Journal of Combinatorial Theory, Series A **95**, 349–359 (2001)
doi:10.1006/jcta.2001.3177, available online at http://www.idealibrary.com on IDEAL®

Parent-Identifying Codes

Noga Alon[1]

Department of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel
E-mail: noga@math.tau.ac.il

Eldar Fischer

NEC Research Institute, 4 Independence Way, Princeton, New Jersey 08540; and DIMACS
E-mail: fischer@research.nj.nec.com

and

Mario Szegedy

School of Mathematics, Institute for Advanced Study, Princeton, New Jersey 08540
E-mail: szegedy@ias.edu

Communicated by the Managing Editors

Received May 12, 2000; published online June 18, 2001

For a set C of words of length 4 over an alphabet of size n, and for $a, b \in C$, let $D(a, b)$ be the set of all descendants of a and b, that is, all words x of length 4 where $x_i \in \{a_i, b_i\}$ for all $1 \leqslant i \leqslant 4$. The code C satisfies the Identifiable Parent Property if for any descendant of two code-words one can identify at least one parent. The study of such codes is motivated by questions about schemes that protect against piracy of software. Here we show that for any $\varepsilon > 0$, if the alphabet size is $n > n_0(\varepsilon)$ then the maximum possible cardinality of such a code is less than εn^2 and yet it is bigger than $n^{2-\varepsilon}$. This answers a question of Hollmann, van Lint, Linnartz, and Tolhuizen. The proofs combine graph theoretic tools with techniques in additive number theory.

[1] Research supported in part by a USA–Israeli BSF grant, by the Israel Science Foundation, and by the Hermann Minkowski Minerva Center for Geometry at Tel Aviv University. Part of this work was done while visiting the School of Mathematics, Institute for Advanced study, Princeton, NJ 08540.

0097-3165/01 $35.00

1. INTRODUCTION

Let $|N| = n$ and $C \subseteq N^4$. For $a, b \in C$ define the set $D(a, b)$ of all descendants of a, b as

$$D(a, b) = \{x \in N^4 \mid x_i \in \{a_i, b_i\} \text{ for } 1 \leqslant i \leqslant 4\}.$$

We say that the code C has the Identifiable Parent Property (IPP) if for every descendant one can always identify at least one of the parents, that is, for every $x \in \bigcup_{a, b \in C} D(a, b)$ there is a $p \in C$ such that if $a, b \in C$ and $x \in D(a, b)$ then $p \in \{a, b\}$. Equivalently, as mentioned in [3], C has the IPP if and only if:

(**IPP1**) For every distinct $a, b, c \in C$ there is an $1 \leqslant i \leqslant 4$ such that a_i, b_i, c_i are all distinct, and

(**IPP2**) for every $a, b, c, d \in C$ with $\{a, b\} \cap \{c, d\} = \emptyset$ there is an $1 \leqslant i \leqslant 4$ such that $\{a_i, b_i\} \cap \{c_i, d_i\} = \emptyset$.

Define

$$f(n) = \max\{|C|: C \subseteq N^4 \text{ has IPP}\}.$$

The study of $f(n)$ is motivated by questions about schemes that protect against piracy of software. The authors of [3] proved that

$$(1 + o(1))\, n^{3/2} \leqslant f(n) \leqslant n^2 \tag{1}$$

and raised the problem of closing the gap between the upper and lower bounds. Here we show that for every $\varepsilon > 0$ there is an $n_0 = n_0(\varepsilon)$ such that for every $n > n_0$,

$$f(n) \leqslant \varepsilon n^2 \tag{2}$$

and yet

$$f(n) \geqslant n^{2-\varepsilon}. \tag{3}$$

2. THE UPPER BOUND

It is convenient to distinguish the alphabets that are used in each coordinate. Let N_i be the alphabet used in coordinate i ($1 \leqslant i \leqslant 4$). $|N_i| = n$, and N_i are pairwise disjoint. Thus $C \subseteq N_1 \times N_2 \times N_3 \times N_4$. By omitting all members of C that have a coordinate that does not belong to any other code word we omit at most $4n$ words, and may assume now that:

($*$) Each letter $l \in N_1 \cup N_2 \cup N_3 \cup N_4$ appears in at least two members of C (or does not appear at all).

