Combinatorics, Probability and Computing (2013) 22, 241–252.
doi:10.1017/S0963548312000557

On the Independence Number of Steiner Systems

ALEX EUSTIS and JACQUES VERSTRAËTE

Department of Mathematics, UCSD, 9500 Gilman Drive, La Jolla, CA 92093-0112, USA
(e-mail: jbaverstraete@gmail.com)

Received 16 February 2012; revised 15 November 2012

A *partial Steiner* (n, r, l)*-system* is an r-uniform hypergraph on n vertices in which every set of l vertices is contained in at most one edge. A partial Steiner (n, r, l)-system is *complete* if every set of l vertices is contained in exactly one edge. In a hypergraph $\mathcal{H}$, the independence number $\alpha(\mathcal{H})$ denotes the maximum size of a set of vertices in $\mathcal{H}$ containing no edge. In this article we prove the following. Given integers r, l such that $r \geqslant 2l - 1 \geqslant 3$, we prove that there exists a partial Steiner (n, r, l)-system $\mathcal{H}$ such that

$$\alpha(\mathcal{H}) \lesssim \left(\frac{l-1}{r-1}(r)_l\right)^{\frac{1}{r-1}} n^{\frac{r-l}{r-1}} (\log n)^{\frac{1}{r-1}} \quad \text{as } n \to \infty.$$

This improves earlier results of Phelps and Rödl, and Rödl and Ŝinajová. We conjecture that it is best possible as it matches the independence number of a random r-uniform hypergraph of the same density. If $l = 2$ or $l = 3$, then for infinitely many r the partial Steiner systems constructed are complete for infinitely many n.

AMS 2010 *Mathematics subject classification*: Primary 05B05
Secondary 05B07, 05B20, 05B25, 05D40

1. Introduction

For integers $1 < l < r < n$, an r-uniform hypergraph on n vertices is called a *partial Steiner* (n, r, l)*-system* or, for short, an (n, r, l)*-system*, if every l-set of vertices is contained in at most one edge of the hypergraph. An (n, r, l)-system is *complete* if every l-set is in exactly one edge. In this paper, we study the independence number of (n, r, l)-systems: this is the size of the largest set of vertices in an r-uniform hypergraph containing no edge. For a hypergraph $\mathcal{H}$, the independence number is denoted $\alpha(\mathcal{H})$. The independence number arises in many applications and is central in extremal hypergraph theory, relative to Turán-type problems and Ramsey Theory, extremal problems in combinatorial geometry [17, 12], and algorithmic complexity. The motivation for this paper is to construct Steiner (n, r, l)-systems which are close to complete and whose independence number is asymptotically the same as the independence number of a random r-uniform hypergraph with the same expected density of edges as $n \to \infty$. We believe that these (n,r,l)-systems have asymptotically the smallest possible independence number amongst all (n,r,l)-systems.

1.1. The independence number of (n,r,l)-systems

We first discuss historical bounds on the independence number of (n,r,l)-systems. An elementary probabilistic argument (see for instance [4] for more precise results) shows that an (n,r,l)-system contains an independent set of size $\Omega\big(n^{\frac{r-l}{r-1}}\big)$. Phelps and Rödl [20] were the first to show that for $(n,3,2)$-systems $\mathcal{H}$, a substantially better bound is possible, namely that if $\mathcal{H}$ is an $(n,3,2)$-system then $\alpha(\mathcal{H}) = \Omega(\sqrt{n\log n})$. More general results were obtained by Duke, Lefmann and Rödl [7] for $(n,r,2)$-systems, building on the paper of Ajtai, Komlós, Pintz, Spencer and Szemerédi [1]. Their result was extended by Rödl and Ŝinajová [22] to cover all (n,r,l)-systems, where Rödl and Ŝinajová showed that if $\mathcal{H}$ is an (n,r,l)-system, then

$$\alpha(\mathcal{H}) = \Omega\left(n^{\frac{r-l}{r-1}}(\log n)^{\frac{1}{r-1}}\right). \tag{1.1}$$

Rödl and Ŝinajová [22] also showed that this is tight up to large constant factors depending on l and r. Shearer's method [23] was used by Kostochka, Mubayi and the second author [18] to obtain better lower bounds when $l = r-1$. In what follows, if f, g are positive-valued functions of n, then we write $f \sim g$ if and only if $\lim_{n\to\infty} f(n)/g(n) = 1$ and $f \lesssim g$ if and only if $\limsup_{n\to\infty} f(n)/g(n) \leqslant 1$.

1.2. Main results

Complete (n,r,l)-systems in general are very difficult to construct; in fact no infinite family is known for $r > l > 3$. The construction of 'near-complete' or 'asymptotic' (n,r,l)-systems constitutes Rödl's solution [21] to a long-standing conjecture of Erdős and Hanani [9] on the existence of asymptotic designs. The contribution of this paper is to construct 'near-complete' (n,r,l)-systems whose independence number is asymptotic to the independence number of a random r-uniform hypergraph with the same density of edges. We shall check in Section 2 that if $\mathcal{H}$ is an r-uniform hypergraph on n-vertices created by sampling the edges of the complete r-uniform hypergraph independently with probability p, and p is chosen so that the expected number of edges of $\mathcal{H}$ equals the number of edges in a complete (n,r,l)-system, then almost surely as $n \to \infty$,

$$\alpha(\mathcal{H}) \sim \left(\frac{l-1}{r-1}(r)_l\right)^{\frac{1}{r-1}} \cdot n^{\frac{r-l}{r-1}}(\log n)^{\frac{1}{r-1}}.$$

It is convenient henceforth to denote the quantity on the right by $A(n,r,l)$. We construct partial Steiner (n,r,l)-systems whose independence number matches this bound for 'more than half' of the pairs $r > l > 1$.

