Combinatorics, Probability and Computing (2017) 26, 52–67.
doi:10.1017/S0963548316000316

Saturated Subgraphs of the Hypercube

J. ROBERT JOHNSON and TREVOR PINTO†

School of Mathematical Sciences, Queen Mary University of London, London E1 4NS, UK
(e-mail: r.johnson@qmul.ac.uk)

Received 1 September 2014; revised 27 February 2016; first published online 19 September 2016

We say a graph is (Q_n, Q_m)-*saturated* if it is a maximal Q_m-free subgraph of the n-dimensional hypercube Q_n. A graph is said to be (Q_n, Q_m)-*semi-saturated* if it is a subgraph of Q_n and adding any edge forms a new copy of Q_m. The minimum number of edges a (Q_n, Q_m)-saturated graph (respectively (Q_n, Q_m)-semi-saturated graph) can have is denoted by $\mathrm{sat}(Q_n, Q_m)$ (respectively $s\text{-sat}(Q_n, Q_m)$). We prove that

$$\lim_{n\to\infty} \frac{\mathrm{sat}(Q_n, Q_m)}{e(Q_n)} = 0,$$

for fixed m, disproving a conjecture of Santolupo that, when $m = 2$, this limit is $1/4$. Further, we show by a different method that $\mathrm{sat}(Q_n, Q_2) = O(2^n)$, and that $s\text{-sat}(Q_n, Q_m) = O(2^n)$, for fixed m. We also prove the lower bound

$$s\text{-sat}(Q_n, Q_m) \geqslant \frac{m+1}{2} \cdot 2^n,$$

thus determining $\mathrm{sat}(Q_n, Q_2)$ to within a constant factor, and discuss some further questions.

2010 *Mathematics subject classification*: Primary 05C35
Secondary 05D05

1. Introduction

Let F be a (simple) graph. We say that a (simple) graph G is F-*free* if it contains no subgraphs isomorphic to F. If G is a maximal F-free subgraph of H, we say that G is (H,F)-*saturated*. In other words, G is F-saturated if it is an F-free subgraph of H and the addition of any edge from $E(H) \setminus E(G)$ forms a copy of F. In this context, H is referred to as the *host graph*, F as the *forbidden graph* and G as a *saturated graph*.

The famous Turán problem in extremal combinatorics can be expressed naturally in the language of saturated graphs. The extremal number of F, $\mathrm{ex}(K_n, F)$ (often written as $\mathrm{ex}(n, F)$), is usually defined as the maximum number of edges in an F-free subgraph of K_n. However, it can

† Supported by an EPSRC doctoral studentship.

equivalently be written as

$$\mathrm{ex}(K_n, F) = \max\{e(G) : G \text{ is } (K_n, F)\text{-saturated}\}.$$

This formulation yields a natural 'opposite' of the Turán problem. We define the *saturation number of* F, $\mathrm{sat}(H,F)$, as

$$\mathrm{sat}(H, F) = \min\{e(G) : G \text{ is } (H, F)\text{-saturated}\}.$$

A variant of this is the *semi-saturation number*, $s\text{-sat}(H,F)$. We say that a graph is (H,F)-*semi-saturated* if G is a subgraph of H and adding any edge from $E(H)\setminus E(G)$ increases the number of copies of F. A graph is (H,F)-saturated if and only if it is (H,F)-semi-saturated and F-free. We define

$$s\text{-sat}(H, F) = \min\{e(G) : G \text{ is } (H, F)\text{-semi-saturated}\}.$$

The most frequently studied host graph is the complete graph, K_n. Since work in the area began with Erdős, Hajnal and Moon [6], many others have studied $s\text{-sat}(K_n,F)$ and $\mathrm{sat}(K_n,F)$: see for instance the survey articles by Pikhurko [11] and J. Faudree, R. Faudree and Schmitt [7], and the references contained therein.

In the literature, $\mathrm{sat}(K_n,F)$ is often written as $\mathrm{sat}(n,F)$, and (K_n,F)-saturated is usually written as F-saturated. Since the results in this paper concern a different host graph, we will reserve this latter abbreviation for a different meaning.

A much studied variant of the Turán problem was initiated by Erdős in [5] and expanded upon by Alon, Krech and Szabó [1]. For a fixed graph F, they ask for $\mathrm{ex}(Q_n,F)$, the maximum number of edges in an F-free subgraph of the n-dimensional hypercube, Q_n. The most natural case is $F = Q_m$, a fixed cube. This is wide open, even for the case $m = 2$. The asymptotic edge density of a maximum Q_2-free graph, that is,

$$\lim_{n\to\infty} \frac{\mathrm{ex}(Q_n, Q_2)}{e(Q_n)},$$

was conjectured by Erdős [5] to be $1/2$. It is still unknown, despite the attention of many authors: see for instance the work of Balogh, Hu, Lidický and Liu [2] and Brass, Harborth and Nienborg [3].

In this paper we focus on the saturation and semi-saturation problems, where the host graph is the hypercube and the forbidden graph is a subcube: *i.e.*, we study $\mathrm{sat}(Q_n,F)$ and $s\text{-sat}(Q_n,F)$. For brevity, we shall often write F-saturated (respectively F-semi-saturated) rather than (Q_n,F)-saturated (respectively (Q_n,F)-semi-saturated) in the remainder of this paper, when the value of n is clear or irrelevant.

The best result along these lines is that of Choi and Guan [4]:

$$\limsup_{n\to\infty} \frac{\mathrm{sat}(Q_n, Q_2)}{e(Q_n)} \leqslant \frac{1}{4}.$$

A conjecture that this is best possible, due to Santolupo, was reported in [7]. The same survey article posed the more general question of determining $\mathrm{sat}(Q_n,Q_m)$.

The main result of this paper, in Section 3, is the construction, for all fixed m, of (Q_n,Q_m)-saturated graphs of arbitrarily low edge density, thus both generalizing and improving the bound of Choi and Guan.

https://doi.org/10.1017/S0963548316000316 Published online by Cambridge University Press

Theorem 1.1. *For fixed m,*

$$\lim_{n\to\infty} \frac{\operatorname{sat}(Q_n, Q_m)}{e(Q_n)} = 0.$$

Slightly more precisely, we show

$$\operatorname{sat}(Q_n, Q_m) \leqslant \frac{c_1}{n^{c_2}} e(Q_n),$$

where c_1 and c_2 are constants depending on m. In the case $m = 2$, $c_2 = 6/7$; it is higher for larger values of m.

In Section 4 we prove a stronger bound for the semi-saturation version of the problem.

