PACIFIC JOURNAL OF MATHEMATICS
Vol 118, No 2, 1985

ASYMPTOTICALLY GOOD COVERINGS

JOEL SPENCER

Dedicated to the memory of Ernst Straus

The Erdős-Hanani conjecture is that for fixed $r < k$ and n large there exists a covering of all r-sets of an n-set by a family of k-sets whose cardinality is asymptotic (in n) to the "counting" lower bound. This conjecture was first proven by Rodl, here we give a more direct argument. We use probabilistic methods, selecting k-sets in large groups, and showing that the hypergraph of uncovered r-sets retains a property we call quasirandomness, meaning that it has the essential (for us) properties of random hypergraph.

0. Introduction. Let $r < k \leq n$ and set $[n] = \{1,\ldots,n\}$, a generic n-set. The covering function $M(n, k, r)$ is defined as the minimal cardinality of a family F of k-sets of $[n]$ such that every r-set of $[n]$ is contained in some $K \in F$. The packing function $m(n, k, r)$ is the maximal cardinality of a family F of k-sets of $[n]$ such that no r-set of $[n]$ is contained in more than one $K \in F$. Elementary counting arguments imply

$$m(n, k, r) \leq \binom{n}{r} \Big/ \binom{k}{r} \leq M(n, k, r) \tag{1}$$

Equality holds if and only if there exists an (n, k, r) tactical configuration —i.e., a collection F of k-sets containing every r-set exactly once. The existance of tactical configurations for various r, k, n (e.g. for $k = 3$, $r = 2$ —Steiner Triple Systems) is a central question of Combinatorial Analysis to which we here do not directly contribute.

In 1963 Paul Erdős and Haim Hanani **[1]** conjectured that for all $r < k$ the inequalities (1) are asymptotically equalities—more precisely, that

$$\lim_{n\to\infty} m(n, k, r)\binom{k}{r} \Big/ \binom{n}{r} = 1 = \operatorname{Lim} M(n, k, r)\binom{k}{r} \Big/ \binom{n}{r}. \tag{2}$$

This was proven for $r = 2$, all k and for $r = 3$, $k = p$ or $p + 1$ where p is a prime power. They also showed that either of the equalities (2) imply the other. These inequalities became known as the Erdős-Hanani Conjecture. In 1983 this conjecture was resolved affirmatively by Vojtech Rodl **[2]** for all values $r < k$. In this paper we present a more direct proof of the

Erdős-Hanani Conjecture. Our argument is based on Rodl's original proof and on personal discussions with Rodl which we gratefully acknowledge.

1. An intuitive view. In this section we present an informal discussion of our proof of the Erdős-Hanani Conjecture. The formal proof is given in the next section.

Let $r < k$ be fixed and let n be very large. Let G_0 be the complete r-graph on vertex set $[n]$. Let δ be a very small positive real. (δ is fixed first and then n is made very large.) Let F_0 be a random collection of $\delta\binom{n}{r}/\binom{k}{r}$ k-cliques from G_0 and let G_1 be the family of r-sets not contained in any $K \in F_0$. As F_0 is δ times the size of a perfect covering of G_0 (if one existed) it would, if there were no overlap, cover a proportion δ of the r-sets in G_0. In fact, overlap is the critical consideration. The typical r-set is covered an average of δ times by F_0. There are many k-sets covering a given r-set and each has only a small chance of being placed in F_0. "Thus" the number of k-sets of F_0 covering a given r-set is given by a Poisson distribution with mean δ. That is, $\delta e^{-\delta}$ of the r-sets are covered exactly once, $(\delta^2/2)e^{-\delta}$ are covered exactly twice, $(\delta^i/i!)e^{-\delta}$ are covered exactly i times and $e^{-\delta}$ are not covered at all and "remain" in G_1. When δ is very small the proportion of r-sets covered twice or more, roughly $\delta^2/2$, is a negligible proportion of the proportion of r-sets, roughly δ, that are covered once. That is, F_0 is an excellent, though not perfect, cover of $G_0 - G_1$.