Theorem 1.1. *If $r \geqslant 2l-1 \geqslant 3$, then there exists an (n,r,l)-system $\mathcal{H}$ such that $\alpha(\mathcal{H}) \sim A(n,r,l)$ as $n \to \infty$.*

For $l = 2$ or $l = 3$, the construction of $\mathcal{H}$ in Theorem 1.1 can actually be made for infinitely many r into a complete $(n,r,2)$-system for infinitely many values of n, using

https://doi.org/10.1017/S0963548312000557 Published online by Cambridge University Press

Wilson's theorem [24] for $l = 2$ and known constructions of inversive planes for $l = 3$. We prove Theorem 1.1 using an iterative algebraic construction, together with some randomness, and the analysis requires a little spectral theory and probability. We believe that Theorem 1.1 extends to all cases $1 < l < r$, but this remains open. In a forthcoming paper [8], we will use Rödl's nibble method to prove that the theorem also holds for $l = r - 1$. We make the following conjecture.

Conjecture 1.2. *Let r, l be integers, where $r > l > 1$. Then for any partial (n, r, l)-system $\mathcal{H}$,*

$$\alpha(\mathcal{H}) \gtrsim A(n, r, l) \qquad \text{as } n \to \infty.$$

For instance, this conjecture predicts that for every partial Steiner triple system on n vertices, the independence number is at least asymptotic to $\sqrt{3n \log n}$ as $n \to \infty$. The current best lower bounds from [18] are $\alpha(\mathcal{H}) \gtrsim 0.458\sqrt{n \log n}$ when $\mathcal{H}$ is any partial $(n, 3, 2)$-system. Perhaps the first interesting case not covered by Theorem 1.1 is $r = 4$ and $l = 3$, where we seek to construct an n-vertex example with independence number asymptotic to $(16n \log n)^{1/3}$ as $n \to \infty$.

1.3. Notation

Hypergraphs. An *r-uniform hypergraph* on a set V is a set $\mathcal{H}$ of r-element subsets of V called *edges*. If $U \subseteq V$, then the *induced r-uniform hypergraph* $\mathcal{H}[U]$ is the subset of edges of $\mathcal{H}$ contained in U. We say that U is an *independent set* if $\mathcal{H}[U] = \emptyset$. The *independence number* $\alpha(\mathcal{H})$ is the maximum size of an independent set in $\mathcal{H}$. Unless otherwise noted, the hypergraphs in this paper all have the same fixed vertex set V of size n. Let $\mathcal{H}_r(n, p)$ denote the *random hypergraph* obtained by independently sampling the edges of a complete r-uniform hypergraph on n vertices with probability p.

Asymptotic notation. The standard limit notation $f = O(g), \Theta(g), \Omega(g), o(g)$ is generally taken with respect to the implicit variable n, unless noted otherwise. Also, if f, g are positive-valued functions of n, then we write $f \sim g$ if and only if $\lim_{n\to\infty} f(n)/g(n) = 1$ and $f \lesssim g$ if and only if $\limsup_{n\to\infty} f(n)/g(n) \leqslant 1$. Given two natural numbers n, r, let $(n)_r = n(n-1)\cdots(n-r+1)$.

2. Independent sets in random hypergraphs

In this section we compute the asymptotic value of $\alpha(\mathcal{H})$ for the random r-uniform hypergraph $\mathcal{H} = \mathcal{H}_r(n, p)$ as $n \to \infty$, for the specific value p defined by

$$p\binom{n}{r} = \mathbb{E}(|\mathcal{H}_r(n, p)|) = \frac{\binom{n}{l}}{\binom{r}{l}}.$$

This states that the expected number of edges of $\mathcal{H}_r(n, p)$ equals the number of edges in a complete (n, r, l)-system. To compute the upper bound on $\alpha(\mathcal{H})$, we first state a technical lemma which will ultimately be used in the proof of Theorem 1.1, and which relies on the first moment method.

Lemma 2.1. *Let r, l be integers with $r > l > 1$, and let λ, β be positive reals such that*

$$\beta > \left(\frac{(l-1)}{(r-1)\lambda}\right)^{\frac{1}{r-1}}. \tag{2.1}$$

Suppose, for infinitely many n, that $\mathcal{H}_n = \mathcal{H}_r(n,p)$ is a random r-uniform hypergraph on n vertices such that, for any $U \subseteq V(\mathcal{H}_n)$ of size

$$u := \lfloor \beta n^{\frac{r-l}{r-1}} (\log n)^{\frac{1}{r-1}} \rfloor,$$

we have

$$-\log \mathbb{P}(\mathcal{H}_n[U] = \emptyset) \gtrsim \lambda \beta^r n^{\frac{r-l}{r-1}} (\log n)^{\frac{r}{r-1}}.$$

Then almost surely as $n \to \infty$, $\alpha(\mathcal{H}_n) < u$.

