Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 8175–8177, November 1989
Mathematics

Quasi-random hypergraphs

(graphs/randomness)

F. R. K. CHUNG* AND R. L. GRAHAM†

*Bell Communications Research, Morristown, NJ 07960; and †AT&T Bell Laboratories, Murray Hill, NJ 07974

Contributed by R. L. Graham, July 14, 1989

ABSTRACT We describe a large equivalence class of properties shared by most hypergraphs, including so-called random hypergraphs. As a result, it follows that many global properties of hypergraphs are actually consequences of simple local conditions.

Hypergraphs form a natural generalization of graphs in which (hyper)edges consist of k-element subsets of the vertices rather than pairs in the case of graphs. In refs. 1 and 2, we introduced the concept of a quasi-random graph property. It was observed there that while the extension of these concepts to hypergraphs would be highly desirable, substantial obstacles to such an extension appeared to exist. In this note, we announce the sought-after generalization. As a consequence, not only do we now have methods for the explicit construction of hypergraphs which simulate many aspects of random hypergraphs but our understanding of quasi-randomness in ordinary graphs has been clarified as well.

Notation

For a fixed positive integer k, a k-uniform hypergraph $G = (V, E)$, or k-graph for short, consists of a set $V = V(G)$, called the *vertex set* of G, and a subset $E = E(G)$ of the set $\binom{V}{k}$ of k-element sets of V, called the *edges* of G. We use the notation $G(n)$ to denote the fact that V has n elements; i.e., $|V| = n$. For $X \subseteq V$, we let $G[X]$ denote the sub-k-graph of G *induced by* X; i.e., $G[X] = \left(X, E \cap \binom{X}{k}\right)$. Let $\chi_G: \binom{V}{k} \to \{0,1\}$ denote the *edge indicator* of G; i.e., for $e \in \binom{V}{k}$,

$$\chi_G(e) = \begin{cases} 1 & \text{if } e \in E, \\ 0 & \text{otherwise.} \end{cases}$$

For another k-graph $G' = (V', E')$, we let $N_G^*(G')$ denote the number of labeled occurrences of G' as an induced subgraph of G. In other words,

$$N_G^*(G') = |\{\lambda: V' \to V| \quad G[\lambda(V)] \cong G'\}|$$

where $\cong$ denotes the union of the natural notion of k-graph isomorphism. If $\mathcal{F}$ is a family of k-graphs then $N_G^*(\mathcal{F})$ denotes $\cup_{G' \in \mathcal{F}} N_G^*(G')$.

Further, we denote the number of copies of G' occurring as (not necessarily induced) sub-k-graphs of G by $N_G(G')$. Thus,

$$N_G(G') = \sum_H N_G^*(H),$$

where the sum is taken over all k-graphs

$$H = (V', E_H), \text{ where } E_H \supseteq E'.$$

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

There is a special k-graph on $2k$ vertices which will be important in what follows. This k-graph is called a k-*octahedron* or just *octahedron*, for short, and is denoted by $\mathcal{O}_k = \mathcal{O}_k(x_1(0), x_1(1), x_2(0), x_2(1), \ldots, x_k(0), x_k(1)) = \mathcal{O}_k(\bar{x}(\bar{\varepsilon}))$. The vertices of $\mathcal{O}_k$ are the $x_i(\varepsilon_i)$, $1 \le i \le k$, $\varepsilon_i \in \{0, 1\}$. The edges of $\mathcal{O}_k$ consist of all k-sets of the form $\{x_1(\varepsilon_1), x_2(\varepsilon_2), \ldots, x_k(\varepsilon_k)|\varepsilon_i \in \{0, 1\}, 1 \le i \le k\}$. Thus, $\mathcal{O}_k$ has 2^k edges.

Finally, we will call a k-graph an *even partial octahedron* if it has the same vertex set as $\mathcal{O}_k$ and has an edge set consisting of an *even number* of the edges of $\mathcal{O}_k$. We let $\mathcal{O}_k^*$ denote the set of all even partial octahedra based on $\mathcal{O}_k$, although occasionally we will let $\mathcal{O}_k^*$ denote an individual one, as well.

The Main Results

We are going to consider various properties which a k-graph $G = G(n)$ might satisfy. Each of the properties will contain occurrences of the asymptotic "little oh" notation $o(1)$. However, the dependence of the different $o(1)$ values on the particular properties they refer to will ordinarily be suppressed. The use of these $o(1)$ values can be viewed in two essentially equivalent ways.

In the first way, suppose we have two properties P and P', each with occurrences of $o(1)$, so that $P = P(o(1))$, $P' = P'(o(1))$. The implication "$P \Rightarrow P'$" then means that for each $\varepsilon > 0$ there is a $\delta > 0$ so that if $G(n)$ satisfies $P(\delta)$ then it must also satisfy $P'(\varepsilon)$, provided $n > n_0(\varepsilon)$.

In the second way, we can think of considering a family $\mathcal{F}$ of k-graphs $G(n)$ with $n \to \infty$. In this case, the interpretation of $o(1)$ is the usual one as $G = G(n)$ ranges over $\mathcal{F}$, with $n \to \infty$. As usual, "almost all" (abbreviated a.a.) denotes a fraction $1 + o(1)$ of the elements of the set in question.