We continue the procedure with G_1. We choose F_1 from among the k-cliques of G_1. This is essential as we do not want any of the $\binom{k}{r}$ r-sets covered by a $K \in F_1$ already covered by F_0. We pick F_1 randomly, choosing the cardinality so that if there were no overlap a proportion δ of G_1 would be covered. We let G_2 be the remaining r-sets—those covered by no $K \in F_1$. Once again (but see below) the number of k-sets covering an r-set of G_1 is given by a Poisson distribution with mean δ and F_1 is an excellent covering of $G_1 - G_2$.

We iterate this procedure—given G_i we find F_i and set G_{i+1} equal to the remaining r-sets—until we reach a G_t with a negligible proportion of r-sets. As each $|G_{i+1}| \sim e^{-\delta}|G_i|$ we let t be large enough so that $e^{-t\delta}$ is very small. At this point the remaining r-sets are covered one by one. Though this is very wasteful (we want k-sets K to cover $\binom{k}{r}$ new r-sets but here we use one k-set to cover one r-set) it is acceptible since $|G_t|$ is small. With δ and $e^{-t\delta}$ very small the total covering has a very small proportion of waste.

To employ this method it is necessary that the G_i retain certain regularity properties. (To illustrate with an extreme case, if a G_i was created that had no k-cliques we could not continue.) Let G be an r-graph with density ρ. We call G quasirandom if for every edge $e \in G$ the proportion of k-sets covering e that are cliques in G is roughly $\rho^{\binom{k}{r}-1}$. (Note that this would be the appropriate proportion for a random graph of density ρ.) The central lemma of §3 states, roughly, that if G is quasirandom then the above method may be employed to find a family of k-sets F so that the remaining graph G^* is also quasirandom. The initial graph—the complete G_0—is certainly quasirandom with unit density. We may thus iterate our procedure—finding a descending sequence of hypergraphs G_t, all of which are quasirandom. One final parameter—to quantify the word "roughly" we say G is quasirandom with tolerance ε if the quasirandom properties hold within a factor of $1 \pm \varepsilon$. When G is quasirandom with tolerance ε the tolerance of the "remaining" G^* will be some higher ε^*. Our lemma allow us to insure that the tolerance remains arbitrarily small even after our procedure has been applied a fixed number t times.

3. The proof. *Throughout this section $2 \leq r < k$ shall be fixed integers.* The term graph shall refer to r-graph (i.e. a collection of r-sets) and edge shall refer to an r-set in the collection. A k-set K is a clique in graph G if $e \in G$ for every edge $e \subset K$. All graphs shall have n vertices.

Special Notation. The term $1 \pm \varepsilon$ refers to a number x satisfying
$$1 - \varepsilon \leq x \leq 1 + \varepsilon.$$
Thus $a = b(1 \pm \varepsilon)$ means
$$b(1 - \varepsilon) \leq a \leq b(1 + \varepsilon).$$
In the Lemma below, $\varepsilon < .01$. Thus, for example,
$$(1 \pm \varepsilon)(1 \pm \varepsilon) = (1 \pm 3\varepsilon) \tag{3}$$
since if $x \leq 1 + \varepsilon$ and $y \leq 1 + \varepsilon$ then $xy \leq (1 + \varepsilon)^2 < 1 + 3\varepsilon$ and similarly $x, y \geq 1 - \varepsilon$ imply $xy \geq 1 - 3\varepsilon$. More generally
$$(1 \pm a\varepsilon)(1 \pm b\varepsilon) = 1 \pm (a + b + 1)\varepsilon \tag{4}$$
for any $1 \leq a, b \leq 10$. Also
$$e^{\pm a\varepsilon} = 1 \pm (a + 1)\varepsilon \tag{5}$$
if $a \leq 5$ as $1 - (a + 1)\varepsilon < e^{-a\varepsilon}$ and $e^{a\varepsilon} < 1 + (a + 1)\varepsilon$ with a so small. These "tolerance estimates" shall often be used tacitly in the proof of the lemma.

