On Hypergraphs of Girth Five*

Felix Lazebnik
Department of Mathematical Sciences
University of Delaware, Newark,
DE 19716, USA
lazebnik@math.udel.edu

Jacques Verstraëte
Theory Group, Microsoft Research
Building 113/2105 Redmond,
WA 98052, USA
jbav@microsoft.com

Submitted: Oct 9, 2002; Accepted: May 8, 2003; Published: May 30, 2003
MR Subject Classifications: 05D05, 05D40, 05D99

To the memory of Dom de Caen

Abstract

In this paper, we study r-uniform hypergraphs $\mathcal{H}$ without cycles of length less than five, employing the definition of a hypergraph cycle due to Berge. In particular, for $r = 3$, we show that if $\mathcal{H}$ has n vertices and a maximum number of edges, then

$$|\mathcal{H}| = \frac{1}{6}n^{3/2} + o(n^{3/2}).$$

This also asymptotically determines the generalized Turán number $T_3(n, 8, 4)$. Some results are based on our bounds for the maximum size of Sidon-type sets in $\mathbb{Z}_n$.

1 Definitions

In this paper, a *hypergraph* $\mathcal{H}$ is a family of distinct subsets of a finite set. The members of $\mathcal{H}$ are called *edges*, and the elements of $V(\mathcal{H}) = \bigcup_{E \in \mathcal{H}} E$ are called *vertices*. If all edges in $\mathcal{H}$ have size r, then $\mathcal{H}$ is called an r-uniform hypergraph or, simply, r-*graph*. For example, a 2-graph is a graph in the usual sense. A vertex v and an edge E are called *incident* if $v \in E$. The *degree* of a vertex v of $\mathcal{H}$, denoted $d(v)$, is the number of edges

*This work was initiated and continued at Microsoft Research during the author's visits, and we are thankful the hosts for the opportunity and support.

of $\mathcal{H}$ incident with v. An r-graph $\mathcal{H}$ is *r-partite* if its vertex set $V(\mathcal{H})$ can be colored in r colors in such a way that no edge of $\mathcal{H}$ contains two vertices of the same color. In such a coloring, the color classes of $V(\mathcal{H})$ – the sets of all vertices of the same color – are called *parts* of $\mathcal{H}$. We refer the reader to Berge [3] or [4] for additional background on hypergraphs.

For $k \geq 2$, a *cycle* in a hypergraph $\mathcal{H}$ is an alternating sequence of vertices and edges of the form $v_1, E_1, v_2, E_2, \ldots, v_k, E_k, v_1$, such that

(i) $v_1, v_2, \ldots, v_k$ are distinct vertices of $\mathcal{H}$
(ii) $E_1, E_2, \ldots, E_k$ are distinct edges of $\mathcal{H}$
(iii) $v_i, v_{i+1} \in E_i$ for each $i \in \{1, 2, \ldots, k-1\}$, and $v_k, v_1 \in E_k$.

We refer to a cycle with k edges as a *k-cycle*, and denote the family of all k-cycles by $\mathcal{C}_k$. For example, a 2-cycle consists of a pair of vertices and a pair of edges such that the pair of vertices is a subset of each edge. The above definition of a hypergraph cycle is the "classical" definition (see, for example, Berge [3], [4], Duchet [11]). For $r = 2$ and $k \geq 3$, it coincides with the definition of a cycle C_k in graphs and, in this case, $\mathcal{C}_k$ is a family consisting of precisely one member. Detailed discussions of alternative definitions of cycles in hypergraphs and the merits of each, as well as their applications in computer science, may be found in Duke [12] and Fagin [18]. The *girth* of a hypergraph $\mathcal{H}$, containing a cycle, is the minimum length of a cycle in $\mathcal{H}$. On a connection between 3-graphs of girth at least five and Greechie diagrams in quantum physics, see McKay, Megill and Pavičić [24].

2 Problems and Results

The topic of this paper falls into the context of Turán-type extremal problems in hypergraphs, on which an excellent survey was given by Füredi [19]. The question we consider is to determine the maximum number of edges in an r-graph on n vertices of girth five. For graphs ($r = 2$), this is an old problem of Erdős [14], which has its origins in a seminal paper of Erdős [13]. The best known lower and upper bounds are $(1/2\sqrt{2})n^{3/2} + O(n)$ and $(1/2)(n-1)^{1/2}n$, respectively. For bipartite graphs, on the other hand, this maximum is $(1/2\sqrt{2})n^{3/2} + O(n)$ as $n \to \infty$. Many attempts at reducing the gap between the constants $1/2\sqrt{2}$ and $1/2$ in the lower and upper bounds have not succeeded thus far (see Garnick, Kwong, Lazebnik [20] for more details). Surprisingly, we are able to obtain upper and lower bounds for the corresponding problem in 3-graphs which have equal leading terms.

Theorem 2.1 *Let $\mathcal{H}$ be a 3-graph on n vertices and of girth at least five. Then*

$$|\mathcal{H}| \leq \tfrac{1}{6}n\sqrt{n-\tfrac{3}{4}}+\tfrac{1}{12}n.$$

For any odd prime power $q \geq 27$, there exist 3-graphs $\mathcal{H}$ on $n = q^2$ vertices, of girth five, with

$$|\mathcal{H}| = \tbinom{q+1}{3} = \tfrac{1}{6}n^{3/2} - \tfrac{1}{6}n^{1/2}.$$

This result is surprising in the sense that Turán-type questions for hypergraphs are generally harder than for graphs. One may formally apply the famous Ray-Chaudhuri and Wilson Theorem [25] to hypergraphs of girth at least three, which states that an r-graph, without a pair of sets intersecting in at least two points, has at most $\binom{n}{2}/\binom{r}{2}$ edges, and the equality is attained for each $r \geq 3$ and infinitely many n.