Proof. Fix $\beta > 0$ and let I denote the number of independent sets of size u in $\mathcal{H}_n$. Then

$$\log \mathbb{E}(I) - \log \binom{n}{u} \leqslant -(1+o(1))\lambda \beta^r n^{\frac{r-l}{r-1}} (\log n)^{\frac{r}{r-1}}.$$

Using standard estimates for binomial coefficients,

$$\begin{aligned} \log \binom{n}{u} &\sim u \log(n/u) \\ &\sim \beta n^{\frac{r-l}{r-1}} (\log n)^{\frac{1}{r-1}} \left(1 - \frac{r-l}{r-1}\right) \log n \\ &= \beta \frac{l-1}{r-1} n^{\frac{r-l}{r-1}} (\log n)^{\frac{r}{r-1}}. \end{aligned}$$

It follows from the inequality on β in the lemma that $\log \mathbb{E}(I) \to -\infty$, so $\mathbb{E}(I) \to 0$. Therefore, by the union bound, almost surely as $n \to \infty$, $I = 0$. □

The main point of the next lemma is to show $\alpha(\mathcal{H}_r(n,p)) \lesssim A(n,r,l)$ as $n \to \infty$. With slightly more work, it is in fact true that $\alpha(\mathcal{H}_r(n,p)) \sim A(n,r,l)$ as $n \to \infty$.

Lemma 2.2. *Let p be chosen such that*

$$\mathbb{E}(|\mathcal{H}_r(n,p)|) = \frac{\binom{n}{l}}{\binom{r}{l}}.$$

Then almost surely as $n \to \infty$,

$$\alpha(\mathcal{H}_r(n,p)) \sim A(n,r,l). \tag{2.2}$$

Proof. Let $\beta > 0$ and let U be a subset of the vertices of $\mathcal{H}_r(n,p)$ with $|U| = u$, where

$$u := \lfloor \beta n^{\frac{r-l}{r-1}} (\log n)^{\frac{1}{r-1}} \rfloor.$$

https://doi.org/10.1017/S0963548312000557 Published online by Cambridge University Press

Since the edges are chosen independently,

$$\begin{aligned}
-\log \mathbb{P}(\mathcal{H}_r(n,p)[U]=\emptyset) &= -\binom{u}{r}\log(1-p)\\
&\sim p\frac{u^r}{r!}\\
&\sim \frac{n^l}{l!}\frac{l!(r-l)!}{r!}\frac{r!}{n^r}\frac{1}{r!}\left(\beta n^{\frac{r-l}{r-1}}(\log n)^{\frac{1}{r-1}}\right)^r\\
&= \frac{\beta^r}{(r)_l}n^{l-r+r(\frac{r-l}{r-1})}(\log n)^{\frac{r}{r-1}}\\
&= \frac{\beta^r}{(r)_l}n^{\frac{r-l}{r-1}}(\log n)^{\frac{r}{r-1}}.
\end{aligned}$$

Let $\lambda = 1/(r)_l$ and take β to satisfy (2.1). Then Lemma 2.1 applies: almost surely as $n \to \infty$,

$$\alpha(\mathcal{H}_r(n,p)) \lesssim \left(\frac{(l-1)}{(r-1)\lambda}\right)^{\frac{1}{r-1}} \cdot n^{\frac{r-l}{r-1}}(\log n)^{\frac{1}{r-1}}.$$

It can be shown that this is also an asymptotic lower bound on $\alpha(\mathcal{H}_r(n,p))$ (see, for instance, Krivelevich and Sudakov [19]). This proves (2.2). □

3. Proof of Theorem 1.1

To prove Theorem 1.1, we give a randomized construction for an (n,r,l)-system $\mathcal{H}$ of low independence number when $r \geqslant 2l-1$. Let q be a prime power and $q > r$, and let $V = \mathbb{F}_q \times \mathbb{F}_q$, where $\mathbb{F}_q$ denotes the finite field of order q. If f is a polynomial over $\mathbb{F}_q$, the graph of f is

$$G_f = \{(x, f(x)) : x \in \mathbb{F}_q\},$$

and let $\mathcal{P} = \mathcal{P}(q,r,l)$ be the hypergraph on V defined by

$$\mathcal{P} = \{G_f : \deg(f) \leqslant l-1\}.$$

Since $|G_f| = q$ for all f, and no two distinct graphs can have l points in common, it follows that $\mathcal{P}$ is a (q^2,q,l)-system (see [2] for the use of this system in the context of the de Bruijn–Erdős problem). Next, let $\mathcal{H}_q$ denote an *asymptotically complete* (q,r,l)-system, such that $|\mathcal{H}_q| \sim \binom{q}{l}/\binom{r}{l}$ as $q \to \infty$. The existence of such asymptotically complete (n,r,l)-systems is given by the semi-random method of Rödl [21]. We assume $\mathcal{H}_q$ has vertex set $[q]$. Independently for each $G_f \in \mathcal{P}$, let $\pi_f : V(\mathcal{H}_q) \to G_f$ be a random bijection, and let

$$\pi_f(\mathcal{H}_q) = \{\{\pi_f(i_1), \pi_f(i_2), \ldots, \pi_f(i_r)\} : \{i_1, i_2, \ldots, i_r\} \in \mathcal{H}_q\}.$$

Thus, independently for each G_f, a randomly permuted copy of $\mathcal{H}_q$ is placed on G_f. Define the hypergraph $\mathcal{H} = \mathcal{H}(q,r,l)$ with vertex set V, and with the (random) edge set

$$\mathcal{H} = \bigcup_{G_f \in \mathcal{P}} \pi_f(\mathcal{H}_q).$$

https://doi.org/10.1017/S0963548312000557 Published online by Cambridge University Press

We observe $\mathcal{H}$ is an (n,r,l)-system, regardless of how the π_f are chosen. Indeed, for any l-set $b \subseteq V$, there can be at most one $G_f \in \mathcal{P}$ containing b, and for this G_f there is at most one $\{i_1, i_2, \dots, i_r\} \in \mathcal{H}_q$ such that $b \subseteq \{\pi_f(i_j) : 1 \leqslant j \leqslant r\}$.