DEFINITION. G is *quasirandom* with density ρ and tolerance ε if

(a) G has $\rho\binom{n}{r}(1 \pm \varepsilon)$ edges

(b) Every edge of G lies in $\rho^{\binom{k}{r}-1}\binom{n-r}{k-r}(1 \pm \varepsilon)$ cliques of size k.

LEMMA. *Let* ρ, ε^*, $\delta > 0$. *Then there exist* $\varepsilon > 0$ *and* n_0 *so that for* $n > n_0$ *the following holds. Let* G *be quasirandom with density* ρ *and tolerance* ε. *The ther exists a family* F *of* k*-cliques of* G *such that*

(i) $|F| = \delta[\rho\binom{n}{r}/\binom{k}{r}](1 \pm \varepsilon^*)$

and so that, letting G^* *be the subgraph of* G *remaining after the deletion of all* k*-cliques* $K \in F$, G^* *is quasirandom with density* $\rho e^{-\delta}$ *and tolerance* ε^*.

In particular

(ii) G^* *has* $\rho e^{-\delta}\binom{n}{r}(1 \pm \varepsilon^*)$ *edges.*

(iii) *Every edge of* G^* *lies in* $(\rho e^{-\delta})^{\binom{k}{r}-1}\binom{n-r}{k-r}(1 \pm \varepsilon^*)$ k*-cliques of* G^*.

Proof. Set

$$p = \delta \Big/ \left[\rho^{\binom{k}{r}-1}\binom{n-r}{k-r}\right] \tag{6}$$

Let F be a *random* collection of k-cliques of G given by placing each k-clique K of G into F with independent probability p. That is,

$$\Pr[K \in F] = p$$

and the events "$K \in F$" are mutually independent over all k-cliques K of G.

For definiteness we fix $\varepsilon > 0$ satisfying

$$10\varepsilon\binom{k}{r} < \varepsilon^*, \qquad \binom{k}{r}\varepsilon < 10^{-4}. \tag{7}$$

(We may think, however, of the tolerance ε of G as being "much much smaller" than the tolerance ε^* required of G^*.)

As G is quasirandom the number of k-cliques of G is

$$\left[\rho\binom{n}{r}(1 \pm \varepsilon)\right]\left[\rho^{\binom{k}{r}-1}\binom{n-r}{k-r}(1 \pm \varepsilon)\right]\Big/\binom{k}{r} = \rho^{\binom{k}{r}}\binom{n}{k}(1 \pm 3\varepsilon). \tag{8}$$

Thus $|F|$ has a binomial distribution with mean

$$p\rho^{\binom{k}{r}}\binom{n}{r}(1 \pm 3\varepsilon) = \left[\delta\rho\binom{n}{r}\Big/\binom{k}{r}\right](1 \pm 3\varepsilon). \tag{9}$$

Chebychev's inequality implies that for any $c > 0$ the probability that $|F| = E[|F|]\,(1 \pm c)$ approaches unity with n. Hence

$$|F| = \left[\delta\rho\binom{n}{r}\Big/\binom{k}{r}\right](1 \pm 4\varepsilon) \tag{10}$$

and so (i) is satisfied, with probability approaching unity with n.

We now consider (ii). For each edge $e \in G$ let $\operatorname{cov}(e)$ denote the number of k-cliques K of G which contain e. Then

$$\Pr(e \in G^*) = (1 - p)^{\operatorname{cov}(e)} \tag{11}$$

as e "survives" if and only if none of these K are selected for F. As p approaches zero with n we may bound

(12) $$1 - p = e^{-p(1\pm\varepsilon)}$$

so that

(13) $$\begin{aligned}(1-p)^{\operatorname{cov}(e)} &= \exp\left[-p(1\pm\varepsilon)\rho^{\binom{k}{r}-1}\binom{n-r}{k-r}(1\pm\varepsilon)\right]\\ &= \exp[-\delta(1\pm 3\varepsilon)] = e^{-\delta}(1\pm 4\varepsilon)\end{aligned}$$

and the expected number of edges in G^* is

(14) $$e^{-\delta}(1\pm 4\varepsilon)\rho\binom{n}{r}(1\pm\varepsilon) = \rho e^{-\delta}\binom{n}{r}(1\pm 6\varepsilon).$$

