Regularity for hypergraphs with bounded VC_2 dimension

Lior Gishboliner[*]   Asaf Shapira[†]   Yuval Wigderson[‡]

Abstract

While Szemerédi's graph regularity lemma is an indispensable tool for studying extremal problems in graph theory, using it comes with a hefty price, since in the worst case, a graph may only have regular partitions of tower-type order. It is thus sensible to ask if there is some natural restriction which forces graphs to have much smaller regular partitions. A celebrated result of this type, due to Alon–Fischer–Newman and Lovász–Szegedy, states that for graphs of bounded VC dimension, one can reduce the tower-type bounds to polynomial.

The graph regularity lemma has been extended to the setting of k-graphs by Gowers, Nagle–Rödl–Schacht–Skokan, and Tao. Unfortunately, these lemmas come with even larger Ackermann-type bounds. Chernikov–Starchenko and Fox–Pach–Suk considered a strong notion of k-graph VC dimension and proved that k-graphs of bounded VC dimension have regular partitions of polynomial size. Shelah introduced a weaker and combinatorially more natural notion of dimension, called VC_2 dimension, which has been extensively studied in the past decade. In particular, Chernikov, Towsner, Terry, and Wolf asked if one can improve the worst case bounds for 3-graph regularity when the 3-graph has bounded VC_2 dimension. Our main result in this paper answers this question positively in the following strong sense: we show that in the setting of bounded VC_2 dimension, one can reduce the bounds for 3-graph regularity by one level in Ackermann hierarchy. Furthermore, our new improved bound is best possible.

Our proof has two key steps. We first introduce a novel method for designing regularity lemmas for graphs of bounded VC dimension, based on the cylinder regularity lemma. We then prove a hypergraph version of the cylinder regularity lemma, which allows us to extend this new method to hypergraphs. We additionally highlight a few other applications of this cylinder regularity lemma, which we expect to find many other uses.

1   Introduction

1.1   Background

1.1.1   Graph regularity

One of the deepest and most powerful results in modern graph theory is Szemerédi's regularity lemma [57]. Roughly speaking, it states that any graph can be approximated as the union of a constant number of "random-like" graphs, where the number of such pieces depends only on the

quality of the approximation and on how random-like we demand the pieces to be. This foundational result has found many applications in graph theory, number theory, theoretical computer science, and beyond.

To state (one version of) the regularity lemma, let X, Y be vertex subsets of some graph G. We denote by $e(X, Y)$ the number of pairs in $X \times Y$ that are edges of G, and by $d(X, Y) := e(X, Y)/(|X||Y|)$ the *edge density* of the pair. We say that the pair (X, Y) is *ε-quasirandom* if, for all $A \subseteq X, B \subseteq Y$, we have that $|e(A, B) - d(X, Y)|A||B|| \leqslant \varepsilon |X||Y|$. Szemerédi's regularity lemma then states that the vertex set of any graph G can be partitioned as $V(G) = V_1 \cup \cdots \cup V_K$ such that at least a $(1 - \varepsilon)$-fraction of the pairs $(x, y) \in V(G)^2$ lie in an ε-quasirandom pair (V_i, V_j), and moreover such that K is a constant depending only on ε.

Szemerédi's original proof shows that K can be upper-bounded as a tower function[1] of $\varepsilon^{-O(1)}$, and a famous construction of Gowers [33] shows that such a tower-type bound is necessary in the worst case. Unfortunately, in many applications, these tower-type losses are far too expensive to invoke. This naturally leads to the following fundamental question.

Question 1.1. *Are there natural conditions which force a graph G to have regular partitions of reasonable size?*

For example, one can ask whether bipartite graphs have smaller ε-regular partitions, and the answer is no: it is well-known that Gowers's [33] construction can be modified to produce a bipartite graph requiring regular partitions of tower-type size. On the other hand, it is known that if G is *semialgebraic of bounded description complexity*—a natural structural assumption which holds for many "geometrically-defined" graphs—then the bound in the regularity lemma can be improved from tower-type to polynomial [26].

1.1.2   VC dimension

One of the most powerful and well-studied notions of "low complexity" for discrete objects is the Vapnik–Chervonenkis (VC) dimension. This notion was originally introduced by Vapnik and Chervonenkis [67] in the context of statistical learning theory, and over the past 50 years has become a central notion in a wide variety of areas within computer science and mathematics. Since the literature on applications of VC dimension is so vast, in what follows we will briefly mention only a handful of results related to our investigation here.

First, in computer science, VC dimension is one of the most important concepts in learning theory. Indeed, the fundamental theorem of PAC learning [8] states that a binary class is learnable in Valiant's PAC model [66] if and only if the class has finite VC dimension. This foundational result lies at the heart of much of learning theory; in particular, it is remarkable that learnability—a fundamentally probabilistic and approximate notion—is characterized by a purely combinatorial notion of dimension. This result has led to a number of extensions, such as [4, 9], connecting notions of learning with combinatorial measures of dimension. Moreover, VC dimension has many applications in other areas of computer science; for example, in computational geometry, VC dimension is of fundamental importance because of its relation to ε-nets (see e.g. [46, Chapter 10]), which are in turn useful for many applications including range searching.

VC dimension also has many connections to and applications within mathematics. In model theory, VC dimension is intimately connected to the notion of stable models [53], and gives an

---

[1]We recall that the *tower function* is defined by $\mathrm{twr}(0) = 1$ and $\mathrm{twr}(x + 1) = 2^{\mathrm{twr}(x)}$ for $x \geqslant 0$.

equivalent characterization of so-called NIP theories. In group theory, VC dimension plays an important role in the proof of the Pillay conjectures on groups definable in o-minimal theories [39] (see also [17]). Finally, in combinatorics, which is the focus of the present paper, there are many different results demonstrating that discrete objects with bounded VC dimension are "structurally simple", and hence many difficult problems become tractable when restricted to objects of bounded VC dimension, e.g. [6, 25, 27, 28, 40, 41, 44]. Many of these deep combinatorial theorems rely, at their core, on simple properties enjoyed by objects of bounded VC dimension, such as the fundamental Sauer–Shelah lemma [52, 53] and its consequences.

The formal definition of VC dimension is as follows. Let $\mathcal{F} \subseteq 2^V$ be a collection of subsets of some universe V. The *VC dimension* of $\mathcal{F}$ is defined to be the largest size of a set $S \subseteq V$ with the property that S is *shattered* by $\mathcal{F}$, meaning that each of the $2^{|S|}$ possible subsets of S are attained as intersections $S \cap F$ for some $F \in \mathcal{F}$. Given a graph G, its VC dimension is then defined as the VC dimension of the set system $\{N(v) : v \in V(G)\} \subseteq 2^{V(G)}$, where $N(v)$ denotes the neighborhood of a vertex v.

1.1.3 Question 1.1 for graphs of bounded VC dimension

With regards to Question 1.1, graphs of bounded VC dimension are particularly well-behaved with respect to the regularity lemma. Indeed, such graphs have regularity partitions with only a *polynomial* number of parts. Moreover, in this regularity partition, we obtain much stronger information than simply saying that almost all pairs are ε-quasirandom. Indeed, let us say that a pair (X, Y) is ε-*homogeneous* if $d(X, Y) \in [0, \varepsilon] \cup [1 - \varepsilon, 1]$; it is easy to see that this is a strictly stronger notion than ε-quasirandomness (up to a polynomial change in ε). The *ultra-strong regularity lemma* for graphs of bounded VC dimension then states the following.

Theorem 1.2. *Let G be a graph with VC dimension d, and let $\varepsilon > 0$. There is a partition $V(G) = V_1 \cup \cdots \cup V_K$, where $K = (1/\varepsilon)^{O(d)}$, such that at least a $(1 - \varepsilon)$-fraction of the pairs $(x, y) \in V(G)^2$ lie in ε-homogeneous pairs (V_i, V_j).*

Such a result was independently proved by Alon–Fischer–Newman [2] and by Lovász–Szegedy [43], and was subsequently quantitatively improved by Fox–Pach–Suk [27]; the result quoted above is due to Fox–Pach–Suk, and is essentially best possible.

In what follows, we shall always view d, the VC dimension of G, as a fixed constant. Thus, Theorem 1.2 states that a graph of bounded VC dimension has a partition into $\mathrm{poly}(1/\varepsilon)$ many parts, such that all but an ε-fraction of the pairs of vertices lie in ε-homogeneous pairs.

We remark here, for future convenience, that there is a more combinatorial formulation of the property that G has bounded VC dimension. Indeed, let F be a bipartite graph with parts A, B. We say that F is a *bipartitely induced subgraph* of G if there exist disjoint $A', B' \subseteq V(G)$ such that a pair $(a, b) \in A' \times B'$ is an edge of G if and only if the corresponding pair in $A \times B$ is an edge of F. Equivalently, this says that G contains an induced subgraph isomorphic to some supergraph of F, obtained by only adding new edges in the parts A, B, without adding or removing any edges across the bipartition.

With this definition, it is clear that G has VC dimension less than d if and only if F_d is not a bipartitely induced subgraph of G, where F_d is the bipartite graph with parts $A = [d]$ and $B = 2^{[d]}$, where every subset of A is the neighborhood of some vertex in B. On the other hand, it is also easy to see that if G has VC dimension greater than $d + \log d$, then for every bipartite graph F with both parts of size d, F is a bipartitely induced subgraph of G.

3

Thus, we conclude that G has bounded VC dimension if and only if there exists some fixed bipartite graph F which is not a bipartitely induced subgraph of G. This formulation will be particularly instructive and useful when we consider the corresponding concepts in hypergraphs.

Let us finally mention that a variant of Theorem 1.2 was obtained by Malliaris and Shelah [45]. In this version, the graph G is assumed not to contain bipartitely induced copies of a specific bipartite graph (the so-called *half graph*), but the conclusion is stronger than the one in Theorem 1.2, since in this case one can guarantee that all pairs in the partition are ε-quasirandom. Variants of the Malliaris–Shelah theorem were also studied in the setting of Green's arithmetic regularity lemma [36], see [3, 64].

1.1.4 Hypergraph regularity and VC$_2$ dimension

After the introduction of Szemerédi's regularity lemma, there was great interest in extending this technique to hypergraphs. There is a natural analogue of the notion of ε-quasirandomness to k-uniform hypergraphs (henceforth k-graphs), and Chung [15] proved (via a natural extension of Szemerédi's method) that every k-graph can be partitioned into a constant number of pieces, such that almost all k-tuples of vertices lie in quasirandom tuples; we discuss this notion in detail in Definition 2.7 and Theorem 2.8. Unfortunately, it was quickly discovered that this notion of hypergraph regularity is not sufficient for many applications (see the discussion before Theorem 6.2). More precisely, this notion is too weak to support a *counting lemma*, which is a key ingredient of most applications of the regularity method. Thus, it remained a major research program to develop a notion of hypergraph regularity that is simultaneously strong enough to support a counting lemma and weak enough to support a regularity lemma.

This was accomplished independently in the mid-2000s, most notably by Gowers [35] and by Rödl et al. [50]. A major insight in the development of hypergraph regularity (going back to the early work of Frankl and Rödl [30]) is that, in uniformity $k \geqslant 3$, it is no longer sufficient to simply partition the vertices. Instead, in uniformity k, one must partition the vertices, the pairs of vertices, the triples of vertices, and so on, up through the $(k-1)$-tuples of vertices. This introduces a large number of conceptual, technical, and notational difficulties. As such, we defer most formal definitions and statements for now, and continue a high-level discussion during this introduction.

Our main focus will be on uniformity 3, so let $\mathcal{H}$ be a 3-graph. The hypergraph regularity lemma requires us to partition its vertices $V(\mathcal{H})$, as well as the pairs $V(\mathcal{H})^2$; these two partitions are called the *vertex partition* and *edge partition*, respectively. Each part of this latter partition can be naturally viewed as a graph, and the quasirandomness notion states very roughly that for most of these graphs G_i, the hyperedges of $\mathcal{H}$ form a "random-like subset" of the set of triangles in G_i. In order for this to be a useful notion, we also need the triangles in G_i to behave in a random-like fashion, hence one needs to impose the assumption that each part G_i of the edge partition is itself a quasirandom graph.

For a graph G, the notion of ε-quasirandomness is only meaningful if ε is much smaller than the edge density of G. Hence, in the 3-graph regularity lemma, we have both a parameter ε—which determines how random-like the hyperedges of $\mathcal{H}$ are with respect to the triangles in G_i—and a function $\psi : (0,1) \to (0,1)$, which determines how quasirandom the graphs G_i are, as a function of their edge densities. In principle ψ can be any function, but in most applications, including the counting and removal lemmas, and thus in the proofs of important consequences such as the multidimensional Szemerédi theorem [35], ψ can be taken to be a polynomial.

