Available online at www.sciencedirect.com

ScienceDirect

Discrete Mathematics 309 (2009) 1730–1735

DISCRETE MATHEMATICS

www.elsevier.com/locate/disc

Note

Bounding the size of square-free subgraphs of the hypercube

Andrew Thomason*, Peter Wagner

DPMMS, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WB, UK

Received 14 May 2007; received in revised form 9 February 2008; accepted 15 February 2008
Available online 24 March 2008

Abstract

We investigate the maximum size of a subset of the edges of the n-cube that does not contain a square, or 4-cycle. The size of such a subset is trivially at most $3/4$ of the total number of edges, but the proportion was conjectured by Erdős to be asymptotically $1/2$. Following a computer investigation of the 4-cube and the 5-cube, we improve the known upper bound from $0.62284\ldots$ to $0.62256\ldots$ in the limit.

Keywords: Hypercube; n-cube; 4-cycle

1. Square-free subgraphs of the hypercube

The n-dimensional hypercube, or n-cube for short, is the graph Q_n whose vertex set is the set $\{0, 1\}^n$ of 0-1 strings of length n, and whose edges join pairs of strings that differ in exactly one place (that is, at Hamming distance one). Thus Q_n has 2^n vertices and $n2^{n-1}$ edges.

In this note we are interested in what is the maximum size of a subset of the edges of the n-cube that does not contain a 4-cycle, or Q_2 — we shall sometimes refer to these 4-cycles as "squares". It is trivial that the size of a square-free subset is at most $3/4$ of the total size, since every edge of the n-cube is in the same number of squares (namely $n - 1$) and at most 3 out of the 4 edges of every square can be retained. However, Erdős [4] long ago conjectured that the proportion drops to $1/2$ as $n \to \infty$. The proportion is certainly never less than $1/2$, since a simple way to select a square-free subset of half the edges is to choose those of the form uv where u has an even number of zeros and v has one less zero than u does.

Writing $e(G)$ for the number of edges of a graph G, we define

$$d_n = \max_{Q_2 \not\subset G \subset Q_n} \frac{e(G)}{e(Q_n)},$$

the maximum being taken over all square-free spanning subgraphs G of the n-cube. Then $d_n \le 3/4$ for all n. Since every edge of the n-cube lies in the same number of $(n-1)$-cubes we have $d_n \le d_{n-1}$ and, in particular, the limit

$$d = \lim_{n\to\infty} d_n = \lim_{n\to\infty} \max_{Q_2 \not\subset G \subset Q_n} \frac{e(G)}{e(Q_n)}$$

exists. Erdős's conjecture then becomes that $d = 1/2$.

* Corresponding author.
E-mail address: a.g.thomason@dpmms.cam.ac.uk (A. Thomason).

0012-365X/$ - see front matter
doi:10.1016/j.disc.2008.02.015

It is not so immediate that $d < 3/4$; indeed, $d_2 = d_3 = d_4 = 3/4$, but $d_5 = 7/10$ (and $d_6 = 11/16$; see Harborth and Nienborg [5]). The best known lower bound for d_n comes from a very elegant construction of Brass, Harborth and Nienborg [2], which has density approximately $(1 + 1/\sqrt{n})/2$. There are reasons for thinking that this construction is best possible. In particular, it is known to have the following Ramsey-theoretical property: if we colour red the edges of Q_n that appear in this construction, and colour the remaining edges blue, then there is neither a red nor a blue square, and the number of red edges is the maximum possible in such a colouring of Q_n. For further discussion of this, and for more information generally about the parameter d_n, the reader is referred to [7]. The corresponding density question for induced subgraphs G is discussed by Johnson and Entringer [6], and more recent Ramsey results about cycles in the cube can be found in Alon, Radoičic, Sudakov and Vondrák [1].

The only published upper bounds on d_n are due to Chung [3], who proved $d \leq 0.62996\ldots$ by means of a simple observation (repeated below in Section 2) about the degrees in square-free subgraphs of the 3-cube, and then improved this to $d \leq 0.62284\ldots$ by a much more elaborate and apparently *ad hoc* argument, but which nevertheless boils down to an assertion about square-free subgraphs of the 3-cube. There are 99 such subgraphs (to within an automorphism of the 3-cube) but it is quite feasible to check the assertion by hand.