To show that $|G^*|$ is nearly always nearly equal to its expectation we bound its variance. For each $e \in G$ let X_e be the indicator random variable for the event "$e \in G^*$" and set

(15) $$X = \sum_{e\in G} X_e$$

so that $X = |G^*|$. Any two distinct $e, e' \in G$ contain at least $r+1$ vertices and therefore there are at most

$$\binom{n-(r+1)}{k-(r+1)} < n^{k-r-1}$$

k-cliques of G containing both of them. We bound

(16) $$\begin{aligned}E[X_eX_{e'}] &\le (1-\rho)^{\operatorname{cov}(e)+\operatorname{cov}(e')-n^{k-r-1}}\\ &\le E[X_e]E[X_{e'}]\left(1+\frac{c}{n}\right)\end{aligned}$$

where c depends only on k, r, δ and ρ. Using general probability methods

(17) $$\operatorname{Var}(X) = \sum_e \operatorname{Var}(X_e) + \sum_{e\neq e'} \operatorname{cov}(X_e, X_{e'}),$$

$$\sum_e \operatorname{Var}(X_e) \le \sum_e E(X_e) = E(X) < n^{-1}E(X)^2,$$

$$\sum_{e\neq e'} \operatorname{cov}(X_e, X_{e'}) \le \left(\frac{c}{n}\right)\sum_{e\neq e'} E(X_e)E(X_{e'}) \le cn^{-1}E(X)^2,$$

so

$$\operatorname{Var}(X) \le (c+1)n^{-1}E(X)^2.$$

(These arguments are quite rough but we only really need $\operatorname{Var}(X) = o(E(X)^2)$.)

Applying Chebychev's inequality

$$X = \rho e^{-\delta}\binom{n}{r}(1 \pm 7\varepsilon) \tag{18}$$

and thus (ii) is satisfied, with probability approaching unity in n.

We now consider (iii). Let e be an edge of G and let $\mathscr{A}_e$ denote the family of k-cliques K of G that contain e. For each $K \in \mathscr{A}_e$ let $\mathscr{S}_K$ denote the family of k-cliques L of G which contain at least one edge of K but do *not* contain e. We define indicator random variables X_K by

$$X_K = \begin{cases} 1, & \text{if } F \cap \mathscr{S}_K = \varnothing, \\ 0, & \text{otherwise.} \end{cases}$$

Then X_K is the indicator random variable for "K is a k-clique of G^*" conditional on "$e \in G^*$". (Conditioning on "$e \in G^*$" is equivalent to assuming that no k-clique L of G which contains e has been placed in F.) Set

$$X = \sum_{K \in \mathscr{A}_e} X_K \tag{19}$$

so that, conditional on $e \in G^*$, X is the number of k-cliques of G^* containing e.

A k-clique $K \in \mathscr{A}_e$ contains $\binom{k}{r} - 1$ edges other than e, each of which lie in $\rho^{\binom{k}{r}-1}\binom{n-r}{k-r}(1 \pm \varepsilon)$ k-cliques L. At most n^{k-r-1} k-cliques (in fact, k-sets) L contain two given distinct edges e, e' (as e, e' have at least $r+1$ points between them) and there are less than k^{2r} such pairs e, e'. Thus

$$\begin{aligned} &\left[\binom{k}{r} - 1\right]\rho^{\binom{k}{r}-1}\binom{n-r}{k-r}(1 \pm \varepsilon) \geq |\mathscr{S}_K| \\ &\quad \geq \left[\binom{k}{r} - 1\right]\rho^{\binom{k}{r}-1}\binom{n-r}{k-r}(1 \pm \varepsilon) - k^{2r}n^{k-r-1}. \end{aligned} \tag{20}$$