The proof of the 3-graph regularity lemma proceeds in an iterative fashion, where we refine

the two partitions at every step. During a single iteration, the quasirandomness of the edge parts G_i may be destroyed, hence at every step of the iteration, one must apply Szemerédi's regularity lemma to further refine the partition and recover the quasirandomness. Because of this, the bounds in the hypergraph regularity lemma deteriorate very quickly: even in the case of uniformity 3, one must iteratively invoke tower-type losses, leading to a final bound of wowzer[2] type. Note that this happens "almost" regardless of what the function ψ is; as long as $1/\psi$ grows slower than the tower function, the contribution of ψ is completely overwhelmed by the other tower-type losses at each step. However, if $1/\psi$ grows extremely quickly, the 3-graph regularity lemma has even worse bounds. For example, if $1/\psi$ grows like the wowzer function, then the hypergraph regularity lemma would yield bounds of A_3 type, where A_3 (the iterated wowzer function) is the next level of the Ackermann hierarchy. For more details, see e.g. [62, Appendix A.1].

Recently, Moshkovitz and Shapira [47] proved that such terrible bounds are in fact necessary. Concretely, in the case of uniformity 3, they proved the existence of a 3-graph $\mathcal{H}$ which has no ε-regular partition (for ψ a polynomial) using fewer than $\mathrm{wow}(\log(1/\varepsilon))$ vertex parts.

Given the prior discussion, it is unsurprising that certain structural assumptions on the hypergraph $\mathcal{H}$ would imply better bounds. For example, it is known that, just as in the graph case, the hypergraph regularity lemma has merely polynomial bounds if $\mathcal{H}$ is semialgebraic of bounded description complexity. As proved by Chernikov–Starchenko [13] and by Fox–Pach–Suk [27], the same conclusion also holds if $\mathcal{H}$ is assumed to have bounded VC dimension, for a strong notion of VC dimension that arises naturally from geometric considerations. For a different notion of VC dimension, called *slicewise VC dimension*, Terry [60] proved that one can obtain a regularity partition with only a double-exponential number of parts, and the authors [32] recently improved her bound to single-exponential, which Terry [61, Theorem 6.8] showed to be best possible.

However, from a purely combinatorial perspective, arguably the most natural extension of VC dimension from graphs to 3-graphs proceeds by generalizing the characterization in terms of forbidden bipartitely induced subgraphs. That is, we say that a 3-graph $\mathcal{H}$ has *bounded VC_2 dimension* if there exists a fixed tripartite 3-graph $\mathcal{V}$ which does not appear as a tripartitely induced subgraph[3] of $\mathcal{H}$. This definition has its origins in model theory, where it was introduced by Shelah [54], and where it was quickly understood to be the "correct" higher-uniformity analogue of VC dimension (see e.g. [7, 11, 12, 38, 55, 58]), in the sense that it is a "low-complexity" notion which picks up on the 3-uniform nature of the setting.

As in the graph case, this definition is equivalent to a definition in terms of shattering.

Definition 1.3. Let $\mathcal{H}$ be a hypergraph and $d \geqslant 1$ an integer. A *shattered $K_{d,d}$* in $\mathcal{H}$ is a pair of disjoint sets $A, B \subseteq V(\mathcal{H})$ with $|A| = |B| = d$ with the property that, for every $S \subseteq A \times B$, there is a vertex $v \in V(\mathcal{H})$ such that $(a, b, v) \in E(\mathcal{H})$ if and only if $(a, b) \in S$. The *VC_2 dimension* of $\mathcal{H}$ is then defined to be the maximum d such that $\mathcal{H}$ contains a shattered $K_{d,d}$.

Clearly, $\mathcal{H}$ contains a shattered $K_{d,d}$ if and only if $\mathcal{H}$ contains $\mathcal{V}_d$ as a tripartitely induced subgraph, where $\mathcal{V}_d$ is the tripartite 3-graph with vertex sets A, B, C, where $A = B = [d]$ and $C = 2^{[d]^2}$, and where a triple (a, b, S) is an edge if and only if $(a, b) \in S$. In the other direction, it is easy to see that if $\mathcal{H}$ has VC_2 dimension at least $d + \log d$, then it contains every tripartite

---

[2]We recall that the *wowzer function* is defined by $\mathrm{wow}(0) = 1$ and $\mathrm{wow}(x+1) = \mathrm{twr}(\mathrm{wow}(x))$ for $x \geqslant 0$.

[3]Extending the graph terminology, we say that $\mathcal{V}$ is a tripartitely induced subgraph of $\mathcal{H}$ if there exist disjoint $A, B, C \subseteq V(\mathcal{H})$, corresponding to the three vertex parts of $\mathcal{V}$, such that each triple $(a, b, c) \in A \times B \times C$ is an edge of $\mathcal{H}$ if and only if the corresponding triple is an edge of $\mathcal{V}$. That is, we only care about edges which go between all three parts, and we don't care if $\mathcal{H}$ has additional edges entirely within one part, or only between two parts.

3-graph with parts of size d as a tripartitely induced subgraph. In this paper, we will mostly use the formulation in terms of forbidden tripartitely induced subgraphs, as this is a natural formulation from a graph-theoretic perspective, and since it is the most convenient when working with the regularity lemma.

1.2 Hypergraph regularity under bounded VC_2 dimension

Following the discussion above, it is natural to ask about the quantitative aspects of the 3-graph regularity lemma if one imposes the restriction that the 3-graph $\mathcal{H}$ has bounded VC_2 dimension; this question was first raised by Chernikov and Towsner [14], and asked explicitly by Terry [59]. In particular, one could expect that the assumption of bounded VC_2 dimension would allow one to improve on the wowzer-type bounds, which are known to be necessary in general.

The first progress in this direction was made by Terry [59], who showed that if $\mathcal{H}$ has bounded VC_2 dimension, then the size of the *edge partition*[4] can be taken to be only polynomial in $1/\varepsilon$. However, her proof uses the hypergraph regularity lemma as a black box, hence only provides wowzer-type upper bounds on the number of vertex parts. More recently, Terry [62, Theorem 1.9] made further progress on this question, proving that wowzer-type bounds are unavoidable in certain extreme cases. More precisely, she showed that for *some* extremely fast-decreasing function $\psi_0 : (0,1) \to (0,1)$, there exists a 3-graph $\mathcal{H}$ with VC_2 dimension equal to 1, and with the property that every (ε, ψ_0)-regular partition of $\mathcal{H}$ uses at least $\text{wow}(\text{poly}(1/\varepsilon))$ parts.

However, this is not the end of the story. The function ψ_0 used by Terry decays extremely quickly, namely $\psi_0(x) \leqslant 1/\text{wow}(\text{poly}(1/x))$. For such rapidly decaying functions, the *upper* bound in the hypergraph regularity lemma is no longer of wowzer type: as discussed above, for this choice of ψ_0, the hypergraph regularity lemma would only deliver upper bounds of A_3 type. Hence, while Terry's result shows that wowzer-type bounds may be necessary even if $\mathcal{H}$ has bounded VC_2 dimension, such a lower bound no longer matches the upper bound.

Additionally, as discussed previously, most of the important applications of the hypergraph regularity lemma do not use such rapidly decaying ψ. Instead, applications usually only use ψ which are polynomial functions, and it is in such cases that the hypergraph regularity lemma yields wowzer-type upper bounds. As such, it is natural to ask (as was done explicitly in [62]) whether the wowzer-type bound can be improved if $\mathcal{H}$ has bounded VC_2 dimension, and if one restricts ψ to be a "sensible" function. Our main result in this paper shows that this is indeed the case.

Theorem 1.4 (Informal statement; see Theorem 2.9). *Let* $\psi : (0,1) \to (0,1)$ *be a polynomial function and let* $\varepsilon > 0$. *If* $\mathcal{H}$ *is a 3-graph with bounded* VC_2 *dimension, then* $\mathcal{H}$ *has an* (ε, ψ)-*regular partition with at most* $\text{twr}(\text{twr}(\text{poly}(1/\varepsilon)))$ *vertex parts.*

A few remarks about Theorem 1.4 are in order. First, we note that the double-tower-type bound in Theorem 1.4 is much better than the wowzer-type bound one has in general, for the same reason that $2^{2^x} \ll \text{twr}(x)$. Just as the double-exponential function 2^{2^x} is of "exponential type" rather than of tower type, so too the bound in Theorem 1.4 is of tower type, rather than wowzer type. Second, we remark that the assumption that ψ is a polynomial function is not really important; the same result holds so long as $1/\psi$ grows much slower than a tower function, e.g. $\psi(x) = 1/2^{2^{2^{2^x}}}$ also works. Essentially, if ψ is such that the usual hypergraph regularity lemma yields wowzer-type

---

[4]Here, and throughout the introduction, we mean the number of edge parts per pair of vertex parts when discussing the size of an edge partition.

bounds, our Theorem 1.4 yields a double-tower bounds under the added assumption of bounded VC_2 dimension. Third, we remark that our proof of Theorem 1.4 also gives a partition where the number of edge parts is tower type, but this is substantially less interesting: as discussed above, the result of Terry [59, Theorem 3.1] shows that we may take the number of edge parts to be polynomial, and moreover this is best possible [63]. For many more results and questions on the sizes of the edge partition, see [63], as well as Conjecture 6.8 in the concluding remarks. Fourth, we recall that Terry [62] proved a wowzer-type lower bound if ψ is an arbitrary function, and Theorem 1.4 gives a tower-type upper bound for "reasonable" ψ; it would be interesting to determine which exact growth rate for ψ determines the transition from tower-type bounds to wowzer-type bounds for regularity partitions of hypergraphs with bounded VC_2 dimension.

Fifth, we remark that the structure given by Theorem 1.4 is stronger than that of a usual regularity partition, in the same way that Theorem 1.2 is both qualitatively and quantitatively stronger than Szemerédi's regularity lemma. Namely, not only do we ensure that the hyperedges in $\mathcal{H}$ form a quasirandom subset of the set of triangles of most graphs G_i in the edge partition, but rather, we show that the hyperedges comprise either almost all or almost none of these triangles. This is a stronger statement, in the same way that ε-homogeneity of a pair of vertex sets is a stronger statement than ε-quasirandomness. It was already proved by Chernikov–Towsner [14] and Terry–Wolf [65] that 3-graphs of bounded VC_2 dimension admit such a decomposition, but their techniques do not give the quantitative improvement on the number of vertex parts over the usual hypergraph regularity lemma[5].

Finally, we recall that $\mathcal{H}$ having bounded VC_2 dimension is equivalent to $\mathcal{H}$ forbidding some fixed tripartite 3-graph $\mathcal{V}$ as a tripartitely induced subgraph. In fact, the proof of Theorem 1.4 does not require that there be *zero* tripartitely induced copies of $\mathcal{V}$; instead, it suffices for there to be *few* such copies, i.e. $o(|\mathcal{H}|^{|\mathcal{V}|})$ such copies. This is quite commonplace in the study of regularity and VC dimension. For example, the regularity lemma for graphs of bounded VC dimension of [2] is explicitly stated under the assumption of few bipartitely induced copies, and moreover, it is proved in [65] that a 3-graph $\mathcal{H}$ admits a homogeneous partition, in the sense above, if and only if $\mathcal{H}$ has few tripartitely induced copies of a fixed tripartite $\mathcal{V}$. That is, the "correct" notion of bounded VC_2 dimension is, in some sense, having few copies rather than having zero copies. However, we opted not to state Theorem 1.4 in this way in order to make the statement cleaner.

Given that Theorem 1.4 improves the upper bound from a wowzer function to a tower-type function, it is natural to ask whether it can be improved even further. As remarked above, for two stronger notions of bounded VC dimension for hypergraphs, Chernikov–Starchenko [13], Fox–Pach–Suk [27], Terry [60], and the authors [32] proved that this is indeed the case, and that one can obtain, respectively, polynomial and single-exponential bounds. However, in the case of VC_2 dimension, this is impossible: tower-type bounds are the truth, as proved by Terry [61].

Proposition 1.5 (Terry [61])**.** *For every $\varepsilon > 0$ and every $\psi : (0,1) \to (0,1)$ with $\psi(x) \leqslant x^C$ for some absolute constant C, there exist infinitely many 3-graphs $\mathcal{H}$ with the following properties. $\mathcal{H}$ has VC_2 dimension at most 1, and every (ε, ψ)-regular partition of $\mathcal{H}$ has at least $\mathrm{twr}(\mathrm{poly}(1/\varepsilon))$ vertex parts.*

In fact, Terry proved something somewhat stronger, namely that every *weak ε-regular* partition of $\mathcal{H}$ has at least $\mathrm{twr}(\mathrm{poly}(1/\varepsilon))$ parts (see Proposition 5.2). We discuss this in more detail

---

[5]In fact, the work of Chernikov–Towsner [14] uses infinitary techniques that give no bound at all; the work of Terry–Wolf [65] is quantitative, but uses the hypergraph regularity lemma as a black box and hence does not improve over its quantitative dependencies.

below, but briefly, a weak ε-regular partition is the structure given by Chung's [15] early work on hypergraph regularity. As its name suggests, this is a strictly weaker structure than that given by the full hypergraph regularity lemma, hence a lower bound on the number of parts in such a partition immediately implies the same lower bound on the number of parts in a full-fledged hypergraph regularity partition. Although Proposition 1.5 follows easily from results of Terry, it has apparently not been explicitly stated in this form before, so we include the details of the derivation in Section 5.

