Equivalent Conditions for Regularity (Extended Abstract)

Y. Kohayakawa[1⋆], V. Rödl[2], and J. Skokan[2]

[1] Instituto de Matemática e Estatística, Universidade de São Paulo,
Rua do Matão 1010, 05508–900 São Paulo, Brazil
yoshi@ime.usp.br

[2] Department of Mathematics and Computer Science,
Emory University, Atlanta, GA, 30322, USA
{rodl,jozef}@mathcs.emory.edu

Abstract. Haviland and Thomason and Chung and Graham were the first to investigate systematically some properties of quasi-random hypergraphs. In particular, in a series of articles, Chung and Graham considered several quite disparate properties of random-like hypergraphs of density $1/2$ and proved that they are in fact equivalent. The central concept in their work turned out to be the so called *deviation* of a hypergraph. Chung and Graham proved that having small deviation is equivalent to a variety of other properties that describe quasi-randomness. In this note, we consider the concept of *discrepancy* for k-uniform hypergraphs with an arbitrary constant density d $(0 < d < 1)$ and prove that the condition of having asymptotically vanishing discrepancy is equivalent to several other quasi-random properties of $\mathcal{H}$, similar to the ones introduced by Chung and Graham. In particular, we give a proof of the fact that having the correct 'spectrum' of the s-vertex subhypergraphs is equivalent to quasi-randomness for any $s \geq 2k$. Our work can be viewed as an extension of the results of Chung and Graham to the case of an arbitrary constant valued density. Our methods, however, are based on different ideas.

1 Introduction and the main result

The usefulness of random structures in theoretical computer science and in discrete mathematics is well known. An important, closely related question is the following: which, if any, of the almost sure properties of such structures suffice for a deterministic object to have to be as useful or relevant?

Our main concern here is to address the above question in the context of hypergraphs. We shall continue the study of quasi-random hypergraphs along the lines initiated by Haviland and Thomason [7, 8] and especially by Chung [2], and Chung and Graham [3, 4]. One of the central concepts concerning hypergraph quasi-randomness, the so called *hypergraph discrepancy*, was investigated

⋆ Partially supported by FAPESP (Proc. 96/04505–2), by CNPq (Proc. 300334/93–1), and by MCT/FINEP (PRONEX project 107/97).

by Babai, Nisan, and Szegedy [1], who found a connection between communication complexity and hypergraph discrepancy. This connection was further studied by Chung and Tetali [5]. Here, we carry out the investigation very much along the lines of Chung and Graham [3, 4], except that we focus on hypergraphs of arbitrary constant density, making use of different techniques.

In the remainder of this introduction, we carefully discuss a result of Chung and Graham [3] and state our main result, Theorem 3 below.

1.1 The result of Chung and Graham

We need to start with some definitions. For a set V and an integer $k \geq 2$, let $[V]^k$ denote the system of all k-element subsets of V. A subset $\mathcal{G} \subset [V]^k$ is called a *k-uniform hypergraph.* If $k = 2$, we have a *graph.* We sometimes use the notation $\mathcal{G} = (V(\mathcal{G}), E(\mathcal{G}))$. If there is no danger of confusion, we shall identify the hypergraphs with their edge sets. Throughout this paper, the integer k is assumed to be a fixed constant.

For any l-uniform hypergraph $\mathcal{G}$ and $k \geq l$, let $\mathcal{K}_k(\mathcal{G})$ be the set of all k-element sets that span a clique $K_k^{(l)}$ on k vertices. We also denote by $K_k(2)$ the complete k-partite k-uniform hypergraph whose every partite set contains precisely two vertices. We refer to $K_k(2)$ as the *generalized octahedron*, or, simply, the *octahedron*.

We also consider a function $\mu_{\mathcal{H}}\colon [V]^k \to \{-1, 1\}$ such that, for all $e \in [V]^k$, we have

$$\mu_{\mathcal{H}}(e) = \begin{cases} -1, & \text{if } e \in \mathcal{H} \\ 1, & \text{if } e \notin \mathcal{H}. \end{cases}$$

Let $[k] = \{1, 2, \ldots, k\}$, let V^{2k} denote the set of all $2k$-tuples $(v_1, v_2, \ldots, v_{2k})$, where $v_i \in V$ $(1 \leq i \leq 2k)$, and let $\Pi_{\mathcal{H}}^{(k)}\colon V^{2k} \to \{-1, 1\}$ be given by

$$\Pi_{\mathcal{H}}^{(k)}(u_1, \ldots, u_k, v_1, \ldots, v_k) = \prod\nolimits_{\varepsilon} \mu_{\mathcal{H}}(\varepsilon_1, \ldots, \varepsilon_k),$$

where the product is over all vectors $\varepsilon = (\varepsilon_i)_{i=1}^k$ with $\varepsilon_i \in \{u_i, v_i\}$ for all i and we understand $\mu_{\mathcal{H}}$ to be 1 on arguments with repeated entries.