Theorem 1.2. *For all n, m,*

$$s\text{-sat}(Q_n, Q_m) < \left(m^2 + \frac{m}{2}\right) 2^n.$$

In the same section, we adapt this proof in the $m = 2$ case to remove all copies of Q_2, and thus prove a bound on $\operatorname{sat}(Q_n, Q_2)$ much stronger than that given by Theorem 1.1.

Theorem 1.3. *For all n,* $\operatorname{sat}(Q_n, Q_2) < 10 \cdot 2^n$.

It is easy to see that both these theorems are best possible up to a constant factor, as all (Q_n, Q_m)-semi-saturated graphs have minimum degree $m - 1$.

In Section 5, we will improve this trivial lower bound, by showing that

$$s\text{-sat}(Q_n, Q_m) \geqslant \frac{m+1}{2} 2^n.$$

In Section 6, we discuss an extension to our zero density upper bound and raise some open questions.

We briefly mention here a somewhat related saturation problem on the cube. Here, Q_n is considered as $\mathcal{P}(X)$, the power set of an n element set, X. Let F be a fixed poset. A family $\mathcal{A} \subseteq \mathcal{P}(X)$ is said to be F-saturated if there is no subfamily of $\mathcal{A}$ with the same poset structure as F, but adding any set to $\mathcal{A}$ destroys this property. Both the maximum and minimum size of such $\mathcal{A}$ have been studied: see for instance Katona and Tarján [8] for the former and Morrison, Noel and Scott [10] for the latter.

2. Preliminaries

In this section we introduce terminology, notation and concepts that will be used frequently in the remainder of the paper.

The hypercube Q_n is the graph with vertex set $\{0,1\}^n$, and with edges between each pair of vertices that differ in exactly one coordinate. Alternatively, the vertex set may be considered as $\mathbb{F}_2^n$, the n-dimensional vector space over the field with two elements. We write $e_1, \ldots, e_n$ for the

https://doi.org/10.1017/S0963548316000316 Published online by Cambridge University Press

canonical basis of $\mathbb{F}_2^n$ (e_i is the vector with a 1 in the ith coordinate, and 0s elsewhere). We can see that x is adjacent to y if and only if $y = x + e_i$, for some $i \in \{1,\ldots,n\}$.

A *subcube* of Q_n is an induced subgraph isomorphic to Q_m, for some $m \leqslant n$. A set S of vertices is the vertex set of a subcube if and only if there is some set of coordinates $J \subseteq [n] = \{1,2,3,\ldots,n\}$, and constants $a_j \in \{0,1\}$ for each $j \in J$, such that $(x_1,\ldots,x_n) \in S$ if and only if, for all $j \in J$, $x_j = a_j$. *Fixed coordinates* are those coordinates in J, whereas *free coordinates* are coordinates that are not fixed. We can thus represent a subcube as an element of $\{0,1,*\}^n$, with stars in the free coordinates and a_j in the fixed coordinates. Since edges can be thought of as Q_1s, we may represent edges as elements of $\{0,1,*\}^n$ in this way. We will say that an edge or subcube lies *along* the directions $i_1,\ldots,i_k$ if these contain all the free coordinates of the edge or subcube. The *weight* of $x \in V(Q_n)$ is the number of coordinates of x that are 1.

We may write $Q_{n_1+n_2}$ as $Q_{n_1}\Box Q_{n_2}$, the graph Cartesian product of Q_{n_1} and Q_{n_2}. In other words, $Q_{n_1+n_2}$ is formed by replacing each vertex of Q_{n_2} with a copy of Q_{n_1}. We call these *principal* Q_{n_1}s. Where there was a Q_{n_2} edge e, we instead put edges between corresponding vertices of the principal Q_{n_1}s placed at the endpoints of e. So we have two types of edges: *internal edges* which have both endpoints in the same principal Q_{n_1} and *external edges* which have endpoints in different principal Q_{n_1}s. Notice that there are n_1 directions along which internal edges lie, and n_2 directions along which external edges lie. This view of $Q_{n_1+n_2}$ is crucial in the proof of Theorem 1.1; we will write $Q_{n_1+n_2}$ as $Q_{n_1}\Box Q_{n_2}$ when we wish to use this viewpoint.

Another way of encapsulating the product nature of Q_n is to write a vertex v as $(v_1|v_2|\cdots|v_t)$, where $v_i \in \{0,1\}^{n_i} = V(Q_{n_i})$ and $n_1+\cdots+n_t = n$. Two vertices $(v_1|v_2|\cdots|v_t)$ and $(u_1|u_2|\cdots|u_t)$ are adjacent if and only if there is a j such that v_j and u_j are adjacent as vertices of Q_{n_j} and for all $i \neq j, v_i = u_i$. We will use this notation heavily in Section 4.

An object we shall use in several of our constructions is the *Hamming code*. The properties of Hamming codes that we require are listed below, but see van Lint [9] for more background. For our purposes, a Hamming code C can be thought of as a subset of $V(Q_n)$, where $n = 2^r - 1$ for some r, with the following properties.

(1) C is a linear subspace of $\mathbb{F}_2^n$. More precisely, C is the kernel of an $r \times n$ matrix H over the field $\mathbb{F}_2$, called a *parity check matrix*. The columns of H are precisely the non-zero vectors in $\mathbb{F}_2^r$.
(2) $|C| = 2^n/(n+1)$.
(3) C has minimum distance 3. In other words, $\min\{d(x,y) : x,y \in C\} = 3$.
(4) C is a dominating set for Q_n. In other words, every vertex of Q_n is either in C or adjacent to a vertex in C.

Property (1) is usually taken as the definition of a Hamming code; the other properties are simple consequences of it.

A subset C with these properties exists only if $n = 2^r - 1$ (and when it exists, it is the largest set with property (3), and the smallest with property (4)). For other values of n, we make do with an *approximate Hamming code*. This is any $C \subset V(Q_n)$ satisfying the following.

(1) C is a linear subspace of $\mathbb{F}_2^n$. More precisely, C is the kernel of an $r = \lceil \log(n+1) \rceil \times n$ matrix H over the field $\mathbb{F}_2$. H has as columns any n distinct binary vectors of length r.
(2) $|C| = 2^n/2^{\lceil \log_2(n+1) \rceil}$.
(3) C has minimum distance 3. In other words, $\min\{d(x,y) : x,y \in C\} = 3$.

https://doi.org/10.1017/S0963548316000316 Published online by Cambridge University Press

3. Zero density bound on $\mathrm{sat}(Q_n, Q_m)$

In this section we shall prove a quantitative version of Theorem 1.1, of which Theorem 1.1 is an immediate consequence.