The first lemma we need states that if T is a large subset of vertices of $\mathcal{H}_q$, chosen uniformly from $V(\mathcal{H}_q)$, then it is very unlikely that T is an independent set of $\mathcal{H}_q$.

Lemma 3.1. *Let $t \in \mathbb{N}$ satisfy $t = o(q^{1-l/r})$ as $q \to \infty$. If $T \subseteq V(\mathcal{H}_q)$ denotes a uniformly chosen set of size t, then*

$$-\log \mathbb{P}(\mathcal{H}_q[T] = \emptyset) \gtrsim \frac{(t)_r \binom{q}{l}}{(q)_r \binom{r}{l}} \qquad \text{as } q \to \infty.$$

Proof. Let $T \subseteq V(\mathcal{H}_q)$ be a uniformly chosen set of size t. Then by inclusion–exclusion,

$$\mathbb{P}\left(\bigcup_{e \in \mathcal{H}_q} \{e \subseteq T\}\right) \geqslant \sum_{e \in \mathcal{H}_q} \mathbb{P}(e \subseteq T) - \sum_{\substack{e,f \in \mathcal{H}_q \\ e \neq f}} \mathbb{P}(e \cup f \subseteq T)$$

$$= \sum_{e \in \mathcal{H}_q} \mathbb{P}(e \subseteq T) - \sum_{k=0}^{r-1} \sum_{\substack{e,f \in \mathcal{H}_q \\ |e \cap f| = k}} \mathbb{P}(e \cup f \subseteq T).$$

Since $\mathcal{H}_q$ is a (q,r,l)-system, all terms in the second sum indexed by $k \geqslant l$ are zero. We consider the contribution of each term from $k = 0$ to $k = l-1$. If we fix a set b of size k, observe that the sets $\{e \setminus b : e \in \mathcal{H}_q, b \subseteq e\}$ form a $(q-k, r-k, l-k)$-system. Therefore the number of edges containing b is at most $\binom{q-k}{l-k}/\binom{r-k}{l-k}$, and so the number of *pairs* of edges whose intersection has size k is at most

$$\binom{q}{k}\binom{\binom{q-k}{l-k}/\binom{r-k}{l-k}}{2} = O(q^{2l-k}).$$

If A is any fixed set of size $s \leqslant t$, then

$$\mathbb{P}(A \subseteq T) = \frac{(t)_s}{(q)_s}.$$

Combining these statements, it follows from the union bound that

$$\mathbb{P}(\mathcal{H}_q[T] = \emptyset) \leqslant 1 - \mathbb{P}\left(\bigcup_{e \in \mathcal{H}_q} \{e \subseteq T\}\right)$$

$$\leqslant 1 - |\mathcal{H}_q| \frac{(t)_r}{(q)_r} + \sum_{k=0}^{l-1} O\left(q^{2l-k} \frac{(t)_{2r-k}}{(q)_{2r-k}}\right)$$

Now, we need to show that each term of the sum is asymptotically irrelevant when $t = o(q^{1-l/r})$. Specifically, we show

$$\lim_{q \to \infty} \left(q^{2l-k} \frac{(t)_{2r-k}}{(q)_{2r-k}}\right) \Big/ \left(|\mathcal{H}_q| \frac{(t)_r}{(q)_r}\right) = 0$$

https://doi.org/10.1017/S0963548312000557 Published online by Cambridge University Press

where the convergence is uniform over $t = o(q^{1-l/r})$. Recall that

$$|\mathcal{H}_q| \sim \binom{q}{l} \bigg/ \binom{r}{l} = O(q^l),$$

so the above limit is zero provided that

$$q^{2l-k} t^{2r-k} q^{k-2r} = o(q^{l-r} t^r).$$

The left side is maximized when $k = 0$, in which case we require

$$q^{2l-2r} t^{2r} = o(q^{l-r} t^r).$$

This is equivalent to $t = o(q^{1-l/r})$, which is precisely the assumption on t in the lemma. So as $q \to \infty$,

$$\begin{aligned} -\log \mathbb{P}(\mathcal{H}_q[T] = \emptyset) &\gtrsim -\log\left(1 - |\mathcal{H}_q| \frac{(t)_r}{(q)_r}\right) \\ &\sim |\mathcal{H}_q| \frac{(t)_r}{(q)_r} \\ &\sim \frac{\binom{q}{l}(t)_r}{\binom{r}{l}(q)_r}. \end{aligned}$$

This completes the proof of Lemma 3.1. □

To be able to apply this lemma to prove Theorem 1.1, we show that if U is any large subset of vertices of $\mathcal{H}$, then U intersects most of the edges G_f in roughly the expected number of vertices, namely $|U||G_f|/|V| \sim |U|/q$. To do this, we consider eigenvalues of an appropriate matrix associated with $\mathcal{P}$.

