-second ACM Symposium on Theory of Computing, 2010.

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

[6] B. Bollobás and V. Nikiforov, Hermitian matrices and graphs: singular values and discrepancy, Discrete Mathematics (2004) **285**, 17–32.

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

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

[9] P. Erdős, M. Goldberg, J. Pach and J. Spencer, Cutting a graph into two dissimilar halves, Journal of graph theory **12**, 121–131 (1988).

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

[11] P. Erdős, L. Lovász, On 3-chromatic hypergraphs, Colloquia Mathematica Sodetatis János Bolyai, North-Holland, Amsterdam (1975), 609–627.

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

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

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

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

[16] R.J. Kang, J. Pach, V. Patel and G. Regts, A precise threshold for quasi-Ramsey numbers, arXiv preprint (2014), arXiv ref:1403.3464.

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

[18] Y. Bilu, N. Linial, Lifts, discrepancy and nearly optimal spectral gap, Combinatorica **26**(5) (2006), 495–519.

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

[20] S. Poljak, Integer Linear Programs and Local Search for Max-Cut, SIAM Journal on Computing **24** (1995), 822–839.

[21] O. Riordan, A. Selby, The maximum degree of a random graph, Combinatorics, Probability and Computing **9** (2000), 549–572.

[22] A. Schäffer and M. Yannakakis, Simple local search problems that are hard to solve,

SIAM Journal on Computing **20** (1991), 56–87.

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

[24] V. Vu. Spectral norm of random matrices, pp 619–626 in: ACM Symposium on Theory of Computing (STOC), 2005.