Following de Caen [10], the generalized Turán number $T_r(n,k,l)$ is defined to be the maximum number of edges in an r-graph on n vertices in which no set of k vertices spans l or more edges (or, equivalently, the union of any l edges contains more than k vertices). To illustrate this definition, the above-mentioned result of Ray-Chadhuri and Wilson is equivalent to the statement $T_r(n, 2r-2, 2) = \binom{n}{2}/\binom{r}{2}$ for each $r \geq 3$ and infinitely many n.

The problem of estimating $T_r(n,k,l)$ in general was first approached by Brown, Erdős, and T. Sós [8], [9], who gave bounds for $T_3(n,k,l)$ for all $k \leq 6$ and $l \leq 9$, and established the asymptotics of the generalized Turán numbers $T_3(n,k,l)$ for $(k,l) \in \{(5,3),(5,4),(6,4)\}$. In the case $(k,l) = (6,3)$, they established $T_3(n,6,3) > cn^{3/2}$ for some constant c. Remarkably precise bounds for $T_3(n,6,3)$ were given by Ruzsa and Szemerédi, who proved that for some constant $c > 0$ and all $\varepsilon > 0$,

$$2^{-c\sqrt{\log n}}n^2 \leq T_3(n,6,3) < \varepsilon n^2.$$

The asymptotic behaviour of the numbers $T_r(n,k,l)$, in general, remains unknown, and seems to be difficult to determine. For example, perhaps one of the most famous problems in extremal combinatorics is to prove or disprove Turán's conjecture, that $T_3(n,4,4) \sim \frac{5}{9}\binom{n}{3}$, $n \to \infty$.

We now continue to relate the problem of estimating the size of hypergraphs of given girth with certain generalized Turán numbers. It is easy to see that $T_3(n,4,2)$ and $T_3(n,6,3)$ are precisely the maximum number of edges in a 3-graph of girth three and four respectively. Similarly, $T_3(n,8,4)$ is precisely the maximum number of edges in a 3-graph of girth five. This is seen by directly checking that any four triples on a set of eight vertices span a hypergraph containing a cycle of length at most four. Together with Theorem 2.1, and results about the density of primes (see Huxley [21]), this implies:

Corollary 2.2 *As $n \to \infty$, $T_3(n,8,4) \sim \frac{1}{6}n^{3/2}$.*

Generalizing to r-graphs, $r \geq 2$, we are able to establish the following:

Theorem 2.3 *Let $\mathcal{H}$ be an r-graph, $r \geq 2$, on n vertices and of girth at least five. Then*

$$|\mathcal{H}| \leq \tfrac{1}{r(r-1)} n^{3/2} + \tfrac{r-2}{2r(r-1)} n + O(n^{-1/2}).$$

Moreover, if $\mathcal{H}$ is r-partite, with n vertices in each part, then

$$|\mathcal{H}| \leq \tfrac{1}{\sqrt{r-1}}\, n^{3/2} + \tfrac{1}{2} n + O(n^{1/2}).$$

Finally, for each $r \geq 2$, there exist r-partite r-graphs of girth at least five, with $n \geq 8r^r$ vertices in each part and $\frac{1}{4} r^{-4r/3} n^{4/3}$ edges.

The proof of Theorem 2.3 for $r = 2$ gives the best known upper bounds for the maximum number of edges for girth five graphs and bipartite graphs, namely $\frac{1}{2} n\sqrt{n-1}$ and $\frac{1}{2} n(1 + \sqrt{4n-3})$, respectively. The latter expression is an upper bound on the Zarankiewicz number – the maximum size of a bipartite graph with each part having n vertices and without cycles of length four (see, Kővári, T. Sós, Turán [22] and Reiman [26]).

The lower bound in Theorem 2.3 is a probabilistic one. Attempts to establish explicit and better lower bounds led us to a generalization of the notion of a Sidon set in $\mathbb{Z}_n$, and to the question of its maximum cardinality. We remind the reader that a *Sidon set* in $\mathbb{Z}_n$ (or in $\mathbb{Z}$) is a set in which no two distinct pairs of elements have the same difference (or, equivalently, the same sum). The reader is referred to Babai and Sós [2] for more details on Sidon sets. Our generalization, roughly, will disallow equality between small integer multiples of such differences, and we present it next.

Let k be a positive integer and let n be relatively prime to all elements of $\{1, 2, \ldots, k\}$. Let a_1, a_2, a_3, a_4 be integers in $\{-k, -k+1, \ldots, 0, \ldots, k-1, k\}$ such that $a_1+a_2+a_3+a_4 = 0$. Let $\mathcal{S}$ be the collection of sets $S \subset \{1, 2, 3, 4\}$ such that $\sum_{i \in S} a_i = 0$ and $a_i \neq 0$ for $i \in S$. Now consider the following equation over $\mathbb{Z}_n$ with respect to $x = (x_1, x_2, x_3, x_4)$:

$$a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 = 0. \tag{1}$$

A solution x of (1) is called *trivial* if there exists a partition of $\{1, 2, 3, 4\}$ into sets $S, T \in \mathcal{S}$ such that $x_i = x_j$ for all $i, j \in S$ and all $i, j \in T$. This is analagous to the definition of a trivial solution (over the integers) to an equation of the form $a_1x_1 + a_2x_2 + \cdots + a_kx_k = 0$ by Ruzsa [27].

For example, consider the equation $x_1 + x_2 - x_3 - x_4 = 0$. Then $\mathcal{S}$ consists of the sets $\{1, 3\}, \{2, 4\}, \{1, 4\}, \{2, 3\}$ and $\{1, 2, 3, 4\}$. Therefore the trivial solutions are those with $x_1 = x_3, x_2 = x_4$, or $x_1 = x_4, x_2 = x_3$, or $x_1 = x_2 = x_3 = x_4$. A set with only trivial solutions to $x_1+x_2-x_3-x_4 = 0$ is precisely a Sidon set. As the second example, consider the equation $2x_1 - 3x_2 + x_4 = 0$. Then $\mathcal{S}$ consists of the set $\{1, 2, 4\}$. The trivial solutions are therefore those for which $x_1 = x_2 = x_4$.