By contrast, the proof of Theorem 1.4 requires substantial new ideas, and so we give a high-level outline of it in the next subsection.

1.3   Outline of the proof of Theorem 1.4

1.3.1   Difficulties with extending the proof of Theorem 1.2

Given that Theorem 1.4 is a uniformity-3 extension of Theorem 1.2, it is natural to attempt to mimic the proof of Theorem 1.2. Unfortunately, this appears to be impossible, because of fundamental differences between the notions of VC dimension and VC_2 dimension. In what follows we elaborate on this issue.

We first briefly discuss the proof of Theorem 1.2, as well as why it does not work in uniformity 3. Although this is not directly relevant to what follows—after all, this is an approach that fails—we believe it is still instructive, as it demonstrates some of the key differences between graphs and 3-graphs that make proving results like Theorem 1.4 so much harder than results like Theorem 1.2. The impatient reader can safely skip to Section 1.3.3, where we explain the approach we actually use in the proof of Theorem 1.4.

As mentioned above, results like Theorem 1.2 were proved by Lovász–Szegedy [43], Alon–Fischer–Newman [2], and Fox–Pach–Suk [27]. These three proofs have certain differences, but a key feature shared by all three is the use of the fact that set systems with bounded VC dimension cannot contain many sets with large pairwise symmetric difference. The strongest version of this fact is Haussler's packing lemma [37], which is one of the central tools in the theory of VC dimension.

Lemma 1.6 (Haussler [37]). *Let $\mathcal{F} \subseteq 2^{[n]}$ be a set system of VC dimension d. If $|S \triangle T| \geqslant \varepsilon n$ for all distinct $S, T \in \mathcal{F}$, then $|\mathcal{F}| = O_d(\varepsilon^{-d})$.*

Using this lemma, Fox, Pach, and Suk proved Theorem 1.2 as follows. Let G be a graph with bounded VC dimension, and let us suppose for simplicity that G is bipartite[6] with parts A, B, where $|A| = |B| = n$. Pick out a maximal collection $a_1, \ldots, a_k \in A$ with the property that $|N(a_i) \triangle N(a_j)| \geqslant \varepsilon n$. By Lemma 1.6, we have that $k \leqslant \mathrm{poly}(1/\varepsilon)$, since the set system $\{N(a_i)\}_{1 \leqslant i \leqslant k} \subseteq 2^B$ has bounded VC dimension. Additionally, by maximality, we find that for every $v \in A$, we have that $|N(v) \triangle N(a_i)| < \varepsilon n$ for some $i \in [k]$. Therefore, we may partition A into $A_1, \ldots, A_k$, where A_i comprises only vertices $v \in A$ with $|N(v) \triangle N(a_i)| < \varepsilon n$. The triangle inequality then implies that $|N(v) \triangle N(v')| < 2\varepsilon n$ for all $v, v' \in A_i$. By interchanging the roles of A and B, we can also partition B into $B_1, \ldots, B_\ell$, where $\ell \leqslant \mathrm{poly}(1/\varepsilon)$, such that $|N(u) \triangle N(u')| < 2\varepsilon n$ for all $u, u' \in B_i$.

In other words, we have partitioned $V(G)$ into $\mathrm{poly}(1/\varepsilon)$ parts, such that every pair of vertices in the same part have the same neighborhood, up to an error of $O(\varepsilon n)$. At this point, a short

---

[6]This proof actually works equally well whether or not G is bipartite. However, it will be more convenient to work with bipartite graphs in what follows.

double-counting argument (see [27, Lemma 2.3]) shows that almost all pairs (A_i, B_j) are $O(\varepsilon)$-homogeneous.

Let us attempt to extend such a proof to the uniformity-3 setting. So let $\mathcal{H}$ be a tripartite 3-graph with parts A, B, C, where $|A| = |B| = |C| = n$, and suppose that $\mathcal{H}$ has bounded VC_2 dimension. For $c \in C$, let us denote by $N(c) \subseteq A \times B$ the *link* of c, i.e. the set of pairs (a, b) such that $(a, b, c) \in E(\mathcal{H})$. Following the proof above, it is natural to select a maximal collection $c_1, \ldots, c_k \in C$ with $|N(c_i) \triangle N(c_j)| \geqslant \varepsilon n^2$, and to partition C accordingly.

However, one immediately encounters a fundamental problem, which is that there is no analogue of Haussler's packing lemma for VC_2 dimension. Concretely, in the attempt above, we cannot get *any* bound on k only in terms of ε, hence we cannot use this technique to partition $V(\mathcal{H})$ into $O_\varepsilon(1)$ parts. To see this, we define a tripartite 3-graph $\mathcal{H}$ with parts A, B, C, where $|A| = |B| = |C| = n$, as follows. For every $c \in C$, we pick uniformly random sets $X_c \subseteq A, Y_c \subseteq B$, making these choices independently. We then define

$$E(\mathcal{H}) := \{(x, y, c) : c \in C, x \in X_c, y \in Y_c\}. \tag{1}$$

In other words, the link of every vertex $c \in C$ is precisely the complete bipartite graph between X_c and Y_c. It is easy to see that, with probability 1, $\mathcal{H}$ does not contain a tripartitely induced copy of $K_{2,2,2}^{(3)} \setminus e$, hence $\mathcal{H}$ has bounded VC_2 dimension (in fact, one can check that $\mathcal{H}$ has no shattered $K_{2,2}$, so its VC_2 dimension equals 1). On the other hand, with high probability, we have that $|X_c \triangle X_{c'}|, |Y_c \triangle Y_{c'}| = \Omega(n)$, for all distinct $c, c' \in C$. Therefore $|N(c) \triangle N(c')| = \Omega(n^2)$ for all $c, c' \in C$, and the approach discussed above would entail partitioning C into singletons.

The fact that there is no analogue of Haussler's packing lemma for the VC_2 dimension demonstrates a fundamental difference between the notions of VC and VC_2 dimension. As remarked earlier, Fox, Pach, and Suk [27] were able to use the same approach to prove a polynomial regularity lemma for hypergraphs which satisfy a much stronger form of bounded VC dimension, precisely because this stronger assumption does support such a packing lemma. However, if we only assume bounded VC_2 dimension, we need to use an alternative approach.

What we infer from the above discussion is that before we attempt to tackle the 3-graph case, we should first prove Theorem 1.2 without relying on Lemma 1.6 (Haussler's packing lemma), hoping that such a proof can then be extended to the setting of 3-graphs. This is precisely the content of the next subsection.

1.3.2 A Haussler-free proof of Theorem 1.2

Our goal in this subsection is to sketch the proof of the following result.

Proposition 1.7. *Let G be a graph with bounded VC dimension, and let $\varepsilon > 0$. There is a partition of $V(G)$ into $K \leqslant 2^{2^{\mathrm{poly}(1/\varepsilon)}}$ parts such that a $(1 - \varepsilon)$-fraction of the pairs $(x, y) \in V(G)^2$ lie in ε-homogeneous pairs of parts.*

Note that this result is much weaker than Theorem 1.2, as the number of parts is double-exponential, rather than the optimal polynomial dependence in Theorem 1.2. However, the proof strategy we now discuss has the advantage of being robust enough that it can be extended to the setting of 3-graphs with bounded VC_2 dimension (with some significant technical difficulties). Furthermore, this double-exponential loss will turn out to be negligible in the 3-uniform setting.

For simplicity, let us first add the assumption that G is bipartite with parts A, B. The key new idea in the proof is to not try to immediately construct a partition of $V(G)$. Instead, we begin by finding a weaker structure, called a *cylinder partition*.

Definition 1.8. Let G be a bipartite graph with parts A, B. A *vertex cylinder* is simply a product set $A' \times B' \subseteq A \times B$, where $A' \subseteq A, B' \subseteq B$. A cylinder $A' \times B'$ is η-quasirandom if the pair (A', B') is η-quasirandom in G. A *(vertex) cylinder partition* of G is a partition of $A \times B$ into vertex cylinders $A_1 \times B_1, \ldots, A_k \times B_k$.

The *cylinder regularity lemma* of Duke, Lefmann, and Rödl[7] [22] states that every bipartite graph has a cylinder partition into $k \leqslant 2^{\text{poly}(1/\eta)}$ parts, such that a $(1 - \eta)$-fraction of the pairs $(x, y) \in A \times B$ lie in an η-quasirandom cylinder $A_i \times B_i$. Let us fix some parameter $\eta = \text{poly}(\varepsilon) > 0$ depending polynomially on ε, and fix some η-quasirandom cylinder partition of G with $k \leqslant 2^{\text{poly}(1/\eta)} = 2^{\text{poly}(1/\varepsilon)}$ parts.

Now consider some η-quasirandom cylinder $A_i \times B_i$. The key observation is that the edge density $d(A_i, B_i)$ must be very close to 0 or 1. Indeed, suppose that $\delta \leqslant d(A_i, B_i) \leqslant 1 - \delta$, for some $\delta > 0$. The *induced counting lemma* (e.g. [1, Lemma 3.2]) implies that if $\eta < \delta^{d^2}$, then the pair (A_i, B_i) contains a bipartitely induced copy of every bipartite graph with both parts of size d. However, since we assumed that G has bounded VC dimension, it does not contain any bipartitely induced copy of some fixed bipartite graph, a contradiction. Thus, we conclude that every η-quasirandom cylinder in our partition must be δ-homogeneous.

We are almost done; all that remains is to convert this cylinder partition into an honest partition of $V(G)$. There is an obvious way of doing this, namely to take the *Venn diagram partition*, namely to partition A according to the common refinement of all the sets A_i, and similarly for B. By doing this, we may exponentiate the number of parts, so we end up partitioning $V(G)$ into $K \leqslant 2^{O(k)} \leqslant 2^{2^{\text{poly}(1/\varepsilon)}}$ parts.

If all we knew about the cylinder partition was that most cylinders are quasirandom, this structure would be completely destroyed in the Venn diagram partition; it is for this reason that one cannot simply prove Szemerédi's regularity lemma by applying the cylinder regularity lemma and then taking the Venn diagram partition. However, we have more information, namely that most cylinders in the partition are δ-homogeneous. A simple application of Markov's inequality now shows that if we pick $\delta = \varepsilon^2$, then this homogeneity property *is* preserved when we pass to the Venn diagram partition. Concretely, if a $(1 - \varepsilon^2)$-fraction of the pairs of vertices lie in ε^2-homogeneous cylinders, then in *any* refinement of the partition, a $(1 - \varepsilon)$-fraction of the pairs of vertices lie in ε-homogeneous pairs of parts. To conclude the proof, we note that we only take δ to depend polynomially on ε and η to depend polynomially on δ (with some dependence on the VC dimension of G, which we treat as a fixed constant), hence we have $\eta = \text{poly}(\varepsilon)$ as promised.

Finally, let us discuss how to lift the assumption that G is bipartite. The cylinder regularity lemma actually holds for t-partite graphs, for any $t \geqslant 2$. Namely, if G is t-partite with parts $X_1, \ldots, X_t$, a cylinder partition is a partition of the product set $X_1 \times \cdots \times X_t$ into cylinders $X_1' \times \cdots \times X_t'$. Duke, Lefmann, and Rödl [22] proved that one can find such a cylinder partition with $2^{\text{poly}(t/\eta)}$ parts, such that a $(1 - \eta)$-fraction of tuples $(x_1, \ldots, x_t)$ lie in cylinders $X_1' \times \cdots \times X_t'$ such that all pairs (X_i', X_j') are η-quasirandom. Now, given a general graph G, we first begin by arbitrarily partitioning it into $t = 1/\varepsilon$ parts of equal sizes, and treating it as a t-partite graph (i.e.

---

[7]In fact, for bipartite graphs, the cylinder regularity lemma essentially goes back to the original work of Szemerédi [56] on arithmetic progressions in dense sets of integers, and precedes the "usual" regularity lemma by several years.

forgetting the edges inside a part). We now run the argument above, noting that since $t = \text{poly}(1/\varepsilon)$ and since the cylinder regularity lemma has a bound of the form $2^{\text{poly}(t/\eta)}$, we still end up with a partition of $V(G)$ into $2^{2^{\text{poly}(1/\varepsilon)}}$ parts, such that almost all pairs are ε-homogeneous. Finally, we note that by the choice of $t = 1/\varepsilon$, the same conclusion holds (up to replacing ε by 2ε) even if we bring back the edges we ignored when we artificially made G t-partite. This concludes the proof sketch of Proposition 1.7.