Following Chung and Graham (see, e.g., [4]), we define the *deviation* $\mathrm{dev}(\mathcal{H})$ of $\mathcal{H}$ by

$$\mathrm{dev}(\mathcal{H}) = \frac{1}{m^{2k}} \sum_{u_i, v_i \in V,\ i \in [k]} \Pi_{\mathcal{H}}^{(k)}(u_1, \ldots, u_k, v_1, \ldots, v_k).$$

For two hypergraphs $\mathcal{G}$ and $\mathcal{H}$, we denote by $\binom{\mathcal{H}}{\mathcal{G}}$ the set of all induced subhypergraphs of $\mathcal{H}$ that are isomorphic to $\mathcal{G}$. We also write $\binom{\mathcal{H}}{\mathcal{G}}^{\mathrm{w}}$ for the number of *weak* (i.e., not necessarily induced) subhypergraphs of $\mathcal{H}$ that are isomorphic to $\mathcal{G}$. Furthermore, we need the notion of the *link* of a vertex.

Definition 1 *Let $\mathcal{H}$ be a k-uniform hypergraph and $x \in V(\mathcal{H})$. We shall call the $(k-1)$-uniform hypergraph*

$$\mathcal{H}(x) = \{e \setminus \{x\}\colon e \in \mathcal{H},\ x \in e\}$$

the link of the vertex x in $\mathcal{H}$. For a subset $W \subset V(\mathcal{H})$, the joint W-link is $\mathcal{H}(W) = \bigcap_{x\in W} \mathcal{H}(x)$. For simplicity, if $W = \{x_1, \dots, x_k\}$, we write $\mathcal{H}(x_1, \dots, x_k)$.

Observe that if $\mathcal{H}$ is k-partite, then $\mathcal{H}(x)$ is $(k-1)$-partite for every $x \in V$. Furthermore, if $k = 2$, then $\mathcal{H}(x)$ may be identified with the ordinary graph neighbourhood of x. Moreover, $\mathcal{H}(x, x')$ may be thought of as the 'joint neighbourhood' of x and x'.

In [3], Chung and Graham proved that if the density of an m-vertex k-uniform hypergraph $\mathcal{H}$ is $1/2$, i.e., $|\mathcal{H}| = (1/2 + o(1))\binom{m}{k}$, where $o(1) \to 0$ as $m \to \infty$, then the following statements are equivalent:

$(Q_1(s))$ for all k-uniform hypergraphs $\mathcal{G}$ on $s \geq 2k$ vertices and automorphism group $\mathrm{Aut}(G)$,

$$\left|\binom{\mathcal{H}}{\mathcal{G}}\right| = (1 + o(1))\binom{m}{s} 2^{-\binom{s}{k}} \frac{s!}{|\mathrm{Aut}(\mathcal{G})|},$$

(Q_2) for all k-uniform hypergraphs $\mathcal{G}$ on $2k$ vertices and automorphism group $\mathrm{Aut}(G)$, we have

$$\left|\binom{\mathcal{H}}{\mathcal{G}}\right| = (1 + o(1))\binom{m}{2k} 2^{-\binom{2k}{k}} \frac{(2k)!}{|\mathrm{Aut}(G)|},$$

(Q_3) $\mathrm{dev}(\mathcal{H}) = o(1)$,

(Q_4) for almost all choices of vertices $x, y \in V$, the $(k-1)$-uniform hypergraph $\overline{\mathcal{H}(x) \triangle \mathcal{H}(y)}$, that is, the *complement* $[V]^{k-1} \setminus (\mathcal{H}(x) \triangle \mathcal{H}(y))$ of the symmetric difference of $\mathcal{H}(x)$ and $\mathcal{H}(y)$, satisfies Q_2 with k replaced by $k-1$,

(Q_5) for $1 \leq r \leq 2k-1$ and almost all $x, y \in V$,

$$\left|\binom{\mathcal{H}(x,y)}{K_r^{(k-1)}}\right| = (1 + o(1))\binom{m}{r} 2^{-\binom{r}{k-1}}.$$

The equivalence of these properties is to be understood in the following sense. If we have two properties $\boldsymbol{P} = \boldsymbol{P}(o(1))$ and $\boldsymbol{P'} = \boldsymbol{P'}(o(1))$, then "$\boldsymbol{P} \Rightarrow \boldsymbol{P'}$" means that for every $\varepsilon > 0$ there is a $\delta > 0$ so that any k-uniform hypergraph $\mathcal{H}$ on m vertices satisfying $\boldsymbol{P}(\delta)$ must also satisfy $\boldsymbol{P'}(\varepsilon)$, provided $m > M_0(\varepsilon)$.