3.1. Incidence matrix and eigenvalues

In this section, we intend to show that if U is a reasonably large set of vertices of $\mathcal{P}(q,r,l)$, then $|U \cap G_f| \sim |U|/q$ for almost every $G_f \in \mathcal{P}$, as follows.

Lemma 3.2 (Main Lemma). *Let $r \geqslant 2l - 1 \geqslant 3$, let $U \subset V$ and suppose $|U|/q \to \infty$ as $q \to \infty$. Then for all but $o(q^l)$ edges $G_f \in \mathcal{P}(q,r,l)$, $|U \cap G_f| \sim |U|/q$, as $q \to \infty$.*

For the rest of this section, we fix r and l and let $\mathcal{P} = \mathcal{P}(q,r,l)$. We recall some facts from linear algebra and use them to obtain spectral information about the hypergraph $\mathcal{P}$. Given any hypergraph $\mathcal{H}$ with vertex set V, the *incidence matrix* of $\mathcal{H}$ is the matrix I whose rows are indexed by V and whose columns are indexed by $\mathcal{H}$ and such that $I_{ve} = 1$ if $v \in e$ and $I_{ve} = 0$ otherwise. If every vertex of $\mathcal{H}$ has degree a then we say that $\mathcal{H}$ is *a-regular*. Define the matrix

$$A(\mathcal{H}) = \begin{pmatrix} 0 & I \\ I^t & 0 \end{pmatrix}.$$

The rows and columns of $A(\mathcal{H})$ are both indexed by $V \cup \mathcal{H}$ in the natural way. We let $\mathbf{1}_S$ denote the characteristic vector of a set $S \subset V \cup \mathcal{H}$, so that $\mathbf{1}_i = 1$ if $i \in S$ and $\mathbf{1}_i = 0$

https://doi.org/10.1017/S0963548312000557 Published online by Cambridge University Press

otherwise. We require two lemmas to prove the Main Lemma. The first can be checked using elementary linear algebra.

Lemma 3.3. *Let $\mathcal{H}$ be a connected a-regular b-uniform hypergraph, where $a, b > 0$. If $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_N$ are the eigenvalues of A, then $\lambda_1 = \sqrt{ab}$ and $\lambda_N = -\sqrt{ab}$ with multiplicity 1, corresponding to eigenvectors $\sqrt{a}\mathbf{1}_V + \sqrt{b}\mathbf{1}_{\mathcal{H}}$ and $\sqrt{a}\mathbf{1}_V - \sqrt{b}\mathbf{1}_{\mathcal{H}}$ respectively.*

We focus our attention on $A(\mathcal{P})$. We observe that $\mathcal{P}$ is a q^{l-1}-regular q-uniform hypergraph: we have $|e| = q$ for all $e \in \mathcal{P}$ and, fixing a pair $(u, v) \in V$, the number of polynomials f of degree at most $l-1$ such that $f(u) = v$ is exactly q^{l-1}. Applying Lemma 3.3, we see that the matrix $A(\mathcal{P})$ has largest and smallest eigenvalues equal to $q^{l/2}$ and $-q^{l/2}$ respectively. The key quantity for our purposes is the maximum absolute value of all the remaining eigenvalues of $A(\mathcal{P})$ – all but the smallest and largest – which we denote by $\lambda(\mathcal{P})$. We determine $\lambda(\mathcal{P})$ exactly, as follows.

Lemma 3.4. $\lambda(\mathcal{P}) = q^{\frac{l-1}{2}}$.

Proof. Let J be the $q^2 \times q^l$ all one matrix, and let

$$K = \left[\begin{array}{c|c} 0 & J \\ \hline J^t & 0 \end{array}\right].$$

We claim that

$$A(\mathcal{P})^3 = (q-1)q^{l-2}K + q^{l-1}A(\mathcal{P}). \tag{3.1}$$

This matrix equation will allow us to compute $\lambda(\mathcal{P})$ using Lemma 3.3. It is convenient to write $v \to e$ if $v \in V$ is contained in $e \in \mathcal{P}$. To prove the lemma, fix $v \in V$ and $e \in \mathcal{P}$, and count the number of walks of length three between e to v, namely the number of choices of e' and v' such that $v \to e' \leftarrow v' \to e$. Let f be the polynomial corresponding to $e \in \mathcal{P}$. First suppose $v \not\to e$. Then there are $q-1$ choices for v', namely $(x, f(x))$, where x differs from the first co-ordinate of v; otherwise v and v' are distinct points with the same x-coordinate, so there is no polynomial passing through both. For each of these choices, there are exactly q^{l-2} choices for e', since we have to choose a polynomial f' of degree at most $l-1$ passing through both v and v' which have different first coordinates. On the other hand, if $v \to e$, then in addition to the above $(q-1)q^{l-2}$ choices of v' and e', we can also choose $v' = v$. In this case, there are q^{l-1} choices for e', namely all the polynomials of degree at most $l-1$ that pass through v. This proves the matrix equation (3.1). Now, if x is an eigenvector corresponding to an eigenvalue $\lambda \notin \{\lambda_1, \lambda_N\}$ of A, then by Lemma 3.3, x is orthogonal to the eigenvectors corresponding to the eigenvalues $\lambda_1 = q^{l/2}$ and $\lambda_N = -q^{l/2}$. By Lemma 3.3, those eigenvectors are $q^{l-1}\mathbf{1}_V + q\mathbf{1}_{\mathcal{P}}$ and $q^{l-1}\mathbf{1}_V - q\mathbf{1}_{\mathcal{P}}$. It follows that $Kx = \mathbf{0}$, and

