COMBINATORICA
Akadémiai Kiadó – Springer-Verlag

COMBINATORICA **12** (4) (1992) 417–423

ON A PROBLEM OF ERDŐS AND LOVÁSZ: RANDOM LINES IN A PROJECTIVE PLANE*

JEFF KAHN

Received May 1, 1990

Let $n(k)$ be the least size of an intersecting family of k-sets with cover number k, and let $\mathcal{P}_k$ denote any projective plane of order $k-1$.
Theorem. *There is a constant A such that if $\mathcal{H}$ is a random set of $m \geq Ak\log k$ lines from $\mathcal{P}_k$ then* $\Pr(\tau(\mathcal{H})<k) \to 0$ $(k \to \infty)$.
Corollary. *If there exists a $\mathcal{P}_k$ then $n(k)=O(k\log k)$.*
These statements were conjectured by P. Erdős and L. Lovász in 1973.

0. Introduction

An old problem of Erdős and Lovász [2] asks, given a positive integer k, what (roughly) is the least $n=n(k)$ for which there exists an n-member intersecting family of k-sets whose cover number is k? (Recall that a family $\mathcal{H}$ is *intersecting* if its members are pairwise nondisjoint; its *cover number*, $\tau(\mathcal{H})$, is the least size of a set meeting all sets of $\mathcal{H}$. For more on this and many related questions see the excellent survey [3].)

The function $n(k)$ was introduced in [2], where it was shown that

(0.1) $n(k)\geq 8k/3-3$, *and*

(0.2) $n(k)\leq 4k^{3/2}\log k$ *if* $k-1$ *is the order of a projective plane.*

Let $\mathcal{P}_k$ denote any projective plane of order $k-1$ (i.e. having k points on a line). The upper bound (0.2) is an immediate consequence of

Theorem 0. ([2]). *If $\mathcal{H}$ is a random set of $m\geq 4k^{3/2}\log k$ lines from $\mathcal{P}_k$ then with high probability $\tau(\mathcal{H})=k$.*

(That is: $\mathcal{H}$ is chosen uniformly at random from m-subsets of the line set of $\mathcal{P}$; *with high probability* means with probability tending to 1 as $k\to\infty$; throughout this paper log denotes natural logarithm.)

Here we prove, as conjectured (with C in place of 22) in [2],

Theorem 1. *If $\mathcal{H}$ is a random set of $m \geq 22k\log k$ lines of $\mathcal{P}_k$ then with high probability $\tau(\mathcal{H})=k$.*

Corollary. *$n(k)=O(k\log k)$ provided there exists a $\mathcal{P}_k$.*

AMS subject classification code 1991: 05 B 40, 05 C 65, 05 D 05, 51 E 15

* Supported in part by NSF–DMS87–83558 and AFOSR grants 89–0066, 89–0512 and 90–0008

Of course $\tau(\mathcal{H}) \leq k$ for *any* set $\mathcal{H}$ of lines of $\mathcal{P}=\mathcal{P}_k$, and equality at least requires that

(0.3) *each point of* $\mathcal{P}$ *is on at least two lines of* $\mathcal{H}$.

It is probably true that if one randomly chooses lines $\ell_1, \ell_2, \ldots$ from $\mathcal{P}$ then with high probability $\tau(\{\ell_1, \ldots, \ell_t\}) = k$ as soon as $\mathcal{H} := \{\ell_1, \ldots, \ell_t\}$ satisfies (0.3) (this happens when t is about $3k \log k$), but we do not see how to prove this.

As for lower bounds, remarkably, nothing is known beyond (0.1). (As mentioned in [2] and again in [1], even $n(k) > 3k$ does not seem easy.) Erdős (e.g. [1]) currently offers \$500 for a proof or disproof of $n(k) = O(k)$.

Curiously, if $n(k) = O(k)$ then the best examples must be quite different from those considered here. Recent results of the author [4] imply that, if we add to the conditions "intersecting family of k-sets with cover number k" the requirement that edge intersection sizes in $\mathcal{H}$ be bounded by some $o(k)$, then indeed $|\mathcal{H}|/k \to \infty$. In particular, $|\mathcal{H}| = \Omega(k\sqrt{\log k/\log\log k})$ — probably improvable to $\Omega(k \log k/\log\log k)$ — whenever $\mathcal{H}$ is a subset of the line set of $\mathcal{P}_k$ with $\tau(\mathcal{H}) = k$.