1.3.3   Proof sketch of Theorem 1.4

In order to use the strategy we described in the previous subsection in uniformity 3, we need a few ingredients. First, we need a 3-graph analogue of the cylinder regularity lemma, which is the main technical contribution of this paper (see Theorem 2.13 for the statement). Crucially, this cylinder regularity lemma yields a partition whose size is only of tower type, much better than the wowzer-type bounds needed to obtain a full regularity partition; this is analogous to the situation in graphs, where the full regularity lemma requires tower-type bounds, but the cylinder regularity lemma requires only exponential bounds. Unsurprisingly, even the notion of a cylinder partition becomes substantially more complicated when we move to hypergraphs. We defer a detailed discussion of these definitions to Section 2.3.

Let us now fix such a cylinder partition of $\mathcal{H}$, as given by the cylinder regularity lemma. The next thing we need is an induced counting lemma, in order to conclude that all the quasirandom cylinders are actually very sparse or very dense. As discussed earlier, the main issue in the development of the hypergraph regularity method is finding notions of quasirandomness that can yield such counting lemmas, while also supporting a corresponding regularity lemma. For our application, it is crucial that the counting lemma only requires polynomial bounds, so that we may set $\eta = \text{poly}(\varepsilon)$ as we did in the graph case (at the end of the penultimate paragraph of Section 1.3.2). While it is plausible that each of the notions of hypergraph regularity does indeed only require polynomial losses in the counting lemma, as far as we are aware, no one has proved this for the regularity notion of Rödl et al. [50]. On the other hand, Gowers's [35] approach to the hypergraph regularity lemma does give a counting lemma with polynomial bounds, hence we use Gowers's notion throughout. However, we remark that the final outcome is essentially agnostic[8] about which notion of regularity is used, because our final partition actually has the stronger ε-homogeneity property, which is easily seen to be stronger than each of the notions of ε-regularity.

We now have a cylinder partition of $\mathcal{H}$ in which almost all cylinders are very sparse or very dense. As before, we now pass to the Venn diagram partition, obtaining a genuine vertex partition and edge partition of $\mathcal{H}$. Unlike in the graph case, this step is essentially "for free": although we exponentiate the number of parts, this loss is negligible given that the number of parts in the cylinder partition is already of tower type. As before, this step maintains the ε-homogeneity property, that is, almost all triples of vertices now lie in parts that are very sparse or very dense.

We are almost done, except that in a hypergraph regularity partition, we require the graphs arising from the edge partition to themselves be quasirandom. This was the case in the cylinder partition, but this property may have been destroyed when passing to the Venn diagram partition. Hence, we now need to apply Szemerédi's regularity lemma in order to further refine the vertex

---

[8]It is known that all the notions of hypergraph regularity are qualitatively equivalent (see e.g. [48]). However, the proofs of these equivalences require applications of the hypergraph regularity lemma itself, and hence are not quantitatively effective. Since the main thrust of Theorem 1.4 is the quantitative improvement, we cannot freely pass between different regularity notions.

partition, and to recover the quasirandomness of the graphs. This final step costs us another tower function, which is why the final bound is of double-tower type. However, we stress that we only apply the graph regularity lemma *once*, whereas it needs to be applied iteratively in the proof of the usual hypergraph regularity lemma[9]. This difference is what ensures that we end up with a final bound of tower type, rather than wowzer type. To summarize, in the proof of the hypergraph regularity lemma, one iteratively applies Szemerédi's regularity lemma, leading to wowzer-type bounds. In our proof, we iteratively apply the cylinder regularity lemma for graphs (incurring tower-type dependencies), and then apply Szemerédi's regularity lemma only once (losing one more tower).

1.3.4 Further directions

The cylinder regularity lemma for graphs is extremely useful and has numerous applications, so we naturally expect our cylinder regularity lemma for 3-graphs to have many applications as well. We discuss two such applications in Section 6.1. In particular, one of these applications is a 3-graph extension of a famous theorem of Rödl [49], closely related to the Erdős–Hajnal conjecture (see [10]), about sparse or dense sets in graphs forbidding a fixed induced subgraph. As observed already in Rödl's original paper [49], the most natural extension of this theorem to 3-graphs is false; nonetheless, our cylinder regularity lemma naturally implies that a subtler statement[10] is true. As our cylinder regularity lemma invokes tower-type bounds, we obtain tower-type bounds in our applications, and we would be very interested to see if these bounds could be improved; see Question 6.4 for details.

Finally, we remark that there are natural generalizations of all questions studied in this paper to hypergraphs of higher uniformity. We plan to address such questions in forthcoming work.

Organization The rest of this paper is organized as follows. In Section 2, we define the relevant notions of hypergraph regularity and give formal statements of our main theorems. We develop the cylinder regularity lemma for hypergraphs in Section 3. We use the cylinder regularity lemma, and the proof strategy outlined above, to prove Theorem 1.4 in Section 4. In Section 5, we prove the lower bound, Proposition 1.5, from the results of Terry [61, 62]. We conclude in Section 6 with some final remarks and open problems.

2 Formal definitions and statements

2.1 Definitions regarding hypergraph regularity

As mentioned in the introduction, we use Gowers's notion of hypergraph regularity [34, 35]. In order to make this paper as accessible as possible to non-experts, we follow the notation of [34] whenever possible. We will give minimal explanation or justification for most of these definitions,

---

[9]That said, the cylinder partition we use *is* required to have its underlying graph be quasirandom, in order to apply the induced counting lemma and conclude that almost all cylinders are very sparse or very dense. As such, in the proof of our cylinder regularity lemma for hypergraphs, we are required to regularize the graph at each step of the iteration. However, as we are working with cylinder partitions, we may apply the cylinder regularity lemma for graphs, rather than Szemerédi's regularity lemma, at every step of the iteration, and thus only pay an exponential loss at every step; this is why our cylinder regularity lemma ends up incurring tower-type bounds.

[10]In fact, the same qualitative statement can be deduced from the full hypergraph regularity lemma, but with worse quantitative bounds.

and defer to [34]; that paper is an excellent resource for those wishing to learn about hypergraph regularity, and includes detailed explanations of the various notions we use.

We begin by defining what it means for a pair of vertex sets in a graph to be quasirandom. We recall that $d_G(X, Y)$ denotes the edge density between two vertex sets in a graph G; we now start including the G subscript (which we omitted before), in order to disambiguate densities of different objects.

Definition 2.1 ([34, Definitions 3.2 and 3.3]). Let X, Y be finite sets, and let $f : X \times Y \to [-1, 1]$ be a function. For a real number $\alpha > 0$, we say that f is α-*quasirandom* if

$$\sum_{x,x' \in X} \sum_{y,y' \in Y} f(x, y) f(x', y) f(x, y') f(x', y') \leqslant \alpha |X|^2 |Y|^2.$$

Let G be a graph, and let $X, Y \subseteq V(G)$. We say that the pair (X, Y) is α-*quasirandom* if the function $f : X \times Y \to [-1, 1]$ defined by

$$f(x, y) \coloneqq \mathbf{1}_{xy \in E(G)} - d_G(X, Y)$$

is α-quasirandom.

If G is a t-partite graph with parts $X_1, \ldots, X_t$, we say that G is α-*quasirandom* if each of the pairs (X_i, X_j) for $1 \leqslant i < j \leqslant t$ is α-quasirandom.

This is a different definition than the definition of α-quasirandomness given in the introduction, but it is not hard to show, via a few applications of the Cauchy–Schwarz inequality, that these two definitions are equivalent[11] up to a polynomial change in the value of α. See [34, Theorem 3.1] for details.

The definition of quasirandomness for 3-graphs will be similar, but as discussed in the introduction, the quasirandomness notion has to do with the way in which the edges of a 3-graph lie within the triangles of a graph on the same vertex set. To discuss this structure, we make the following definition.

Definition 2.2 ([34, Definition 6.1]). Given a graph G, let $\Delta(G)$ denote the set of triangles in G. A *chain* is a pair $(G, \mathcal{H})$, where G is a graph and $\mathcal{H}$ is a 3-graph, such that G and $\mathcal{H}$ have the same vertices, and $E(\mathcal{H}) \subseteq \Delta(G)$.

Similarly, if G is a tripartite graph with parts X, Y, Z, we say that a function $f : X \times Y \times Z \to \mathbb{R}$ is *supported on* $\Delta(G)$ if $f(x, y, z) \neq 0$ only if $xyz \in \Delta(G)$.

Given $t \geqslant 3$, a t-*partite chain* is a chain in which G is a t-partite graph, $\mathcal{H}$ is a t-partite 3-graph, both with respect to the same partition $V(G) = V(\mathcal{H}) = X_1 \sqcup \cdots \sqcup X_t$.

If $(G, \mathcal{H})$ is a chain, the *relative density* of $\mathcal{H}$ in G is defined as

$$d(\mathcal{H} \mid G) \coloneqq \frac{|E(\mathcal{H})|}{|\Delta(G)|},$$

with the convention that $0/0 = 0$.

---

[11]This equivalence is (essentially) one of the many equivalences proven in the seminal work of Chung, Graham, and Wilson [16], namely that quasirandomness can be characterized by C_4 counts. Indeed, the sum appearing in Definition 2.1 is, up to lower-order terms, equal to the count of 4-cycles in $X \times Y$, minus the expected number in a random graph of the same density.

We are now ready to define quasirandomness[12] for a 3-graph relative to a graph.

Definition 2.3 ([34, Definition 6.3]). Let G be a tripartite graph with parts X, Y, Z, and let $f : X \times Y \times Z \to [-1, 1]$ be a function supported on $\Delta(G)$. Given $\eta > 0$, we say that f is η-*quasirandom relative to* G if

$$\sum_{\substack{x,x' \in X \\ y,y' \in Y \\ z,z' \in Z}} f(x,y,z)f(x,y,z')f(x,y',z)f(x,y',z')f(x',y,z)f(x',y,z')f(x',y',z)f(x',y',z')$$

$$\leqslant \eta[d_G(X,Y)d_G(X,Z)d_G(Y,Z)]^4|X|^2|Y|^2|Z|^2.$$

If $(G, \mathcal{H})$ is a tripartite chain, we say that $\mathcal{H}$ is η-*quasirandom relative to* G if the function $f : X \times Y \times Z \to [-1, 1]$ defined by

$$f(x,y,z) := (\mathbf{1}_{xyz \in E(\mathcal{H})} - d(\mathcal{H} \mid G))\mathbf{1}_{xyz \in \Delta(G)}$$

is η-quasirandom.

Finally, if $(G, \mathcal{H})$ is a t-partite chain for some $t \geqslant 3$, then we say that $\mathcal{H}$ is η-*quasirandom relative to* G if this holds for each of the $\binom{t}{3}$ tripartite subchains.

As discussed in the introduction, in our regularity decompositions, we require not only that $\mathcal{H}$ is quasirandom with respect to the graphs in the edge partition, but also that these graphs are themselves highly quasirandom. This idea is captured in the following definition.

Definition 2.4 ([34, Definition 8.5]). Let $(G, \mathcal{H})$ be a t-partite chain with parts $X_1, \ldots, X_t$. The *product density* of G is defined as

$$\delta(G) := \prod_{1 \leqslant i < j \leqslant t} d_G(X_i, X_j).$$

Given a number $\eta > 0$ and a function $\psi : (0, 1) \to (0, 1)$, we say that the chain $(G, \mathcal{H})$ is (η, ψ)-*quasirandom* if $\mathcal{H}$ is η-quasirandom relative to G, and if G is $\psi(\delta(G))$-quasirandom.

Finally, in order to state the hypergraph regularity lemma, we need to define chain partitions of hypergraphs.

Definition 2.5 ([34, Definition 8.7]). Let $\mathcal{H}$ be a 3-graph. A *chain partition* of $\mathcal{H}$ is a pair $\mathcal{Q} = (\mathcal{Q}_V, \mathcal{Q}_E)$, where $\mathcal{Q}_V$ is a partition of $V(\mathcal{H})$ into parts, and $\mathcal{Q}_E$ consists, for all distinct $Y_i, Y_j \in \mathcal{Q}_V$, of a partition $\mathcal{Q}_E(Y_i \times Y_j)$ of the complete bipartite graph $Y_i \times Y_j$ into subgraphs.