Theorem 1.1′. *For all $m \geqslant 1$, there exist constants, c_m and a_m, such that*

$$\mathrm{sat}(Q_n, Q_m) \leqslant \frac{c_m}{n^{a_m}} e(Q_n).$$

More precisely, $a_1 = 1$ and $a_m = 1/(7 \cdot 3^{m-2})$, for all $m > 1$.

Before discussing the proof of Theorem 1.1′, we sketch a proof of the $(1/4 + o(1))$ bound of Choi and Guan, as this contains the main ideas of the proof of Theorem 1.1′. This proof is significantly different from that of Choi and Guan, which may be considered more direct. However, our approach, which uses $1/3 + o(1)$ density-saturated graphs to build $1/4 + o(1)$ density-saturated graphs, naturally gives rise to an iterative approach for proving Theorem 1.1′.

We assume that there exist three (Q_n, Q_2)-saturated graphs, A_1, A_2 and A_3 of $1/3 + o(1)$ density, such that every edge of Q_n lies in one of them. We will use these to produce a $1/4 + o(1)$ density (Q_{n+3}, Q_2)-saturated graph B'. These A_i are relatively easy to construct; we will require a generalization of them in our proof of Theorem 1.1′.

We first construct an 'almost' (Q_{n+3}, Q_2)-saturated graph B. We consider Q_{n+3} as $Q_n \square Q_3$. We leave two principal Q_ns corresponding to antipodal vertices of Q_3 empty. Around each of these empty Q_n, we arrange copies of A_1, A_2, A_3, as in Figure 1. We also add all external edges with one endpoint in either of the two empty principal Q_ns (as indicated by the bold edges in the figure).

The graph constructed has the property that for any edge e of an empty Q_n, the corresponding edge e' is present in one of the A_i. So adding e forms a Q_2 comprising e, e' and the two external edges that connect corresponding endpoints of e and e'. Since the A_i are themselves Q_2-saturated graphs, adding any internal edge forms a copy of Q_2.

It is easy to see that B is still Q_2-free, and a quick calculation shows that B has edge density $1/4 + o(1)$. We now prove a simple lemma that allows us to extend B to a Q_2-saturated graph.

Lemma 3.1. *Fix $m \geqslant 2$. Suppose that G is a Q_m-free subgraph of Q_n and $S \subseteq E(Q_n)$. Then we can form a Q_m-free graph G' by adding no more than $|S|$ edges to G with the property that adding any edge in $S \setminus E(G)$ forms a copy of Q_m.*

Proof. We order the edges in S arbitrarily. Consider these edges in this order and add them to G if and only if doing so does not form a copy of Q_m. Since only edges of S are added by the process, we are done. □

We apply this lemma to B, with S being the set of external edges that have not already been added, that is, those represented by the thin edges in Figure 1. This forms a Q_2-saturated graph, B'. Since there are $(3/(n+3))e(Q_{n+3})$ external edges, the asymptotic edge density is still $1/4$.

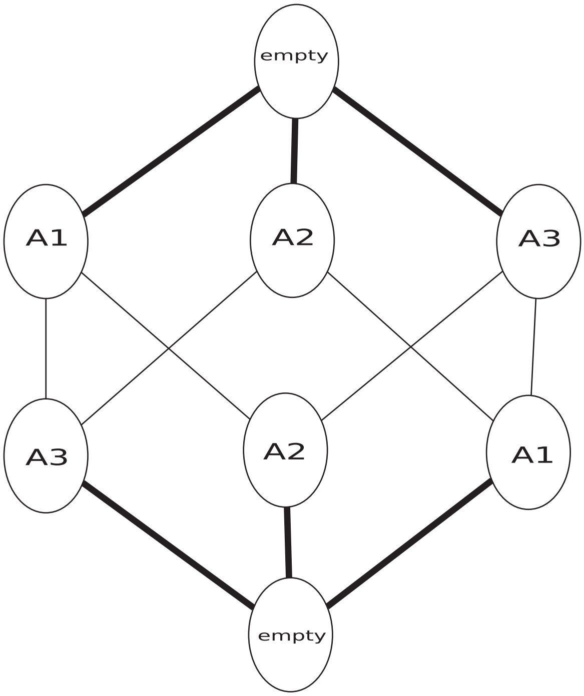

Figure 1. The 'almost' saturated graph, B.

The proof of Theorem 1.1′ uses a similar method multiple times to produce (Q_n, Q_m)-saturated graphs of arbitrarily low density. In the case where $m = 2$, we assume that we have a collection of Q_2-saturated graphs $A_1, \ldots, A_k$ of edge density at most ρ, such that every edge of Q_n is contained in at least one of the A_i. We will view Q_{n+k} as $Q_n \square Q_k$ and leave several principal Q_ns empty. We shall ensure that each empty Q_n is adjacent, for every i, to a principal Q_n filled with A_i, and add every external edge leaving these empty Q_n. This ensures that adding an edge within the empty Q_n forms a copy of Q_2. The constraint on the empty principal Q_n is that the set of vertices that we replace with empty Q_ns must have minimum distance 3, and so we employ a Hamming code, enabling us to produce a graph with a lower density, ρ'. Of course, to apply this method again, we need several (Q_{n+k}, Q_m)-saturated graphs of density ρ', which between them cover the edges of Q_{n+k}. This turns out to be not much harder, using cosets of the Hamming code.

In the general m case we adapt this method. We would like to use a collection of A_is that cover all the copies of Q_{m-1} in Q_n. Such a collection seems hard to construct, but a modification of the argument shows that it suffices to cover almost all copies of Q_{m-1}. The other modification is that instead of using empty principal Q_ns, we fill them with low density Q_{m-1}-saturated graphs, which we may assume exist by induction on m. We will use the following claim as a key part of the inductive step in proving the theorem.

Claim 1. *Suppose we have a collection $A_1, \ldots, A_k$ of (Q_n, Q_m)-saturated graphs, each of density at most ρ, and some n_0 such that every Q_{m-1} that lies along the first n_0 directions is within one*

https://doi.org/10.1017/S0963548316000316 Published online by Cambridge University Press

of these A_i. Suppose also that there is a $(Q_n, Q_m - 1)$-saturated graph G with no more than $(c_{m-1}/(n^{a_{m-1}}))e(Q_n)$ edges. Then there is a collection of $k+1$ (Q_{n+k}, Q_m)-saturated graphs, $B_0, \ldots, B_k$, such that every Q_{m-1} that lies along the first n_0 directions is in one of these B_i. Further, each of the B_i has density at most

$$\left(1 - \frac{1}{2k}\right)\rho + f(n, n_0),$$

where f is a function that tends to zero whenever n, $n_0 \to \infty$ in such a way that $n_0/n \to 1$.