Proof of Theorem 1

A few conventions. We write S and L for the point and line sets of $\mathcal{P}=\mathcal{P}_k$. For X, $Y \subseteq S$, $L(X)$ denotes the set of lines meeting X, $\overline{L}(X)$ the complementary set, and $L(X,Y)$ the set of lines meeting both X and Y. For $x \in S$ we shorten $L(\{x\})$ to $L(x)$, etc., except that when $x \in X$, we use $L(x,X)$ for $L(\{x, X \setminus \{x\}\})$. We set $Q = |S| = |L| = k^2 - k + 1$.

In what follows, A, B, C, D, $E > 0$ and $\delta \in (0,1)$ are constants whose values will be set later. We must show that for suitable A (eventually about 22), if $\mathcal{H}$ is a random subset of L of size $m \geq Ak \log k$, then

(1.1) *with high probability* $\tau(\mathcal{H}) = k$.

We assume throughout that k is large enough to support our assertions.

As in [2], we use the "counting sieve", proving the somewhat stronger

$$\sum \left\{ \Pr(\mathcal{H} \subseteq L(X)) \ : \ X \in \binom{S}{k-1} \right\} = o(1). \tag{1.2}$$

As mentioned in [2], X's for which $\overline{L}(X)$ is very small are easily handled:

$$\sum \left\{ \Pr(\mathcal{H} \subseteq L(X)) \ : \ X \in \binom{S}{k-1},\ |\overline{L}(X)| < k^{3/2} - k \right\} = o(1). \tag{1.3}$$

Proof. (Sketch.) As observed in [2] (see Lemma on p. 625), $|\overline{L}(X)| < k^{3/2} - k$ implies there is some $\ell \in L$ such that $|\ell \setminus X| < k^{1/2}$. Noting that $|\ell \setminus X| = t$ implies $|\overline{L}(X)| \geq t(k-t)$, we find that the left-hand side of (1.3) is less than

$$\sum_{1 \leq t < \sqrt{k}} Q \binom{k}{t} \binom{Q}{t-1} \left(1 - \frac{t(k-t)}{Q}\right)^{Ak \log k},$$

which is $o(1)$ if $A > 3$. ∎

If $|\overline{L}(X)| = k^2/\gamma$ then

$$\Pr(\mathcal{H} \subseteq L(X)) < \left(1 - \frac{1}{\gamma}\right)^{Ak \log k} < e^{-(Ak/\gamma)\log k}.$$

So in view of (1.3), (1.2) follows from

$$\textbf{(1.4)} \qquad \sum_{\gamma < \sqrt{k}+2} \left| \left\{ X \in \binom{S}{k-1} \; : \; |\overline{L}(X)| = \frac{k^2}{\gamma} \right\} \right| e^{-(Ak/\gamma)\log k} = o(1).$$

(Of course we only consider γ for which $k^2/\gamma \in \mathbb{N}$.)

Let $B < A$ constant. If $\gamma \le B$ then the γ^{th} summand in (1.4) is less than

$$\binom{Q}{k-1} e^{-(Ak/\gamma)\log k} < e^{k(\log k + 1) - (Ak/B)\log k}.$$

So (1.4) follows from

(1.5) *For* $B < \gamma < k^{1/2} + 2$,

$$\left| \left\{ X \in \binom{S}{k-1} \; : \; |\overline{L}(X)| = \frac{k^2}{\gamma} \right\} \right| < e^{(Bk/\gamma)\log k}.$$

We assume henceforth that $B < \gamma < k^{1/2} + 2$.

The basic idea of the proof of (1.5) is that when $\overline{L}(X)$ is small, $L(x, X)$ tends to be small for $x \in X$ (most lines are tangent to X), and large for $x \notin X$, the discrepancy between "small" and "large" being great enough that even for an appropriately small, randomly chosen $X_0 \subseteq X$, the value $|L(x, X_0)|$ will usually determine whether x is in X. This is made precise in (1.6)–(1.8) below.

We will need the following more or less standard fact.

Lemma. *If $R \subseteq S$ and $M \subseteq L$ are such that each $x \in R$ is on at most $t < k|M|/Q$ lines of M, then*

$$|R| \le \left(\frac{k|M|}{Q} - t\right)^{-2} (k-1)|M| \left(1 - \frac{|M|}{Q}\right).$$

Proof. Writing $d_M(x)$ for the number of lines of M containing x, we have

$$\sum_{x \in S} d_M(x) = k|M|,$$

and

$$\sum_{x \in S} d_M(x)(d_M(x) - 1) = |M|(|M| - 1),$$

whence a little calculation gives

$$|R| \left(\frac{k|M|}{Q} - t\right)^2 \le \sum_{x \in S} \left(d_M(x) - \frac{k|M|}{Q}\right)^2 = (k-1)|M| \left(1 - \frac{|M|}{Q}\right). \qquad \blacksquare$$