For a vertex $x \in V(\mathcal{H})$, we denote by $\mathcal{Q}_V[x]$ the unique part of $\mathcal{Q}_V$ containing x. For distinct $x_1, x_2 \in V(\mathcal{H})$, we denote by $\mathcal{Q}_E[x_1, x_2]$ the unique part of $\mathcal{Q}_E(\mathcal{Q}_V[x_1] \times \mathcal{Q}_V[x_2])$ containing the edge (x_1, x_2). Note that this notation is only well-defined in case x_1 and x_2 lie in distinct parts of $\mathcal{Q}_V$.

Given disjoint sets $Y_1, Y_2, Y_3 \subseteq V(\mathcal{H})$ and $Z_{12} \subseteq Y_1 \times Y_2, Z_{13} \subseteq Y_1 \times Y_3, Z_{23} \subseteq Y_2 \times Y_3$, we denote by $\mathcal{H}[Y_1 \cup Y_2 \cup Y_3; Z_{12} \cup Z_{13} \cup Z_{23}]$ the chain $(G, \mathcal{H}')$, where $V(G) = Y_1 \cup Y_2 \cup Y_3$, $E(G) = Z_{12} \cup Z_{13} \cup Z_{23}$, and where $\mathcal{H}'$ is the tripartite subgraph of $\mathcal{H}$ with $V(\mathcal{H}') = Y_1 \cup Y_2 \cup Y_3$ and $E(\mathcal{H}') = E(\mathcal{H}) \cap \Delta(G)$.

---

[12]We remark that, similarly to how Definition 2.1 captures the count of $C_4 = K_{2,2}$ in a bipartite graph, Definition 2.3 captures the count of the octahedron hypergraph $K_{2,2,2}^{(3)}$.

Given a chain partition $\mathcal{Q}$ and vertices $x_1, x_2, x_3 \in V(\mathcal{H})$, we denote by $\mathcal{H}[\mathcal{Q}; x_1, x_2, x_3]$ the chain $\mathcal{H}[\mathcal{Q}_V[x_1] \cup \mathcal{Q}_V[x_2] \cup \mathcal{Q}_V[x_3]; \mathcal{Q}_E[x_1, x_2] \cup \mathcal{Q}_E[x_1, x_3] \cup \mathcal{Q}_E[x_2, x_3]]$. Again, this notation is only well-defined if x_1, x_2, x_3 lie in different parts of $\mathcal{Q}_V$.

Finally, a chain partition $\mathcal{Q} = (\mathcal{Q}_V, \mathcal{Q}_E)$ of $\mathcal{H}$ is said to be (η, ψ)-*quasirandom* if, for at least a $(1-\eta)$-fraction of the triples $(x_1, x_2, x_3) \in V(\mathcal{H})^3$, the chain $\mathcal{H}[\mathcal{Q}; x_1, x_2, x_3]$ is (η, ψ)-quasirandom as in Definition 2.4.

Recall that the notation $\mathcal{H}[\mathcal{Q}; x_1, x_2, x_3]$ is only defined for triples (x_1, x_2, x_3) lying in three different parts of $\mathcal{Q}$. Thus, when we say that the quasirandomness holds for at least a $(1-\eta)$-fraction of triples, we mean that this notation is defined, and is (η, ψ)-quasirandom, for at least $(1-\eta)|V(\mathcal{H})|^3$ triples.

We are finally ready to state the regularity lemma for 3-graphs.

Theorem 2.6 ([34, Theorem 8.10]). *For all $\eta > 0$ and $\psi : (0,1) \to (0,1)$, there exists a constant $K = K(\eta, \psi)$ such that the following holds. For every 3-graph $\mathcal{H}$, there is a chain partition $\mathcal{Q} = (\mathcal{Q}_V, \mathcal{Q}_E)$ which is (η, ψ)-quasirandom.*

Additionally, $|\mathcal{Q}_V| \leqslant K$ and $|\mathcal{Q}_E(Y_i \times Y_j)| \leqslant K$ for all distinct $Y_i, Y_j \in \mathcal{Q}_V$. Moreover, K is bounded by $\mathrm{poly}(1/\eta)$ iterations of the function $\mathrm{twr}(\mathrm{poly}(1/\psi(\eta)))$.

The explicit bound on K is discussed but not proved in [34], and a formal proof is given in [62, Appendix A].

To conclude this section, we briefly discuss weak regularity for hypergraphs, as this will be relevant in the proof of Proposition 1.5. The following definition is a direct extension of the definition of graph quasirandomness discussed in the introduction. It is not quite the definition introduced by Chung [15], but it is easily seen to be equivalent up to a polynomial change in η.

Definition 2.7. Let $\mathcal{H}$ be a 3-graph, and let $X_1, X_2, X_3 \subseteq V(\mathcal{H})$. We let $e_{\mathcal{H}}(X_1, X_2, X_3)$ denote the number of triples in $X_1 \times X_2 \times X_3$ that are edges of $\mathcal{H}$, and let $d_{\mathcal{H}}(X_1, X_2, X_3) := e_{\mathcal{H}}(X_1, X_2, X_3)/(|X_1||X_2||X_3|)$ be the edge density. We say that the triple (X_1, X_2, X_3) is *weakly η-quasirandom* if, for all $X_i' \subseteq X_i$, we have that

$$|e_{\mathcal{H}}(X_1', X_2', X_3') - d_{\mathcal{H}}(X_1, X_2, X_3)|X_1'||X_2'||X_3'|| \leqslant \eta|X_1||X_2||X_3|.$$

A partition $\mathcal{Q}_V$ of $V(\mathcal{H})$ is called *weakly η-quasirandom* if, for at least a $(1-\eta)$-fraction of the triples $(x_1, x_2, x_3) \in V(\mathcal{H})^3$, the triple $(\mathcal{Q}_V[x_1], \mathcal{Q}_V[x_2], \mathcal{Q}_V[x_3])$ is weakly η-quasirandom.

With this definition, we may state Chung's weak hypergraph regularity lemma [15].

Theorem 2.8 ([15, Theorem 2.1]). *For all $\eta > 0$, there exists a constant $K = K(\eta)$ such that the following holds. For every 3-graph $\mathcal{H}$, there is a weakly η-quasirandom partition of $V(\mathcal{H})$ into at most K parts. Moreover, $K \leqslant \mathrm{twr}(\mathrm{poly}(1/\eta))$.*

As the names suggest, weak quasirandomness is a weaker notion than quasirandomness (see Proposition 5.1 for details). In particular, Theorem 2.8 follows from Theorem 2.6, except that the latter yields a worse bound on K.

2.2 Formal statement of our main theorem

We now formally state Theorem 1.4, our strengthened regularity lemma for 3-graphs of bounded VC_2 dimension.

Theorem 2.9. *Fix an integer d. For all $\eta > 0$ and all increasing polynomial functions $\psi : (0,1) \to (0,1)$ with $\psi(x) \leqslant x$, the following holds. Let $\mathcal{H}$ be a sufficiently large 3-graph with VC_2 dimension at most d. There is a chain partition $\mathcal{Q} = (\mathcal{Q}_V, \mathcal{Q}_E)$ of $\mathcal{H}$ such that, for at least a $(1 - \eta)$-fraction of the triples $(x_1, x_2, x_3) \in V(\mathcal{H})^3$, the chain $(G, \mathcal{H}') := \mathcal{H}[\mathcal{Q}; x_1, x_2, x_3]$ satisfies the following properties. The graph G is $\psi(\delta(G))$-quasirandom, and the relative density $d(\mathcal{H}' \mid G)$ satisfies $d(\mathcal{H}' \mid G) \in [0, \eta] \cup [1 - \eta, 1]$.*

Additionally,

$$|\mathcal{Q}_V| \leqslant \mathrm{twr}(\mathrm{twr}(\psi(\eta)^{-C})),$$

where C is a constant depending only on d.

As indicated in the introduction, the assumption that $\mathcal{H}$ has bounded VC_2 dimension allows us to strengthen Theorem 2.6 in two ways. First, and most importantly, the bound on the number of parts is improved from wowzer-type to tower-type. Secondly, as in Theorem 1.2, we obtain a stronger conclusion than most triples of vertices lying in (η, ψ)-quasirandom chains: most triples now lie in chains whose graph is $\psi(\delta(G))$-quasirandom, but where the relative density of the hypergraph is very close to 0 or 1. As in the graph case, it is straightforward to show that this is indeed a strictly stronger property than quasirandomness; for a proof, see e.g. [59, Proposition 2.24].

2.3 Cylinder regularity for hypergraphs

The goal of this section is to state our cylinder regularity lemma for 3-graphs, which is the key new tool we develop in this paper towards the proof of Theorem 1.4. In order to do so, we first need to introduce some further definitions and notation.

Definition 2.10 (Cylinders and partitions)**.** Let $X_1, \ldots, X_t$ be finite sets, which we think of as the vertex parts of some t-partite (hyper)graph. A *vertex subcylinder* is a product set $Y := Y_1 \times \cdots \times Y_t$, where $Y_i \subseteq X_i$ for all i. If Y is a vertex subcylinder, we also denote by $Y(X_i)$ the set Y_i.

A *vertex cylinder partition* is a collection $\mathcal{P}_V = \{Y^1, \ldots, Y^m\}$, where each $Y^j = Y^j(X_1) \times \cdots \times Y^j(X_t)$ is a vertex subcylinder, such that the subcylinders Y^j partition the vertex cylinder $X_1 \times \cdots \times X_t$. In other words, for every $(x_1, \ldots, x_t) \in X_1 \times \cdots \times X_t$, there is a unique j such that $(x_1, \ldots, x_t) \in Y^j$. We denote this unique part by $\mathcal{P}_V[x_1, \ldots, x_t]$. We denote by $|\mathcal{P}_V|$ the number of parts in $\mathcal{P}_V$.

An *edge partition*[13] $\mathcal{P}_E(Y)$ of a vertex cylinder $Y = Y_1 \times \cdots \times Y_t$ consists of, for all $1 \leqslant i < j \leqslant t$, a partition $\mathcal{P}_E(Y_i \times Y_j)$ of the complete bipartite graph $Y_i \times Y_j$ into subgraphs $Z_{ij}^1, \ldots, Z_{ij}^\ell$. Note that if we select one such bipartite graph for each $1 \leqslant i < j \leqslant t$, we obtain a t-partite graph; in particular, given $\mathcal{P}_E(Y)$, each t-tuple $(x_1, \ldots, x_t) \in Y_1 \times \cdots \times Y_t$ defines a unique such t-partite graph, namely the one containing all edges $x_i x_j$. We denote this t-partite graph by $\mathcal{P}_E[x_1, \ldots, x_t]$. We also define $|\mathcal{P}_E(Y)| := \max_{i<j} |\mathcal{P}_E(Y_i \times Y_j)|$ to be the maximum number of bipartite graphs we use in any of our $\binom{t}{2}$ edge partitions.

---

[13]We stress that the partition of the edges is not "cylindrical": this notion is identical to that of the edge partition appearing in Definition 2.5.

Finally, a *cylinder chain partition* is a pair $\mathcal{P} = (\mathcal{P}_V, \mathcal{P}_E)$, where $\mathcal{P}_V$ is a vertex cylinder partition, and $\mathcal{P}_E = \{\mathcal{P}_E(Y) : Y \in \mathcal{P}_V\}$ is a set of edge partitions, one for each part of $\mathcal{P}_V$. We denote by $|\mathcal{P}_E|$ the maximum of $|\mathcal{P}_E(Y)|$ over all $Y \in \mathcal{P}_V$.

Definition 2.11 (Subhypergraphs on chains)**.** Let $\mathcal{H}$ be a t-partite 3-graph on vertex parts $X_1, \ldots, X_t$. For a vertex subcylinder $Y = Y_1 \times \cdots \times Y_t$, we denote by $\mathcal{H}[Y]$ the induced sub-hypergraph on vertex set $Y_1 \cup \cdots \cup Y_t$. For a t-partite graph Z on $Y_1 \cup \cdots \cup Y_t$, we further denote by $\mathcal{H}[Y; Z]$ the subhypergraph of $\mathcal{H}[Y]$ comprising all edges supported on $\Delta(Z)$.

Note that $\mathcal{H}[Y; Z]$ can itself be viewed as a union of $\binom{t}{3}$ chains. Namely, for every $\{i, j, k\} \in \binom{[t]}{3}$, we have the chain with vertex sets Y_i, Y_j, Y_k, edge sets Z_{ij}, Z_{ik}, Z_{jk}, and the hyperedges of $\mathcal{H}$ which are supported on $\Delta(Z_{ij} \cup Z_{ik} \cup Z_{jk})$. We denote this chain by $\mathcal{H}[Y; Z; \{i, j, k\}]$.

We now extend the quasirandomness notions discussed above to cylinders and cylinder partitions.