We absorb the overlap term $k^{2r}n^{k-r-1}$ into the main term and deduce

$$|\mathscr{S}_K| = \left[\binom{k}{r} - 1\right]\rho^{\binom{k}{r}-1}\binom{n-r}{k-r}(1 \pm 2\varepsilon). \tag{21}$$

thus

$$\begin{aligned} E[X_K] &= (1-\rho)^{|\mathscr{S}_K|} = \exp\left[-\varepsilon\left[\binom{k}{r} - 1\right](1 \pm 3\varepsilon)\right] \\ &= \exp\left[-\delta\left[\binom{k}{r} - 1\right]\right]\left(1 \pm 4\varepsilon\binom{k}{r}\right). \end{aligned} \tag{22}$$

Here we have approximated $(1 - p)$ by $\exp(-p)$ and $\exp(\pm 3\varepsilon\delta\binom{k}{r})$ as

$$1 \pm 4\varepsilon\delta\binom{k}{r} = 1 \pm 4\varepsilon\binom{k}{r}.$$

Summing over all $K \in \mathscr{A}_e$.

(23) $$E(X) = \rho^{\binom{k}{r}-1}\binom{n-r}{k-r}(1 \pm \varepsilon)\exp\left[-\delta\left[\binom{k}{r}-1\right]\right]\left(1 \pm 4\binom{k}{r}\varepsilon\right)$$
$$= \left(\rho e^{-\delta}\right)^{\binom{k}{r}-1}\binom{n-r}{k-r}\left(1 \pm 5\binom{k}{r}\varepsilon\right).$$

Once again we must show X is nearly always nearly equal to its expectation. Our requirements this time are far more stringent since there are cn^r variables X (one for each edge) each of which must be nearly equal its expectation. In fact we shall show that the probability $X \neq E(X)(1 \pm \varepsilon)$ is exponentially small. To do this we shall require a strong sense of mutual independence of the X_K, $K \in \mathscr{A}_e$. Note, however, that when $K', K'' \in \mathscr{A}_e$ intersect in more than e the corresponding $X_{K'}$, $X_{K''}$ are highly correlated. Our first task, then, is to break $\mathscr{A}_e$ into classes in which that does not occur.

We call a subfamily $\mathscr{C} \subseteq \mathscr{A}_e$ neardisjoint if $K' \cap K'' = e$ for all distinct $K', K'' \in \mathscr{C}$. For each $K' \in \mathscr{A}_e$ there are at most k

$$\binom{n}{k-r-1} \leq n^{k-r-1}$$

cliques $K'' \in \mathscr{A}_e$ with $K' \cap K'' = e$ (k choices for $x \in K' - e$, $\binom{n}{k-r-1}$ choices for K'' containing e, x). We partition $\mathscr{A}_e$ into

(24) $$\mathscr{A}_e = \bigcup_{\alpha \in I} \mathscr{C}_\alpha \cup \mathscr{D}$$

(I an index set) where each $|\mathscr{C}_\alpha| = n^{.3}$, each $\mathscr{C}_\alpha$ is neardisjoint, and $|\mathscr{D}| < n^{k-r-.7}$. To do this we pull $\mathscr{C}_\alpha$ from $\mathscr{A}_e$ as long as possible until we get stuck. At that point we have a family $\mathscr{D}$ of remaining sets and a neardisjoint family $\mathscr{C} \subseteq \mathscr{D}$, $|\mathscr{C}| < n^{.3}$, which cannot be extended. There are at most $|\mathscr{C}|n^{k-r-1} < n^{k-r-.7}$ sets $K'' \in \mathscr{D}$ which intersect some $K' \in \mathscr{C}$ in more than e. As this must be all of $\mathscr{D}$, $|\mathscr{D}| < n^{k-r-.7}$.

Let $\mathscr{C} \subseteq \mathscr{A}_e$, $|\mathscr{C}| = n^{.3}$ be neardisjoint and set

(25) $$X_{\mathscr{C}} = \sum_{K \in \mathscr{C}} X_K.$$

Suppose K', K'' are distinct elements of $\mathscr{C}$ and $L \in \mathscr{S}_{K'} \cap \mathscr{S}_{K''}$. Then L contains edges $e' \in K'$. $e'' \in K''$, $e' \neq e \neq e''$. As $K' \cap K'' = e$, $e' \neq e''$ so L contains at least $r + 1$ points from $K' \cup K''$. Thus