In [3] Chung and Graham stated that "it would be profitable to explore quasi-randomness extended to simulating random k-uniform hypergraphs $G_p(n)$ for $p \neq 1/2$, or, more generally, for $p = p(n)$, especially along the lines carried out so fruitfully by Thomason [13, 14]." Our present aim is to explore quasi-randomness from this point of view. In this paper, we concentrate on the case in which p is an arbitrary constant. In certain crucial parts, our methods are different from the ones of Chung and Graham. Indeed, it seems to us that the fact that the density of $\mathcal{H}$ is $1/2$ is essential in certain proofs in [3] (especially those involving the concept of deviation).

1.2 Discrepancy and subgraph counting

The following concept was proposed by Frankl and Rödl and later investigated by Chung [2] and Chung and Graham in [3, 4]. For an m-vertex k-uniform hypergraph $\mathcal{H}$ with vertex set V, we define the *density* $d(\mathcal{H})$ and the *discrepancy* $\mathrm{disc}_{1/2}(\mathcal{H})$ of $\mathcal{H}$ by letting $d(\mathcal{H}) = |\mathcal{H}|\binom{m}{k}^{-1}$ and

$$\mathrm{disc}_{1/2}(\mathcal{H}) = \frac{1}{m^k} \max_{\mathcal{G}\subset[V]^{k-1}} \big||\mathcal{H}\cap\mathcal{K}_k(\mathcal{G})| - |\bar{\mathcal{H}}\cap\mathcal{K}_k(\mathcal{G})|\big|, \tag{1}$$

where the maximum is taken over all $(k-1)$-uniform hypergraphs $\mathcal{G}$ with vertex set V, and $\bar{\mathcal{H}}$ is the complement $[V]^k \setminus \mathcal{H}$ of $\mathcal{H}$.

To accommodate arbitrary densities, we extend the latter concept as follows.

Definition 2 *Let $\mathcal{H}$ be a k-uniform hypergraph with vertex set V with $|V| = m$. We define the* discrepancy $\mathrm{disc}(\mathcal{H})$ *of $\mathcal{H}$ as follows:*

$$\mathrm{disc}(\mathcal{H}) = \frac{1}{m^k} \max_{\mathcal{G}\subset[V]^{k-1}} \big||\mathcal{H}\cap\mathcal{K}_k(\mathcal{G})| - d(\mathcal{H})|\mathcal{K}_k(\mathcal{G})|\big|, \tag{2}$$

where the maximum is taken over all $(k-1)$-uniform hypergraphs $\mathcal{G}$ with vertex set V.

Observe that if $d(\mathcal{H}) = 1/2$, then $\mathrm{disc}(\mathcal{H}) = (1/2)\,\mathrm{disc}_{1/2}(\mathcal{H})$, so both notions are equivalent. Following some initial considerations by Frankl and Rödl, Chung and Graham investigated the relation between discrepancy and deviation. In fact, Chung [2] succeeded in proving the following inequalities closely connecting these quantities:

(i) $\mathrm{dev}(\mathcal{H}) < 4^k(\mathrm{disc}_{1/2}(\mathcal{H}))^{1/2^k}$,

(ii) $\mathrm{disc}_{1/2}(\mathcal{H}) < (\mathrm{dev}(\mathcal{H}))^{1/2^k}$.

For simplicity, we state the inequalities for the density 1/2 case. For the general case, see Section 5 of [2].

Before we proceed, we need to introduce a new concept. If the vertex set of a hypergraph is totally ordered, we say that we have an *ordered* hypergraph. Given two ordered hypergraphs $\mathcal{G}_\leq$ and $\mathcal{H}_{\leq'}$, where $\leq$ and $\leq'$ denote the orderings on the vertex sets of $\mathcal{G} = \mathcal{G}_\leq$ and $\mathcal{H} = \mathcal{H}_{\leq'}$, we say that a function $f\colon V(\mathcal{G}) \to V(\mathcal{H})$ is an *embedding of ordered hypergraphs* if (i) it is an injection, (ii) it respects the orderings, i.e., $f(x) \leq' f(y)$ whenever $x \leq y$, and (iii) $f(g) \in \mathcal{H}$ if and only if $g \in \mathcal{G}$, where $f(g)$ is the set formed by the images of all the vertices in g. Furthermore, if $\mathcal{G} = \mathcal{G}_\leq$ and $\mathcal{H} = \mathcal{H}_{\leq'}$, we write $\binom{\mathcal{H}}{\mathcal{G}}_{\mathrm{ord}}$ for the number of such embeddings.

As our main result, we shall prove the following extension of Chung and Graham's result.