Definition 2.12 (Quasirandomness of cylinders)**.** Let $Y = Y_1 \times \cdots \times Y_t$ be a vertex cylinder. We say that an edge partition $\mathcal{P}_E(Y)$ is α-quasirandom if, whenever $(x_1, \ldots, x_t) \in Y_1 \times \cdots \times Y_t$ is picked uniformly at random, then the t-partite graph $\mathcal{P}_E[x_1, \ldots, x_t]$ is α-quasirandom with probability at least $1 - \alpha/2$.

Let $\mathcal{H}$ be a t-partite 3-graph on vertex parts $X_1, \ldots, X_t$. Given a cylinder chain partition $\mathcal{P} = (\mathcal{P}_V, \mathcal{P}_E)$, and given $\eta > 0$ and $\psi : (0, 1) \to (0, 1)$, we say that $\mathcal{H}$ is (η, ψ)-quasirandom relative to $\mathcal{P}$ if the following holds. When $(x_1, \ldots, x_t) \in X_1 \times \cdots \times X_t$ is picked uniformly at random, with probability at least $1 - \eta$ we have that $\mathcal{H}[\mathcal{P}_V[x_1, \ldots, x_t]; \mathcal{P}_E[x_1, \ldots, x_t]]$ is (η, ψ)-quasirandom, in the sense of Definition 2.4.

We are now ready to state our cylinder regularity lemma for 3-graphs, which will be the main technical tool used in the proof of Theorem 2.9. In the statement, we will use the following simple extension of the tower function used above, where we denote by $\mathrm{twr}(\tau, x)$ a tower of twos of height τ with an x as the top exponent. Formally, we recursively define this function by $\mathrm{twr}(0, x) = x$ and $\mathrm{twr}(\tau + 1, x) = 2^{\mathrm{twr}(\tau, x)}$. Note that $\mathrm{twr}(\tau, 1) = \mathrm{twr}(\tau)$, that is, this agrees with the previous definition upon setting $x = 1$.

Theorem 2.13. *Fix an integer t, a parameter $\eta > 0$, and an increasing polynomial function $\psi : (0, 1) \to (0, 1)$ satisfying $\psi(x) \leqslant 2^{-100} x^{28}$. Let $\mathcal{H}$ be a t-partite 3-graph. Then there exists a cylinder chain partition $\mathcal{P} = (\mathcal{P}_V, \mathcal{P}_E)$, such that $\mathcal{H}$ is (η, ψ)-quasirandom relative to $\mathcal{P}$, and such that*

$$|\mathcal{P}_V| \leqslant \mathrm{twr}\left(\frac{10^5 t^6}{\eta^3}, \psi\left(\frac{\eta}{t^2}\right)^{-20}\right)$$

and

$$|\mathcal{P}_E| \leqslant \mathrm{twr}\left(\frac{10^5 t^6}{\eta^3}, \left(\frac{t^2}{\eta}\right)^{10}\right).$$

3 Cylinder regularity

Our goal in this section is to prove Theorem 2.13. In the first subsection below we recall some basic facts and tools. We then use these tools to prove Theorem 2.13 in Section 3.2. We stress that Theorem 2.13 holds for all 3-graphs, and in particular, in this section we do not assume that the 3-graphs we are dealing with have bounded VC_2 dimension.

3.1 Mean-squared density

In this subsection, we record various facts about the mean-squared density of a cylinder chain partition, which we now define.

First, we extend our notion of edge partition (from Definition 2.10) as follows. Let G be a tripartite graph with parts Y_1, Y_2, Y_3, comprising three bipartite graphs $G_{12} \subseteq Y_1 \times Y_2, G_{13} \subseteq Y_1 \times Y_3, G_{23} \subseteq Y_2 \times Y_3$. An *edge partition* $\mathcal{P}_E(G)$ consists of a partition of each of the three bipartite graphs G_{12}, G_{13}, G_{23} into subgraphs. We denote by $\mathcal{P}_E(G; Y_1 \times Y_2)$ the partition of G_{12}, and similarly for the other two pairs of parts. As before, $|\mathcal{P}_E(G)| \coloneqq \max\{|\mathcal{P}_E(G; Y_1 \times Y_2)|, |\mathcal{P}_E(G; Y_1 \times Y_3)|, |\mathcal{P}_E(G; Y_2 \times Y_3)|\}$ denotes the maximum number of parts partitioning the bipartite graphs. Note that all of these notions extend our previous notion of an edge partition $\mathcal{P}_E(Y)$ of a tripartite cylinder; the previous case corresponds to the case when G is the complete tripartite graph on $Y_1 \cup Y_2 \cup Y_3$. As we will be dealing with such complete tripartite graphs repeatedly in what follows, we also establish the following notation: $K(Y_1, Y_2, Y_3)$ denotes the complete tripartite graph on $Y_1 \cup Y_2 \cup Y_3$.

We recall that if $(G, \mathcal{H})$ is a tripartite chain, we denote by

$$d(\mathcal{H} \mid G) = \frac{|E(\mathcal{H})|}{|\Delta(G)|}$$

the relative density of $\mathcal{H}$ on G, with the convention that $0/0 = 0$. In case G' is a tripartite subgraph of G, we similarly define

$$d(\mathcal{H} \mid G') \coloneqq \frac{|E(\mathcal{H}) \cap \Delta(G')|}{|\Delta(G')|}$$

to be the relative density of the chain $(G', \mathcal{H}')$, where $\mathcal{H}'$ is the subhypergraph of $\mathcal{H}$ supported on $\Delta(G')$.

Definition 3.1 (Mean-squared density[14])**.** Let $(G, \mathcal{H})$ be a tripartite chain with vertex parts $Y_i \cup Y_j \cup Y_k$, and let $\mathcal{P}_E(G)$ be an edge partition of G. The *mean-squared density* of $\mathcal{P}_E(G)$ is defined to be

$$q(\mathcal{P}_E(G)) \coloneqq \sum_{\substack{Z_{ij} \in \mathcal{P}_E(G; Y_i \times Y_j) \\ Z_{ik} \in \mathcal{P}_E(G; Y_i \times Y_k) \\ Z_{jk} \in \mathcal{P}_E(G; Y_j \times Y_k)}} \frac{|\Delta(Z_{ij} \cup Z_{ik} \cup Z_{jk})|}{|\Delta(G)|} d(\mathcal{H} \mid Z_{ij} \cup Z_{ik} \cup Z_{jk})^2. \tag{2}$$

In words, we sum over all tripartite subgraphs G' of G obtained by picking one element of each of the partitions of the three bipartite subgraphs. For each such G', we square the relative density of $\mathcal{H}$, and then weight this summand by the fraction of triangles of G that appear in G'. Note that if $\mathcal{P}_E(G)$ is the trivial edge partition, then $q(\mathcal{P}_E(G)) = d(\mathcal{H} \mid G)^2$.

Let $\mathcal{H}$ now be a t-partite 3-graph with vertex set $X_1 \cup \cdots \cup X_t$. Suppose $Y = Y_1 \times \cdots \times Y_t$ is a vertex cylinder and $\mathcal{P}_E(Y)$ is an edge partition as in Definition 2.10. The mean-squared density of the cylinder Y is then defined to be

$$q(\mathcal{P}_E(Y)) \coloneqq \sum_{1 \leqslant i < j < k \leqslant t} q(\mathcal{P}_E(K(Y_i, Y_j, Y_k))), \tag{3}$$

---

[14]We remark that there are several different reasonable definitions of the mean-squared density; for example, rather than weighting the summands in (2) by $|\Delta(Z_{12} \cup Z_{13} \cup Z_{23})|$, we could weight by its approximation $\delta(Z_{12} \cup Z_{13} \cup Z_{23})|Z_{12}||Z_{13}||Z_{23}|$ (recall Definition 2.4). The definition we use is a generalization to cylinder partitions of the one used by Gowers [34], and is more technically convenient for our purposes.

where we recall that $K(Y_i, Y_j, Y_k)$ is the complete tripartite graph on $Y_i \cup Y_j \cup Y_k$. Finally, given a cylinder chain partition $\mathcal{P} = (\mathcal{P}_V, \mathcal{P}_E)$ of $X_1 \times \cdots \times X_t$, we define its mean-squared density to be

$$q(\mathcal{P}) := \sum_{Y \in \mathcal{P}_V} \left(\prod_{i=1}^{t} \frac{|Y_i|}{|X_i|} \right) q(\mathcal{P}_E(Y)).$$

The basic observations we need about the mean-squared density are standard, namely that it is bounded and monotonic with respect to refinement. To formally state these properties, let us say that a vertex cylinder partition $\mathcal{P}'_V$ *refines* another vertex cylinder partition $\mathcal{P}_V$ if, for every $Y' \in \mathcal{P}'_V$, there is a unique $Y \in \mathcal{P}_V$ containing Y'. That is, if $Y' = Y'_1 \times \cdots \times Y'_t$ and $Y = Y_1 \times \cdots \times Y_t$, then $Y'_i \subseteq Y_i$ for all i. Similarly, an edge partition $\mathcal{P}'_E(G)$ *refines* another edge partition $\mathcal{P}_E(G)$ of the same t-partite graph if for all pairs of parts Y_i, Y_j, each part of $\mathcal{P}'_E(G; Y_i \times Y_j)$ lies in a unique part of $\mathcal{P}_E(G; Y_i \times Y_j)$. For two edge partitions $\mathcal{P}_E(Y), \mathcal{P}'_E(Y)$ of a vertex cylinder, we say that $\mathcal{P}'_E(Y)$ refines $\mathcal{P}_E(Y)$ if it refines it in the sense above, where we view it as an edge partition of the complete t-partite graph. Finally, let us say that one cylinder chain partition $\mathcal{P}' = (\mathcal{P}'_V, \mathcal{P}'_E)$ *refines* another cylinder chain partition $\mathcal{P} = (\mathcal{P}_V, \mathcal{P}_E)$ if $\mathcal{P}'_V$ refines $\mathcal{P}_V$ in the sense above, and if $\mathcal{P}'_E$ refines $\mathcal{P}_E$ in the following sense. If $Y' \in \mathcal{P}'_V$, let Y be the part of $\mathcal{P}_V$ containing Y'. For all $i < j$, the partition $\mathcal{P}_E(Y_i \times Y_j)$ naturally restricts to a partition $\mathcal{P}_E(Y_i \times Y_j)|_{Y'_i \times Y'_j}$ of the complete bipartite graph $Y'_i \times Y'_j$. We then say that $\mathcal{P}'_E$ refines $\mathcal{P}_E$ if, for all $Y' \in \mathcal{P}'_V$ and for all $i < j$, the partition $\mathcal{P}'_E(Y'_i \times Y'_j)$ refines $\mathcal{P}_E(Y_i \times Y_j)|_{Y'_i \times Y'_j}$.

We may now state the properties we will use of the mean-squared density.

Lemma 3.2. *Let $(G, \mathcal{H})$ be a tripartite chain.*

(a) *If $\mathcal{P}_E(G)$ is an edge partition of G, then $0 \leqslant q(\mathcal{P}_E(G)) \leqslant 1$.*

(b) *If $\mathcal{P}'_E(G)$ is a refinement of $\mathcal{P}_E(G)$, then $q(\mathcal{P}'_E(G)) \geqslant q(\mathcal{P}_E(G))$.*

Similarly, let $\mathcal{H}$ be a t-partite 3-graph.

(c) *If $\mathcal{P} = (\mathcal{P}_V, \mathcal{P}_E)$ is a cylinder chain partition, then $0 \leqslant q(\mathcal{P}) \leqslant \binom{t}{3}$.*

(d) *If $\mathcal{P}'$ is a refinement of $\mathcal{P}$, then $q(\mathcal{P}') \geqslant q(\mathcal{P})$.*

Lemma 3.2 is proved by several elementary but tedious applications of the Cauchy–Schwarz inequality, and we thus defer the proof to Appendix A.1.

As in the majority of proofs of regularity lemmas, our proof of Theorem 2.13 will proceed by an energy increment approach. Concretely, given a cylinder chain partition of $\mathcal{H}$, we will either find that it is sufficiently quasirandom, or else find a refinement of it that increases the mean-squared density by a constant amount (independent of $\mathcal{H}$ or of the given partition). Each single step of this iteration will be obtained by the following lemma of Gowers [34]. It states that if a tripartite chain is not η-quasirandom, then we can find an edge partition of it for which the mean-squared density increases appreciably, while the number of parts in the new edge partition is not too large.

Lemma 3.3 ([34, Lemma 8.4])**.** *Consider a tripartite graph G with vertex sets Y_1, Y_2, Y_3 and edge sets Z_{12}, Z_{13}, Z_{23}. Let $\delta = \delta(G)$. Suppose that each of the three bipartite graphs is α-quasirandom, where $\alpha \leqslant 2^{-100} \delta^{28}$.*

Let $(G, \mathcal{H})$ be a tripartite chain, and suppose that $\mathcal{H}$ is not η-quasirandom relative to G. Let $d = d(\mathcal{H} \mid G)$ be the relative density of $\mathcal{H}$. Then there is an edge partition $\mathcal{P}_E(G)$ such that $q(\mathcal{P}_E(G)) \geqslant d^2 + 2^{-10}\eta^2$. Moreover, $|\mathcal{P}_E(G)| \leqslant 3^{\delta^{-4}}$.