FACT 2.1. *No two members of C have three common coordinates.*

Proof. If $a, b \in C$ with $a_1 = b_1$, $a_2 = b_2$, $a_3 = b_3$, then by assumption ($*$) there is a $c \in C$, $c \neq a$ such that $c_4 = a_4$. But then $\{a, b, c\}$ violate (IPP1). ∎

FACT 2.2. *If there are distinct $i_1, i_2 \in \{1, 2, 3, 4\}$ and two distinct words $a, c \in C$ with $a_{i_1} = c_{i_1}$, $a_{i_2} = c_{i_2}$, then there are no distinct words $b, d \in C$ such that $b_{j_1} = d_{j_1}$, $b_{j_2} = d_{j_2}$, where $\{j_1, j_2\} = \{1, 2, 3, 4\} \setminus \{i_1, i_2\}$.*

Proof. Assume the opposite. Then a, b, c, d violate (IPP2) if all words are distinct. If, say, $a = b$, then $\{a, c, d\}$ violate (IPP1). ∎

FACT 2.3. *For every distinct $i_1, i_2 \in \{1, 2, 3, 4\}$,*

$$|\{x \in C \mid (\exists y \in C)((y \neq x) \wedge (y_{i_1} = x_{i_1}) \wedge (y_{i_2} = x_{i_2}))\}| \leqslant 2n - 1.$$

Proof. Assume the fact does not hold for say, $i_1 = 1$, $i_2 = 2$. Construct a bipartite graph G with color classes N_3 and N_4 as follows: for each $x \in \{x \in C \mid (\exists y \in C)((y \neq x) \wedge (x_1 = y_1) \wedge (x_2 = y_2))\}$ the pair $x_3 x_4$ is an edge of G. By assumption and Fact 2.2, G has more than $2n - 1$ edges, hence it has a cycle. Therefore, since it is bipartite, it contains a path of length 3. Let $x_4 (x_3 = y_3)(y_4 = z_4)\, z_3$ be that path, where these coordinates arise from appropriate $x, y, z \in C$. Let $x' \in C$ be such that $x'_1 = x_1$, $x'_2 = x_2$, $x' \neq x$. If $x' = z$ then $\{x, y, z\}$ violate (IPP1), otherwise $\{x', y, x, z\}$ violate (IPP2) with the grouping $\{x, z\}$, $\{x', y\}$. ∎

To prove the upper bound, we also need the following result, proved in Alon *et al.* [1] by applying the regularity lemma of Szemerédi [5].

LEMMA 2.4 [1, Proposition 4.4]. *For every $\gamma > 0$ and every integer k there exists a $\delta = \delta(k, \gamma) > 0$ such that every graph G on n vertices containing less than δn^k copies of the complete graph K_k on k vertices, contains a set of less than γn^2 edges whose deletion destroys all copies of K_k in G.*

We can now prove the required upper bound for $f(n)$.

THEOREM 2.5. *For every $\varepsilon > 0$ there exists $n_0 = n_0(\varepsilon)$ such that $f(n) < \varepsilon n^2$ for every $n > n_0$.*

Proof. Let $C \subseteq N_1 \times N_2 \times N_3 \times N_4$ have the IPP, $|C| = f(n)$, with N_i being pairwise disjoint and satisfying $|N_1| = |N_2| = |N_3| = |N_4| = n$. By Facts 2.1 and 2.3 we can omit from C at most $6 \cdot 2n + 4n = 16n$ members to get a code C', $|C'| \geqslant f(n) - 16n$ that has IPP in which no two code words share more than one coordinate. Let H be the 4-partite graph on the classes of vertices N_1, N_2, N_3, N_4 obtained by taking the edge-disjoint union of all K_4 copies $\{x_1, x_2, x_3, x_4\}$ for every $x \in C'$.

This graph has at least $(f(n) - 16n)\, 6$ edges, and as it is the edge-disjoint union of $f(n) - 16n$ copies of K_4, one has to delete at least $f(n) - 16n$ of its edges to destroy all copies of K_4 contained in the graph. If we assume that $f(n) > \varepsilon n^2$, this implies, for sufficiently large n, that we have to delete at least $\frac{\varepsilon}{2} n^2$ edges of H to destroy all copies of K_4.