Since the 3-cube is a subgraph of the 4-cube, these arguments could be viewed as ones involving square-free subgraphs of the 4-cube, and it ought in principle to be possible to obtain better bounds by using different assertions. Our purpose here is to investigate what bounds can be obtained either by using general information about square-free subgraphs of the 4-cube, or by using information about the degrees in square-free subgraphs of the n-cube for $n \geq 4$. Arguments of varying complexity can be constructed, but the simplest argument we have that improves on the existing bound, yields the following result.

Theorem 1. *The limit d satisfies $d \leq \beta$, where $\beta = 0.62256280\ldots$ is a root of $3 - 6\beta + 4\beta^2 - 4\beta^3 + \beta^4 = 0$.*

Investigations of this kind based on subgraphs of the n-cube for $n \geq 4$ inevitably require computer assistance. For the 4-cube, though, there are only about three million square-free subgraphs, and it is a very straightforward matter to find these and then to check any given assertion about them, as we shall indicate. The best bound obtainable by an argument involving just the degrees of subgraphs of the 4-cube is $d \leq 0.62581\ldots$, better than the bound obtained by considering degrees in the 3-cube but not as good as the more general 3-cube argument. We outline later a more general 4-cube argument that appears to yield $d \leq 0.62083\ldots$; this is better than the bound in Theorem 1 but the argument would be noticeably more complex.

There are far more subgraphs of the 5-cube than could be listed even by computer, so it is not possible to take the general argument further. But it is possible to list the degree sequences of these subgraphs, and thus to give a simple argument based just on the degrees of subgraphs of the 5-cube. It is this argument that lies behind Theorem 1. In particular, the only point where a computer is needed is to verify the very simple Lemma 2 below. However, even just listing the degree sequences of subgraphs of the 5-cube is no longer a straightforward computational task, so we shall be obliged to say a word or two about how it was done (in Section 4).

The improvements in the upper bounds for d obtained by making use of computer information are not dramatic, though we shall suggest reasons (in Section 5) why the kinds of methods used here, even if fully developed theoretically, cannot hope to yield Erdős's conjecture itself.

2. Subgraphs and stars

Let $k \leq n$. The number of copies of the k-cube Q_k within Q_n is readily seen to be $2^{n-k}\binom{n}{k}$. In particular, there are $n2^{n-1}$ edges in Q_n. Given a subgraph $G \subset Q_n$ and a vertex u of Q_n, we define the number δ_u to be such that $\delta_u n$ is the degree of u in G. Let $1 \leq \ell \leq k$. An *ℓ-star* in a subgraph $G \subset Q_n$ is a set of ℓ edges of G with a common end-vertex. The number of ℓ-stars with end-vertex u is therefore $\binom{\delta_u n}{\ell}$. Each k-star lies in precisely one k-cube of Q_n and, more generally, each ℓ-star lies in precisely $\binom{n-\ell}{k-\ell}$ k-cubes.

Given a square-free subgraph G of Q_n and a 3-cube $Q_3 \subset Q_n$, the edges of G induce a square-free subgraph of Q_3. Now it is readily verified that this subgraph can contain at most two 3-stars. So, by counting on the one hand all 3-stars in G, and by counting on the other hand all 3-stars induced in 3-cubes of Q_n, it follows that $\sum_u \binom{\delta_u n}{3} \leq 2 \times 2^{n-3}\binom{n}{3}$. Writing δ for the average value of the δ_u, that is, $\delta = 2^{-n}\sum_u \delta_u$, we then have $\binom{\delta n}{3} \leq 2 \times 2^{-3}\binom{n}{3}$, or