Using Lemma 3.3, we are able to take an irregular cylinder chain partition and further partition each of the irregular edge parts in order to increase the mean-squared density. However, we cannot then directly iterate Lemma 3.3, as the assumption of Lemma 3.3 requires that each bipartite graph is α-quasirandom, and this condition may be destroyed by the previous partitioning step. To get around this, we will apply the cylinder regularity lemma for graphs [22] at each step. We will actually use the following, slightly generalized, version of the cylinder regularity lemma for graphs, which allows us to simultaneously regularize a family of graphs on the same vertex set.

Lemma 3.4. *Let $G_1, \ldots, G_m$ be t-partite graphs on the same vertex set. For any $\alpha > 0$, there exists a vertex cylinder partition $\mathcal{P}_V$ such that the following holds. If $(x_1, \ldots, x_t)$ is chosen uniformly at random from $X_1 \times \cdots \times X_t$, then with probability at least $1 - \alpha/2$, we have that for all $i \in [m]$, the subgraph of G_i induced on the vertex set $\mathcal{P}_V[x_1, \ldots, x_t]$ is α-quasirandom. Moreover, $|\mathcal{P}_V| \leqslant 2^{mt^2 \alpha^{-20}}$.*

The only differences between Lemma 3.4 and the cylinder regularity lemma proved by Duke, Lefmann, and Rödl [22] are that we are regularizing multiple graphs, and that we use a different notion of quasirandomness, leading to the different power of α (since these two notions of graph quasirandomness are polynomially related). It is well known, and used frequently in hypergraph regularity theory, that one can simultaneously regularize multiple graphs on the same vertex set; see e.g. [34, Theorem 7.8] or [30, Lemma 3.7]. For a proof of the cylinder regularity lemma for multiple graphs, see [23, Theorem 7.2].

3.2 Proof of Theorem 2.13

With these preliminaries, we are ready to prove Theorem 2.13. In fact, for future convenience (see Lemma 4.2), we prove the following slight strengthening of Theorem 2.13.

Theorem 3.5. *Fix an integer t, a parameter $\eta > 0$, and an increasing polynomial function $\psi : (0,1) \to (0,1)$ satisfying $\psi(x) \leqslant 2^{-100} x^{28}$. Let $\mathcal{H}$ be a t-partite 3-graph. Then there exists a cylinder chain partition $\mathcal{P} = (\mathcal{P}_V, \mathcal{P}_E)$, such that $\mathcal{H}$ is (η, ψ)-quasirandom relative to $\mathcal{P}$, and such that*

$$|\mathcal{P}_V| \leqslant \operatorname{twr}\left(\frac{10^5 t^6}{\eta^3}, \psi\left(\frac{\eta}{t^2} \right)^{-20} \right)$$

and

$$|\mathcal{P}_E| \leqslant \operatorname{twr}\left(\frac{10^5 t^6}{\eta^3}, \left(\frac{t^2}{\eta} \right)^{10} \right).$$

Furthermore, there is a constant $\delta > 0$, depending only on t, η, and ψ, such that a $(1 - \eta)$-fraction of tuples $(x_1, \ldots, x_t)$ lie in a chain with product density at least δ.

The only difference between Theorems 2.13 and 3.5 is the final sentence, stating that most t-tuples lie in chains whose product density is not too small. This will be useful for us later, and arises naturally from the proof.

Proof of Theorem 3.5. We begin by defining functions $A, B : \mathbb{N} \to \mathbb{N}$ as follows. We set $A(0) = B(0) = 1$, and recursively define

$$B(\tau + 1) = 2^{2^{(t^2 B(\tau)/\eta)^{2t^2}}},$$

$$A(\tau + 1) = 2^{B(\tau+1) t^4 \psi(\eta/(t^2 B(\tau+1)))^{-20t^2}} A(\tau).$$

20

The fact that ψ is a polynomial function implies that

$$A(\tau) \leqslant \mathrm{twr}\left(3\tau, \psi\left(\frac{\eta}{t^2}\right)^{-20}\right) \qquad \text{and} \qquad B(\tau) \leqslant \mathrm{twr}\left(3\tau, \left(\frac{t^2}{\eta}\right)^{10}\right). \tag{4}$$

We remark that similar bounds would hold if ψ were any function such that $1/\psi$ is of constant tower height; the only difference is that the tower height of $A(\tau)$ would be at most $C\tau$, for some constant C depending on ψ. In case ψ is a polynomial, we may take $C = 3$, as stated in (4).

We will define a sequence $\mathcal{P}^{(0)}, \mathcal{P}^{(1)}, \ldots$ of cylinder chain partitions, and inductively maintain the following properties, where $X_1, \ldots, X_t$ are the t vertex parts of $\mathcal{H}$.

(i) $|\mathcal{P}_V^{(\tau)}| \leqslant A(\tau)$.

(ii) $|\mathcal{P}_E^{(\tau)}| \leqslant B(\tau)$.

(iii) $q(\mathcal{P}^{(\tau)}) \geqslant 2^{-10}\eta^3\tau/t^3$.

(iv) Let $\delta = (\eta/(t^2 B(\tau)))^{\binom{t}{2}}$ and $\alpha = \psi(\delta)$. If we sample $(x_1, \ldots, x_t) \in X_1 \times \cdots \times X_t$ uniformly at random, the following happens with probability at least $1 - \eta/2$: the t-partite graph $Z := \mathcal{P}_E^{(\tau)}[x_1, \ldots, x_t]$ satisfies $\delta(Z) \geqslant \delta$ and Z is α-quasirandom.

(v) Either $\mathcal{H}$ is (η, ψ)-quasirandom with respect to $\mathcal{P}^{(\tau)}$, or else we can define $\mathcal{P}^{(\tau+1)}$ satisfying the above properties.

To begin the induction, we let $\mathcal{P}_V^{(0)}$ be the trivial cylinder partition into one part $Y^0 = X_1 \times \cdots \times X_t$, and let the corresponding edge partition $\mathcal{P}_E(Y^0(X_i) \times Y^0(X_j))$ also be the trivial partition into one part $X_i \times X_j$, for all $1 \leqslant i < j \leqslant t$. Then it is clear that the claimed properties (i)–(iv) hold.

Inductively, suppose we have defined $\mathcal{P}^{(\tau)}$ satisfying properties (i)–(iv). We are done if $\mathcal{H}$ is (η, ψ)-quasirandom with respect to $\mathcal{P}^{(\tau)}$, so we may assume that this is not the case.

As in property (iv), let $\delta = (\eta/(t^2 B(\tau)))^{\binom{t}{2}}$ and let $\alpha = \psi(\delta)$. The fact that $\mathcal{H}$ is not (η, ψ)-quasirandom with respect to $\mathcal{P}^{(\tau)}$ means that if we pick $(x_1, \ldots, x_t) \in X_1 \times \cdots \times X_t$ uniformly at random, then with probability at least η, the chain $\mathcal{H}[\mathcal{P}_V^{(\tau)}[x_1, \ldots, x_t]; \mathcal{P}_E^{(\tau)}[x_1, \ldots, x_t]]$ is not (η, ψ)-quasirandom. That, in turn, means that with probability at least η, either the chain $\mathcal{H}[\mathcal{P}_V^{(\tau)}[x_1, \ldots, x_t]; \mathcal{P}_E^{(\tau)}[x_1, \ldots, x_t]]$ is not η-quasirandom, or that the t-partite graph $\mathcal{P}_E^{(\tau)}[x_1, \ldots, x_t]$ is not $\psi(\delta(\mathcal{P}_E^{(\tau)}[x_1, \ldots, x_t]))$-quasirandom.

In this random experiment, we know by property (iv) that with probability at least $1 - \eta/2$, the t-partite graph $\mathcal{P}_E^{(\tau)}[x_1, \ldots, x_t]$ satisfies $\delta(\mathcal{P}_E^{(\tau)}[x_1, \ldots, x_t]) \geqslant \delta$ and is α-quasirandom. Since ψ is increasing, this in turn implies that $\mathcal{P}_E^{(\tau)}[x_1, \ldots, x_t]$ is $\psi(\delta(\mathcal{P}_E^{(\tau)}[x_1, \ldots, x_t]))$-quasirandom. Therefore, with probability at least $\eta - \eta/2 = \eta/2$, we have that $\mathcal{P}_E^{(\tau)}[x_1, \ldots, x_t]$ is $\psi(\delta(\mathcal{P}_E^{(\tau)}[x_1, \ldots, x_t]))$-quasirandom, whereas the chain $\mathcal{H}[\mathcal{P}_V^{(\tau)}[x_1, \ldots, x_t]; \mathcal{P}_E^{(\tau)}[x_1, \ldots, x_t]]$ is not η-quasirandom. By the pigeonhole principle, there exist distinct indices $a, b, c \in [t]$ such that for at least an $\eta/(2\binom{t}{3})$-fraction of the tuples $(x_1, \ldots, x_t) \in X_1 \times \cdots \times X_t$, the tripartite chain

$$\mathcal{H}[\mathcal{P}_V^{(\tau)}[x_1, \ldots, x_t]; \mathcal{P}_E^{(\tau)}[x_1, \ldots, x_t]; \{a, b, c\}] \tag{5}$$

is not η-quasirandom but does satisfy that $\delta(\mathcal{P}_E^{(\tau)}[x_1, \ldots, x_t]) \geqslant \delta$ and that $\mathcal{P}_E^{(\tau)}[x_1, \ldots, x_t]$ is $\psi(\delta(\mathcal{P}_E^{(\tau)}[x_1, \ldots, x_t]))$-quasirandom. We call such t-tuples $(x_1, \ldots, x_t)$ *useful*. Extending this terminology, we also call the tripartite chain in (5) *useful*.

Fix a useful t-tuple $(x_1, \ldots, x_t)$, and let $Y = \mathcal{P}_V^{(\tau)}[x_1, \ldots, x_t]$ and $Z = \mathcal{P}_E^{(\tau)}[x_1, \ldots, x_t]$. Let G be the tripartite graph with parts $Y_a \cup Y_b \cup Y_c$ and edge sets $Z_{ab} \cup Z_{ac} \cup Z_{bc}$. Note that $\delta(G) \geqslant \delta(Z) \geqslant \delta$ and that G is $\psi(\delta(G))$-quasirandom, as ψ is increasing and by the definition of a useful tuple. We may thus apply Lemma 3.3 to G, using our assumptions that ψ is increasing and satisfies $\psi(x) \leqslant 2^{-100} x^{28}$. Doing this gives us, for each such useful chain, an edge partition of G into at most $3^{\delta(G)^{-4}} \leqslant 3^{\delta(Z)^{-4}} \leqslant 3^{\delta^{-4}}$ parts.

We now do this process for all useful chains, and use this step to define a new edge partition $\mathcal{P}_E^{(\tau+\frac{1}{2})}$ (for the moment, we will not change $\mathcal{P}_V^{(\tau)}$). To do so, let us fix a vertex cylinder $Y \in \mathcal{P}_V^{(\tau)}$, and define the new edge partition $\mathcal{P}_E^{(\tau+\frac{1}{2})}(Y)$ as follows. Let the bipartite graphs in $\mathcal{P}_E(Y_a \times Y_b)$ be $Z_{ab}^1, \ldots, Z_{ab}^\ell$. For each $1 \leqslant \mu \leqslant \ell$, we consider the set of all tripartite graphs arising by adding to Z_{ab}^μ some part of $\mathcal{P}_E^{(\tau)}(Y_a \times Y_c)$ and $\mathcal{P}_E^{(\tau)}(Y_b \times Y_c)$. Some of these tripartite chains are useful and hence subpartitioned by Lemma 3.3, and we then let Z_{ab}^μ be partitioned as the common refinement of all of these partitions. By doing this for all $1 \leqslant \mu \leqslant \ell$, we obtain a new edge partition $\mathcal{P}_E^{(\tau+\frac{1}{2})}(Y_a \times Y_b)$. We do the same to obtain the edge partitions $\mathcal{P}_E^{(\tau+\frac{1}{2})}(Y_a \times Y_c), \mathcal{P}_E^{(\tau+\frac{1}{2})}(Y_b \times Y_c)$. For all $ij \notin \{ab, ac, bc\}$, we set $\mathcal{P}_E^{(\tau+\frac{1}{2})}(Y_i \times Y_j) = \mathcal{P}_E^{(\tau)}(Y_i \times Y_j)$. Finally, doing this for all $Y \in \mathcal{P}_V^{(\tau)}$ defines our new edge partition $\mathcal{P}_E^{(\tau+\frac{1}{2})}$.