Theorem 3 *Let $\mathcal{H} = (V, E)$ be a k-uniform hypergraph of density $0 < d < 1$. Then the following statements are equivalent:*

(P_1) $\operatorname{disc}(\mathcal{H}) = o(1)$,

(P_2) $\operatorname{disc}(\mathcal{H}(x)) = o(1)$ *for all but* $o(m)$ *vertices* $x \in V$ *and* $\operatorname{disc}(\mathcal{H}(x,y)) = o(1)$ *for all but* $o(m^2)$ *pairs* x, $y \in V$,

(P_3) $\operatorname{disc}(\mathcal{H}(x,y)) = o(1)$ *for all but* $o(m^2)$ *pairs* x, $y \in V$,

(P_4) *the number of octahedra* $K_k(2)$ *in* $\mathcal{H}$ *is asymptotically minimized among all* k*-uniform hypergraphs of density* d*; indeed,*

$$\left|\binom{\mathcal{H}}{K_k(2)}^{\mathrm{w}}\right| = (1+o(1))\frac{m^{2k}}{2^k k!}d^{2^k},$$

(P_5) *for any* $s \geq 2k$ *and any* k*-uniform hypergraph* $\mathcal{G}$ *on* s *vertices with* $e(\mathcal{G})$ *edges and automorphism group* $\operatorname{Aut}(\mathcal{G})$,

$$\left|\binom{\mathcal{H}}{\mathcal{G}}\right| = (1+o(1))\binom{m}{s}d^{e(\mathcal{G})}(1-d)^{\binom{s}{k}-e(\mathcal{G})}\frac{s!}{|\operatorname{Aut}(\mathcal{G})|},$$

(P'_5) *for any ordering* $\mathcal{H}_\leq$ *of* $\mathcal{H}$ *and for any fixed integer* $s \geq 2k$*, any ordered* k*-uniform hypergraph* $\mathcal{G}_\leq$ *on* s *vertices with* $e(\mathcal{G})$ *edges is such that*

$$\left|\binom{\mathcal{H}}{\mathcal{G}}_{\mathrm{ord}}\right| = (1+o(1))\binom{m}{s}d^{e(\mathcal{G})}(1-d)^{\binom{s}{k}-e(\mathcal{G})},$$

(P_6) *for all* k*-uniform hypergraphs* $\mathcal{G}$ *on* $2k$ *vertices with* $e(\mathcal{G})$ *edges and automorphism group* $\operatorname{Aut}(\mathcal{G})$,

$$\left|\binom{\mathcal{H}}{\mathcal{G}}\right| = (1+o(1))\binom{m}{2k}d^{e(\mathcal{G})}(1-d)^{\binom{2k}{k}-e(\mathcal{G})}\frac{(2k)!}{|\operatorname{Aut}(\mathcal{G})|}.$$

(P'_6) *for any ordering* $\mathcal{H}_\leq$ *of* $\mathcal{H}$*, any ordered* k*-uniform hypergraph* $\mathcal{G}_\leq$ *on* $2k$ *vertices with* $e(\mathcal{G})$ *edges is such that*

$$\left|\binom{\mathcal{H}}{\mathcal{G}}_{\mathrm{ord}}\right| = (1+o(1))\binom{m}{2k}d^{e(\mathcal{G})}(1-d)^{\binom{2k}{k}-e(\mathcal{G})}.$$

Some of the implications in Theorem 3 are fairly easy or are by now quite standard. There are, however, two implications that appear to be more difficult.

The proof of Chung and Graham that $\operatorname{dev}_{1/2}(\mathcal{H}) = o(1)$ implies P_5 (the 'subgraph counting formula') is based on an approach that has its roots in a seminal paper of Wilson [15]. This beautiful proof seems to make non-trivial use of the fact that $d(\mathcal{H}) = 1/2$. Our proof of the implication that small discrepancy implies the subgraph counting formula ($P_1 \Rightarrow P'_5$) is based on a different technique, which works well in the arbitrary constant density case (see Section 2.2).

Our second contribution, which is somewhat more technical in nature, lies in a novel approach for the proof of the implication $P_2 \Rightarrow P_1$. Our proof is based on a variant of the Regularity Lemma of Szemerédi [12] for hypergraphs [6] (see Section 2.1).

2 Main steps in the proof of Theorem 3

2.1 The first part

The first part of the proof of Theorem 3 consists of proving that properties $P_1, \ldots, P_4$ are mutually equivalent. As it turns out, the proof becomes more transparent if we restrict ourselves to k-partite hypergraphs. In the next paragraph, we introduce some definitions that will allow us to state the k-partite version of $P_1, \ldots, P_4$ (see Theorem 15). We close this section introducing the main tool in the proof of Theorem 15, namely, we state a version of the Regularity Lemma for hypergraphs (see Lemma 20).

Definitions for partite hypergraphs. For simplicity, we first introduce the term *cylinder* to mean partite hypergraphs.