Table 1
The convex hull of the 64 935 5-tuples of stars in subgraphs of Q_5

$\mathbf{s}_0$	0	0	0	0	0	$\mathbf{s}_{17}$	96	116	69	17	0
$\mathbf{s}_1$	32	0	0	0	0	$\mathbf{s}_{18}$	96	116	71	20	1
$\mathbf{s}_2$	36	40	40	20	4	$\mathbf{s}_{19}$	96	116	76	28	4
$\mathbf{s}_3$	48	40	40	20	4	$\mathbf{s}_{20}$	96	118	74	25	4
$\mathbf{s}_4$	48	52	40	20	4	$\mathbf{s}_{21}$	96	96	32	0	0
$\mathbf{s}_5$	58	64	56	24	4	$\mathbf{s}_{22}$	98	118	69	17	0
$\mathbf{s}_6$	60	60	40	20	4	$\mathbf{s}_{23}$	98	120	74	25	4
$\mathbf{s}_7$	60	64	56	24	4	$\mathbf{s}_{24}$	102	124	72	22	4
$\mathbf{s}_8$	64	32	0	0	0	$\mathbf{s}_{25}$	102	126	76	24	4
$\mathbf{s}_9$	64	68	44	20	4	$\mathbf{s}_{26}$	104	124	68	20	4
$\mathbf{s}_{10}$	76	68	40	20	4	$\mathbf{s}_{27}$	104	128	76	24	4
$\mathbf{s}_{11}$	80	96	64	16	0	$\mathbf{s}_{28}$	104	130	76	18	0
$\mathbf{s}_{12}$	84	100	68	24	4	$\mathbf{s}_{29}$	106	133	79	23	3
$\mathbf{s}_{13}$	84	100	72	28	4	$\mathbf{s}_{30}$	108	136	79	22	3
$\mathbf{s}_{14}$	90	110	76	28	4	$\mathbf{s}_{31}$	110	140	80	20	2
$\mathbf{s}_{15}$	92	108	72	28	4	$\mathbf{s}_{32}$	112	144	80	16	0
$\mathbf{s}_{16}$	96	112	64	16	0						

$\delta \leq 2^{-2/3} + O(1/n)$. So, if G is maximal square-free, we have $d \leq d_n = \delta \leq 0.62996\ldots + O(1/n)$. This is the simple argument of Chung [3] referred to in Section 1. The more general argument referred to there comes down to estimating the numbers of some other subgraphs of G, not stars, and counting their possible occurrences in square-free subgraphs of the 3-cube. As already stated, we shall restrict our attention for the time being to just stars, and in order to prove Theorem 1 we count their occurrences in subgraphs of the 5-cube.

Let H be a square-free subgraph of the 5-cube Q_5. Thus H has 32 vertices and somewhere between 0 and $3e(Q_5)/4 = 60$ edges (in fact, it is known that 56 is the maximum number of edges — see [2], though this will follow in any case from the computer results to be described soon). We associate with H a 5-tuple $\mathbf{s}_H = (s_H(1), s_H(2), s_H(3), s_H(4), s_H(5))$ of numbers, where $s_H(\ell)$ is the number of ℓ-stars in H. Thus the empty graph has $\mathbf{s} = (0, 0, 0, 0, 0)$. Note that a 1-star is just an edge with a distinguished end-vertex, so $s_H(1)$ is exactly twice the number of edges. More generally, knowing the numbers in $\mathbf{s}_H$ is equivalent to knowing the multiset of degrees of vertices in H; for example, the unique 56-edge square-free graph has $\mathbf{s} = (112, 144, 80, 16, 0)$, from which it can be inferred, if desired, that this graph has 16 vertices of degree 3 and 16 of degree 4. But all that we need about these 5-tuples is expressed in the following lemma.

Lemma 2. *Let H be a square-free subgraph of the 5-cube and let $s_H(\ell)$ be the number of ℓ-stars in H. Then*

$$3s_H(1) - 3s_H(2) + 2s_H(3) - 4s_H(4) + 5s_H(5) \geq 0.$$

It is necessary to employ a computer to show that this inequality holds for every 5-tuple $\mathbf{s}_H$ that can arise from a square-free graph H. It turns out that there are 64 935 such 5-tuples altogether. In fact, it would be enough to check just 33 of these 5-tuples (listed in Table 1), because the others are convex combinations of these 33. Lemma 2 can readily be verified by hand for these few 5-tuples, which implies its correctness for every H. Thus, if the reader is willing to believe the accuracy of Table 1, no further appeal to a computer is necessary. But our proof does not rely on Table 1.

3. The proof of Theorem 1

After this preamble, it is straightforward to prove the theorem.