A precise upper bound on the densities of the B_i is required for the quantitative part of the theorem; this will be stated at the end of the proof of this claim.

Proof of Claim 1. We start by constructing a $k+1$ colouring c_0 of Q_k, with the colours $0, 1, \ldots, k$. Fix C_0, an approximate Hamming code in Q_k. We set $c_0(x) = 0$ for all $x \in C_0$, and for all $j \in \{1, \ldots, k\}$ and all $x \in C_0$ we set $c_0(x + e_j) = j$. Note that when $k+1$ is not a power of 2 (*i.e.* when we do not have a genuine Hamming code), this colouring is not fully defined, since C_0 is not dominating. For now we assign arbitrary colours other than 0 to these vertices, but we will later decide on these colours.

We write $Q_{n+k} = Q_n \square Q_k$. We induce from c_0 a colouring on the set of principal Q_ns in the natural way. We start forming the graph B_0 by placing a copy of A_j in each principal Q_n coloured j, for each $j \neq 0$. Also, we add to the graph B_0 every external edge with one endpoint in a principal Q_n coloured 0.

We place a graph isomorphic to G in each Q_n that is coloured 0 (we will choose which isomorphism later).

Notice that, so far, B_0 is Q_m-free. Indeed, suppose that B_0 does contain a Q_m. This Q_m cannot lie entirely within a single principal Q_n, by our assumption that the A_i are saturated. As we have only added external edges that leave Q_n coloured 0, the Q_m may contain an edge between two principal Q_ns only if one of them is coloured 0. Since the Hamming code has minimum distance 3, the Q_m must contain edges in exactly two principal Q_ns, one of which is coloured 0. But such Q_n are Q_{m-1}-saturated and thus contain no Q_{m-1}, yielding a contradiction.

So far, B_0 is not quite Q_m-saturated. For instance, adding an external edge may not create a copy of Q_m. However, we use Lemma 3.1 to remedy this. We add at most $(k/(n+k))e(Q_{n+k})$ edges to B_0 and we now only need to consider adding internal edges.

Adding an edge within a Q_n coloured $j \neq 0$ forms a Q_m, as each A_j is Q_m-saturated. Adding an edge within a principal Q_n coloured 0 will form a Q_{m-1} within that Q_n. If that Q_{m-1} only uses edges in the first n_0 directions, it lies within one of the A_j by the hypothesis of Claim 1. Since every principal Q_n coloured zero is adjacent to a principal Q_n of every non-zero colour, a Q_m will be formed. Therefore, we only need to worry about adding edges to G if the Q_{m-1} formed does not lie exclusively along the first n_0 directions: we call such edges *bad edges*. We will now show that we may assume there are not very many bad edges.

Apply a random automorphism of Q_n to G, our low density Q_{m-1}-saturated graph. We call the graph formed $G' \subseteq Q_n$, which is to be placed within a principal Q_n coloured 0. Let e be a fixed edge of this principal Q_n. Then

$$\begin{aligned}\mathbb{P}(e \text{ is a bad edge}) &\leqslant 1 - \frac{n_0}{n}\cdot\frac{n_0-1}{n-1}\cdot\dots\cdot\frac{n_0-m+2}{n-m+2}\\ &\leqslant 1-\frac{(n_0-m)^{m-1}}{n^{m-1}}\\ &= \frac{n^{m-1}-(n_0-m)^{m-1}}{n^{m-1}}.\end{aligned}$$

This tells us that the expected number of bad edges, in each principal Q_n coloured 0, is no more than

$$\left(\frac{n^{m-1}-(n_0-m)^{m-1}}{n^{m-1}}\right)e(Q_n).$$

We now choose the automorphism of G that we left unspecified earlier; we can do this such that we get no more bad edges than the expected number. We use Lemma 3.1, with S being the set of bad edges, to form a graph that we also call B_0, which is Q_m-saturated.

We now construct the other B_i to cover the required Q_{m-1}s. To construct B_i, we repeat the same method used for constructing B_0, except that we use $C_i := \{c+e_i : c \in C_0\}$ instead of C_0. Note that we can make the arbitrary choices of colours to ensure each principal Q_n is filled with each of the graphs $A_1,\dots,A_k$, in one of the B_i.

It is easy to see that the B_i satisfy the necessary Q_{m-1} condition. Indeed any $Q_m \subseteq Q_{n+k}$ along the first n_0 directions must lie within a principal Q_n. When considered as a subgraph of this Q_n, it must lie in a copy of one of the A_i – say A_j. This principal Q_n is filled with A_j in one of the B_i, so we are done.

It remains only to bound the number of edges in each saturated subgraph, B_i. Let

$$e(A)=\max\{e(A_i)\},\quad e(B)=\max\{e(B_i)\},\quad \rho(A)=\frac{e(A)}{n2^{n-1}},\quad \rho(B)=\frac{e(B)}{(n+k)2^{n+k-1}}.$$

In the calculations that follow, we write $a = a_{m-1}$ and $c = c_{m-1}$ for brevity.

Recall that edges were added to each B_j in four ways: from copies of A_i, from adding external edges, from the Q_{m-1}-saturated graphs, and from adding bad edges.

Thus we have

$$\begin{aligned}e(B) \leqslant{}& 2^k\left(1-\frac{1}{2^{\lceil\log(k+1)\rceil}}\right)e(A)+\frac{k}{n+k}e(Q_{n+k})\\ &+\frac{2^k}{2^{\lceil\log(k+1)\rceil}}e(Q_n)\left(c_{m-1}n^{-a}+\frac{n^{m-1}-(n_0-m)^{m-1}}{n^{m-1}}\right).\end{aligned}$$

Therefore,

$$\begin{aligned}\rho(B) \leqslant{}& \left(1-\frac{1}{2^{\lceil\log(k+1)\rceil}}\right)\rho(A)+\frac{k}{n+k}\\ &+\frac{1}{2^{\lceil\log(k+1)\rceil}}\left(c_{m-1}n^{-a}+\frac{n^{m-1}-(n_0-m)^{m-1}}{n^{m-1}}\right)\\ \leqslant{}& \left(1-\frac{1}{2k}\right)\rho(A)+\frac{k}{n}+\frac{1}{k}\left(c_{m-1}n^{-a}+\frac{n^{m-1}-(n_0-m)^{m-1}}{n^{m-1}}\right).\end{aligned}$$

https://doi.org/10.1017/S0963548316000316 Published online by Cambridge University Press

Clearly, if n_0 is large enough and $n = (1+o(1))n_0$, the last two terms can be arbitrarily small, thus concluding the proof of the claim. □

We now return to prove Theorem 1.1′.