A *k-fold Sidon set* is a set $A \subset \mathbb{Z}_n$ such that the equation (1) has only trivial solutions in A. For example, a 1-fold Sidon set is a Sidon set in the usual sense. For a 2-fold Sidon set A, each of the equations below admits only trivial solutions with $x_1, x_2, x_3, x_4 \in A$:

$$x_1 - x_2 + x_3 - x_4 = 0, \quad x_1 + x_2 - 2x_3 = 0, \quad x_1 - x_2 + 2x_3 - 2x_4 = 0$$

The definition of a k-fold Sidon set also extends to the set $\{1, 2, \ldots, n\} \subset \mathbb{Z}$, in which case the condition that n is relatively prime to all integers in $\{1, 2, \ldots, k\}$ may be dropped.

How large can a k-fold Sidon set A in $\mathbb{Z}_n$ be? Let us first present an elementary upper bound. To each pair $\{a, a'\}$ of distinct elements of A, we can associate the set $\{i(a - a') \mid i \in \{1, 2, \ldots, k\}\}$. Note that each set has k elements and, for distinct pairs, the corresponding sets are disjoint. It follows immediately that $k\binom{|A|}{2} \le n$ and therefore $|A| < (2n/k)^{1/2} + 1$. To improve this bound we will use Theorem 2.3 in a way described below.

Let A be a subset of $\mathbb{Z}_n$, and let B be a set of r integers. Define $\mathcal{H}(A, B)$ to be the r-partite r-graph with parts $V_b = \mathbb{Z}_n$, $b \in B$. For each $x \in \mathbb{Z}_n$ and each $a \in A$, an edge of $\mathcal{H}(A, B)$ is the set of r vertices $\{x + ba : b \in B\}$, where $x + ba \in V_b$. Hence $\mathcal{H}(A, B)$ contains rn vertices and $|A|n$ edges. The following proposition establishes a connection between r-partite r-graphs of girth five and k-fold Sidon sets.

Proposition 2.4 *Let n, k, r be positive integers, and n be odd. Let $B \subset \mathbb{Z}$ be a Sidon set of integers of size r such that all differences of distinct elements of B are relatively prime to n and do not exceed k. Let A be a k-fold Sidon set in $\mathbb{Z}_n$. Then $\mathcal{H}(A, B)$ is an r-partite r-graph of girth at least five, with $|A|n$ edges.*

Theorem 2.3 and Proposition 2.4 sometimes lead to a better constant in the upper bound for the size of a k-fold Sidon set of $\mathbb{Z}_n$. For example, let $k = 3$, $\gcd(n, 6) = 1$, and $B = \{-1, 0, 2\}$ (a Sidon set). Then, applying Theorem 2.3 (with $r = 3$) and Proposition 2.4, we can reduce the bound $(2n/3)^{1/2}$ on a 3-fold Sidon set to $(n/2)^{1/2}$.

Next, for infinitely many n, we provide a lower bound within 2 factor of the upper bound on the size of a 2-fold Sidon set:

Theorem 2.5 *Let t be a positive integer, and let $n = 2^{2^{t+1}} + 2^{2^t} + 1$. Then, there exists a 2-fold Sidon set A in $\mathbb{Z}_n$, such that*

$$|A| \ge \tfrac{1}{2} n^{1/2} - 3.$$

It seems likely that for each integer $k \geq 3$, there exists a k-fold Sidon set in $\mathbb{Z}_n$ (or in $\{1, 2, \ldots, n\} \subset \mathbb{Z}$) of size $cn^{1/2}$ for some $c > 0$ depending only on k.

By Theorem 2.5 and Theorem 2.3, we immediately obtain the following result:

Theorem 2.6 *Let $\mathcal{H}$ be a 3-partite 3-graph with $n \geq 3$ vertices in each of its parts and of girth at least five. Then*

$$|\mathcal{H}| \leq \tfrac{1}{\sqrt{2}} n^{3/2} + n.$$

Let i be a positive integer and let $n = 2^{2^{i+1}} + 2^{2^i} + 1$. Then there exists a 3-partite 3-graph $\mathcal{H}$, with n vertices in each part, of girth at least five, such that

$$|\mathcal{H}| \geq \tfrac{1}{2} n^{3/2} - 3n.$$

We remark that from the second part of Theorem 2.1, we obtain a weaker lower bound of $(1+o(1))\frac{\sqrt{3}}{9} n^{3/2}$, by applying the Erdős-Kleitmann Lemma [15] in the case $r = 3$: every r-graph $\mathcal{H}$ on rn vertices contains an r-partite r-graph with n vertices in each part and at least $\frac{r!}{r^r}|\mathcal{H}|$ edges.

3 Proofs

Here we will prove the results stated in the previous section.

Proof of Theorem 2.1

Upper Bound. The upper bound of Theorem 2.1 is obtained by setting $r = 3$ in the upper bound of Theorem 2.3. Therefore it is sufficient to prove the latter. Let $\mathcal{H}$ be an r-graph of girth at least five, and let $m = |\mathcal{H}|$. For each vertex $v \in V(\mathcal{H})$ and for each unordered pair of edges A, B incident with v, associate the set $v(A, B)$ of unordered pairs of vertices in $A \cup B \setminus \{v\}$ which are not contained in A or B. We first note that $v(A, B) \cap v(C, D) = \emptyset$ whenever $\{A, B\} \neq \{C, D\}$, otherwise $\mathcal{H}$ contains a 2-cycle. Hence $|v(A, B)| = (r - 1)^2$. Now we define