Proof of Theorem 1. Let G be a square-free subgraph of Q_n of maximal size. Then G has $N = 2^n$ vertices and $d_n n 2^{n-1} = d_n nN/2$ edges. Let the 64 935 5-tuples of stars discussed in Section 2 be labelled $\mathbf{s}_j = (s_j(1), s_j(2), s_j(3), s_j(4), s_j(5)), 0 \leq j < 64\,935$. There are $q = \frac{1}{32}\binom{n}{5} N$ sub-cubes isomorphic to Q_5 in Q_n, and in each of these G induces a square-free subgraph H with $\mathbf{s}_H = \mathbf{s}_j$ for some $j, 0 \leq j < 64935$. Let the number of H with $\mathbf{s}_H = \mathbf{s}_j$ be $x_j q, 0 \leq j < 64935$; so $0 \leq x_j \leq 1$ and $\sum_{0 \leq j < 64935} x_j = 1$.

Recall from Section 2 that the degree of the vertex u is $\delta_u n$, and that each ℓ-star appears in $\binom{n-\ell}{5-\ell}$ 5-cubes. Counting appearances of ℓ-stars in 5-cubes we therefore have

$$\sum_{0\le j<64935} x_j q s_j(\ell) = \sum_u \binom{\delta_u n}{\ell}\binom{n-\ell}{5-\ell}, \quad \text{for } 1 \le \ell \le 5.$$

Thus, on dividing through by q, we obtain the equations

$$\sum_{0\le j<64935} x_j s_j(\ell) = 32\binom{5}{\ell}\frac{1}{N}\sum_u \delta_u^\ell + o(1), \quad \text{for } 1 \le \ell \le 5,$$

where $o(1)$ signifies a term tending to zero as $n \to \infty$.

We now define the 5-tuple $\mathbf{t} = (3, -3, 2, -4, 5) = (t(1), t(2), t(3), t(4), t(5))$. If we write

$$L = \frac{160\alpha}{N}\sum_u \delta_u + \sum_{\ell=1}^{5} t(\ell)\left[32\binom{5}{\ell}\frac{1}{N}\sum_u \delta_u^\ell - \sum_{0\le j<64935} x_j s_j(\ell)\right],$$

where α is some positive number (a function of β, to be specified later), then we see that $L = 160\alpha d_n + o(1)$. We shall now show that $L \le 160\alpha\beta$. This will imply that $d_n \le \beta + o(1)$, which in turn implies $d \le \beta$, and so the proof of Theorem 1 will be complete.

Observe that

$$L = \frac{160}{N}\sum_u \left[(\alpha+3)\delta_u - 6\delta_u^2 + 4\delta_u^3 - 4\delta_u^4 + \delta_u^5\right] - \sum_{0\le j<64935} x_j\, \mathbf{t}\cdot\mathbf{s}_j,$$

where $\mathbf{t}\cdot\mathbf{s}_j = \sum_{\ell=1}^5 t(\ell)s_j(\ell)$. But Lemma 2 shows that $\mathbf{t}\cdot\mathbf{s}_j \ge 0$ for every j. So, if we write

$$f(z) = (\alpha+3)z - 6z^2 + 4z^3 - 4z^4 + z^5,$$

then to finish the proof it will suffice to verify that $f(z) \le \alpha\beta$ for every real z with $0 \le z \le 1$. Certainly $f(\beta) = \alpha\beta$, by the definition of β. Moreover, $f''(z) = -4(3(1-z)^2 + 5z^2(1-z) + 4z^2) < 0$ for $0 \le z \le 1$. We have not yet defined α; we choose $\alpha = -3 + 12\beta - 12\beta^2 + 16\beta^3 - 5\beta^4 = 2.92\ldots$, so that $f'(\beta) = 0$. Then $f(z) \le f(\beta)$ for $0 \le z \le 1$, and the theorem is proved. $\square$

The proof of the theorem as written looks a bit like a rabbit pulled out of a hat, but the reader will readily recognize that the method used is the natural one of maximizing $\frac{1}{N}\sum_u \delta_u$ subject to the constraints imposed by the equations for $\sum_j x_j s_j(\ell)$, via the technique of Lagrange multipliers. The solution was discovered by first replacing the $\mathbf{s}_j$ by the extreme points of their convex hull, thus dramatically reducing the number of variables, and then experimenting by computer to see which $\mathbf{s}_j$ were important (that is, had $x_j > 0$) in the maximizing solution. There were only six such $\mathbf{s}_j$ and, given the clue as to which they were, it was then straightforward to find the appropriate multipliers (that is, to find the vector $\mathbf{t}$). This in turn showed that the appropriate inequality to use in the proof of Theorem 1 is the one expressed in Lemma 2.