Proof of Theorem 1.1′. We use induction on m.

Base case: $m = 1$. This is trivial – the subgraph of Q_n with no edges is Q_1-saturated.

Inductive step. Take $m > 1$ and assume that the theorem holds for $m-1$. That is, there is a $(Q_n, Q_m - 1)$-saturated graph G with no more than $(c_{m-1}/(n^{a_{m-1}}))e(Q_n)$ edges.

We first find a collection of subgraphs $A_1, \ldots, A_{m+1}$ of Q_{n_0} that satisfy the hypothesis of Claim 1, with $\rho = 1$. To do this, let A_i initially consist of all edges whose lowest weight endpoint has weight in $\{i, \ldots, i+m-2\}$ mod $m+1$, and then extend greedily until A_i is Q_m saturated. Each A_i contains every Q_{m-1} whose lowest weight vertex has weight i mod $m+1$, so every Q_{m-1} is contained in one of these A_i. Trivially, we may bound the density of these A_i above by 1, and it is easy to see this is best possible up to a constant.

We now apply Claim 1 repeatedly, t times. We write k_i and n_i for the value of k and n after the ith iteration. Clearly, $k_{i+1} = k_i + 1, k_0 = m+1, n_{i+1} = n_i + k_i$ and

$$n_t = n_0 + \sum_{i=m}^{m+t} i = n_0 + O(t^2).$$

After t steps, we end with saturated graphs of density ρ:

$$\begin{aligned}
\rho &\leqslant \prod_{i=0}^{t-1}\left(1 - \frac{1}{2k_i}\right) + \sum_{i=0}^{t-1}\left(\frac{k_i}{n_i} + \frac{c_{m-1}}{k_i}\cdot n_i^{-a} + \frac{n_i^{m-1} - (n_0-m)^{m-1}}{k_i n_i^{m-1}}\right)\\
&\leqslant c\prod_{m=1}^{m+t}\left(1 - \frac{1}{2i}\right) + \frac{t(m+t+1)}{n_0} + \frac{tc_{m-1}}{m}\cdot n_0^{-a} + \frac{t}{m}\,\frac{n_t^{m-1} - (n_0-m)^{m-1}}{n_0^{m-1}}\\
&= c'\cdot\exp\left(-\frac{1}{2}\sum_{i=1}^{t+m}\frac{1}{i}\right) + O(t^2 n_0^{-1}) + O(tn_0^{-a}) + O\left(\frac{t^3}{n_0}\right)\\
&= c'' t^{-1/2} + O(tn_0^{-a}) + O(t^3 n_0^{-1}).
\end{aligned}$$

Here, c, c' and c'' are constants dependent on m. If $m = 2$ it is optimal to take $t = n_0^{2/7}$; otherwise, for $a < 3/7$, it is optimal to take $t = n_0^{2a/3}$.

This gives the required bound. □

Note that the better bound for $\mathrm{sat}(Q_n, Q_2)$ in the next section can be fed into the induction in the theorem to produce the slightly better bound of $a_m = 1/(7\cdot 3^{m-3})$.

4. Bounded average degree constructions

4.1. Semi-saturation

In this section we will prove Theorem 1.2, by constructing for each m a family of Q_m-semi-saturated graphs with bounded average degree. Although it seems difficult in general to make these graphs Q_m-free, in the $m = 2$ case we will use similar ideas to prove Theorem 1.3.

https://doi.org/10.1017/S0963548316000316 Published online by Cambridge University Press

In what follows it will be useful to write $n = m(2^t - 1) + r$, where $0 \leqslant r < m2^t$, and to let $n_0 = 2^t - 1$. We write a vertex of Q_n as $(v_1|v_2|\cdots|v_m|v_{m+1})$, where $v_i \in \{0,1\}^{n_0}$ for $i \leqslant m$ and $v_{m+1} \in \{0,1\}^r$. The final section of the vector is only included to make the number of coordinates exactly n but otherwise has no importance in the construction.

Proof of Theorem 1.2. Let $C \subseteq \{0,1\}^{n_0}$ be a Hamming code. We define

$$A = \{(v_1|\cdots|v_m|v_{m+1}) \in V(Q_n) : \exists i \in \{1,m\} \text{ such that } v_i \in C\}.$$

We form $E(G)$ by picking all edges with at least one endpoint in A. Note that vertices in A have degree n in G; all other vertices have degree m. Therefore

$$e(G) = \frac{1}{2}((n-m)|A| + m2^n) \leqslant \frac{m}{2}\left(n\frac{2^n}{(n_0+1)} + 2^n\right).$$

As $n/n_0 < 2m$, $e(G)$ satisfies the bounds of the theorem.

We now show that G is Q_m-semi-saturated. Assume $e \in E(Q_n) \setminus E(G)$ is along a direction i in $\{1, n_0\}$ (all other cases can be dealt with similarly). We will write the endpoints of the edges as $(v_1|v_2|\cdots|v_m|v_{m+1})$ and $(v_1'|v_2|\cdots|v_m|v_{m+1})$, where v_1' and all of the v_i do not lie in C. Thus for $i = 2,3,\ldots,m$ there exists $c_i \in C$ adjacent to v_i. Consider the 2^m points of the form $(x_1|\cdots|x_m|v_{m+1})$, where $x_1 \in \{v_1, v_1'\}$, and for $i = 2,3,\ldots,m$, $x_i \in \{v_i, c_i\}$. These vertices form a subcube of Q_n and all but the endpoints of e are in A. Thus, when the edge e is added, a copy of Q_m is formed, concluding our proof. □

Remark. Clearly, when $n = m(2^t - 1)$ for some t, we get the slightly stronger bound

$$s\text{-sat}(Q_n, Q_m) \leqslant \left(\frac{m^2}{2} + \frac{m}{2}\right)2^n.$$

4.2. Improved bound for $\mathrm{sat}(Q_n, Q_2)$

In the $m = 2$ case, the Q_2-semi-saturated graph constructed above consists of all edges incident with vertices in

$$A = \{(v_1|v_2|v_3) \in V(Q_n) : v_1 \in C \text{ or } v_2 \in C\}.$$

It is easy to see that this contains large subcubes, of the form

$$(c|*,\ldots,*|*,\ldots,*) \quad \text{or} \quad (*,\ldots,*|c|*,\ldots,*), \quad \text{for } c \in C.$$

