Quasirandomness in hypergraphs

Elad Aigner-Horev [1]

Department of Mathematics and Computer Science, Ariel University, Israel

David Conlon [2,3]

Mathematical Institute, Oxford OX2 6GG, United Kingdom

Hiệp Hàn [4,5]

Departmento de Matemática Ciencia Computación, Universidad de Santiago Chile

Yury Person [6,7]

Goethe-Universität, Institut für Mathematik, Robert-Mayer-Str. 10, 60325 Frankfurt am Main, Germany

Mathias Schacht [8]

Fachbereich Mathematik, Universität Hamburg, Bundesstraße 55, D-20146 Hamburg, Germany

Abstract

A graph G is called quasirandom if it possesses typical properties of the corresponding random graph $G(n,p)$ with the same edge density as G. A well-known theorem of Chung, Graham and Wilson states that, in fact, many such 'typical' properties are asymptotically equivalent and, thus, a graph G possessing one property immediately satisfies the others.

In recent years, more quasirandom graph properties have been found and extensions to hypergraphs have been explored. For the latter, however, there exist several distinct notions of quasirandomness. A complete description of these notions has been provided recently by Towsner, who proved several central equivalences using an analytic framework. The purpose of this paper is to give short purely combinatorial proofs of most of Towsner's results.

Keywords: Hypergraphs, quasirandomness.

1 Introduction

Quasirandomness studies typical properties that a random structure satisfies with high probability. It has found numerous applications in combinatorics and theoretical computer science. We refer the reader to the excellent survey [4]. A prime example, which has received particular attention, is the notion of *quasirandom graphs*. Let $(G_n)_{n\in\mathbb{N}}$ be a sequence of graphs, where G_n is a graph on n vertices. For a fixed $p \in [0,1]$, we say that (G_n) is p-quasirandom if G_n has *uniform edge distribution*:

$$e(G_n[S]) = p\binom{|S|}{2} + o(n^2) \text{ for every } S \subseteq V(G_n), \tag{1}$$

where $e(G_n[S])$ denotes the number of edges in the induced subgraph $G_n[S]$. The property above is often referred to as *discrepancy*. The seminal result of Chung, Graham and Wilson [1] states that (1) is a *quasirandom property* in the sense that satisfying property (1) is asymptotically equivalent to several other properties typically satisfied by the random graph $G(n,p)$. In particular, having uniform edge distribution is asymptotically equivalent to the property:

[1] Email: horev@ariel.ac.il
[2] Email: David.Conlon@maths.ox.ac.uk
[3] The second author was supported by a Royal Society University Research Fellowship and by ERC Starting Grant 676632.
[4] Email: hiep.han@pucv.cl
[5] The third author was supported by the FONDECYT Iniciación grant 11150913 and by Millenium Nucleus Information and Coordination in Networks.
[6] Email: person@math.uni-frankfurt.de
[7] The fourth author was supported by DFG grant PE 2299/1-1.
[8] Email: schacht@math.uni-hamburg.de

$$e(G_n) = p\binom{n}{2} + o(n^2) \text{ and } N_{G_n}(C_4) = p^4n^4 + o(n^4), \tag{2}$$

where $N_{G_n}(C_4)$ denotes the number of labeled copies of C_4, the cycle of length 4, in G_n. This is somewhat surprising, as (2) seems at first glance to be a much weaker condition. It is not difficult to show that *any* graph G_n on n vertices of edge density p contains at least $p^4n^4 + o(n^4)$ labeled copies of C_4. Thus, a graph sequence (G_n) is quasirandom if it is a minimiser for the number of copies of C_4. Another quasirandom property for graphs is that of knowing all densities of graphs on $\ell \geq 4$ vertices, for a fixed ℓ:

$$N_{G_n}(F) = p^{e(F)}n^{v(F)} + o(n^{v(F)}) \text{ for all graphs } F \text{ on } \ell \geq 4 \text{ vertices}, \tag{3}$$

where again $N_{G_n}(F)$ denotes the number of labeled copies of F and $v(F)$ and $e(F)$ are the number of vertices and edges in F, respectively.