(26) $$|\mathscr{S}_{K'} \cap \mathscr{S}_{K''}| < \binom{2k}{r+1}$$

where c depends only on k, r. Set

(27) $$\mathscr{T} = \bigcup_{K', K'' \in \mathscr{C}} \mathscr{S}_{K'} \cap \mathscr{S}_{K''}$$

so that

$$|\mathscr{T}| \le |\mathscr{C}|^2 c n^{k-r-1} < n^{k-r-.39}. \tag{28}$$

For $K \in \mathscr{C}$ define random variables Y_K by

$$Y_K = \begin{cases} 1, & \text{if } F \cap (\mathscr{S}_K - \mathscr{T}) = \varnothing \\ 0, & \text{otherwise.} \end{cases} \tag{29}$$

and define

$$Y_{\mathscr{C}} = \sum_{K \in \mathscr{C}} Y_K. \tag{30}$$

Since the sets $\mathscr{S}_K - \mathscr{T}$ are mutually disjoint the variables Y_K are mutually independent. We require a classic result (for explicit reference see the appendix of **[3]**) on the sum of independent random variables.

Fact. If $Z_1, \ldots, Z_m$ are mutually independent zero-one random variables, $Z = \sum_{i=1}^m Z_i$, and $\alpha > 0$ then

$$\Pr\left[|Z - E(Z)| > \alpha\right] < 2e^{-\alpha^2/m}. \tag{31}$$

Applying this result with $m = |\mathscr{C}| = n^{.3}$, $\alpha = n^{.2}$

$$\Pr\left[|Y_{\mathscr{C}} - E(Y_{\mathscr{C}})| > n^{.2}\right] < 2e^{-n^{.1}}. \tag{32}$$

(This was the critical step as we have the probability exponentially small.) Now we need show that $Y_{\mathscr{C}}$ provides a good approximation to $X_{\mathscr{C}}$. Set

$$W = |F \cap \mathscr{T}|. \tag{33}$$

For all K, $Y_K \le X_K$ and thus $Y_{\mathscr{C}} \le X_{\mathscr{C}}$. If $Y_K = 0$ and $X_K = 1$ then $F \cap \mathscr{S}_K \cap \mathscr{T} \ne \varnothing$. Each $L \in F \cap \mathscr{T}$ lies in at most k families $\mathscr{S}_K$, $K \in \mathscr{C}$. (L must have at least one point in $K - e$, $|L| = k$, and the sets $K - e$, $K \in \mathscr{C}$, are disjoint.) Thus

$$X_{\mathscr{C}} - kW \le Y_{\mathscr{C}} \le X_{\mathscr{C}}. \tag{34}$$

Therefore

$$E\left[|X_{\mathscr{C}} - Y_{\mathscr{C}}|\right] \le kE[W] \le kn^{k-r-.39}p < n^{-.38}. \tag{35}$$

Moreover, W has binomial distribution $B(|\mathscr{T}|, p)$ so

$$\Pr\left[X_{\mathscr{C}} - Y_{\mathscr{C}} > n^{.2}\right] < \Pr\left[W > n^{.2}/k\right] < \left[|\mathscr{T}|p\right]^{n^{.2}/k} < \left[n^{-.38}\right]n^{.2}/k < e^{-n^{.2}} \tag{36}$$

$$|X_{\mathscr{C}} - E(X_{\mathscr{C}})| \le |X_{\mathscr{C}} - Y_{\mathscr{C}}| + |Y_{\mathscr{C}} - E(Y_{\mathscr{C}})| + |E(Y_{\mathscr{C}}) - E(X_{\mathscr{C}})| \tag{37}$$

we combine (32), (35), (36) to derive

$$\Pr\left[|X_{\mathscr{C}} - E(X_{\mathscr{C}})| > 3n^{.2}\right] < 2e^{-n^{.1}}. \tag{38}$$

From the decomposition (24) we decompose X into

$$X = \sum_{\alpha \in I} X_{\mathscr{C}_\alpha} + X_{\mathscr{D}}. \tag{39}$$

Thus

$$|X - E(X)| \leq \sum_{\alpha \in I} \left|X_{\mathscr{C}_\alpha} - E\left(X_{\mathscr{C}_\alpha}\right)\right| + |X_{\mathscr{D}} - E(X_{\mathscr{D}})|. \tag{40}$$