We next state a series of properties for k-graphs $G(n)$ which are shared by almost all random k-graphs $G_{1/2}(n)$ on n vertices. [For $G_{1/2}(n)$, each possible k-set is chosen to be an edge independently with probability 1/2.]

$Q_1(s)$: For all k-graphs $G'(s)$ on s vertices,
$$N_{G(n)}^*(G'(s)) = (1 + o(1))n^s/2^{\binom{s}{k}}.$$

Q_2: For all k-graphs $G'(2k)$,
$$N_{G(n)}^*(G'(2k)) = (1 + o(1))n^{2k}/2^{\binom{2k}{k}}.$$

Q_3:
$$N_{G(n)}^*(\mathcal{O}_k^*) \le (1 + o(1))n^{2k}/2.$$

In other words, the number of induced even partial octahedra $\mathcal{O}_k^*$ occurring in $G(n)$ is at most $(1 + o(1))n^{2k}/2$, which is just the expected number occurring in the random k-graph $G_{1/2}(n)$.

For our final properties, we need another definition. Let $G = (V, E)$ be a k-graph and let x, y be elements of V. The *sameness* (k − 1)-*graph* $G(x, y)$ is defined to be the $(k - 1)$-graph $G' = (V', E')$ with $V' = V \backslash \{x, y\}$ and edge set

$$E' = \left\{e' \in \binom{V'}{k-1} \middle| \chi_G(e' \cup \{x\}) = \chi_G(e' \cup \{y\})\right\}.$$

Q_4: For almost all choices of x, y elements of V, the sameness $(k-1)$-graph $G(x, y)$ of $G(n) = (V, E)$ satisfies Q_2, with k replaced by $k-1$.

Q_5: For $1 \le r \le 2k-1$ and almost all $x, y \in V$,

$$N_{G(x,y)}(K_r^{(k-1)}) = (1 + o(1))n^r/2^{\binom{r}{k-1}},$$

where $G(x, y)$ is the sameness $(k-1)$-graph of $G(n) = (V, E)$ and $K_r^{(k-1)}$ denotes the complete $(k-1)$-graph on r vertices, i.e., having all the possible $\binom{r}{k-1}$ edges (and we use the convention that any set of $r < k-1$ vertices forms a $K_r^{(k-1)}$).

Several implications among these properties are immediate or easily proved. For example,

$$Q_1(2k) = Q_2 \Rightarrow Q_3 \text{ and } Q_1(s+1) \Rightarrow Q_1(s).$$

Our main result asserts that for $s \ge 2k$, all of these properties are in fact *equivalent*.

THEOREM 1. *For* $s \ge 2k$,

$$Q_1(s) \Rightarrow Q_2 \Rightarrow Q_3 \Rightarrow Q_4 \Rightarrow Q_5 \Rightarrow Q_1(s).$$

Hypergraphs which satisfy any one (and therefore all) of these properties we call *quasi-random*. The proofs of the various implications in *Theorem 1* involve applications of the second moment method, linear algebra, extremal hypergraph theory, and results from ref. 2 and will be given elsewhere.

A useful consequence of *Theorem 1* is the following characterization.

COROLLARY. *A* k*-graph* G = (V, E) *is quasi-random if and only if for almost all choices of* x, y *elements of* V, *the sameness* (k − 1)*-graph* G(x, y) *is quasi-random.*

The relevance of condition Q_3 stems from the following fact.

LEMMA 1. *For any* k*-graph* F = F(n),

$$N_F^*(\mathbb{O}_k^e) \ge (1 + o(1))n^{2k}/2.$$

A condition which is a consequence of quasi-randomness is the following.

Q_6: For all $X \subseteq V$,

$$|E(G[X])| = \frac{1}{2}\binom{|X|}{k} + o(n^k).$$

In ref. 2, it was shown that for $k = 2$, Q_6 implies quasi-randomness. However, as was pointed out by Rödl (3), this fails dramatically for $k \ge 3$. In fact, the following k-graph $G_0 = G_0(n)$ is a strong example of this failure. For a vertex set V of size n, select a random $(k-1)$-graph $G = G_{1/2}(n) = (V, E)$. Then $G_0 = (V, E_0)$ is the k-graph defined by choosing $e_0 \in \binom{V}{k}$ to be an edge of G_0 if and only if

$$\sum_{e \in \binom{e_0}{k-1}} \chi_G(e) \equiv 0 \pmod 2.$$

THEOREM 2. *For almost all choices of* $G_{1/2}(n)$,
(a) *for any* $X \subset V$,

$$|E(G_0[X])| = \frac{1}{2}\binom{|X|}{k} + o(n^k),$$

and
(b) G_0 *does not contain an induced copy of the* k*-graph having* k + 1 *vertices and* k *edges.*

In fact, J. H. Spencer (personal communication) has strengthened (a) for the case $k = 3$ by showing for almost all choices of $G_{1/2}(n)$,

$$\left| |E(G_0[X])| - \frac{1}{2}\binom{|X|}{3} \right| < 200|X|^2$$

holds for all $X \subset V$. By results of Erdös and Spencer (4), this is best possible (up to the value of the constant 200) for any 3-graph.