By Lemma 2.4 (with $k = 4$, $\gamma = \varepsilon/2$), this implies that H contains at least δn^4 distinct copies of K_4 for a constant $\delta = \delta(\varepsilon) > 0$. Among these K_4 copies, only $f(n) \leqslant n^2$ correspond each to one $x \in C$. Similarly, the number of K_4 copies that contain at least two edges arising from the same $x \in C$ is at most $O(n^3)$, since there are at most n^2 ways to choose x, at most 15 ways to choose two of its edges, and this determines already at least three vertices of the K_4. It follows that H contains a copy of K_4 in which every edge comes from a different $x \in C$. In particular, if a_1, a_2, a_3, a_4 are the vertices of this K_4, then there exist distinct $x, y, z, w \in C$ such that

$$\begin{array}{llll} x_1 = a_1, & y_3 = a_3, & z_2 = a_2, & w_1 = a_1, \\ x_2 = a_2, & y_4 = a_4, & z_3 = a_3, & w_4 = a_4. \end{array}$$

But then x, y, z, w violate (IPP2), contradicting the fact that C has IPP. Thus $f(n) \leqslant \varepsilon n^2$ for $n > n_0(\varepsilon)$, completing the proof. ∎

Remark 2.6. The proof and the known bounds in the proof of the regularity lemma actually show that

$$f(n) = O\left(\frac{n^2}{(\log^* n)^{1/5}}\right),$$

where

$$\log^* n = \min\{k \mid \underbrace{\log_2 \log_2 \cdots \log_2}_{k \text{ times}} n \leqslant 1\}.$$

3. THE LOWER BOUND

Our main tool here is an arithmetic lemma proven using the method of Behrend [2], and its extension by Ruzsa [4], with some modifications.

A linear equation with integer coefficients

$$\sum a_i x_i = 0 \tag{4}$$

in the unknowns x_i is homogeneous if $\sum a_i = 0$. If $X \subseteq N = \{1, 2, ..., n\}$, we say that X has no non-trivial solution to (4), if whenever $x_i \in X$ and $\sum a_i x_i = 0$, it follows that all x_i are equal.

Note that if X has no non-trivial solution to (4), then the same holds for any shift $(X + u) \cap N$ (where u is positive, negative or zero).

We need the following simple fact, which follows from the convexity of the function $g(t) = t^2$.

FACT 3.1. *Let $p_1, p_2, ..., p_k$ be k strictly positive reals whose sum is* 1, *and suppose $\sum_{i=1}^{k} p_i r_i = r$, where $r_1, r_2, ..., r_k$ are reals. Then*

$$\sum_{i=1}^{k} p_i r_i^2 \geqslant r^2,$$

and the inequality is strict unless $r_1 = r_2 = \cdots = r_k = r$.

Proof. Put $r_i = r + \varepsilon_i$, then

$$\sum_{i=1}^{k} p_i (r + \varepsilon_i) = r + \sum_{i=1}^{k} p_i \varepsilon_i = r$$

and hence $\sum_{i=1}^{k} p_i \varepsilon_i = 0$. It thus follows that

$$\sum_{i=1}^{k} p_i r_i^2 = \sum_{i=1}^{k} p_i (r + \varepsilon_i)^2 = \sum_{i=1}^{k} p_i r^2 + 2r \sum_{i=1}^{k} p_i \varepsilon_i + \sum_{i=1}^{k} p_i \varepsilon_i^2 = r^2 + \sum_{i=1}^{k} p_i \varepsilon_i^2 \geqslant r^2,$$

and the last inequality is strict unless all numbers ε_i are 0. ∎

LEMMA 3.2 (Main Lemma). *For $q = \lceil 2^{\sqrt{\log n}} \rceil$ there exist:*

(1) *a set $X_1 \subseteq N$, $|X_1| \geqslant n / 2^{O(\log^{3/4} n)}$ with no non-trivial solution to*

$$2x + 3y + qz - (q + 5)\, w = 0; \tag{5}$$

(2) *a set $X_2 \subseteq N$, $|X_2| \geqslant n / 2^{O(\log^{3/4} n)}$ with no non-trivial solution to*

$$5x + (q + 3)\, y - 3z - (q + 5)\, w = 0; \tag{6}$$

(3) *a set* $X_3 \subseteq N$, $|X_3| \geqslant n/2^{O(\log^{3/4} n)}$ *with no non-trivial solution to*

$$5x+qy-2z-(q+3)\,w=0. \tag{7}$$

Proof. To prove part (1) we apply the method of Behrend [2]. Let d be an integer (to be chosen later) and define