Definition 4 *Let $k \geq l \geq 2$ be two integers. We shall refer to any k-partite l-uniform hypergraph $\mathcal{H} = (V_1 \cup \cdots \cup V_k, E)$ as a k-partite l-cylinder or (k,l)-cylinder. If $l = k-1$, we shall often write $\mathcal{H}_i$ for the subhypergraph of $\mathcal{H}$ induced on $\bigcup_{j \neq i} V_j$. Clearly, $\mathcal{H} = \bigcup_{i=1}^{k} \mathcal{H}_i$. We shall also denote by $K_k^{(l)}(V_1, \ldots, V_k)$ the complete (k,l)-cylinder with vertex partition $V_1 \cup \cdots \cup V_k$.*

Definition 5 *For a (k,l)-cylinder $\mathcal{H}$, we shall denote by $\mathcal{K}_j(\mathcal{H})$, $l \leq j \leq k$, the (k,j)-cylinder whose edges are precisely those j-element subsets of $V(\mathcal{H})$ that span cliques of order j in $\mathcal{H}$.*

When we deal with cylinders, we have to measure density according to their natural vertex partitions.

Definition 6 *Let $\mathcal{H}$ be a (k,k)-cylinder with k-partition $V = V_1 \cup \cdots \cup V_k$. We define the* k-partite density *or simply the* density $d(\mathcal{H})$ *of $\mathcal{H}$ by*

$$d(\mathcal{H}) = \frac{|\mathcal{H}|}{|V_1| \ldots |V_k|}.$$

To be precise, we should have a distinguished piece of notation for the notion of k-partite density. However, the context will always make clear which notion we mean when we talk about the density of a (k,k)-cylinder.

We should also be careful when we talk about the discrepancy of a cylinder.

Definition 7 *Let $\mathcal{H}$ be a (k,k)-cylinder with vertex set $V = V_1 \cup \cdots \cup V_k$. We define the* discrepancy $\mathrm{disc}(\mathcal{H})$ *of $\mathcal{H}$ as follows:*

$$\mathrm{disc}(\mathcal{H}) = \frac{1}{|V_1| \ldots |V_k|} \max_{\mathcal{G}} \big| |\mathcal{H} \cap \mathcal{K}_k(\mathcal{G})| - d(\mathcal{H}) |\mathcal{K}_k(\mathcal{G})| \big|, \tag{3}$$

where the maximum is taken over all $(k,k-1)$-cylinders $\mathcal{G}$ with vertex set $V = V_1 \cup \cdots \cup V_k$.

We now introduce a simple but important concept concerning the "regularity" of a (k,k)-cylinder.

Definition 8 *Let $\mathcal{H}$ be a (k,k)-cylinder with k-partition $V = V_1 \cup \cdots \cup V_k$ and let $\delta < \alpha$ be two positive real numbers. We say that $\mathcal{H}$ is (α,δ)-regular if the following condition is satisfied: if $\mathcal{G}$ is any $(k,k-1)$-cylinder such that $|\mathcal{K}_k(\mathcal{G})| \geq \delta |V_1| \dots |V_k|$, then*

$$(\alpha - \delta)|\mathcal{K}_k(\mathcal{G})| \leq |\mathcal{H} \cap \mathcal{K}_k(\mathcal{G})| \leq (\alpha + \delta)|\mathcal{K}_k(\mathcal{G})|.$$

Lemma 9 *Let $\mathcal{H}$ be an (α,δ)-regular (k,k)-cylinder. Then* $\operatorname{disc}(\mathcal{H}) \leq 2\delta$.

Lemma 10 *Suppose $\mathcal{H}$ is a (k,k)-cylinder with k-partition $V = V_1 \cup \cdots \cup V_k$. Put $\alpha = d(\mathcal{H})$ and assume that* $\operatorname{disc}(\mathcal{H}) \leq \delta$. *Then $\mathcal{H}$ is $(\alpha, \delta^{1/2})$-regular.*

The k-partite result. Suppose $\mathcal{H}$ is a k-uniform hypergraph and let $\mathcal{H}'$ be a 'typical' k-partite spanning subhypergraph of $\mathcal{H}$. In this section, we relate the discrepancies of $\mathcal{H}$ and $\mathcal{H}'$.

Definition 11 *Let $\mathcal{H} = (V,E)$ be a k-uniform hypergraph with m vertices and let $\mathcal{P} = (V_i)_1^k$ be a partition of V. We denote by $\mathcal{H}_{\mathcal{P}}$ the (k,k)-cylinder consisting of the edges $h \in \mathcal{H}$ satisfying $|h \cap V_i| = 1$ for all $1 \leq i \leq k$.*

The following lemma holds.

Lemma 12 *For any partition $\mathcal{P} = (V_i)_1^k$ of V, we have*

(i) $\operatorname{disc}(\mathcal{H}) \geq |d(\mathcal{H}_{\mathcal{P}}) - d(\mathcal{H})||V_1| \dots |V_k|/m^k$,
(ii) $\operatorname{disc}(\mathcal{H}_{\mathcal{P}}) \leq 2 \operatorname{disc}(\mathcal{H}) m^k / |V_1| \dots |V_k|$.

An immediate consequence of the previous lemma is the following.