$$\lambda^3 = q^{l-1}\lambda.$$

So if $\lambda \notin \{\lambda_1, \lambda_N\}$ is an eigenvalue of $A(\mathcal{P})$, then $|\lambda| = q^{(l-1)/2}$ or $\lambda = 0$. It is straightforward to see that $\lambda(\mathcal{P}) = 0$ is impossible, and therefore $\lambda(P) = q^{(l-1)/2}$, as required. □

The reason for considering $\lambda(\mathcal{P})$ is that it is strongly connected to the pseudorandomness properties of $\mathcal{P}$, in the following sense. For any hypergraph $\mathcal{H}$, we can define the matrix $A(\mathcal{H})$ and let $\lambda(\mathcal{H})$ denote the maximum absolute value of all but the largest and smallest eigenvalues of $A(\mathcal{H})$. If $\mathcal{H}$ is a hypergraph and $S \subset \mathcal{H}$ and $T \subset V(\mathcal{H})$, let $e(S,T)$ denote the number of pairs $(v,e) \in T \times S$ such that $v \in e$. For completeness, we give a proof of the following lemma (see [13, 15] for further details).

Lemma 3.5. *Let $\mathcal{H}$ be a hypergraph for which $A(\mathcal{H})$ has row and column sums equal to a and b respectively. Then for any $S \subset \mathcal{H}$ and $T \subset V = V(\mathcal{H})$,*

$$\left| e(S,T) - \frac{a}{|V|}|S||T| \right| \leqslant \lambda(\mathcal{H})\sqrt{|S||T|}.$$

Proof. Let $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_N$ be the eigenvalues of $A(\mathcal{H})$. Let χ_S and χ_T denote the characteristic vectors of S and T. Let $x_1, x_2, \ldots, x_N$ be an orthonormal basis of eigenvectors, where x_i is the eigenvector corresponding to λ_i, and let

$$\chi_S = \sum_{i=1}^{N} s_i x_i, \quad \chi_T = \sum_{i=1}^{N} t_i x_i.$$

We may express $e(S,T)$ in linear algebra terms:

$$e(S,T) = \langle A\chi_S, \chi_T \rangle = \lambda_1 s_1 t_1 + \lambda_N s_N t_N + \sum_{i=2}^{N-1} \lambda_i s_i t_i.$$

The values of s_1, t_1, s_N and t_N are recovered from the knowledge of the first and last eigenvectors, x_1 and x_N, as given by Lemma 3.3. Noting that $\|\chi_S\|^2 = |S|$ and $\|\chi_T\|^2 = |T|$, and using $\lambda_1 = \sqrt{ab}$ and $\lambda_2 = -\sqrt{ab}$, it is straightforward to see

$$e(S,T) = \frac{a}{|V|}|S||T| + \sum_{i=2}^{N-1} \lambda_i s_i t_i.$$

Finally, by Cauchy–Schwarz,

$$\sum_{i=2}^{N-1} \lambda_i s_i t_i \leqslant \lambda(\mathcal{H}) \left(\sum_{i=1}^{N} s_i^2 \right)^{1/2} \left(\sum_{i=1}^{N} t_i^2 \right)^{1/2}$$

and the sums are $\|\chi_S\| = \sqrt{|S|}$ and $\|\chi_T\| = \sqrt{|T|}$ respectively. □

We now apply this lemma in the case $\mathcal{H} = \mathcal{P}$ to prove Lemma 3.2.

Proof of Main Lemma. Fix $\varepsilon > 0$ and let

$$S = S(\varepsilon) = \{G_f \in \mathcal{P} : |G_f \cap U| < (1-\varepsilon)|U|/q\}.$$

Suppose $|S| = \delta q^l$. According to the preceding lemma with $\mathcal{H} = \mathcal{P}$ and $T = U$, together with Lemma 3.4 we obtain

$$\left| e(S,U) - \frac{1}{q}|S||U| \right| \leqslant q^{\frac{l-1}{2}} \sqrt{|S||U|}.$$

https://doi.org/10.1017/S0963548312000557 Published online by Cambridge University Press

In particular,

$$e(S,U) \geqslant \frac{1}{q}|S||U| - q^{\frac{l-1}{2}}\sqrt{|S||U|} \geqslant (\delta|U| - \sqrt{\delta|U|q})q^{l-1}.$$

On the other hand, by definition of S,

$$e(S,U) < \frac{(1-\varepsilon)}{q}|S||U| \leqslant (1-\varepsilon)\delta q^{l-1}|U|.$$

Comparing the bounds, we get

$$\varepsilon^2\delta|U| < q.$$

Since $|U|/q \to \infty$, and $\varepsilon > 0$ is fixed, we conclude $\delta \to 0$ as $q \to \infty$. This is valid for any $\varepsilon > 0$, so we conclude $|S(\varepsilon)| = o(q^l)$ for all $\varepsilon > 0$, which proves the lemma. □