$$D_v = \bigcup_{\{A,B\}\,:\, v \in A \cap B} v(A, B).$$

Then $D_v \cap D_w = \emptyset$ whenever v and w are distinct vertices of $\mathcal{H}$, otherwise it is easy to check that $\mathcal{H}$ contains a cycle of length at most four. Also, no pair in D_v is contained in an edge of $\mathcal{H}$, otherwise $\mathcal{H}$ contains a 3-cycle. Since $\mathcal{H}$ contains no 2-cycle, and the number of pairs of vertices contained in edges is precisely $\binom{r}{2}m$, we have

$$\begin{aligned}
\binom{n}{2} - \binom{r}{2}m &\geq \sum_{v \in V(\mathcal{H})} |D_v| \\
&= \sum_{v \in V(\mathcal{H})} \sum_{\{A,B\}\,:\, v \in A \cap B} |v(A,B)| \\
&= \sum_{v \in V(\mathcal{H})} \binom{d(v)}{2} |D_v| \\
&= (r-1)^2 \sum_{v \in V(\mathcal{H})} \binom{d(v)}{2} \\
&\geq (r-1)^2 \cdot \left(\frac{r^2 m^2}{2n} - \frac{rm}{2} \right).
\end{aligned}$$

In the last inequality, we used the fact that $\sum_{v \in V(\mathcal{H}} d(v) = \binom{r}{2}m$, and Jensen's inequality for function $\binom{x}{2}$ on $[1, \infty]$. Multiplying by $2n$ we obtain,

$$r^2(r-1)^2 m^2 + r(r-1)(2-r)n\,m - n^2(n-1) \leq 0.$$

This gives

$$\begin{aligned}
m &\leq \frac{1}{r(r-1)} n \sqrt{n + \frac{r^2 - 4r}{4}} + \frac{r-2}{2r(r-1)} n \\
&\leq \frac{1}{r(r-1)} n^{3/2} + \frac{r-2}{2r(r-1)} n + O(n^{-1/2}),
\end{aligned}$$

as required. For $r = 3$, we get

$$m \leq \frac{1}{6} n \sqrt{n - \frac{3}{4}} + \frac{1}{12} n < \frac{1}{6} n^{3/2} + \frac{1}{12} n + \frac{1}{16} n^{-1/2}. \qquad \square$$

Lower Bound. We provide an explicit construction by using the so-called *polarity graph* of the projective plane $PG(2, q)$, which we denote by Pol_q (see Erdős and Rényi [16], Erdős, Rényi and T. Sós [17], Brown [6]). We start by a brief description of this graph along with its properties.

Let $\mathbb{F}_q$ denote the finite field of odd characteristic. We consider a nondegenerate orthogonal geometry on $V = \mathbb{F}_q^3$ corresponding to the bilinear form $x \cdot y = x_1 y_1 + x_2 y_2 + x_3 y_3$. The nondegeneracy means that no nonzero vector of V is orthogonal to all vectors of V. It implies that $\dim(U) + \dim(U^\perp) = 3$ for each subspace U of V, where $U^\perp$ denotes the orthogonal complement of U. We define the vertex set of Pol_q to be the set of all lines (1-dimensional subspaces) of this space. Clearly, Pol_q has $\frac{q^3-1}{q-1} = q^2 + q + 1$ vertices. The

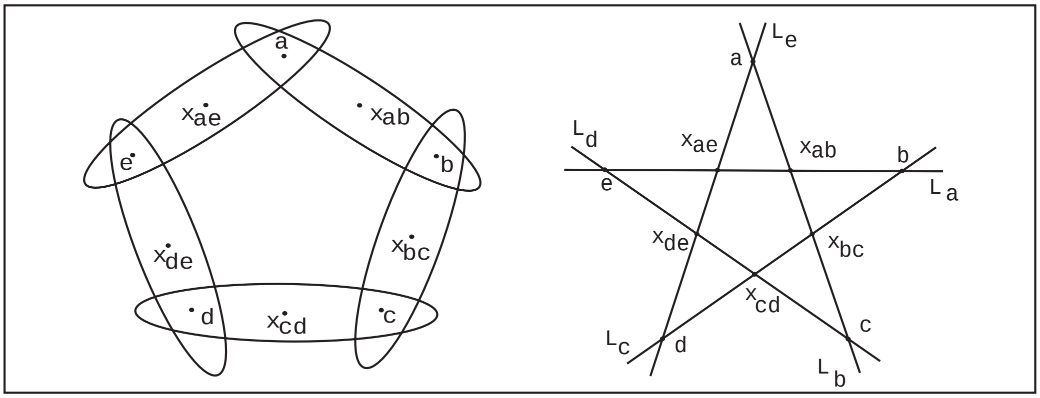

Figure 1: a 5-cycle in $\mathcal{H}_q$

edges of Pol_q are formed by all (unordered) pairs of distinct orthogonal lines. A line is called *isotropic*, if it is spanned by a vector x such that $x \cdot x = 0$. Since the geometry is nondegenerate, the orthogonal complement to a line is a plane (a 2-dimensional subspace). Each plane contains $\frac{q^2-1}{q-1} = q+1$ lines, hence the degree of a vertex in our graph is $q+1$ for a nonisotropic line and q for an isotropic one. It is a well known fact that the number of isotropic lines in the geometry is $q+1$. Therefore the number of nonisotropic lines is q^2, and Pol_q has $\frac{1}{2}((q+1)q^2 + q(q+1)) = (q+1)^2q/2$ edges.

For each pair of distinct lines there exists a unique line orthogonal to both of them, namely the orthogonal complement of the plane defined by the lines. Therefore Pol_q contains no 4-cycles, and every edge formed by two nonisotropic lines belongs to exactly one triangle (3-cycle). Next we observe that an isotropic line cannot be a vertex of a triangle: if it were, the orthogonal complement to this line would be 3-dimensional, a contradiction with nondegeneracy. As the number of edges spanned by all nonisotropic lines is $(q+1)^2q/2 - q(q+1) = (q+1)q(q-1)/2$, the number of triangles in Pol_q is $\binom{q+1}{3}$.