Given $X \in \binom{S}{k-1}$ with $|\overline{L}(X)| = k^2/\gamma$, set

$$Z = Z(X) = \{x \in X \ : \ |L(x,X)| \geq Ck/\gamma\}$$
$$V = V(X) = \{u \notin X \ : \ |L(u,X)| \leq \delta k\}.$$

(1.6) $|Z| < 2k/C.$

(1.7) $$|V| < (1+o(1))\frac{k}{\gamma}\left(1-\frac{1}{\gamma}\right)(1-\frac{1}{\gamma}-\delta)^{-2}.$$

Proof of (1.6). For $x \in X$ set $s(x) = |L(x,X)|$. We have

$$Q - k^2/\gamma = |L(X)| \leq \sum_{x\in X}(k - s(x)) + \frac{1}{2}\sum_{x\in X} s(x)$$
$$= k^2 - k - \frac{1}{2}\sum_{x\in X} s(x).$$

A little rearranging gives

$$2k^2/\gamma > 2(k^2/\gamma - 1) \geq \sum_{x\in X} s(x) \geq |Z|\frac{Ck}{\gamma},$$

and (1.6) follows. ∎

Proof of (1.7). Apply the lemma with $R = V$, $M = L(X)$ and $t = \delta k$. ∎

(1.8) *If $\delta D > C$ and k is sufficiently large, then there exists $X_0 \subseteq X$ with $|X_0| = \lfloor Dk/\gamma \rfloor$ and*

(1.9) $|L(u,X_0)| > Ck/\gamma$ *for all* $u \in S \setminus (V \cup X)$.

Proof. For $u \in S\setminus(V\cup X)$ let $\{\ell_1,\ldots,\ell_m\} \subseteq L(u,X)$ where $m = \lceil \delta k \rceil$, and let $x_i \in \ell_i \cap X$. Then $|L(u,X_0)| \geq |\{x_1,\ldots,x_m\}\cap X_0|$ (for any X_0). Now take X_0 random of size $\lfloor Dk/\gamma \rfloor$ from X. The expected value of $|\{x_1,\ldots,x_m\}\cap X_0|$ is $m\lfloor Dk/\gamma\rfloor/|X| \sim \delta Dk/\gamma$. Since $C < \delta D$ (and these quantities are constants), (1.8) follows from standard large deviation results. ∎

Suppose that with each X as above we associate some $X_0 := \varphi(X)$ as in (1.8). Of course,

(1.10) *there are at most* $\binom{Q}{\lfloor Dk/\gamma \rfloor}$ *possibilities for* X_0

(as X varies), a number compatible with an upper bound $e^{O((k/\gamma)\log k)}$. We accordingly fix X_0 and try to bound $|\varphi^{-1}(X_0)|$.

Let

$$U = \{u \in S \setminus X_0 \ : \ |L(u,X_0)| \leq Ck/\gamma\}.$$

If $\varphi(X)=X_0$, then

$$|U \setminus X| \le |V(X)| < \varepsilon k$$

where

$$\varepsilon = (1+o(1))\gamma^{-1}\left(1-\frac{1}{\gamma}\right)\left(1-\frac{1}{\gamma}-\delta\right)^{-2}$$

(by (1.9) and (1.7)). Thus $|U|<(1+\varepsilon)k$ and

(1.11) *there are at most*

$$\sum_{i\le\varepsilon k}\binom{\lfloor(1+\varepsilon)k\rfloor}{i} \quad \textit{possibilities for} \quad U\setminus X, \quad \textit{or equvalently, for} \quad U\cap X.$$

We now fix $U\cap X$ and estimate the number of ways to choose $X\setminus U$. Set $T=(U\cap X)\cup X_0$ and $W=\{u\notin T : |L(u,T)|<k/E\}$.

(1.12) $|X\setminus(W\cup T)|<Ek/\gamma$.

Proof. This follows from

$$Q-k^2/\gamma = |L(X)| \le k|X| - (k/E)|X\setminus(W\cup T)|. \qquad ∎$$

Thus

(1.13) *there are at most* $\displaystyle\sum_{i<Ek/\gamma}\binom{Q}{i}$ *choices for* $X\setminus(W\cup T)$.

Set $\beta=1/E+2/C$ and $\alpha=1-\beta-\sqrt{(1-\beta)^2-2/\gamma}$. (For large γ, α will be about $(1-\beta)^{-1}\gamma^{-1}$.)

(1.14) *If*

$$(1-\beta)^2-2/\gamma > (2k)^{-2}, \tag{1.15}$$

then $|W\setminus X|\le\alpha k$.