There are other Q_2s in this graph, but those within these large subcubes are hardest to deal with. We prevent subcubes of the first type by only adding edges of the form $\{(c|v),(c|v')\}$, where $c \in \{0,1\}^{n_0}$ and $v \in \{0,1\}^{n-n_0}$ and v has lower weight than v', if v_1 has even weight. Of course, doing just this alteration means the graph is no longer semi-saturated; we get around this by picking a subset D of $V(Q_{n_0})$ with similar properties to C, and adding edges starting at $(d|v_2|v_3)$ if $(v_2|v_3)$ contains an odd number of 1s and if $d \in D$. We make use of the following claim, which allows us to choose a D with the required properties.

https://doi.org/10.1017/S0963548316000316 Published online by Cambridge University Press

Claim 2. *There exists a Q_2-free spanning subgraph, H, of Q_{n_0}, that has two independent dominating sets, $C,D \subset V(H) = \{0,1\}^{n_0}$, with C disjoint from D, where $|C| = 2^{n_0}/(n_0+1)$ and $|D| = 3 \cdot 2^{n_0}/(n_0+1)$. Further, H only contains edges incident with $C \cup D$ and $e(H) \leqslant 2^{n_0+1}$.*

We shall prove this claim later, but first we show why it implies the theorem.

Proof of Theorem 1.3. Similarly to before, we write $n = 2(2^t-1)+r$, where $0 \leqslant r < 2^{t+1}$, and let $n_0 = 2^t - 1$. We write an element, x, of $\{0,1,*\}^n$ as $(x_1|x_2|x_3)$, where $x_1, x_2 \in \{0,1,*\}^{n_0}$ and $x_3 \in \{0,1,*\}^r$. We refer to x_1 as the first part of x, x_2 as the second part, and so on. We will use this notation particularly when x represents a vertex or an edge of Q_n (it contains no stars or one star).

We start by constructing a graph G that is Q_2-free and will then use Lemma 3.1 add a 'few' edges ($o(2^n)$ edges) to form G', a Q_2-saturated graph. As in the proof of Theorem 1.2, we will define a subset A of the vertices, which will be dominating in G:

$$A = \{(v_1|v_2|v_3) \in \{0,1\}^n : v_1 \in C \cup D \text{ or } v_2 \in C \cup D\}.$$

The definition of G is slightly more complicated. We add edges to $E(G)$ in three stages, and then delete some of these edges to ensure G is Q_2-free.

First, we add all edges e where $e_1 \in C$ and the remainder $(e_2|e_3)$ contains an even number of 1s and a single star, as well as edges where $e_2 \in C$ and the remainder $(e_1|e_3)$ contains an even number of 1s and a single star. We call these type 1 edges. There are

$$2|C|(n-n_0)2^{n-n_0-2} \leqslant \frac{(n-n_0)}{2(n_0+1)}2^n$$

type 1 edges.

Similarly, we add those edges e where $e_1 \in D$ and the remainder $(e_2|e_3)$ contains an odd number of 1s and a single star, as well as edges where $e_2 \in D$ and the remainder contains an odd number of 1s and a single star. We call these type 2 edges. There are

$$2(n-n_0)|D|2^{n-n_0-2} \leqslant \frac{3(n-n_0)}{2(n_0+1)}2^n$$

type 2 edges.

Finally, we add all edges e where e_1 or e_2 is an edge of H. There are

$$2 \cdot 2^{n-n_0} e(H) \leqslant 4 \cdot 2^n$$

type 3 edges.

We now delete all edges e which have an endpoint $(v_1|v_2|v_3)$ such that both v_1 and v_2 lie in $C \cup D$. Thus

$$e(G) \leqslant \left(\frac{2(n-n_0)}{n_0+1} + 4\right)2^n - \frac{n2^n}{(n_0+1)^2}.$$

Suppose, for contradiction, that G contains a Q_2. Note that as all edges of G are incident with a vertex of A, this Q_2 must contain a vertex $(v_1|v_2|v_3) \in A$, where, without loss of generality, $v_1 \in C \cup D$. Note that none of the vertices can have their second part in $C \cup D$, or there is a vertex of the Q_2 with both first and second part in $C \cup D$, impossible by our deletion step.

Let s be the number of stars of the Q_2 that are in the first part of its vector representation. If $s = 2$, all four edges are type 3 edges, impossible as H is Q_2-free.

If instead $s = 1$, suppose the other star is in the second part (the other case is identical). Then we may write the vertices of the Q_2 as $(v_1|v_2|v_3)$, $(v_1'|v_2|v_3)$, $(v_1'|v_2'|v_3)$ and $(v_1|v_2'|v_3)$, where $v_1 \in C \cup D$ and $v_2, v_2' \notin C \cup D$. It is easy to see that $v_1' \in C \cup D$. By a parity argument, v_1 and v_1' are both in C or both in D. But this is impossible as C and D are each H_0-independent sets.

Finally, if $s = 0$, then we can have only type 1 edges or only type 2 edges (depending on whether $v_1 \in C$ or $v_1 \in D$). But this is impossible by a simple parity argument.

We now show that while G is not quite saturated, it is 'almost' saturated. Suppose e is a Q_n-edge not incident with A. Without loss of generality, the endpoints are $(v_1|v_2|v_3)$ and $(v_1'|v_2|v_3)$, where $v_1, v_1' v_2, v_3 \notin C \cup D$. This is an element of $E(Q_n) \setminus E(G)$. Assume that $(v_1|v_3)$ is even (the other case is very similar) and v_1' has higher weight than v_1. Then pick $c \in C$ adjacent to v_2; $\{(v_1'|v_2|v_3), (v_1'|c|v_3)\}$ and $\{(v_1|v_2|v_3), (v_1|c|v_3)\}$ are type 3 edges. Also, $\{(v_1|c|v_3), (v_1'|c|v_3)\}$ is a type 1 edge as $(x|y)$ is even. Thus a Q_2 would be formed by adding the edge.

All Q_n-edges with exactly one endpoint in A are edges of G, so we only need to consider edges where one endpoint $(v_1|v_2|v_3)$ has v_1 and $v_2 \in C \cup D$. There are $2^n/n$ edges of this type, and so we may use Lemma 3.1 and add them greedily to G to form a Q_2-saturated graph G', which has no more edges than the bound in the theorem. □

Remark. Again, we get a stronger bound for some values of n; when $n = 2(2^t - 1)$ for some t, it is easy to see that $\mathrm{sat}(Q_n, Q_2) \leqslant 6 \cdot 2^n$.