Consider a 3-graph $\mathcal{H}_q$ with vertex set being the set of all $n = q^2$ nonisotropic lines, and the edge set formed by the sets of vertices of each triangle. Then

$$|\mathcal{H}_q| = \binom{q+1}{3} = \frac{1}{6}n^{3/2} - \frac{1}{6}n^{1/2}.$$

We claim that the girth of $\mathcal{H}_q$ is five. As no two triangles of Pol_q share an edge, $\mathcal{H}_q$ contains no 2-cycles and no 3-cycles. If $\mathcal{H}_q$ has a 4-cycle, then Pol_q contains a 4-cycle with exactly same vertices, a contradiction. Therefore the girth of $\mathcal{H}_q$ is at least five.

Representing a line by a nonzero vector in it, one can easily check that the following sequence of vertices and edges (see Figure 1) determines a 5-cycle $aE_{ab}bE_{bc}cE_{cd}dE_{de}eE_{ea}a$ in $\mathcal{H}_q$, for q odd and not divisible by 3.

$$a = \langle(1,0,0)\rangle,\ b = \langle(0,0,1)\rangle,\ c = \langle(1,1,0)\rangle,\ d = \langle(1,-1,1)\rangle,\ e = \langle(0,1,1)\rangle,$$

$$x_{a,b} = \langle(0,1,0)\rangle,\ x_{bc} = \langle(1,-1,0)\rangle,\ x_{cd} = \langle(-1,1,2)\rangle,$$
$$x_{de} = \langle(2,1,-1)\rangle,\ \ x_{ea} = \langle(0,1,-1)\rangle.$$
$$E_{ab} = \{a, x_{ab}, b\},\ E_{bc} = \{b, x_{bc}, c\},\ E_{cd} = \{c, x_{cd}, d\},$$
$$E_{de} = \{d, x_{de}, e\},\ E_{ea} = \{e, x_{ea}, a\}.$$

In general, it is easy to show show that there exists q_0 such $\mathcal{H}_q$ contains a 5-cycle for all odd prime powers $q \geq q_0$. We did not try to determine the smallest q_0 with this property, but it is easy to show that $q_0 = 27$ will suffice. $\square$

Remark We also would like to mention another explicit construction of a 3-graph $\mathcal{G}_q$ of order $n = q(q-1)$ which may have at least as many edges as any subhypergraph of $\mathcal{H}_q$ of the same order. Let $\mathbb{F}_q$ denote the finite field of odd characteristic, and let C_q denote the set of points on the curve $2x_2 = x_1^2$, where $(x_1, x_2) \in \mathbb{F}_q \times \mathbb{F}_q$. Define a hypergraph $\mathcal{G}_q$ as follows. The vertex set of $\mathcal{G}_q$ is $\mathbb{F}_q \times \mathbb{F}_q \setminus C_q$. Three distinct vertices $a = (a_1, a_2)$, $b = (b_1, b_2)$ and $c = (c_1, c_2)$ form an edge $\{a, b, c\}$ of $\mathcal{G}_q$ if and only if the following three equations are satisfied:

$$\begin{aligned} a_2 + b_2 &= a_1 b_1 \\ b_2 + c_2 &= b_1 c_1 \\ c_2 + a_2 &= c_1 a_1. \end{aligned}$$

It is not difficult to check that $\mathcal{G}_q$ has girth at least five for all odd q and girth five for all sufficiently large q. The number of edges in $\mathcal{G}_q$ is precisely $\binom{q}{3} \approx \frac{n\sqrt{n}}{6}$, since there are $\binom{q}{3}$ choices for distinct numbers a_1, b_1 and c_1, which uniquely specify a_2, b_2 and c_2 such that a, b, c are not on the curve $2y = x^2$ and $\{a, b, c\}$ is an edge. It is interesting to understand whether $\mathcal{G}_q$ is a subhypergraph of $\mathcal{H}_q$, but we have not been able to resolve this question yet.

Proof of Theorem 2.3

Upper Bounds The first upper bound has been established in the proof of Theorem 2.1, and our argument for the second upper bound is a modification of the one used there.

Let $\mathcal{H}$ be an r-partite r-graph of girth at least five, with n vertices in each part. Let A_i, $i \in [r]$, be the parts of $\mathcal{H}$. We estimate the cardinality of the set $S = \{(v, \{x, y\})\}$, where $v \in V(\mathcal{H})$ and x and y are distinct from v, belong to the same part and are in different edges on v. Let E_1, E_2 be two distinct edges incident to $v \in A_i$. There are exactly $r-1$ sets $\{x, y\}$ such that $(v, \{x, y\}) \in S$, since each such pair $\{x, y\}$ is the intersection of $E_1 \cup E_2$ with a part A_j, $j \neq i$. On the other hand, the absence of cycles of length less than five in $\mathcal{H}$ implies that $|S| \leq \sum_{i \in [r]} \binom{|A_i|}{2} = r\binom{n}{2}$. Therefore we obtain

$$|S| = (r-1)\sum_{v\in V(\mathcal{H})}\binom{d(v)}{2} = (r-1)\sum_{i\in[r]}\sum_{v\in A_i}\binom{d(v)}{2} \;\le\; \sum_{i\in[r]}\binom{|A_i|}{2} = r\binom{n}{2}.$$

Again, as in the proof of Theorem 2.3, applying Jensen's inequality to $\sum_{v\in A_i}\binom{d(v)}{2}$ and using the fact that $\sum_{v\in A_i} d(v) = m = |\mathcal{H}|$ for each i, we get

$$(r-1)rn\binom{m/n}{2} \le |S| \le r\binom{n}{2}.$$

This implies $(r-1)m^2 - (r-1)nm - n^2(n-1) \le 0$, or

$$|\mathcal{H}| \;\le\; \frac{n}{2} + n\sqrt{\frac{n-1}{r-1}+\frac{1}{4}} \le \frac{1}{\sqrt{r-1}}n^{3/2} + \frac{1}{2}n + O(n^{1/2}),$$

as required for the upper bound in Theorem 2.3. □

Lower Bound. It is sufficient to establish the lower bound for r-partite r-graphs with n vertices in each part. Also we assume that $r \ge 3$, since for $r = 2$ a better lower bound $\Omega(n^{3/2})$ is provided by the point-line incidence graph of a projective plane.