Generalising the above results to hypergraphs is somewhat delicate. For example, Rödl [8] observed that straightforward generalisations of (1) and (3) are not equivalent, while a generalisation of (2) is anything but clear.

Let $(H_n)_{n\in\mathbb{N}}$ be a sequence of k-uniform hypergraphs, i.e. pairs (V_n, E_n) with $E_n \subseteq \binom{V_n}{k}$. The straightforward generalisation of (3), namely:

$$N_{G_n}(F) = p^{e(F)}n^{v(F)} + o(n^{v(F)}) \text{ for all graphs } F \text{ on } \ell \geq 2k \text{ vertices}, \tag{4}$$

was shown to require a *stronger notion* of discrepancy for (H_n) than merely:

$$e(H_n[S]) = p\binom{|S|}{k} + o(n^k) \text{ for every } S \subseteq V(H_n). \tag{5}$$

Instead, one needs to count edges with respect to all '$(k-1)$-uniform hypergraphs' G:

$$e(H_n[S]) = p|\mathcal{K}_k(G)| + o(n^k) \text{ for every } (k-1)\text{-uniform } G \text{ on } V(H_n), \tag{6}$$

where $\mathcal{K}_k(G)$ is the number of cliques on k vertices that are present in G. A significant body of work regarding the relationship between (4), (5), and (6) exists, though we omit references due to space limitations.

We pick up the trail of results here with [2], where the equivalence of (5) and (4) is established if the requirement of the latter is weakened to linear hypergraphs F. Several additional quasirandom properties equivalent to (5) were established in [2]. In particular, an analogue of (2) is found. Furthermore, in [2], some guesses are made regarding other possible notions of hypergraph discrepancy of intermediate strength and their corresponding minimising hypergraphs. Subsequently, Lenz and Mubayi [5,6,7] extended these results by adding an unexpected *spectral property* and providing additional equivalences for notions of hypergraph quasirandomness of intermediate strength.

Finally, Towsner [9] proved a general theorem which provides, for *any reasonable notion* of hypergraph quasirandomness, generalisations of (1), (2) and (3) that are all equivalent. This he accomplishes using the language of non-standard analysis and hypergraph limits. By generalising constructions of Lenz and Mubayi [7], he also showed that these quasirandomness notions are all distinct, again using analytic language. Towsner remarks that finitizing his arguments appears rather difficult. The purpose of this paper is to do just that and to provide short combinatorial proofs for most of Towsner's results.

2 Definitions and the main result

For a finite set X, we write $\overrightarrow{X}$ to denote the set of all orderings of the members of X. For an integer $k \geq 1$ and a set V, the set of all k-element subsets of V is denoted by $\binom{V}{k}$ and we write $\binom{V}{k}_<$ to denote $\overrightarrow{\binom{V}{k}}$. Given a set (of indices) $Q \subseteq [k]$ we write V^Q to denote $V^{|Q|}$ and refer to its members as *Q-tuples*. Unlike the members of $\binom{V}{k}_<$, Q-tuples may contain non-distinct entries. By a *Q-directed hypergraph*, we mean a pair (V, E) where $E \subseteq V^Q$.

Let $\mathcal{Q} \subseteq 2^{[k]}$ be a set system. For a collection (with elements possibly repeated) $\mathcal{G} = (G_Q)_{Q \in \mathcal{Q}}$ of Q-directed hypergraphs G_Q, let $\mathcal{K}_k(\mathcal{G}) \subseteq \binom{V}{k}_<$ consist of all ordered k-tuples $\boldsymbol{v} = (v_1, \ldots, v_k)$ satisfying $\boldsymbol{v}_Q := (v_i \colon i \in Q) \in E(G_Q)$ for all $Q \in \mathcal{Q}$. Such tuples are said to be *supported* by $\mathcal{G}$ and $\mathcal{K}_k(\mathcal{G})$ is thus referred to as the *support* of $\mathcal{G}$.

Below we define the properties that will be of interest to us. We begin with a generalized notion of discrepancy.