$$X_1=\left\{\sum_{i=0}^{k} x_i d^i \mid x_i<\frac{d}{q+5}\ (0\leqslant i\leqslant k) \wedge \sum_{i=0}^{k} x_i^2=B\right\},$$

where $k=\lfloor \log n/\log d\rfloor-1$ and B is chosen to maximize the cardinality of X_1. If $x, y, z, w \in X_1$ satisfy (5) and

$$x=\sum_{i=0}^{k} x_i\, d^i, \qquad y=\sum_{i=0}^{k} y_i\, d^i, \qquad z=\sum_{i=0}^{k} z_i\, d^i, \qquad w=\sum_{i=0}^{k} w_i\, d^i,$$

then

$$2x_i+3y_i+qz_i=(q+5)\,w_i$$

for every $0\leqslant i\leqslant k$. But then, by Fact 3.1 (with $k=3$, $p_1=\frac{2}{q+5}$, $p_2=\frac{3}{q+5}$, and $p_3=\frac{q}{q+5}$),

$$2x_i^2+3y_i^2+qz_i^2\geqslant(q+5)\,w_i^2$$

for every $0\leqslant i\leqslant k$, and each such inequality is strict unless $x_i=y_i=z_i=w_i$. As $\sum x_i^2=\sum y_i^2=\sum z_i^2=\sum w_i^2$, this implies that $x_i=y_i=z_i=w_i$ for $0\leqslant i\leqslant k$, showing that X_1 has no non-trivial solution to (5). The size of X_1 satisfies

$$|X_1|\geqslant\frac{n}{d^2(q+5)^{k+1}\,(k+1)(d^2/(q+5)^2)}\geqslant\frac{n}{(q+5)^{\log n/\log d}\,d^4\log n}.$$

Take $d=\lfloor 2^{\sqrt{\log n\log q}}\rfloor$ ($\gg q$) to conclude that

$$|X_1|\geqslant\frac{n}{2^{O(\sqrt{\log n\log q})}}. \tag{8}$$

In order to prove Part (2) we apply the method of Ruzsa [4]. By Behrend's method (that is, by an obvious modification of the constants in the argument given in the proof of Part (1) above) there exists $Q \subseteq \{1, 2, ..., q/5\}$ satisfying $|Q| \geqslant q/2^{O(\sqrt{\log q})}$ with no non-trivial solution to $5x = y + 3z + w$. Define

$$X_2 = \left\{ \sum_{i=0}^{k} x_i (q+4)^i \mid x_i \in Q \right\},$$

where $k = \lfloor \log n / \log(q+4) \rfloor - 1$. Note that

$$|X_2| = |Q|^{k+1} \geqslant \frac{n}{2^{O(\log n / \sqrt{\log q})}}. \tag{9}$$

Suppose now that there is a non-trivial solution $x, y, z, w \in X_2$ of (6), where

$$x = \sum_{i=0}^{k} x_i (q+4)^i, \qquad y = \sum_{i=0}^{k} y_i (q+4)^i,$$
$$z = \sum_{i=0}^{k} z_i (q+4)^i, \qquad w = \sum_{i=0}^{k} w_i (q+4)^i.$$

Then

$$\sum_{i=0}^{k} 5x_i (q+4)^i + (q+3) \sum_{i=0}^{k} y_i (q+4)^i = \sum_{i=0}^{k} 3z_i (q+4)^i + (q+5) \sum_{i=0}^{k} w_i (q+4)^i.$$

Let j be the minimum index such that not all $\{x_i, y_i, z_i, w_i\}$ are equal. Then

$$\sum_{i=j}^{k} 5x_i (q+4)^i + (q+3) \sum_{i=j}^{k} y_i (q+4)^i = \sum_{i=j}^{k} 3z_i (q+4)^i + (q+5) \sum_{i=j}^{k} w_i (q+4)^i.$$

Reducing modulo $(q+4)^{j+1}$ we conclude that

$$5x_j (q+4)^j \equiv y_j (q+4)^j + 3z_j (q+4)^j + w_j (q+4)^j \pmod{(q+4)^{j+1}}.$$

But both sides are less than $(q+4)^{j+1}$, as $x_j, y_j, z_j, w_j \leqslant \frac{1}{5} q$, hence this is an equality (and not only a modular equality):

$$5x_j (q+4)^j = y_j (q+4)^j + 3z_j (q+4)^j + w_j (q+4)^j.$$

Dividing by $(q+4)^j$ we get $5x_j = y_j + 3z_j + w_j$, contradicting the assumption that Q has no non-trivial solution to this equation. Thus X_2 has no non-trivial solution to (6), as needed.