4. Computing subgraphs of the 5-cube

This section provides a brief description of the manner in which the necessary computations were performed.

There are 32 edges in Q_4 and so 2^{32} spanning subgraphs, which can be labelled in a natural way by the integers $0, 1, \ldots, 2^{32} - 1$. The automorphism group of Q_4 has order $16 \times 4! = 384$. This group acts on the set of labelled subgraphs, and we seek a representative from each orbit that contains square-free subgraphs. This is a relatively trivial exercise that can be completed in a way similar to the sieve of Eratosthenes, by writing down a list of 2^{32} ones and setting to zero those corresponding to subgraphs isomorphic to subgraphs earlier in the list. The C-program we wrote to do this job took about ten minutes on a standard computer to produce the 3212 821 square-free spanning subgraphs of Q_4, to within an automorphism of Q_4.

It is entirely impractical, however, to perform the same computation to discover the square-free subgraphs of Q_5, which has 2^{80} labelled subgraphs. But there are not so many possible 5-tuples of stars. Indeed, $\binom{37}{5} = 435897$ is an obvious upper bound, insofar as this is the number of multisets of 32 numbers chosen from the set $\{0, 1, 2, 3, 4, 5\}$, and so is a bound on the number of degree sequences (which determine the 5-tuples).

The procedure used to actually compute the possible 5-tuples was as follows. The 5-cube Q_5 can be viewed as two 4-cubes joined by a set of 16 edges; we think of these 4-cubes as the "left and right halves" Q_L and Q_R of Q_5. Every spanning subgraph H of Q_5 can therefore be formed in the following way: take a subset S of the vertices of the 4-cube, take subgraphs H_L of Q_L and H_R of Q_R, and join the sets S_L and S_R, which correspond in Q_L and Q_R to S, by a set of $|S|$ edges. It is evident that H is square-free if and only if the following condition holds: both H_L and H_R are square-free, and no edge of the 4-cube whose ends are both in S appears in both H_L and H_R. All square-free subgraphs of the 5-cube can be constructed, in principle, by finding all triples (S, H_L, H_R) that satisfy the described condition. Remember, though, that it is only the 5-tuples of stars that we want, not all the subgraphs.

The reason for constructing the 5-tuples in this way is that the calculations for the different sets S are independent, and the calculation for any specific S is, if done carefully, feasible. There are 2^{16} subsets of the vertices of Q_4, but since H can have at most 60 edges and since, by embedding H suitably in Q_5, we can assume $|S| \le e(H)/5$, we need only consider $|S| \le 12$. Under the 384 isomorphisms of Q_4 there are then only 402 choices for S. Let such a choice now be fixed.

Given S, there are $3212\,821 \times 384$ choices for H_L. For each of these we compute both the set D_L of degrees contributed to H by H_L (after the edges in S have been taken into account), and also the set F_L of edges of H_L whose ends both lie in S. It turns out that, for a fixed S, there are never more than 4000 different sets D_L, and for each set D_L we store all possibilities for F_L allowed by D_L — that is, for which there is an H_L realizing both D_L and F_L. This turns out to be quite practicable. The possibilities for D_R and F_R are, of course identical.

To find all the possibilities for the 5-tuples, all that need be done now is to examine all pairs (D_L, D_R) to see whether they are realizable by some square-free H, and this means precisely that there is a set F_L allowed by D_L and a set F_R allowed by D_R for which $F_L \cap F_R = \emptyset$. Given the information that has been saved, this too is readily checked. In this way it is feasible to determine the list of all possible 5-tuples, and the final calculation took around 17 hours. The algorithm as laid out above proceeds in several distinct and relatively simple stages, and the program code can be debugged in similar stages. The final program was further checked by using it to compute the corresponding star-sequences for subgraphs of Q_3 and Q_4, since these numbers were already known by other means.