Definition 2.1 [$\mathrm{DISC}_{\mathcal{Q},d}$] For an integer $k \geq 2$, a set system $\mathcal{Q} \subseteq 2^{[k]}$, and reals $\varepsilon > 0$, $d \in [0, 1]$, we say a k-uniform hypergraph $H = (V, E)$ with $|V| = n$ satisfies $\mathrm{DISC}_{\mathcal{Q},d}(\varepsilon)$ if, for every sequence $\mathcal{G} = (G_Q)_{Q \in \mathcal{Q}}$,

$$\Big| |\overrightarrow{E} \cap \mathcal{K}_k(\mathcal{G})| - d|\mathcal{K}_k(\mathcal{G})| \Big| \leq \varepsilon n^k,$$

where G_Q is a Q-directed hypergraph (for every $Q \in \mathcal{Q}$).

We also consider the following weighted version of $\mathrm{DISC}_{\mathcal{Q},d}$, where the sequence of directed hypergraphs $\mathcal{G}$ is replaced by an ensemble of functions

$\mathcal{W} = \big(w_Q\colon V^Q \to [-1,1]\big)_{Q\in\mathcal{Q}}$ and the set of supported tuples $\mathcal{K}_k(\mathcal{G})$ is replaced with the function $\mathcal{W}\colon V^k \to [-1,1]$ given by $\mathcal{W}(\boldsymbol{v}) = \prod_{Q\in\mathcal{Q}} w_Q(\boldsymbol{v}_Q)$, where we set $w_Q(\boldsymbol{v}_Q)$ to be zero whenever $\boldsymbol{v}_Q$ is not a proper set, i.e., whenever it has any non-distinct entries.

Definition 2.2 [$\mathrm{WDISC}_{\mathcal{Q},d}$] For an integer $k \geq 2$, a set system $\mathcal{Q} \subseteq 2^{[k]}$, and reals $\varepsilon > 0$, $d \in [0,1]$, we say a k-uniform hypergraph $H = (V,E)$ with $|V| = n$ satisfies $\mathrm{WDISC}_{\mathcal{Q},d}(\varepsilon)$ if, for every ensemble of (weight) functions $\mathcal{W} = (w_Q)_{Q\in\mathcal{Q}}$ with $w_Q\colon V^Q \to [-1,1]$ for every $Q \in \mathcal{Q}$,

$$\left| \sum_{\boldsymbol{v}\in V^k} \big(\mathbb{1}_{\overrightarrow{E}}(\boldsymbol{v}) - d\big)\mathcal{W}(\boldsymbol{v}) \right| \leq \varepsilon n^k \,,$$

where $\mathbb{1}_{\overrightarrow{E}}\colon V^k \to \{0,1\}$ denotes the indicator function of $\overrightarrow{E}$.

When $w_Q = \mathbb{1}_{G_Q}$ for every $Q \in \mathcal{Q}$, the difference between $\sum_{\boldsymbol{v}\in V^k} \big(\mathbb{1}_{\overrightarrow{E}}(\boldsymbol{v}) - d\big)\mathcal{W}(\boldsymbol{v})$ and $|\overrightarrow{E} \cap \mathcal{K}_k(\mathcal{G})| - d|\mathcal{K}_k(\mathcal{G})|$ is exactly the number of $\boldsymbol{v}$ which have some non-distinct entries, yet are supported by $\mathcal{G}$ (times d). However, this difference has order of magnitude $O_k(n^{k-1})$, so hypergraphs H satisfying $\mathrm{WDISC}_{\mathcal{Q},d}(\varepsilon)$ must also satisfy $\mathrm{DISC}_{\mathcal{Q},d}(2\varepsilon)$. The opposite implication follows by an averaging argument similar to one of Gowers [3, Section 3].