The proof of Part (3) is analogous to that of Part (2). Here we start with $Q \subset \{1, 2, ..., \frac{1}{5}q\}$ having no non-trivial solution to $5x = y + 2z + 2w$ and satisfying $|Q| \geqslant q/2^{O(\sqrt{\log q})}$. Then we take $X_3 = \{\sum_{i=0}^{k} x_i (q+1)^i \mid x_i \in Q\}$ where $k = \lfloor \log n / \log(q+1) \rfloor - 1$.

As before,

$$|X_3| \geqslant \frac{n}{2^{O(\log n/\sqrt{\log q})}}. \tag{10}$$

If we assume that $x, y, z, w \in X_3$ form a non-trivial solution to (7), and define x_i, y_i, z_i, w_i and j as before, we conclude, by reducing modulo $(q+1)^{j+1}$, that

$$5x_j(q+1)^j \equiv y_j(q+1)^j + 2z_j(q+1)^j + 2w_j(q+1)^j \qquad (\mathrm{mod}\ (q+1)^{j+1}).$$

As before, this is actually an equality, implying that $5x_j = y_j + 2z_j + 2w_j$ and supplying the desired contradiction.

This completes the proof of the lemma. Since

$$q = \lceil 2^{\sqrt{\log n}} \rceil$$

we obtain, from (8), (9), and (10), that

$$|X_1|, |X_2|, |X_3| \geqslant \frac{n}{2^{O((\log n)^{3/4})}}. \quad \blacksquare$$

COROLLARY 3.3. *There exists a set* $X \subset \{1, ..., n\}$ *satisfying*

$$|X| \geqslant \frac{n}{2^{O((\log n)^{3/4})}}$$

such that X *has no non-trivial solution to* (5), *no non-trivial solution to* (6), *and no non-trivial solution to* (7).

Proof. Take two integers $-n \leqslant u_2 \leqslant n$ and $-n \leqslant u_3 \leqslant n$ randomly, uniformly and independently. $X = X_1 \cap (X_2 + u_2) \cap (X_3 + u_3)$ has no non-trivial solution to any of the above equations, and each $x \in X_1$ has probability $\Omega(2^{-O((\log n)^{3/4})})$ to lie in the intersection. The result thus follows from the linearity of the expectation. $\blacksquare$

THEOREM 3.4. *The function $f(n)$ satisfies*

$$f(n) \geqslant \frac{n^2}{2^{O((\log n)^{3/4})}}. \tag{11}$$

Proof. It is more convenient to show that

$$f(n2^{\sqrt{\log n}} + 6n) \geqslant \frac{n^2}{2^{O((\log n)^{3/4})}},$$

which clearly gives (11).

Put $q = \lceil 2^{\sqrt{\log n}} \rceil$ and let X be as in the corollary. Define

$$C = \{(p, p+2x, p+5x, p+(q+5)\,x) \mid 1 \leqslant p \leqslant n, x \in X\}.$$

Then $C \subset N^4$ for $N = \{1, 2, ..., (q+6)\,n\}$. Clearly

$$|C| \geqslant \frac{n^2}{2^{O((\log n)^{3/4})}}.$$

We claim that C has the IPP. Indeed, no two words in C share more than one coordinate. Thus, if $a, b, c \in C$ are distinct they cannot violate (IPP1) since otherwise for every $1 \leqslant i \leqslant 4$ there exists a pair among a, b, c sharing the same coordinate in place i, implying by the pigeonhole principle that some pair of words shares at least 2 coordinates, which is impossible.

It remains to check (IPP2). Suppose that

$$a = (p_1, p_1 + 2x, p_1 + 5x, p_1 + (q+5)\,x),$$

$$b = (p_2, p_2 + 2y, p_2 + 5y, p_2 + (q+5)\,y),$$

$$c = (p_3, p_3 + 2z, p_3 + 5z, p_3 + (q+5)\,z),$$

$$d = (p_4, p_4 + 2w, p_4 + 5w, p_4 + (q+5)\,w)$$

satisfy $\{a, b\} \cap \{c, d\} = \emptyset$ and yet $\{a_i, b_i\} \cap \{c_i, d_i\} \neq \emptyset$ for all $1 \leqslant i \leqslant 4$. Choose $g_i \in \{a_i, b_i\} \cap \{c_i, d_i\}$ for each i. No word can share 3 coordinates with $g = (g_1, g_2, g_3, g_4)$. Indeed, if for example, $a_1 = g_1$, $a_2 = g_2$ and $a_3 = g_3$ then, as $g_i \in \{c_i, d_i\}$ for every i, either c or d have to agree with a on at least 2 coordinates, which is impossible.