As $0 \leq X_{\mathscr{D}} \leq |\mathscr{D}| < n^{k-r-.7}$ always,

$$|X_{\mathscr{D}} - E(X_{\mathscr{D}})| < n^{k-r-.7} \tag{41}$$

with probability one. Now assume

$$\left|X_{\mathscr{C}_\alpha} - E\left(X_{\mathscr{C}_\alpha}\right)\right| \leq 3n^{.2} \tag{42}$$

for *every* $\alpha \in I$. Summing over $\alpha \in I$, and noting $|I| \leq |\mathscr{A}_e|/n^{.3} \leq n^{k-r-.3}$

$$\sum_{\alpha \in I} \left|X_{\mathscr{C}_\alpha} - E\left(X_{\mathscr{C}_\alpha}\right)\right| \leq 3n^{.2}n^{k-r-.3} \tag{43}$$

so

$$|X - E(X)| < 4n^{k-r-.1}. \tag{44}$$

The probability that this does not occur is at most $|I|(2e^{-n^{.1}}) < n^k e^{-n^{.1}}$. We know from (23) that $E(X) > cn^{k-r}$ where c is a constant dependent on k, r, ρ, δ and ε but not on n. Thus $4n^{k-r-.1} < \varepsilon E(X)$ and so

$$\Pr[X \neq E(X)(1 \pm \varepsilon)] < n^k e^{-n^{.1}} < e^{-n^{.09}} \tag{45}$$

by the dominance of the exponential term. Combining (23) and (45)

$$\Pr\left[X \neq \left(\rho e^{-\delta}\right)^{\binom{k}{r}-1}\binom{n-r}{k-r}\left(1 \pm 6\binom{k}{r}\varepsilon\right)\right] < e^{-n^{.09}}. \tag{46}$$

Recall that X represents the number of k-cliques of G^* containing e, conditional on $e \in G^*$, for a given e. There are less than n^r different e. Thus the probability that *some* $e \in G^*$ does not lie in the appropriate number, i.e.

$$\left(pe^{-\delta}\right)^{\binom{k}{r}-1}\binom{n-r}{k-r}(1 \pm \varepsilon^*),$$

of k-cliques of G^* is bounded from above by $n^r e^{-n^{.09}}$. Once again the exponential term dominates. The probability of (iii) holding approaches infinity.

We return to the beginning of the proof. Having fixed ε we let n_0 be such that for $n > n_0$ conditions (i), (ii), (iii) all hold with probability at least .9. Then with probability at least .7 all three conditions hold simultaneously. Thus there exists a specific F for which all three conditions hold. This completes the proof of the Lemma.

THEOREM. *Let $2 \le r < k$ and $a > 0$ be fixed. Then for n sufficiently large*

$$M(n, k, r) < \left[\binom{n}{r} \Big/ \binom{k}{r}\right](1 + a). \tag{47}$$

Proof. We first select $a > 0$ so that

$$\frac{\delta}{1 - e^{-\delta}} < 1 + a. \tag{48}$$

This may be done as $\operatorname{Lim}_{\delta \to 0} \delta/(1 - e^{-\delta}) = 1$. We then select $\varepsilon > 0$ so that

$$\frac{\delta(1 + \varepsilon)}{1 - e^{-\delta} - 2\varepsilon} < 1 + a \tag{49}$$

which may be done as

$$\operatorname*{Lim}_{\varepsilon \to 0} \frac{\delta(1 + \varepsilon)}{1 - e^{-\delta} - 2\varepsilon} = \frac{\delta}{1 - e^{-\delta}}.$$

We then select a positive integer t so that

$$\frac{\delta(1 + \varepsilon)}{1 - e^{-\delta} - 2\varepsilon} + \binom{k}{r} e^{-t\delta}(1 + \varepsilon) < 1 + a \tag{50}$$

which may be done as

$$\operatorname*{Lim}_{t \to \infty} \binom{k}{r} e^{-t\delta}(1 + \varepsilon) = 0.$$