To state the analogues of (2) and (3) corresponding to $\mathrm{DISC}_{\mathcal{Q},d}$, we require some further notation. Given $\mathcal{Q} \subseteq 2^{[k]}$, we say a k-uniform hypergraph $F = (V_F, E_F)$ is *$\mathcal{Q}$-simple* if there exists an ordering of its edges $E_F = \{e_1, \ldots, e_m\}$ such that for every $i = 1, \ldots, m$ there exists an ordering of $e_i = \{v_{i_1}, \ldots, v_{i_k}\}$ such that for every $h < i$ there exists a set $Q \in \mathcal{Q}$ such that $\{r\colon\ v_{i_r} \in e_h \cap e_i\} \subseteq Q$. Note that the orderings of the vertices for every edge of F can be chosen independently and may be incompatible with each other. The analogue of (3) is now as follows.

Definition 2.3 [$\mathrm{CL}_{\mathcal{Q},d}$] For an integer $k \geq 2$, a subset $\mathcal{Q} \subseteq 2^{[k]}$, reals $\varepsilon > 0$, $d \in [0,1]$, and a $\mathcal{Q}$-simple k-uniform hypergraph $F = (V_F, E_F)$, we say a k-uniform hypergraph $H = (V,E)$ with $|V| = n$ satisfies $\mathrm{CL}_{\mathcal{Q},d}(F,\varepsilon)$ if the number $N_H(F)$ of labeled copies of F in H satisfies

$$\left| N_H(F) - d^{|E_F|} n^{|V_F|} \right| \leq \varepsilon n^{|V_F|} \,.$$

For a k-partite k-uniform hypergraph F with vertex partition $V(F) = X_1 \dot\cup \ldots \dot\cup X_k$ and a set $Q \subseteq [k]$, the *Q-doubling of F* is the hypergraph $\mathrm{db}_Q(F)$

obtained by taking two copies of F and identifying the vertex classes indexed by elements in Q, i.e., the vertex set of the Q-doubling is

$$V(\mathrm{db}_Q(F)) = Y_1\dot\cup\ldots\dot\cup Y_k \quad \text{where} \quad Y_q = \begin{cases} X_q & \text{if } q\in Q\,, \\ X_q\times\{0,1\} & \text{if } q\notin Q\,, \end{cases}$$

and the edges of the Q-doubling are given by the set $\{x_q\colon\ q\in Q\}\dot\cup\{(x_r,a)\colon\ r\in[k]\setminus Q, a\in\{0,1\}\}$ where $\{x_1,\ldots,x_k\}\in E(F)$.

It is easy to check that for two sets Q, $R\subseteq[k]$ and a k-partite, k-uniform hypergraph F the ordering of the doubling operations does not matter, i.e., $\mathrm{db}_Q(\mathrm{db}_R(F)) = \mathrm{db}_R(\mathrm{db}_Q(F))$. Hence, for $\mathcal{Q}\subseteq 2^{[k]}\setminus\{[k]\}$ (the operation $\mathrm{db}_{[k]}$ leaves the hypergraph unchanged), we may define the k-partite k-uniform hypergraph $M_{\mathcal{Q}}$ recursively by setting $M_\emptyset = K_k^{(k)}$, the k-uniform hypergraph consisting of one edge (viewed as k-partite), and, for any $Q\in\mathcal{Q}$, define $M_{\mathcal{Q}} = \mathrm{db}_Q(M_{\mathcal{Q}\setminus\{Q\}})$. It follows from these definitions that $M_{\mathcal{Q}}$ has $2^{|\mathcal{Q}|}$ hyperedges and $\sum_{i=1}^k 2^{|\mathcal{Q}|-\deg_{\mathcal{Q}}(i)}$ vertices, where $\deg_{\mathcal{Q}}(i)$ denotes the number of sets of $\mathcal{Q}$ containing the element i. In analogy with the C_4-case for graphs, we may apply the Cauchy–Schwarz inequality for every $Q\in\mathcal{Q}$ to show that every k-uniform hypergraph H on n vertices with density $d>0$ contains at least $(d^{|E(M_{\mathcal{Q}})|}-o(1))n^{|V(M_{\mathcal{Q}})|}$ labeled copies of $M_{\mathcal{Q}}$. The analogue of (2) is now as follows.