Lemma 13 *If* $\operatorname{disc}(\mathcal{H}) = o(1)$, *then* $\operatorname{disc}(\mathcal{H}_{\mathcal{P}}) = o(1)$ *for $(1-o(1))k^m$ partitions $\mathcal{P} = (V_i)_1^k$ of V.*

With some more effort, one may prove a converse to Lemma 13.

Lemma 14 *Suppose there exists a real number $\gamma > 0$ such that* $\operatorname{disc}(\mathcal{H}_{\mathcal{P}}) = o(1)$ *for γk^m partitions $\mathcal{P} = (V_i)_1^k$ of V. Then* $\operatorname{disc}(\mathcal{H}) = o(1)$.

We now state the k-partite version of a part of our main result, Theorem 3.

Theorem 15 *Suppose $V = V_1 \cup \cdots \cup V_k$, $|V_1| = \cdots = |V_k| = n$, and let $\mathcal{H} = (V,E)$ be a (k,k)-cylinder with $|\mathcal{H}| = dn^k$. Then the following four conditions are equivalent:*

(C_1) *$\mathcal{H}$ is $(d, o(1))$-regular;*
(C_2) *$\mathcal{H}(x)$ is $(d, o(1))$-regular for all but $o(n)$ vertices $x \in V_k$ and $\mathcal{H}(x,y)$ is $(d^2, o(1))$-regular for all but $o(n^2)$ pairs x, $y \in V_k$;*

(C_3) *$\mathcal{H}(x,y)$ is $(d^2, o(1))$-regular for all but $o(n^2)$ pairs x, $y \in V_k$;*
(C_4) *the number of copies of $K_k(2)$ in $\mathcal{H}$ is asymptotically minimized among all such (k,k)-cylinders of density d, and equals $(1+o(1))n^{2k}d^{2^k}/2^k$.*

Remark 1. The condition $|V_1| = \cdots = |V_k| = n$ in the result above has the sole purpose of making the statement more transparent. The immediate generalization of Theorem 15 for $V_1, \ldots, V_k$ of arbitrary sizes holds.

Remark 2. The fact that the minimal number of octahedra in a (k,k)-cylinder is asymptotically $(1+o(1))n^{2k}d^{2^k}/2^k$ is not difficult to deduce from a standard application of the Cauchy–Schwarz inequality for counting "cherries" (paths of length 2) in bipartite graphs.

We leave the derivation of the equivalence of properties $P_1, \ldots, P_4$ from Theorem 15 to the full paper.

A regularity lemma. The hardest part in the proof of Theorem 15 is the implication $C_2 \Rightarrow C_1$. In this paragraph, we discuss the main tool used in the proof of this implication. It turns out that, in what follows, the notation is simplified if we consider $(k+1)$-partite hypergraphs.

Throughout this paragraph, we let $\mathcal{G}$ be a fixed $(k+1,k)$-cylinder with vertex set $V(\mathcal{G}) = V_1 \cup \cdots \cup V_{k+1}$. Recall that $\mathcal{G} = \bigcup_{i=1}^{k+1} \mathcal{G}_i$, where $\mathcal{G}_i$ is the corresponding (k,k)-cylinder induced on $\bigcup_{j \neq i} V_j$. In this section, we shall focus on "regularizing" the (k,k)-cylinders $\mathcal{G}_1, \ldots, \mathcal{G}_k$, ignoring $\mathcal{G}_{k+1}$. Alternatively, we may assume that $\mathcal{G}_{k+1} = \emptyset$.

Definition 16 *Let $\mathcal{F} = \bigcup_{i=1}^{k} \mathcal{F}_i$ be a $(k,k-1)$-cylinder with vertex set $V_1 \cup \cdots \cup V_k$. For a vertex $v \in V_{k+1}$, we define the $\mathcal{G}$-link $\mathcal{G}_{\mathcal{F}}(x)$ of x with respect to $\mathcal{F}$ to be the $(k,k-1)$-cylinder $\mathcal{G}_{\mathcal{F}}(x) = \mathcal{G}(x) \cap \mathcal{F}$.*

Definition 17 *Let $W \subset V_{k+1}$ and let $\mathcal{F} = \bigcup_{i=1}^{k} \mathcal{F}_i$ be as above. We shall say that the pair $(\mathcal{F}, W)$ is (ε, d)-regular if*

$$\left| \frac{|\mathcal{K}_k(\mathcal{G}_{\mathcal{F}}(x))|}{|\mathcal{K}_k(\mathcal{F})|} - d \right| < \varepsilon \tag{4}$$

for all but at most $\varepsilon|W|$ vertices $x \in W$, and

$$\left| \frac{|\mathcal{K}_k(\mathcal{G}_{\mathcal{F}}(x)) \cap \mathcal{K}_k(\mathcal{G}_{\mathcal{F}}(y))|}{|\mathcal{K}_k(\mathcal{F})|} - d^2 \right| < \varepsilon \tag{5}$$

for all but at most $\varepsilon|W|^2$ pairs x, $y \in W$.