Proof. Notice that $|X\setminus T|<2k/C$ (by (1.6) since $U\supseteq X\setminus Z \Rightarrow T\supseteq X\setminus Z \Rightarrow X\setminus T\subseteq Z$). It follows that for $w\in W$,

$$|L(w,X)| \le |L(w,T)| + |X\setminus T| < \beta k.$$

Now if $\{w_0,\dots,w_{m-1}\}\subseteq W\setminus X$ we have

$$\begin{aligned} k^2/\gamma &= |\overline{L}(X)| \\ &\ge |L(\{w_0,\dots,w_{m-1}\})\setminus L(X)| \\ &= \sum_{i=0}^{m-1}|L_i(w_i)\setminus L(X\cup\{w_0,\dots,w_{i-1}\})| \\ &> \sum_{i=0}^{m-1}((1-\beta)k-i) \\ &> (1-\beta)km-\frac{m^2}{2}. \end{aligned}$$

Since this holds for every $m \le |W\backslash X|$ it follows that $|W\backslash X| \le \alpha k$ where α, as above, is the smaller root of

$$\frac{1}{2}x^2 - (1-\beta)x + \frac{1}{\gamma} = 0.$$

(Note the hypothesis (1.15) guarantees the existence of $m \in \mathbb{N}$ for which $\frac{m^2}{2} - (1-\beta)km + \frac{k^2}{\gamma} < 0$.)

Since also $|W \cap X| \le |X \backslash T| < 2k/C$ we have

(1.16) *If* (1.15) *holds then there are at most* $\sum_{i \le \alpha k} \binom{\lfloor(\alpha+2/C)k\rfloor}{i}$ *choices for* $W\backslash X$, *or, equivalently, for* $W \cap X$.

In sum, the number of possibilities for $X \in \binom{S}{k-1}$ with $|\overline{L}(X)| = k^2/\gamma$ is at most the product of the bounds in (1.10), (1.11), (1.13) and (1.16), namely,

$$(1.17) \qquad \binom{Q}{\lfloor \frac{Dk}{\gamma} \rfloor} \sum_{i \le \varepsilon k} \binom{\lfloor(1+\varepsilon)k\rfloor}{i} \sum_{i < Ek/\gamma} \binom{Q}{i} \sum_{i \le \alpha k} \binom{\lfloor(\alpha+2/C)k\rfloor}{i}$$

(again, assuming $\delta D > C$ and (1.15) holds). It is now easy, using the estimate $\binom{b}{a} < e^{a(\log(b/a)+1)}$, to choose parameters so that this product is at most, say,

$$e^{(21.5+o(1))(k/\gamma)\log k}.$$

For large γ the log of the product (1.17) is asymptotically at most

$$\frac{k}{\gamma}\left[(D+E)\log k + \left(D + (1-\delta)^{-2} + E + (1-\beta)^{-1}\right)\log\gamma\right]$$
$$\le \frac{k}{\gamma}\log k\left[\frac{3}{2}(D+E) + \frac{1}{2}(1-\delta)^{-2} + \frac{1}{2}(1-1/E-2/C)^{-1}\right]$$

(using $\gamma < \sqrt{k}+2$). Then if $C=4$, $D=7$, $E=3$ and $\gamma=3/5$ (not the best possible values), the expression of brackets of $21\frac{1}{8}$.

For small γ ($\gamma = k^{o(1)}$ is enough), things are even easier. Here the log of (1.17) is just

$$(1+o(1))(D+E)(k/\gamma)\log k.$$

(This requires $E/\gamma < 1/3$ (say) — so that the third factor in (1.17) is less than $\binom{Q}{\lceil Ek/\gamma\rceil}$ — which will be true since we will have $(\gamma >)B > 3E$), and we can easily choose legal parameters with $D+E<21$.

So (with $A=22$) we have Theorem 1. ∎

Added in proof: The author recently proved $n(k) = O(k)$.

References

[1] P. Erdős: On the combinatorial problems I would most like to see solved, *Combinatorica* **1** (1981), 25–42.

[2] P. Erdős, and L. Lovász: Problems and results on 3-chromatic hypergraphs and some related questions, in: *Infinite and Finite Sets* (Proc. Colloq. Math. Soc. J. Bolyai 10, Keszthely, Hungary, 1973), A Hajnal et. al. (eds.), North Holland, Amsterdam, 1975, 609–627.

[3] Z. Füredi: Matching and covers in hypergraphs, *Graphs and Combinatorics* **4** (1988), 115–206.

[4] J. Kahn: On a theorem of Frankl and Rödl, in preparation.

Jeff Kahn

Department of Mathematics
and Center for O. R.
Rutgers University, New Brunswick
NJ 08903, U.S.A.
jkahn@math.rutgers.edu