Definition 2.4 [$\mathrm{MIN}_{\mathcal{Q},d}$] For an integer $k\geq 2$, a subset $\mathcal{Q}\subseteq 2^{[k]}$, and reals $\varepsilon>0$, $d\in[0,1]$, we say a k-uniform hypergraph $H=(V,E)$ with $|V|=n$ satisfies $\mathrm{MIN}_{\mathcal{Q},d}(\varepsilon)$ if

(i) the density $d(H) = |E|/\binom{n}{k}$ satisfies $d(H)\geq d-\varepsilon$ and

(ii) the number $N_H(M_{\mathcal{Q}})$ of labeled copies of $M_{\mathcal{Q}}$ in H satisfies

$$N_H(M_{\mathcal{Q}}) \leq (d^{|E(M_{\mathcal{Q}})|}+\varepsilon)n^{|V(M_{\mathcal{Q}})|}\,.$$

It is sometimes more convenient to work with the following weighted version of $\mathrm{MIN}_{\mathcal{Q},d}$.

Definition 2.5 [$\mathrm{DEV}_{\mathcal{Q},d}$] For an integer $k\geq 2$, a subset $\mathcal{Q}\subseteq 2^{[k]}$, and reals $\varepsilon>0$, $d\in[0,1]$, we say a k-uniform hypergraph $H=(V,E)$ with $|V|=n$ satisfies $\mathrm{DEV}_{\mathcal{Q},d}(\varepsilon)$ if

$$\sum_M \prod_{e\in E(M)} (\mathbb{1}_E(e)-d) \leq \varepsilon n^{|V(M_{\mathcal{Q}})|}\,,$$

where the sum ranges over all labeled copies of $M_{\mathcal{Q}}$ in the complete k-uniform hypergraph $K_V^{(k)}$ on the vertex set V.

Our main result, relating all of the notions described above, is as follows.

Theorem 2.6 (Main result) *For every $k \geq 2$, every set system $\mathcal{Q} \subseteq 2^{[k]} \setminus \{[k]\}$, and $d \in [0,1]$, the properties $\mathrm{DISC}_{\mathcal{Q},d}$, $\mathrm{WDISC}_{\mathcal{Q},d}$, $\mathrm{CL}_{\mathcal{Q},d}$, and $\mathrm{DEV}_{\mathcal{Q},d}$ are all equivalent.*

As mentioned above, Towsner [9, Section 9] also provides constructions distinguishing the notions of quasirandomness provided by distinct choices of $\mathcal{Q}$ from one another. We again address this issue in a purely combinatorial manner.

References

[1] Chung, F. R. K., Graham, R. L., and Wilson, R. M., Quasi-random graphs, Combinatorica 9 (1989) no. 1 345-124.

[2] Conlon, D., Hàn, H., Person, Y., and Schacht, M., Weak quasi-randomness for uniform hypergraphs, Random Structures & Algorithms 40 (2012) no. 1 1-38.

[3] Gowers, W. T., Quasirandomness, counting and regularity for 3-uniform hypergraphs, Combin. Probab. Comput. 15 (2006) no. 1-2 143-184.

[4] Krivelevich, M. and Sudakov, B., Pseudo-random graphs, in: E. Györa, G. O. H. Katona, L. Lovásg, T. Fleiner (Eds.), More sets, graphs and numbers, Springer, Berlin, 2006, pp. 199-262.

[5] Lenz, J. and Mubayi, D., Eigenvalues of non-regular linear quasirandom hypergraphs, Discrete Math. 340 (2017) no. 2 145-153.

[6] Lenz, J. and Mubayi, D., Eigenvalues and linear quasirandom hypergraphs, Forum Math. Sigma 3 (2015) e2-26.

[7] Lenz, J. and Mubayi, D., The poset of hypergraph quasirandomness, Random Structures & Algorithms 46 (2015) no. 4 762-800.

[8] Rödl, V., On universality of graphs with uniformly distributed edges, Discrete Math. 59 (1986) no. 1-2 125-134.

[9] Towsner, H., σ-algebras for quasirandom hypergraphs, Random Structures & Algorithms 50 (2017) no. 1 1098-2418.