We now return to prove the claim.

Proof of Claim 2. Let C be a Hamming code in Q_{n_0}. For $i = 1, \ldots, n_0$, let v_i be the image of the basis vector e_i under the parity check matrix M of the Hamming code. We may assume that $v_1 = (1, 0, \ldots, 0)$, $v_2 = (0, 1, 0, \ldots, 0)$ and $v_3 = (1, 1, 0, \ldots, 0)$, as every vector in $\mathbb{F}_2^t$ occurs as a column of M . We shall construct H in four stages, and then prove that it has the required properties.

(1) Add to $E(H)$ every Q_{n_0}-edge adjacent to an element of C.
(2) Add to $E(H)$ every Q_{n_0}-edge of the form $\{c + e_1 + e_k, c + e_1\}$, where $c \in C$, and where $k \in [4, n_0]$ is such that v_k has a 0 in the first coordinate.
(3) Add to $E(H)$ every Q_{n_0}-edge of the form $\{c + e_1 + e_k, c' + e_2\}$, where $c, c' \in C$, and where $k \in [4, n_0]$ is such that v_k has a 1 in the first coordinate and a 0 in the second coordinate.
(4) Add to $E(H)$ every Q_{n_0}-edge of the form $\{c + e_1 + e_k, c' + e_3\}$, where $c, c' \in C$, and where $k \in [4, n_0]$ is such that v_k has a 1 in the first coordinate and a 1 in the second coordinate.

Since C is a Hamming code, it is an independent, dominating set and $|C| = 2^{n_0}/(n_0 + 1)$. We write $C_i = \{c + e_i : c \in C\}$; in other words, $C_i = M^{-1}(v_i)$. Let $D = C_1 \cup C_2 \cup C_3$. It is easy to see every edge of H is incident with $C \cup D$. Since the C_i are disjoint translates of C, a Hamming code, $|D| = 3 \cdot 2^{n_0}/(n_0 + 1)$.

Again using that C_1 is a translate of a Hamming code, every $x \in V(Q_{n_0}) \setminus C_1$ can be written uniquely in the form $c + e_1 + e_k$ for $c \in C$ and $k \in [1, n_0]$. The restriction $k \neq 1$ is equivalent to

https://doi.org/10.1017/S0963548316000316 Published online by Cambridge University Press

$x \notin C$. The restriction $k \neq 2$ is equivalent to $x \notin C_3$. This is because

$$M(c+e_1+e_2) = M(c)+M(e_1)+M(e_2) = v_1+v_2 = v_3.$$

Similarly, $k=3$ if and only if $x \in C_2$. Thus steps (2), (3) and (4) ensure D is independent and dominating in H.

Notice also that each $x \notin C \cup D$ is H-adjacent to exactly one element in D. Hence $e(H) \leqslant 2|Q_{n_0}|$, as required. It remains only to show that H is Q_2-free. Suppose not. Since we have only added edges with at least one endpoint in $C \cup D$, the Q_2 must contain two opposite vertices in $C \cup D$. Since C has minimum distance 3, and since every $x \notin C \cup D$ is adjacent to only one element in D, one of these vertices is in D, and one is in C. Thus the vertices of the Q_2 may be written in the form $c \in C, c+e_i, c+e_j$ and $c+e_j+e_i \in C_k$, where $i, j \in [4, n_0]$ are such that $v_i + v_j = v_k$, and $k \in \{1,2,3\}$. But it is impossible for all the edges of this Q_2 to lie in $e(H)$. Indeed, suppose for example that $k=3$. Then v_i and v_j must both have 1 in the first coordinate and 1 in the second coordinate, impossible if they sum to v_k. This concludes the proof of the claim. □

5. Lower bounds

All the lower bounds in this section are for s-sat; easily $s\text{-sat}(Q_n, Q_m) \leqslant \text{sat}(Q_n, Q_m)$, so the bounds are also valid for sat.

If a graph is (Q_n, Q_m)-semi-saturated for $m \geqslant 2$, it must be connected. Thus it contains a spanning tree for Q_n and so $s\text{-sat}(Q_n, Q_m) \geqslant 2^n - 1$. This shows that Theorems 1.2 and 1.3 are best possible up to a constant factor.

Another trivial observation improves this for $m \geqslant 3$: if a graph is (Q_n, Q_m)-semi-saturated, it has minimum degree $m-1$. Thus

$$s\text{-sat}(Q_n, Q_m) \geqslant \frac{m-1}{2} 2^n.$$

We do better than both trivial bounds for all m.

Theorem 5.1. *If $m \geqslant 2$,*

$$s\text{-sat}(Q_n, Q_m) \geqslant \left(\frac{m+1}{2} - o(1)\right) 2^n.$$

Proof. Let G be a (Q_n, Q_2)-semi-saturated graph with minimum degree $m-1$; note that this contains all (Q_n, Q_m)-semi-saturated graphs. We call a pair (v, e), where $v \in V(Q_n)$, $e \in E(Q_n) \setminus E(G)$, *good* if there is a path of length 3 in G linking the endpoints of e, that passes through v, meaning v is not a start- or end-vertex of the path.

Note that every non-edge of G is in at least two good pairs, whereas each vertex v is in at most $\binom{d(v)}{2}$ good pairs.

Therefore

$$\sum_{v \in V(Q_n)} \binom{d(v)}{2} \geqslant 2(e(Q_n) - e(G)).$$

https://doi.org/10.1017/S0963548316000316 Published online by Cambridge University Press

Subject to fixed $\sum_v d(v)$, the left-hand side is maximized when the degrees are as different as possible. But no degree can be larger than n or smaller than $m-1$. Note that $2e(G) = \sum_v d(v)$, so we have

$$(2e(G) - 2^n)/(n-1)$$

vertices of degree n in this extreme case.

So certainly

$$\frac{2e(G) - (m-1)2^n}{n-1}\binom{n}{2} \geqslant n2^n - 2e(G),$$
$$(n+2)e(G) - n(m-1)2^{n-1} \geqslant n2^n,$$
$$e(G) \geqslant \left(\frac{m+1}{2} - o(1)\right)2^n. \qquad \square$$

6. Further questions

Having seen that

$$\lim_{n\to\infty} \frac{\mathrm{sat}(Q_n, Q_m)}{n2^{n-1}} = 0,$$

it is natural to ask for a more precise bound. In Section 4 we determined $\mathrm{sat}(Q_n, Q_m)$ up to a constant, for $m = 2$, but there is still a wide gap between the best upper and lower bounds for general m. In particular, we do not know whether families of Q_m-saturated graphs of bounded average degree exist for all m.