Let $\mathcal{H} = \mathcal{H}_{r,n,p}$ denote a random r-partite r-graph with n vertices in each part in which edges are present uniformly and independently with probability p. Let $X = X(\mathcal{H})$ be the number of edges in $\mathcal{H}$. Then X is a binomial random variable with probability p and mean $\mu = pn^r$. Let us choose $p = r^{-4r/3}n^{4/3-r}$. We will use a version of the Chernoff bound implied by the one from Alon and Spencer [1] (page 238, Theorem A.13): for binomial random variables X with mean $\mu \ge 0$ and probability p, $\Pr[X < \frac{1}{2}\mu] < \exp(-\frac{1}{8}\mu)$. Hence

$$\Pr[X < \frac{1}{2}pn^r] < \exp(-\frac{1}{8}pn^r).$$

Therefore, the number of edges in $\mathcal{H}$ is at least $\frac{1}{2}r^{-4r/3}n^{4/3}$ with probability greater than $\frac{1}{2}$, as $n \ge 8r^r$. The numbers of 2-cycles, 3-cycles and 4-cycles in the complete r-partite r-graph are, respectively, at most $(rn)^{2(r-1)}$, $(rn)^{3(r-1)}$ and $(rn)^{4(r-1)}$. The expected number of cycles of length at most four in $\mathcal{H}$ is therefore at most

$$p^2(rn)^{2(r-1)} + p^3(rn)^{3(r-1)} + p^4(rn)^{4(r-1)}.$$

As $r \ge 3$, and by our choice of p, this is at most $3\cdot r^{-4}r^{-4r/3}n^{4/3} < \frac{1}{8}r^{-4r/3}n^{4/3}$. By Markov's Inequality, the probability that the number of cycles of length at most four in $\mathcal{H}$ is at least $\frac{1}{4}r^{-4r/3}n^{4/3}$ is less than $\frac{1}{2}$. Therefore, with positive probability, $\mathcal{H}$ has at least $\frac{1}{2}r^{-4r/3}n^{4/3}$ edges and at most $\frac{1}{4}r^{-4r/3}n^{4/3}$ cycles of length at most four. Deleting an edge from each copy of a cycle of length at most four in $\mathcal{H}$, we obtain an r-partite r-graph of girth at least five with at least $\frac{1}{4}r^{-4r/3}n^{4/3}$ edges, as required. □

Proof of Proposition 2.4

Suppose $A \subset \mathbb{Z}_n$ is a k-fold Sidon set, and let $\mathcal{H}(A, B)$ be the r-partite r-graph defined before Proposition 2.4. Let us verify that $\mathcal{H}(A, B)$ has girth at least five. For convenience, set $\mathcal{H}(A, B) = \mathcal{H}$.

It is clear that $\mathcal{H}$ contains no 2-cycle. For if $\mathcal{H}$ contains a 2-cycle, comprising edges $E_1 = \{x + bx_1 : b \in B\}$ and $E_2 = \{y + bx_2 : b \in B\}$, then $x + ix_1 = y + ix_2$ and $x + jx_1 = y + jx_2$ for some distinct $i, j \in B$. Therefore $(i - j)(x_1 - x_2) = 0$. Since n is relatively prime to all differences of distinct elements of B, n is relatively prime to $i - j$, and so $x_1 = x_2$ and $x = y$. Therefore $E_1 = E_2$, a contradiction. Thus $\mathcal{H}$ contains no 2-cycle.

Suppose $\mathcal{H}$ contains a 3-cycle, formed by the following edges:

$$\begin{aligned} E_1 &= \{x + bx_1 : b \in B\}, \\ E_2 &= \{y + bx_2 : b \in B\}, \\ E_3 &= \{z + bx_3 : b \in B\}. \end{aligned}$$

Suppose these edges intersect in the following way:

$$\begin{aligned} E_1 \cap E_2 &= \{x + hx_1\} = \{y + hx_2\}, \\ E_2 \cap E_3 &= \{y + ix_2\} = \{z + ix_3\}, \\ E_3 \cap E_1 &= \{z + jx_3\} = \{x + jx_1\}. \end{aligned}$$

Note that h, i, j are pairwise distinct. This implies

$$(h - j)x_1 + (i - h)x_2 + (j - i)x_3 = 0.$$

With $a_1 = h - j$, $a_2 = i - h$ and $a_3 = j - i$, we have $a_1 + a_2 + a_3 = 0$, $1 \leq |a_i| \leq k$ for all $i \in \{1, 2, 3\}$, and $a_1x_1 + a_2x_2 + a_3x_3 = 0$. By the definition of a k-fold Sidon set, we must have $x_1 = x_2 = x_3$. This implies that $E_1 = E_2 = E_3$, a contradiction. Therefore $\mathcal{H}$ contains no 3-cycle.