Definition 18 *Let t be a positive integer and let $V_{k+1} = W_1 \cup \cdots \cup W_t$ be an arbitrary partition of V_{k+1}. For every $i \in [k]$, consider a t-partition $P_i^{(t)} = \{\mathcal{E}_1^{(i)}, \ldots, \mathcal{E}_t^{(i)}\}$ of $V_1 \times \cdots \times V_{i-1} \times V_{i+1} \times \cdots \times V_k = \bigcup_{\alpha=1}^{t} \mathcal{E}_\alpha^{(i)}$. Put $P^{(t)} = (P_1^{(t)}, \ldots, P_k^{(t)})$. We shall write $\boldsymbol{\mathcal{E}}(P^{(t)})$ for the collection of all $(k,k-1)$-cylinders $\mathcal{E}$ of the form $\mathcal{E}_{\alpha_1}^{(1)} \cup \cdots \cup \mathcal{E}_{\alpha_k}^{(k)}$, where $\mathcal{E}_{\alpha_i}^{(i)} \in P_i^{(t)}$ for all $1 \leq i \leq k$.*

Clearly, with the notation as above, we have $|\boldsymbol{\mathcal{E}}(P^{(t)})| = t^k$. Moreover, observe that each of the t^{k+1} pairs $(\mathcal{E}, W_i)$, where $\mathcal{E} \in \boldsymbol{\mathcal{E}}(P^{(t)})$ and $1 \le i \le t$, may be classified as ε-regular or ε-irregular (i.e., not ε-regular), according to Definition 17. Also, notice that each $v = (v_1, \dots, v_{k+1}) \in V_1 \times \cdots \times V_{k+1}$ is 'covered' by exactly one such pair, that is, $v \in \mathcal{K}_k(\mathcal{E}) \times W_i$ for a unique pair $(\mathcal{E}, W_i)$.

Definition 19 *Let $P^{(t)} = \left(P_i^{(t)}\right)_1^k$ and $\left(W_i\right)_1^t$ be as in Definition 18. We shall say that the system of partitions $\left\{P_1^{(t)}, \dots, P_k^{(t)}, \{W_1, \dots, W_t\}\right\}$ is ε-regular if the number of $(k+1)$-tuples $(v_1, \dots, v_{k+1}) \in V_1 \times \cdots \times V_{k+1}$ that are not covered by the family of ε-regular pairs $(\mathcal{E}, W_i)$ with $\mathcal{E} \in \boldsymbol{\mathcal{E}}(P^{(t)})$ and $1 \le i \le t$ is at most $\varepsilon |V_1| \dots |V_{k+1}|$.*

The main tool in the proof of $C_2 \Rightarrow C_1$ is the following result (see [9] for the details).

Lemma 20 *For every $\varepsilon > 0$ and $t_0 \ge 1$, there exist integers n_0 and T_0 such that every $(k+1, k)$-cylinder $\mathcal{G} = \bigcup_{i=1}^{k+1} \mathcal{G}_i$ with vertex set $V_1 \cup \cdots \cup V_{k+1}$, where $|V_i| \ge n_0$ for all $1 \le i \le k+1$, admits an ε-regular system of partitions $\{P_1^{(t)}, \dots, P_k^{(t)}, \{W_1, \dots, W_t\}\}$ with $t_0 < t < T_0$.*

2.2 The subgraph counting formula

In this section, we shall state the main result that may be used to prove the implication $P_1 \Rightarrow P_5'$. To this end, we need to introduce some notation. Throughout this section, $s \ge 2k$ is some fixed integer.

If $\mathcal{H}$ and $\mathcal{G}$ are, respectively, k-uniform and ℓ-uniform ($k \ge \ell$), then we say that $\mathcal{H}$ is *supported* on $\mathcal{G}$ if $\mathcal{H} \subset \mathcal{K}_k(\mathcal{G})$.

Suppose we have pairwise disjoint sets $W_1, \dots, W_s$, with $|W_i| = n$ for all i. Suppose further that we have a sequence $\mathcal{G}^{(2)}, \dots, \mathcal{G}^{(k)}$ of s-partite cylinders on $W_1 \cup \cdots \cup W_s$, with $\mathcal{G}^{(i)}$ an (s,i)-cylinder and, moreover, such that $\mathcal{G}^{(i)}$ is supported on $\mathcal{G}^{(i-1)}$ for all $3 \le i \le k$. Suppose also that, for all $2 \le i \le k$ and for all $1 \le j_1 < \cdots < j_i \le s$, the (i,i)-cylinder $\mathcal{G}[j_1, \dots, j_i] = \mathcal{G}^{(i)}[W_{j_1} \cup \cdots \cup W_{j_i}]$ is (γ_i, δ)-*regular with respect to* $\mathcal{G}^{(i-1)}[j_1, \dots, j_i] = \mathcal{G}^{(i-1)}[W_{j_1} \cup \cdots \cup W_{j_i}]$, that is, whenever $\mathcal{G} \subset \mathcal{G}^{(i-1)}[j_1, \dots, j_i]$ is such that $|\mathcal{K}_i(\mathcal{G})| \ge \delta |\mathcal{K}_i(\mathcal{G}^{(i-1)}[j_1, \dots, j_i])|$, we have