Since $g_i \in \{a_i, b_i\}$ and $g_i \in \{c_i, d_i\}$ for every i, each of the 4 words a, b, c, d agrees with $g = (g_1, g_2, g_3, g_4)$ on exactly 2 coordinates. Moreover, the indices of those common coordinates of a and g, and those of b and g, are disjoint (as together they have to cover all 4 coordinates); and the same occurs with those of c and g with respect to those of d and g. It follows that up to symmetry there are 3 possible cases.

Case 1.

$$a_1 = g_1, \quad b_3 = g_3, \quad c_2 = g_2, \quad d_1 = g_1,$$
$$a_2 = g_2, \quad b_4 = g_4, \quad c_3 = g_3, \quad d_4 = g_4.$$

Case 2.

$$a_1 = g_1, \quad b_2 = g_2, \quad c_2 = g_2, \quad d_1 = g_1,$$
$$a_3 = g_3, \quad b_4 = g_4, \quad c_3 = g_3, \quad d_4 = g_4.$$

Case 3.

$$a_1 = g_1, \quad b_2 = g_2, \quad c_1 = g_1, \quad d_3 = g_3,$$
$$a_3 = g_3, \quad b_4 = g_4, \quad c_2 = g_2, \quad d_4 = g_4.$$

In Case 1, by noting that

$$(g_2 - g_1) + (g_3 - g_2) + (g_4 - g_3) - (g_4 - g_1) = 0$$

and that

$$g_2 - g_1 = a_2 - a_1 = 2x, \qquad g_3 - g_2 = c_3 - c_2 = 3z,$$
$$g_4 - g_3 = b_4 - b_3 = qy, \qquad g_4 - g_1 = d_4 - d_1 = (q+5)\,w,$$

we conclude that

$$2x + 3z + qy - (q+5)\,w = 0.$$

Thus $x = y = z = w$ by the construction of X that has no non-trivial solution to (5). But then it follows that $a = d$, in contradiction to $\{a, b\} \cap \{c, d\} = \emptyset$.

Similarly, Case 2 leads by the fact that X has no non-trivial solution to (6), to the fact that $x = y = z = w$ and hence again to the contradiction $a = d$. Case 3 leads to $x = y = z = w$ as X has no non-trivial solution to (7), giving the contradiction $a = c$. This completes the proof. ∎

ACKNOWLEDGMENTS

We thank Benny Sudakov, Sergei Konyagin, and Endre Szemerédi for helpful discussions.

Note added in proof. As observed by S. Konyagin, the lower bound given in Theorem 3.4 can be slightly improved to

$$f(n) \geqslant \frac{n^2}{2^{O((\log n)^{2/3})}}$$

by proving the first part of Lemma 3.2 using the method applied in the proofs of its second and third part.

REFERENCES

1. N. Alon, R. A. Duke, H. Lefmann, V. Rödl, and R. Yuster, The algorithmic aspects of the Regularity Lemma, *in* "Proceedings of the 33rd IEEE FOCS at Pittsburgh, 1992," pp. 473–481; *J. Algorithms* **16** (1994), 80–109.
2. F. A. Behrend, On sets of integers which contain no three terms in arithmetic progression, *Proc. Nat. Acad. Sci. U.S.A.* **32** (1946), 331–332.
3. H. D. L. Hollmann, J. H. Van Lint, J.-P. Linnartz, and L. M. G. M. Tolhuizen, On codes with the identifiable parent property, *J. Combin. Theory Ser. A* **82** (1998), 121–133.
4. I. Ruzsa, Solving a linear equation in a set of integers, I, *Acta Arith.* **65** (1993), 259–282.
5. E. Szemerédi, Regular partitions of graphs, *in* "Proc. Colloque Inter. CNRS" (J. C. Bermond, J. C. Fournier, M. Las Vergnas, and D. Sotteau, Eds.), Vol. 260, pp. 399–401, 1978.