Finally, suppose $\mathcal{H}$ contains a 4-cycle. Then, in the same way as above, we consider its subsequent edges

$$\begin{aligned} E_1 &= \{x + bx_1 : b \in B\}, \\ E_2 &= \{y + bx_2 : b \in B\}, \\ E_3 &= \{z + bx_3 : b \in B\}, \\ E_4 &= \{w + bx_4 : b \in B\} \end{aligned}$$

and vertices

$$\begin{aligned} E_1 \cap E_2 &= \{x + gx_1\} = \{y + gx_2\}, \\ E_2 \cap E_3 &= \{y + hx_2\} = \{z + hx_3\}, \\ E_3 \cap E_4 &= \{z + ix_3\} = \{w + ix_4\}, \\ E_4 \cap E_1 &= \{w + jx_4\} = \{x + jx_1\}. \end{aligned}$$

Then $g-j, h-g, i-h, j-i$ are all non-zero, otherwise two of the edges E_i are identical. Similarly to the case of 3-cycles, here we obtain an equation of the form:

$$(g-j)x_1 + (h-g)x_2 + (i-h)x_3 + (j-i)x_4 = 0,$$

so that $a_1 = g-j, a_2 = h-g, a_3 = i-h, a_4 = j-i$ are non-zero. The above equation is

$$a_1x_1 + a_3x_3 + a_2x_2 + a_4x_4 = 0.$$

Now if $x_i = x_{i+1}$ for some i modulo four, then two of E_1, E_2, E_3, E_4 are equal. Therefore, as A is a k-fold Sidon set, the only possible trivial solutions are $x_1 = x_3$ and $x_2 = x_4$ and $x_1 \neq x_2$. By the definition of a trivial solution, this implies $a_1 + a_3 = 0$ and $a_2 + a_4 = 0$. Both of these equations imply $i + g - j - h = 0$. As B is a Sidon set, we must have $i = j$ and $g = h$, or $i = h$ and $g = j$. In the first case we obtain $x + gx_1 = y + hx_2 = z + hx_3$ and and in the second case we obtain $w + jx_4 = x + hx_1 = y + gx_2$, so three of E_1, E_2, E_3, E_4 intersect. This contradiction completes the proof. $\square$

Proof of Theorem 2.5

Lower Bound. For a positive integer i, let $t = 2^i$ and $n = 2^{2t} + 2^t + 1$. Let D be a Singer difference set (see Singer [29]) in $\mathbb{Z}_n$ of $2^t + 1$ elements with multiplier 2. Since every nonzero element of $\mathbb{Z}_n$ can be written uniquely as the difference of two members of D, D is a Sidon set and $D = 2D = \{2d : d \in D\}$. This implies that D is formed by members of a cycle or a union of cycles of the permutation $\pi : x \mapsto 2x$ of $\mathbb{Z}_n$ (the map is a permutation since $\gcd(2, n) = 1$). Only one cycle of D (and of π) has length 1, namely $\{0\}$. For each m-cycle C of D, we label its consecutive vertices by $c_1, c_2, \ldots, c_m$, and then delete all vertices c_j with j odd. Finally, deleting the element zero from D, we obtain a Sidon set $S \subset D$ such that $a \neq 2b$ for every $a, b \in S$. We now verify that S is the required 2-fold Sidon set in $\mathbb{Z}_n$.

Since S is a Sidon set, the equations $u + v = x + y$, $2u + 2v = 2x + 2y$ and $u + v = 2x$ have only trivial solutions. So the only thing we need to check is that for $u, v, x, y \in S$, $u + 2v = x + 2y$ implies $(u, v) = (x, y)$.

If $u = x$, then $v = y$, and we are done. Suppose $u \neq x$. Since $u, v, x, y \in S$ and $S \subset D = 2D$, we have $u, 2v, x, 2y \in D$. Therefore $u = 2y$ and $2v = x$. Hence S contains both y and $2y$, as well as v and $2v$, and this contradicts our construction of S.

Finally, how large is $|S|$? When we delete every second element from each even cycle of D, we delete exactly half of its elements. When we do this for an odd cycle, we delete one more than half of its elements. Let $\mathrm{odd}(D)$ and $\mathrm{odd}(\pi)$ represent the number of odd cycles of length greater than 1 of D and π, respectively. Then, since $0 \in D \setminus S$,

$$|S| = (|D|-1)/2 - \mathrm{odd}(D) = 2^{t-1} - \mathrm{odd}(D).$$

Therefore we need to estimate $\mathrm{odd}(D)$. If $\{x, 2x, 2^2x, \ldots, 2^{e-1}x\}$ is a cycle of π of length e, then e is the smallest positive integer such that $2^e x \equiv x \bmod n$, or equivalently,

$$\frac{n}{\gcd(x,n)} \mid (2^e - 1).$$

Let $a \in \mathbb{Z}_n^*$ (the group of units of ring $\mathbb{Z}_n$), and let $\mathrm{ord}(a,n)$ denote the order of a in $\mathbb{Z}_n^*$. Then $e = \mathrm{ord}(2, n/\gcd(x,n))$. As $2^{3t} - 1 = (2^t-1)n$, it follows that $n \,|\, (2^{3t}-1)$, and therefore $\mathrm{ord}(2,n) \,|\, 3t$. Clearly, $\mathrm{ord}(2,n) > 2t$. So $\mathrm{ord}(2,n) = 3t$. Therefore $e \,|\, 3t$. Since $e > 1$ and e is odd, and $t = 2^i$, $e = 3$. So all odd cycles of π have length three. Therefore $\mathrm{odd}(\pi) = \frac{1}{3}c$, where c is the number of $x \in \mathbb{Z}_n \setminus \{0\}$ such that $n \mid x(2^3-1) = 7x$. Since $7 \mid n$, $c = 6$, and the number of cycles of length 3 in π is 2. Consequently,

$$|S| = 2^{t-1} - \mathrm{odd}(D) \geq 2^{t-1} - \mathrm{odd}(\pi) \geq 2^{t-1} - 2 = \frac{1}{2}n^{1/2} + O(1),\ i \to \infty. \quad \square$$

Remarks. If t is not a power of 2, the magnitude of S is not so clear. The number of odd cycles in π can be rather large. For example, if $t = 9$, then $n = 262657$, $\mathrm{odd}(\pi) = 9728$ and all odd cycles (but $\{0\}$) are of length 27. For $t = 11$, $n = 4196353$, $\mathrm{odd}(\pi) = 127164$, there are 2 cycles of length 3 and all other odd cycles are of length 33.