Set $\varepsilon_t = \varepsilon$. By reverse induction on i we find $\varepsilon_t > \varepsilon_{t-1} > \cdots > \varepsilon_0$ so that the Lemma applies with $\rho = e^{-i\delta}$, $\varepsilon^* = \varepsilon_i$, δ as itself, and ε_{i-1} as the "ε" given by the Lemma. Now let n be sufficiently large so that the Lemma holds in all t cases. Set G_0 equal the complete r-graph on n vertices. Then G_0 is quasirandom with density 1 and tolerance ε_0–in fact, with tolerance zero. Applying the Lemma we find, for $0 \le i < t$, families F_i and graphs G_{i+1} so that

(i) $|F_i| < \delta[e^{-i\delta}\binom{n}{r}/\binom{k}{r}]\,(1 + \varepsilon_i)$

(ii) G_{i+1} is G_i with all cliques of F_i deleted.

(iii) G_{i+1} is quasirandom with density $\rho e^{-\delta}$ and tolerance ε_{i+1}

As all $\varepsilon_i \leq \varepsilon$ we simplify (i), (iii) to

(i′) $|F_i| < \delta[e^{-i\delta}\binom{n}{r}/\binom{k}{r}](1+\varepsilon)$

(iii′) G_{i+1} is quasirandom with density $\rho e^{-\delta}$ and tolerance ε.

For each i, $0 \leq i < t$

$$|G_i| > e^{-i\delta}\binom{n}{r}(1-\varepsilon) \qquad (51)$$
$$|G_{i+1}| < e^{-(i+1)\delta}\binom{n}{r}(1+\varepsilon)$$

so

$$|G_i| - |G_{i+1}| > e^{-i\delta}\binom{n}{r}(1 - e^{-\delta} - 2\varepsilon) \qquad (52)$$

and

$$\frac{|F_i|\binom{k}{r}}{|G_i| - |G_{i+1}|} < \frac{\delta(1+\varepsilon)}{1 - e^{-\delta} - 2\varepsilon}. \qquad (53)$$

The families $F_0, F_1, \dots, F_{t-1}$ cover all r-sets except G_t. For each $e \in G_t$ let K_e be an arbitrary k-set containing e (not necessarily a clique in G_t) and let F_∞ denote the family of those K_e. Then

$$|F_\infty| \leq |G_t| < e^{-t\delta}\binom{n}{r}(1+\varepsilon). \qquad (54)$$

Now the set

$$F = F_0 \cup F_1 \cup \cdots \cup F_{t-1} \cup F_\infty \qquad (55)$$

covers all r-sets on n vertices. Summing (53) for $0 \leq i < t$

$$\left[\sum_{i=0}^{t-1} |F_i|\right]\binom{k}{r} < \left[\sum_{i=0}^{t-1} |G_i|\right]\frac{\delta(1+\varepsilon)}{1 - e^{-\delta} - 2\varepsilon} \qquad (56)$$
$$< \binom{n}{r}\frac{\delta(1+\varepsilon)}{1 - e^{-\delta} - 2\varepsilon}.$$

Adding F_∞;

$$\binom{k}{r}M(n,k,r) \leq |F|\binom{k}{r} \qquad (57)$$
$$< \binom{n}{r}\left[\frac{\delta(1+\varepsilon)}{1 - e^{-\delta} - 2\varepsilon} + \binom{k}{r}e^{-t\delta}(1+\varepsilon)\right]$$
$$< \binom{n}{r}(1+a)$$

by our propitious choices of δ, ε and t—completing the proof.

References

[1] P. Erdős and H. Hanani, *On a Limit Theorem in Combinatorial Analysis*, Publ. Math. Debrecen, **10** (1963), 10–13.
[2] V. Rődl, *Solution of the Erdős—Hanani conjecture*, (in preparation).
[3] J. Spencer, *Sequences with small discrepency relative to n events*, Composition Math., **47** (1982), 365–392.

Received May 31, 1984.

SUNY—Stony Brook
Stony Brook, NY 11794