Question 1. For which m does there exist a constant c_m such that for all n, $\mathrm{sat}(Q_n, Q_m) \leqslant c_m 2^n$?

In Section 4 we were able to produce better bounds on s-$\mathrm{sat}(Q_n, Q_2)$ than $\mathrm{sat}(Q_n, Q_2)$. Further, the construction we had for s-sat contained many copies of Q_2. This small amount of evidence may suggest that, in general, the two are different, even asymptotically.

Question 2. Is $\mathrm{sat}(Q_n, Q_2) = s$-$\mathrm{sat}(Q_n, Q_2)$ for all n? Does equality hold for all sufficiently large n? If not, is

$$\liminf \frac{\mathrm{sat}(Q_n, Q_2)}{2^n} > \limsup \frac{s\text{-}\mathrm{sat}(Q_n, Q_2)}{2^n}?$$

Recall that all our lower bounds are for s-sat – it seems hard to bound sat more strongly.

Another version of sat that has been studied in the literature (see Section 10 of [7], where the host graph is K_n) could be studied for this problem. We say that a graph $G \subseteq Q_n$ is *weakly* (Q_n, Q_m)*-saturated* if we can add the edges in $E(Q_n) \setminus E(G)$ one at a time (in some order) such that every new edge creates at least one new copy of F. We write w-$\mathrm{sat}(Q_n, Q_m)$ for the minimum number of edges a weakly (Q_n, Q_m)-saturated graph can have. Clearly, w-$\mathrm{sat}(Q_n, Q_m) \leqslant s$-$\mathrm{sat}(Q_n, Q_m) \leqslant \mathrm{sat}(Q_n, Q_m)$. It is not hard to see, by induction on n, that there are many weakly

(Q_n, Q_2)-saturated trees and so w-sat$(Q_n, Q_2) = 2^n - 1$. Indeed, given any G_1, G_2, possibly different weakly (Q_{n-1}, Q_2)-saturated trees, we place them in complementary Q_{n-1}s, and connect any one pair of corresponding vertices. This forms a weakly (Q_n, Q_2)-saturated tree. However, w-sat(Q_n, Q_m) is in general not known.

Question 3. For $m \geqslant 3$, what is w-sat(Q_n, Q_m)?

In [1], Alon, Krech and Szabó discuss an interesting hypergraph-type generalization of the Turán problem on the hypercube. We write Q_n^t for the 2^t-uniform hypergraph with vertex set $\{0,1\}^n$ and edge set consisting of all t-dimensional subcubes of Q_n. We say that a subhypergraph H of Q_n^t is Q_m^t-free if it contains no subhypergraph isomorphic to Q_m^t. As in the usual ($t = 1$) case of this Turán problem, they ask how many edges H can have – in particular asking for the limit

$$\lim_{n\to\infty} \max\left\{ \frac{e(H)}{\binom{n}{t} 2^{n-t}} \right\}.$$

This question is still open, but it is interesting to know that the corresponding saturation problem can be attacked by the same method as the proof of Theorem 1.1′.

Let H be a subhypergraph of Q_n^t. We say that G is (Q_n^t, Q_m^t)-saturated if G is Q_m^t-free but adding another 2^t-edge to G forms a subhypergraph isomorphic to Q_m^t. In other words, G is a maximal Q_m^t-free subgraph of Q_n^t. We write sat(Q_n^t, Q_m^t) for the smallest number of edges a (Q_n^t, Q_m^t)-saturated H can have. We can show by the same method as the proof of Theorem 1.1′ that, for $t \geqslant 1$ and $s \geqslant 0$,

$$\lim_{n\to\infty} \frac{\mathrm{sat}(Q_n^t, Q_{t+s}^t)}{\binom{n}{t} 2^{n-t}} = 0.$$

As in the proof of Theorem 1.1′ we proceed by induction on s with the $s = 0$ case being trivial. The iteration step analogous to Claim 1 is based on the same colouring of principal Q_ns. In each principal Q_n with colour 0 we place a low density Q_{t+s-1}^t-saturated subgraph of Q_n^t. We also add all those 2^t-edges which contain 2^{t-1} points in some principal Q_n with colour 0. The remainder of the proof is a straightforward generalization and the details are left to the reader.

Acknowledgements

The second author was supported by an EPSRC doctoral studentship.

References

[1] Alon, N., Krech, A. and Szabó, T. (2007) Turán's theorem in the hypercube. *SIAM J. Discrete Math.* **21** 66–72.

[2] Balogh, J., Hu, P., Lidický, B. and Liu, H. (2014) Upper bounds on the size of 4- and 6-cycle-free subgraphs of the hypercube. *European J. Combin.* **35** 75–85.

[3] Brass, P., Harborth, H. and Nienborg, H. (1995) On the maximum number of edges in a C_4-free subgraph of Q_n. *J. Graph Theory* **19** 17–23.

[4] Choi, S. and Guan, P. (2008) Minimum critical squarefree subgraph of a hypercube. *Congressus Numerantium* **189** 57–64.

https://doi.org/10.1017/S0963548316000316 Published online by Cambridge University Press

[5] Erdős, P. (1984) Some problems in graph theory, combinatorial analysis and combinatorial number theory. In *Graph Theory and Combinatorics* (B. Bollobás, ed.), Academic Press, pp. 1–17.

[6] Erdős, P., Hajnal, A. and Moon, J. W. (1964) A problem in graph theory. *Amer. Math. Monthly* **71** 1107–1110.

[7] Faudree, J. R., Faudree, R. J. and Schmitt, R. (2011) A survey of minimum saturated graphs. *Electron. J. Combin.* **18**.

[8] Katona, G. O. H. and Tarján, T. G. (1983) Extremal problems with excluded subgraphs in the n-cube. In *Graph Theory: Łagów, Poland*, Vol. 1018 of *Lecture Notes in Mathematics*, Springer, pp. 84–93.

[9] van Lint, J. H. (1999) *Introduction to Coding Theory*, Springer.

[10] Morrison, N., Noel, J. A. and Scott, A. (2014) On saturated k-Sperner systems. *Electron. J. Combin.* **21**, Paper #P3.22.

[11] Pikhurko, O. (2004) Results and open problems on minimum saturated hypergraphs. *Ars Combinatorica* **72** 435–451.

https://doi.org/10.1017/S0963548316000316 Published online by Cambridge University Press