3.2. Proof of Theorem 1.1

We now use Lemma 3.2 combined with Lemma 2.1 to prove Theorem 1.1 for the random hypergraph $\mathcal{H} = \mathcal{H}(q,r,l)$. According to Lemma 2.1, to prove Theorem 1.1 it is sufficient to show that for any $\beta > 0$ and for any set $U \subseteq V$ with

$$|U| = u := \lfloor \beta n^{\frac{r-l}{r-1}}(\log n)^{\frac{1}{r-1}} \rfloor,$$

we have

$$-\log \mathbb{P}(\mathcal{H}[U] = \emptyset) \gtrsim \frac{\beta^r}{(r)_l} n^{\frac{r-l}{r-1}}(\log n)^{\frac{r}{r-1}}. \tag{3.2}$$

Here $n = q^2$ and the theorem imposes the condition $r \geqslant 2l-1 \geqslant 3$. Define $U_f = \pi_f^{-1}(G_f \cap U)$, so $\mathcal{H}[U] = \emptyset$ if and only if $\mathcal{H}_q[U_f] = \emptyset$ for all G_f. Now according to Lemma 3.2, $|U_f| \sim |U|/q$ for $q^l - o(q^l)$ of the edges $G_f \in \mathcal{P}$ provided $|U|/q \to \infty$. Since $r \geqslant 2l-1$, it is indeed the case that

$$\frac{|U|}{q} = \frac{\lfloor \beta q^{\frac{2(r-l)}{r-1}}(\log q^2)^{\frac{1}{r-1}} \rfloor}{q} \to \infty,$$

so Lemma 3.2 does apply. By construction, the events $\{\mathcal{H}_q[U_f] = \emptyset\}$ are independent over all $G_f \in \mathcal{P}$. This implies that

$$\mathbb{P}(\mathcal{H}[U] = \emptyset) = \prod_{G_f \in \mathcal{P}} \mathbb{P}(\mathcal{H}_q[U_f] = \emptyset).$$

Also, if we fixed $|U_f| = u$, then U_f has the uniform distribution among all sets of size u in $V(\mathcal{H}_q)$. By Lemma 3.1, applied to each $T = U_f$,

$$\begin{aligned} -\log \mathbb{P}(\mathcal{H}[U] = \emptyset) &= \sum_{G_f \in \mathcal{P}} -\log \mathbb{P}(\mathcal{H}_q[U_f] = \emptyset) \\ &\geqslant \sum_{|U_f| \sim |U|/q} -\log \mathbb{P}(\mathcal{H}_q[U_f] = \emptyset) \\ &\gtrsim \frac{\binom{q}{l}}{(q)_r \binom{r}{l}} \sum_{|U_f| \sim |U|/q} (|U_f|)_r \end{aligned}$$

https://doi.org/10.1017/S0963548312000557 Published online by Cambridge University Press

$$\sim \frac{q^{l-r}}{(r)_l} \cdot (q^l - o(q^l)) \cdot \left(\frac{|U|}{q}\right)^r$$
$$\sim \frac{\beta^r}{(r)_l} n^{\frac{r-l}{r-1}} (\log n)^{\frac{r}{r-1}}.$$

Now Lemma 2.1 shows that almost surely as $q \to \infty$, $\alpha(\mathcal{H}) \lesssim A(r,n,l)$. This completes the proof of Theorem 1.1.

3.3. Complete (n,r,l)-systems for $l = 2$ and $l = 3$

The construction of Section 3 based on the polynomial system $\mathcal{P}$ gives an asymptotically complete (n,r,l)-system with low independence number. In this section, we show that for $l = 2$ and $r-1$ a prime power, we can construct complete $(n,r,2)$-systems with independence number asymptotic to $A(n,r,2)$ as $n \to \infty$. For $l = 3$ and r a prime power, we can construct complete $(n,r,3)$-systems with independence number asymptotic to $A(n,r,3)$ as $n \to \infty$.

For $l = 2$, we let $\mathcal{P}$ be a projective plane of order q instead of the polynomial system described in the last section. Now $\mathcal{P}$ is a complete $(q^2+q+q, q+1, 2)$-system, and the matrix $A(\mathcal{P})$ has row and column sums equal to $q+1$ and $\lambda_1 = q+1$, $\lambda_2 = -(q+1)$, and it is well known that $\lambda(\mathcal{P}) = \sqrt{q}$ (for example, see [13] or follow the proof of Lemma 3.4 for details). If we then choose $\mathcal{H}_q$ to be a complete $(q+1,r,2)$-system, then the resulting r-uniform hypergraph $\mathcal{H}$, after randomly 'filling in' each line of $\mathcal{P}$ with $\mathcal{H}_q$, is a *complete* $(q^2+q+1,r,2)$-system. However, this requires the simultaneous existence of a projective plane $\mathcal{P}$ of order q and complete $(q+1,r,2)$-system $\mathcal{H}_q$. The choice of q as an odd power of $r-1$ ensures that the projective plane $\mathcal{P}$ exists, since $r-1$ is a prime power. Wilson's theorem [24] states that a complete $(q+1,r,2)$-system exists for all large enough q such that $q \equiv 0 \bmod r-1$ and $q(q+1) \equiv 0 \bmod r(r-1)$. The choice of q as an odd power of $r-1$ ensures that both of these congruences are satisfied, and if q is large enough, we are done. In particular, this gives for infinitely many n a Steiner triple system on n vertices with independence number asymptotic to $\sqrt{3n\log n}$ as $n \to \infty$.