Acknowledgement

The authors are thankful to László Lovász [23] for his suggestion to consider a 3-graph with edges formed by the sets of vertices of triangles in the polarity graph Pol_q. Taking a subgraph $\mathcal{H}_q$ in this graph allowed the authors to improve their original asymptotics for the number of edges in 3-graph on n vertices and of girth five of $\frac{1}{6}n^{3/2} - \frac{1}{4}n + o(n^{1/2})$ (based on 3-graph $\mathcal{G}_q$) to $\frac{1}{6}n^{3/2} - \frac{1}{6}n^{1/2} + o(n^{1/2})$ (based on 3-graph $\mathcal{H}_q$). We also are thankful to Qing Xiang for helpful discussions concerning difference sets, to Brendan McKay for bringing reference [24] to our attention, and to an anonymous referee for many useful suggestions.

References

[1] Alon, N., Spencer, J. H., The Probabilistic Method, A Wiley-Interscience Publication, John Willey & Sons, Inc., 1992.

[2] Babai, L., Sós, V., Sidon sets in groups and induced subgraphs of Cayley graphs. European J. Combin. 6 (1985), no. 2, 101–114.

[3] Berge, C., Hypergraphs, Combinatorics of Finite Sets. North–Holland, Amsterdam, 1989.

[4] Berge, C., Hypergraphs, in Selected Topics in Graph Theory 3, edited by Lowell W. Beineke and Robin.J. Wilson, Academic Press Limited, 1998, 189-207.

[5] Beutelspacher, A., Rosenbaum, U., Projective geometry: from foundations to applications. Cambridge University Press, Cambridge, 1998.

[6] Brown, W., On graphs that do not contain a Thomsen graph, Canad. Math. Bull 9 (1966), 281–285.

[7] Bollobás, B., Extremal Graph Theory, L.M.S. Monographs, No. 11, Academic Press, London, 1978.

[8] Brown, W., Erdős, P., Sós, V., Some extremal problems on r-graphs, in New Directions in the Theory of Graphs, *Proc. Third Ann Arbor Conf., Univ. of Mich. 1971*, pp. 53-63, Academic Press, New York, 1973.

[9] Brown, W., Erdős, P., Sós, V., On the existence of triangulated spheres in 3-graphs, Periodica Math. Hungar. 3 (1973) 221–228.

[10] de Caen, D., The current status of Turán's problem on hypergraphs, Extremal problems for finite sets (Visegrád, 1991), 187–197, Bolyai Soc. Math. Stud., 3, János Bolyai Math. Soc., Budapest, 1994.

[11] Duchet, P., Hypergraphs, Handbook of Combinatorics, Volume 1, North–Holland, Amsterdam, 1995, 381–432.

[12] Duke, R., Types of cycles in hypergraphs. Cycles in graphs, 399–417, North-Holland Math. Stud., 115, North-Holland, Amsterdam, 1985.

[13] Erdős, P., On sequences of integers no one of which divides the product of two others and on some related problems, Mitt. Forsch.-Inst. Math. Mech. Univ. Tomsk 2, (1938), 74–82.

[14] Erdős, P., Some recent progress on extremal problems in graph theory, Congr. Numer. 14 (1975), 3–14.

[15] Erdős, P., Kleitman, D., On coloring graphs to maximize the proportion of multicolored k-edges. J. Combinatorial Theory 5 (1968) 164–169.

[16] Erdős, P., Rényi, A., On a problem in the theory of graphs (in Hungarian), Publ. Math. Inst. Hungar. Acad. Sci. 7, (1962).

[17] Erdős, P., Rényi, A., T. Sós, V., On a problem of graph theory, Studia Sci.Math. Hungar. 1 (1966), 51–57.

[18] Fagin, R., Degrees of acyclicity for hypergraphs and relational data-base schemes, J. ACM Vol. 30 No. 3 (July 1983), 514–550.

[19] Füredi, Z., Turán Type Problems. Survey's in Combinatorics, 1991. Cambrtidge University Press, Cambridge, 1991.

[20] Garnick, D.K., Kwong, Y.H.H., and Lazebnik, F., Extremal graphs without three-cycles and four-cycles, J. Graph Theory 17 (1993), 633–645.

[21] Huxley, M. On the difference between consecutive primes. Invent. Math. 15 (1972), 164–170.

[22] Kővári T., T. Sós, V., Turán, P., On a problem of Zarankiewicz, Colloq. Math. 3, (1954), 50-57.

[23] Lovász, L., private communication, 2002.

[24] McKay, B.D., Megill N.D., Pavičić, M., Algorithms for Greechie Diagrams, Internat. J. Theor. Phys. 39, (2000), 2393-2417.

[25] Ray-Chaudhuri, D., Wilson, R., The existence of resolvable block designs. Survey of combinatorial theory (Proc. Internat. Sympos., Colorado State Univ., Fort Collins, Colo., 1971), pp. 361–375. North-Holland, Amsterdam, 1973.

[26] Reiman, I., Über ein Problem von K. Zarankiewicz, Acta. Math. Acad. Sci. Hungar. 9, (1958), 269–279.

[27] Ruzsa, I., Solving a linear equation in a set of integers. I. Acta Arith. 65 (1993), no. 3, 259–282.

[28] Ruzsa, I., Szemerédi. E., Triple systems with no six points carrying three triangles. Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976), Vol. II, pp. 939–945, Colloq. Math. Soc. János Bolyai, 18, North-Holland, Amsterdam-New York, 1978.

[29] Singer J., A theorem in finite projective geometry and some applications to number theory, Trans. Amer. Math. 43 1938 377–385.