If $G = (V, E)$ is a k-graph and $v \in V$, define the *projection* $G^{(v)}$ to be the $(k-1)$-graph with vertex set $V \backslash \{v\}$ and edge set

$$\left\{ e' \in \binom{V}{k-1} \middle| e' \cup \{v\} \in E \right\}.$$

Theorem 1 implies the following.

COROLLARY. *If* G = (V, E) *is a quasi-random* k*-graph then almost all projections* $G^{(v)}$, v *an element of* V, *are quasi-random* (k − 1)*-graphs.*

The reverse implication does not hold, however, as shown by the k-graph G_0 just described.

COROLLARY. *For almost all choices of* $G_{1/2}(n)$, *all the projections* $G_0^{(v)}$, $v \in V$, *are quasi-random* (k − 1)*-graphs.*

Examples. We now give two simple explicit examples of quasi-random k-graphs. Of course, almost all random k-graphs are quasi-random.

The "even intersection" k-graph $I_k(n) = (V, E)$ is defined as follows. For V we take $2^{[n]}$, the class of all subsets of $[n] = \{1, 2, \ldots, n\}$. A k-set $\{X_1, \ldots, X_k\} \in E$ if and only if

$$|X_1 \cap \cdots \cap X_k| \equiv 0 \pmod 2.$$

FACT. $I_k(n)$ *is quasi-random.*

For a prime p, define the (generalized) Paley k-graph $P_k(p)$ as follows. The vertex set of $P_k(p)$ is the finite field $GF(p)$. A k-set $\{i_1, \ldots, i_k\}$ is an edge of $P_k(p)$ if and only if $i_1 + \cdots + i_k$ is a quadratic residue modulo p.

FACT. $P_k(p)$ *is quasi-random.*

Not surprisingly, $I_k(n)$ and $P_k(p)$ have many relatives which are also quasi-random—e.g., requiring the sets X_i used in defining edges of $I_k(n)$ to have a fixed cardinality (such as $[n/2]$), or requiring the cardinality of the intersection to belong to a fixed set $Y \subseteq \{0, 1, \ldots, 2m-1\}$ of size m of residues modulo $2m$ [instead of being $0 \pmod 2$)].

Questions. Perhaps the most natural open question we have at present is to know the least value of s so that a k-graph satisfying property $Q_1(s)$ is forced to be quasi-random. By *Theorem 1*, $Q_1(2k)$ guarantees quasi-randomness. On the other hand, it is not difficult to give examples of non-quasi-random k-graphs which satisfy $Q_1(k+1)$. The first gap occurs for $k = 3$. In particular, are there non-quasi-random 3-graphs which contain all (ordered) 3-graphs on five vertices asymptotically equally often?

The following problem arises in connection with property Q_5. For each fixed t, is there a 2-graph $G(n)$ so that $N_{G(n)}(K_r) = (1 + o(1))n^r 2^{-\binom{r}{2}}$, $0 \le r \le t$, but for which $N_{G(n)}(K_{t+1}) \ne (1 + o(1))n^{t+1}2^{-\binom{t+1}{2}}$ (where K_r denotes the complete 2-graph on r vertices)? It seems quite likely (to us) that such graphs exist.

Finally, we mention a related intriguing problem of P. Frankl and V. Rödl (personal communication). Suppose $G = G(n) = (V, E)$ is a 3-graph so that for every 2-graph $H = (V, E')$, we have

$$\left| \left\{ e \in E \middle| \binom{e}{2} \subset E' \right\} \right| = \frac{1}{2} N_H(K_3) + o(n^3).$$

In other words, for any H which has cn^3 triangles, about half of them correspond to edges in G. Does this property imply (or is it implied by) quasi-randomness? (See *Note Added in Proof.*)

It seems to us that it would be profitable to explore quasi-randomness extended to simulating random k-graphs $G_p(n)$ for $p \ne 1/2$ or, more generally, for $p = p(n)$, especially along the lines carried out so fruitfully for 2-graphs by Thomason (5, 6) and by Haviland and Thomason (7).

Note Added in Proof. It is now known that the condition of Frankl and Rödl is in fact equivalent to quasi-randomness.

1. Chung, F. R. K., Graham, R. L. & Wilson, R. M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 969–970.
2. Chung, F. R. K., Graham, R. L. & Wilson, R. M. (1989) *Combinatorica*, in press.
3. Rödl, V. (1986) *Discrete Math.* **59**, 125–134.
4. Erdös, P. & Spencer, J. H. (1972) *Networks* **1**, 379–385.
5. Thomason, A. (1987) *London Math. Soc. Lect. Note Ser.* **123**, 173–196.
6. Thomason, A. (1987) *Ann. Discrete Math.* **33**, 307–331.
7. Haviland, J. & Thomason, A. (1989) *Discrete Math.* **75**, 255–278.