For $l = 3$ and $r \geqslant 5$, one could consider using an *inversive plane* $\mathcal{P}$ of order q which (among other properties) is a complete $(q^2+1, q+1, 3)$-system, instead of the polynomial system (see [6]). Again the eigenvalue computations could be repeated for an inversive plane, and in each circle of the inversive plane one inserts a complete $(q+1,r,3)$-system. However, for $r \geqslant 5$, there are no necessary and sufficient conditions on q and r for such systems to exist. Infinite families of complete $(n,r,3)$-systems are known to exist (see p. 67 in [5]), and similar computations could be carried out as for the case $l = 2$. However, we do not discuss the technical details here. We do, however, mention a very simple construction: if $n = q^2+1$, where $q = p^{2^k}$ for some prime power p and $k \geqslant 0$, there exists a complete $(q+1,p,3)$-system and a complete $(n,q+1,3)$-system. Applying the method of Theorem 1.1, this yields an $(n,p,3)$-system $\mathcal{H}_n$ for any prime power $p \geqslant 5$ and $n \in \{p^2+1, p^4+1, p^8+1, \ldots\}$, such that $\alpha(\mathcal{H}_n) \sim A(n,p,3)$ as $n \to \infty$. Since only finitely many complete (n,r,l)-systems are known when $r > l > 3$, the cases $l > 3$ seem much more challenging due to this key obstruction. In general, the method works well whenever there is an (n,q,l)-system and a (q,p,l)-system to produce a random (n,p,l)-system with low independence number.

References

[1] Ajtai, M., Komlós, J., Pintz, J., Spencer, J. and Szemerédi, E. (1982) Extremal uncrowded hypergraphs. *J. Combin. Theory Ser. A* **32** 321–335.

[2] Alon, N., Mellinger, K., Mubayi, D. and Verstraëte, J. (2011) The de Bruijn–Erdős theorem for hypergraphs. Submitted.

[3] Alon, N. and Spencer, J. (2000) *The Probabilistic Method*, second edition, Wiley.

[4] Caro, Y. and Tuza, Z. (1991) Improved lower bounds on k-independence. *J. Graph Theory* **15** 99–107.

[5] Colbourn, C. and Dinitz, J., eds (1996) *The CRC Handbook of Combinatorial Designs*, CRC Press.

[6] Dembowski, P. (1996) *Finite Geometries*, Springer. Reprint of 1968 edition.

[7] Duke, R., Lefmann, H. and Rödl, V. (1995) On uncrowded hypergraphs. *Random Struct. Alg.* **6** 209–212.

[8] Eustis, A. and Verstraëte, J. (2012) Independent sets in randomized construction of Steiner $(n,r,r-1)$-systems. Preprint.

[9] Erdős, P. and Hanani, H. (1963) On a limit theorem in combinatorial analysis. *Publ. Math. Debrecen* **10** 10–13.

[10] Frieze, A. and Mubayi, D. (2008) On the chromatic number of simple triangle-free triple systems. *Electron. J. Combin.* **15** R121.

[11] Frieze, A. and Mubayi, D. (2008) Coloring simple hypergraphs. Preprint.

[12] Füredi, Z. (1991) Maximal independent subsets in Steiner systems and in planar sets. *SIAM J. Discrete Math.* **4** 196–199

[13] Haemers, W. (1980) Eigenvalue techniques in design and graph theory. PhD thesis, Technical University Eindhoven. Math. Centre Tract 121, Mathematical Centre, Amsterdam.

[14] Hajnal A. and Szemerédi, E. (1970) Proof of a conjecture of Erdős. *Combin. Theory Appl.* **2** 601–623.

[15] Keevash, P., Sudakov, B. and Verstraëte, J. (2011) On a conjecture of Erdős and Simonovits: Even Cycles. *Combinatorica.* (accepted).

[16] Kirkman, T. (1847) On a problem in combinations. *The Cambridge and Dublin Mathematical Journal* (Macmillan, Barclay, and Macmillan) **II** 191–204.

[17] Komlós, J., Pintz, J. and Szemerédi, E. (1982) A lower bound for Heilbronn's problem. *J. London Math. Soc.* (2) **25** 13–24.

[18] Kostochka, A., Mubayi, D. and Verstraëte, J. (2011) On independent sets in hypergraphs. *Random Struct. Alg.* (accepted).

[19] Krivelevich, M. and Sudakov, B. (1998) The chromatic numbers of random hypergraphs. *Random Struct. Alg.* **12** 381–403.

[20] Phelps, K. and Rödl, V. (1986) Steiner triple systems with minimum independence number. *Ars Combin.* **21** 167–172.

[21] Rödl, V. (1985) On a packing and covering problem. *Europ. J. Combin.* **6** 69–78.

[22] Rödl, V. and Ŝinajová, V. (1994) Note on independent sets in Steiner systems. *Random Struct. Alg.* **5** 183–190.

[23] Shearer, J. (1983) A note on the independence number of triangle-free graphs. *Discrete Math.* **46** 83–87.

[24] Wilson, R. (1975) An existence theorem for pairwise balanced designs III: Proof of the existence conjecture. *J. Combin. Theory Ser. A* **18** 71–79.

https://doi.org/10.1017/S0963548312000557 Published online by Cambridge University Press