5. Final remarks

The argument of Chung that gave the bound $0.62284\ldots$ can readily be fitted into the framework of Section 3. It involves looking at some subgraphs of Q_3 other than stars, listing their occurrences in each of the 99 square-free subgraphs of Q_3, estimating the number of their occurrences in a square-free subgraph G of Q_n, and then maximizing $\frac{1}{N}\sum_u \delta_u$ subject to the implied constraints. This more general approach can be extended to subgraphs of Q_4, since the list of all $3212\,821$ square-free subgraphs of Q_4 is available. Computer experiments suggest the bound $0.62083\ldots$ is achievable in this way but we have not pursued the details. The nonstar subgraphs involved in the argument consist of, for example, two or three squares with a common edge in Q_n but having no edge in G. Variables would be introduced such as, say, ζ_e for each edge e of Q_n, counting the number of copies of Q_2 containing e and having no edge in G. The number of subgraphs of G comprising two or three such Q_2's with a common edge would then correspond to ζ_e^2 or ζ_e^3 and so on. Appropriate Lagrange multipliers could be determined experimentally and then an argument along the lines of the proof above could be constructed. The computational effort would be less but the theoretical argument would be considerably more complicated, and we felt the simple proof given was the best compromise between the difficulty and the strength of the result obtained.

It would seem that methods of this kind, using computer generated data, are unlikely to lead to dramatic improvements in the bound on d. It could be wondered, though, whether it might not be possible to extend theoretically the present argument, of using a bound on the number of stars in relatively small cubes to lever a stronger bound in large cubes. More precisely, the principle behind the arguments in the main part of the paper, based on k-cubes where $k \le 5$, is that it is necessary on the one hand for many of the Q_k inside Q_n to contain lots of vertices of large degree (large stars), but that, on the other hand, a square-free subgraph of Q_k of around the maximum possible density d_k is constrained in the number of large stars it can have (when $k \le 5$): thus d_n, which is the average density of the Q_ks, is significantly less than d_k. Such an argument could be made to work more generally (with k growing large but remaining small relative to n), if it could be shown that large degrees in a square-free subgraph of Q_k forced the density down.

The following observation suggests that this approach too is unlikely to work. Suppose we have an almost $d_n n$-regular square-free subgraph G of Q_n. Consider a randomly chosen sub-k-cube Q_k of G and consider the degree of a randomly chosen vertex of it; equivalently, choose randomly a vertex of Q_n and choose k incident edges, and consider the number of these edges that lie in G. This random number follows (approximately) a binomial distribution $\mathrm{Bin}(k, d_n)$, so this distribution is the distribution of degrees in sub-k-cubes of G. We will make progress only if it is impossible for a square-free subgraph of Q_k to have large degrees following this distribution whilst still having density close to d_k. But such subgraphs can exist as k grows large. To see this, take an almost $d_k k$-regular square-free subgraph H. Each vertex has at most k^j other vertices within distance j in Q_k. Pick a subset V of $2^k/k^2$ vertices so that vertices within V are pairwise distance at least three apart, and let W be the neighbours of V in Q_k. Remove all edges in H that meet $V \cup W$, and then add k-stars at the vertices in V, to obtain a subgraph H'. Then the density of H' is still nearly d_k, it is square-free, and it contains many stars, certainly more than required by the distribution $\mathrm{Bin}(k, d_n)$.

References

[1] N. Alon, R. Radoičic, B. Sudakov, J. Vondrák, A Ramsey-type result for the hypercube, J. Graph Theory 53 (2006) 196–208.

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

[3] F.R.K. Chung, Subgraphs of a hypercube containing no small even cycles, J. Graph Theory 16 (1992) 273–286.

[4] P. Erdős, On some problems in graph theory, combinatorial analysis and combinatorial number theory, in: Graph Theory and Combinatorics, Academic Press, London, 1984, pp. 1–17.

[5] H. Harborth, H. Nienborg, Maximum number of edges in a six-cube without four-cycles, Bull. Inst. Combin. Appl. 12 (1994) 55–60.

[6] K.E. Johnson, R. Entringer, Largest induced subgraphs of the n-cube that contain no 4-cycles, J. Combin. Theory (Ser. B) 46 (1989) 346–355.

[7] R.H. Schelp, A. Thomason, On quadrilaterals in layers of the cube and extremal problems for directed and oriented graphs, J. Graph Theory 33 (2000) 66–82.