$$(\gamma_i - \delta)|\mathcal{K}_i(\mathcal{G})| \le |\mathcal{G}[j_1, \dots, j_i] \cap \mathcal{K}_i(\mathcal{G})| \le (\gamma_i + \delta)|\mathcal{K}_i(\mathcal{G})|.$$

Finally, let us say that a copy of $K_s^{(k)}$ in $W_1 \cup \cdots \cup W_s$ is *transversal* if $|V(K_s^{(k)}) \cap W_i| = 1$ for all $1 \le i \le s$.

Our main result concerning counting subhypergraphs is then the following.

Theorem 21 *For any $\varepsilon > 0$ and any $\gamma_2, \dots, \gamma_k > 0$, there is $\delta_0 > 0$ such that if $\delta < \delta_0$, then the number of transversal $K_s^{(k)}$ in $\mathcal{G}^{(k)}$ is $(1 + O(\varepsilon))\gamma_k^{\binom{s}{k}} \dots \gamma_2^{\binom{s}{2}} n^s$.*

Theorem 21 above is an instance of certain counting lemmas developed by Rödl and Skokan for such *complexes* $\boldsymbol{\mathcal{G}} = \left(\mathcal{G}^{(i)}\right)_{2 \le i \le k}$ (see, e.g., [11]).

3 Concluding remarks

We hope that the discussion above on our proof approach for Theorem 3 gives some idea about our methods and techniques. Unfortunately, because of space limitations and because we discuss the motivation behind our work in detail, we are unable to give more details. We refer the interested reader to [9].

It is also our hope that the reader will have seen that many interesting questions remain. Probably, the most challenging of them concerns developing an applicable theory of sparse quasi-random hypergraphs. Here, we have in mind such lemmas for sparse quasi-random graphs as the ones in [10].

References

1. L. Babai, N. Nisan, and M. Szegedy, *Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs*, J. Comput. System Sci. **45** (1992), no. 2, 204–232, Twenty-first Symposium on the Theory of Computing (Seattle, WA, 1989).
2. F.R.K. Chung, *Quasi-random classes of hypergraphs*, Random Structures and Algorithms **1** (1990), no. 4, 363–382.
3. F.R.K. Chung and R.L. Graham, *Quasi-random hypergraphs*, Random Structures and Algorithms **1** (1990), no. 1, 105–124.
4. ———, *Quasi-random set systems*, Journal of the American Mathematical Society **4** (1991), no. 1, 151–196.
5. F.R.K. Chung and P. Tetali, *Communication complexity and quasi randomness*, SIAM J. Discrete Math. **6** (1993), no. 1, 110–123.
6. P. Frankl and V. Rödl, *The uniformity lemma for hypergraphs*, Graphs and Combinatorics **8** (1992), no. 4, 309–312.
7. J. Haviland and A.G. Thomason, *Pseudo-random hypergraphs*, Discrete Math. **75** (1989), no. 1–3, 255–278, Graph theory and combinatorics (Cambridge, 1988).
8. ———, *On testing the "pseudo-randomness" of a hypergraph*, Discrete Math. **103** (1992), no. 3, 321–327.
9. Y. Kohayakawa, V. Rödl, and J. Skokan, *Equivalent conditions for regularity*, in preparation, 1999.
10. Y. Kohayakawa, V. Rödl, and E. Szemerédi, *The size-Ramsey number of graphs of bounded degree*, in preparation, 1999.
11. V. Rödl and J. Skokan, *Uniformity of set systems*, in preparation, 1999.
12. E. Szemerédi, *Regular partitions of graphs*, Problèmes Combinatoires et Théorie des Graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976) (Paris), Colloques Internationaux CNRS n. 260, 1978, pp. 399–401.
13. A.G. Thomason, *Pseudorandom graphs*, Random graphs '85 (Poznań, 1985), North-Holland Math. Stud., vol. 144, North-Holland, Amsterdam-New York, 1987, pp. 307–331.
14. ———, *Random graphs, strongly regular graphs and pseudorandom graphs*, Surveys in Combinatorics 1987 (C. Whitehead, ed.), London Mathematical Society Lecture Note Series, vol. 123, Cambridge University Press, Cambridge–New York, 1987, pp. 173–195.
15. R.M. Wilson, *Cyclotomy and difference families in elementary abelian groups*, J. Number Theory **4** (1972), 17–47.