We now claim that

$$q((\mathcal{P}_V^{(\tau)}, \mathcal{P}_E^{(\tau+\frac{1}{2})})) \geqslant q((\mathcal{P}_V^{(\tau)}, \mathcal{P}_E^{(\tau)})) + 2^{-10}\eta^3/t^3. \tag{6}$$

Indeed, we first note that as we did not change the vertex partition,

$$q((\mathcal{P}_V^{(\tau)}, \mathcal{P}_E^{(\tau+\frac{1}{2})})) - q((\mathcal{P}_V^{(\tau)}, \mathcal{P}_E^{(\tau)})) = \sum_{Y \in \mathcal{P}_V^{(\tau)}} \left(\prod_{1 \leqslant i \leqslant t} \frac{|Y_i|}{|X_i|} \right) \left(q(\mathcal{P}_E^{(\tau+\frac{1}{2})}(Y)) - q(\mathcal{P}_E^{(\tau)}(Y)) \right). \tag{7}$$

To estimate the inner difference, let us fix some $Y \in \mathcal{P}_V^{(\tau)}$. We then have by definition that

$$q(\mathcal{P}_E^{(\tau+\frac{1}{2})}(Y)) - q(\mathcal{P}_E^{(\tau)}(Y)) = \sum_{1 \leqslant i < j < k \leqslant t} \left[q(\mathcal{P}_E^{(\tau+\frac{1}{2})}(K(Y_i, Y_j, Y_k))) - q(\mathcal{P}_E^{(\tau)}(K(Y_i, Y_j, Y_k))) \right].$$

For each triple $(i, j, k) \neq (a, b, c)$, we recall that the edge partition $\mathcal{P}_E^{(\tau+\frac{1}{2})}(K(Y_i, Y_j, Y_k))$ refines $\mathcal{P}_E^{(\tau)}(K(Y_i, Y_j, Y_k))$, hence $q(\mathcal{P}_E^{(\tau+\frac{1}{2})}(K(Y_i, Y_j, Y_k))) - q(\mathcal{P}_E^{(\tau)}(K(Y_i, Y_j, Y_k))) \geqslant 0$ by Lemma 3.2(b). Therefore,

$$q(\mathcal{P}_E^{(\tau+\frac{1}{2})}(Y)) - q(\mathcal{P}_E^{(\tau)}(Y)) \geqslant q(\mathcal{P}_E^{(\tau+\frac{1}{2})}(K(Y_a, Y_b, Y_c))) - q(\mathcal{P}_E^{(\tau)}(K(Y_a, Y_b, Y_c))). \tag{8}$$

By definition, we have that

$$q(\mathcal{P}_E^{(\tau+\frac{1}{2})}(K(Y_a, Y_b, Y_c))) = \sum_{\substack{Z'_{ab} \in \mathcal{P}_E^{(\tau+\frac{1}{2})}(Y_a \times Y_b) \\ Z'_{ac} \in \mathcal{P}_E^{(\tau+\frac{1}{2})}(Y_a \times Y_c) \\ Z'_{bc} \in \mathcal{P}_E^{(\tau+\frac{1}{2})}(Y_b \times Y_c)}} \frac{|\Delta(Z'_{ab} \cup Z'_{ac} \cup Z'_{bc})|}{|Y_a||Y_b||Y_c|} d(\mathcal{H} \mid Z'_{ab} \cup Z'_{ac} \cup Z'_{bc})^2. \tag{9}$$

22

Recall that $\mathcal{P}_E^{(\tau+\frac{1}{2})}(Y_a \times Y_b)$ refines $\mathcal{P}_E^{(\tau)}(Y_a \times Y_b)$, and similarly for $Y_a \times Y_c$ and $Y_b \times Y_c$. This means that we may rearrange the sum in (9) to first sum over parts in $\mathcal{P}_E^{(\tau)}$ and thus find that $q(\mathcal{P}_E^{(\tau+\frac{1}{2})}(K(Y_a, Y_b, Y_c)))$ equals

$$
\sum_{\substack{Z_{ab} \in \mathcal{P}_E^{(\tau)}(Y_a \times Y_b) \\ Z_{ac} \in \mathcal{P}_E^{(\tau)}(Y_a \times Y_c) \\ Z_{bc} \in \mathcal{P}_E^{(\tau)}(Y_b \times Y_c)}} \quad \sum_{\substack{Z'_{ab} \in \mathcal{P}_E^{(\tau+\frac{1}{2})}(Y_a \times Y_b) \\ Z'_{ac} \in \mathcal{P}_E^{(\tau+\frac{1}{2})}(Y_a \times Y_c) \\ Z'_{bc} \in \mathcal{P}_E^{(\tau+\frac{1}{2})}(Y_b \times Y_c) \\ Z'_{ab} \subseteq Z_{ab}, Z'_{ac} \subseteq Z_{ac}, Z'_{bc} \subseteq Z_{bc}}} \frac{|\Delta(Z'_{ab} \cup Z'_{ac} \cup Z'_{bc})|}{|Y_a||Y_b||Y_c|} d(\mathcal{H} \mid Z'_{ab} \cup Z'_{ac} \cup Z'_{bc})^2.
$$

This, in turn, is equal to

$$
\sum_{\substack{Z_{ab} \in \mathcal{P}_E^{(\tau)}(Y_a \times Y_b) \\ Z_{ac} \in \mathcal{P}_E^{(\tau)}(Y_a \times Y_c) \\ Z_{bc} \in \mathcal{P}_E^{(\tau)}(Y_b \times Y_c)}} \frac{|\Delta(Z_{ab} \cup Z_{ac} \cup Z_{bc})|}{|Y_a||Y_b||Y_c|} \sum_{\substack{Z'_{ab} \in \mathcal{P}_E^{(\tau+\frac{1}{2})}(Y_a \times Y_b) \\ Z'_{ac} \in \mathcal{P}_E^{(\tau+\frac{1}{2})}(Y_a \times Y_c) \\ Z'_{bc} \in \mathcal{P}_E^{(\tau+\frac{1}{2})}(Y_b \times Y_c) \\ Z'_{ab} \subseteq Z_{ab}, Z'_{ac} \subseteq Z_{ac}, Z'_{bc} \subseteq Z_{bc}}} \frac{|\Delta(Z'_{ab} \cup Z'_{ac} \cup Z'_{bc})|}{|\Delta(Z_{ab} \cup Z_{ac} \cup Z_{bc})|} d(\mathcal{H} \mid Z'_{ab} \cup Z'_{ac} \cup Z'_{bc})^2.
$$

Crucially, we now note that this inner sum is exactly the type of quantity appearing in (2). More precisely, suppose we fix some $Z_{ab} \in \mathcal{P}_E^{(\tau)}(Y_a \times Y_b), Z_{ac} \in \mathcal{P}_E^{(\tau)}(Y_a \times Y_c), Z_{bc} \in \mathcal{P}_E^{(\tau)}(Y_b \times Y_c)$, and define an edge partition $\widetilde{\mathcal{P}_E}(Z_{ab} \cup Z_{ac} \cup Z_{bc})$ of the tripartite graph they form by simply restricting $\mathcal{P}_E^{(\tau+\frac{1}{2})}$ to this tripartite graph. Then the inner sum above is exactly $q(\widetilde{\mathcal{P}_E}(Z_{ab} \cup Z_{ac} \cup Z_{bc}))$, as defined in (2). Therefore, we can rewrite the above as

$$
q(\mathcal{P}_E^{(\tau+\frac{1}{2})}(K(Y_a, Y_b, Y_c))) = \sum_{\substack{Z_{ab} \in \mathcal{P}_E^{(\tau)}(Y_a \times Y_b) \\ Z_{ac} \in \mathcal{P}_E^{(\tau)}(Y_a \times Y_c) \\ Z_{bc} \in \mathcal{P}_E^{(\tau)}(Y_b \times Y_c)}} \frac{|\Delta(Z_{ab} \cup Z_{ac} \cup Z_{bc})|}{|Y_a||Y_b||Y_c|} q(\widetilde{\mathcal{P}_E}(Z_{ab} \cup Z_{ac} \cup Z_{bc})).
$$

For such Z_{ab}, Z_{ac}, Z_{bc}, we also define another edge partition $\mathcal{P}'_E(Z_{ab} \cup Z_{ac} \cup Z_{bc})$ of the same tripartite graph as follows. If the chain $(Z_{ab} \cup Z_{ac} \cup Z_{bc}, \mathcal{H})$ is not useful, we let $\mathcal{P}'_E(Z_{ab} \cup Z_{ac} \cup Z_{bc})$ be the trivial edge partition where each bipartite graph is partitioned into a single part. On the other hand, if this chain is useful, then $\mathcal{P}'_E(Z_{ab} \cup Z_{ac} \cup Z_{bc})$ is defined to be the edge partition given to us by Lemma 3.3. Note that in either case, $\widetilde{\mathcal{P}_E}(Z_{ab} \cup Z_{ac} \cup Z_{bc})$ refines the edge partition $\mathcal{P}'_E(Z_{ab} \cup Z_{ac} \cup Z_{bc})$; indeed, $\mathcal{P}_E^{(\tau+\frac{1}{2})}$ was defined to be the common refinement of all partitions arising in this way. Hence, applying Lemma 3.2(b) to each tripartite graph $Z_{ab} \cup Z_{ac} \cup Z_{bc}$, we find that

$$
q(\mathcal{P}_E^{(\tau+\frac{1}{2})}(K(Y_a, Y_b, Y_c))) \geqslant \sum_{\substack{Z_{ab} \in \mathcal{P}_E^{(\tau)}(Y_a \times Y_b) \\ Z_{ac} \in \mathcal{P}_E^{(\tau)}(Y_a \times Y_c) \\ Z_{bc} \in \mathcal{P}_E^{(\tau)}(Y_b \times Y_c)}} \frac{|\Delta(Z_{ab} \cup Z_{ac} \cup Z_{bc})|}{|Y_a||Y_b||Y_c|} q(\mathcal{P}'_E(Z_{ab} \cup Z_{ac} \cup Z_{bc})).
$$

23

Plugging in the definition of $q(\mathcal{P}_E^{(\tau)}(K(Y_a, Y_b, Y_c)))$, we conclude that

$$q(\mathcal{P}_E^{(\tau+\frac{1}{2})}(K(Y_a, Y_b, Y_c))) - q(\mathcal{P}_E^{(\tau)}(K(Y_a, Y_b, Y_c)))$$
$$\geqslant \sum_{\substack{Z_{ab} \in \mathcal{P}_E^{(\tau)}(Y_a \times Y_b) \\ Z_{ac} \in \mathcal{P}_E^{(\tau)}(Y_a \times Y_c) \\ Z_{bc} \in \mathcal{P}_E^{(\tau)}(Y_b \times Y_c)}} \frac{|\Delta(Z_{ab} \cup Z_{ac} \cup Z_{bc})|}{|Y_a||Y_b||Y_c|} \left[q(\mathcal{P}'_E(Z_{ab} \cup Z_{ac} \cup Z_{bc})) - d(\mathcal{H} \mid Z_{ab} \cup Z_{ac} \cup Z_{bc})^2 \right].$$

In the sum above, if the choice of Z_{ab}, Z_{ac}, Z_{bc} is such that $(Z_{ab} \cup Z_{ac} \cup Z_{bc}, \mathcal{H})$ is not useful, the difference inside the sum equals zero, as $\mathcal{P}'_E(Z_{ab} \cup Z_{ac} \cup Z_{bc})$ is then a trivial partition, whose mean-squared density equals $d(\mathcal{H} \mid Z_{ab} \cup Z_{ac} \cup Z_{bc})^2$ (recall Definition 3.1). For the remaining summands, $\mathcal{P}'_E(Z_{ab} \cup Z_{ac} \cup Z_{bc})$ is defined to be the output of Lemma 3.3, and hence the difference inside the sum is at least $2^{-10}\eta^2$. Combining this with (7) and (8), we conclude that

$$q((\mathcal{P}_V^{(\tau)}, \mathcal{P}_E^{(\tau+\frac{1}{2})})) - q((\mathcal{P}_V^{(\tau)}, \mathcal{P}_E^{(\tau)}))$$
$$\geqslant 2^{-10}\eta^2 \cdot \sum_{Y \in \mathcal{P}_V^{(\tau)}} \left(\prod_{1 \leqslant i \leqslant t} \frac{|Y_i|}{|X_i|} \right) \sum_{\substack{Z_{ab} \in \mathcal{P}_E^{(\tau)}(Y_a \times Y_b) \\ Z_{ac} \in \mathcal{P}_E^{(\tau)}(Y_a \times Y_c) \\ Z_{bc} \in \mathcal{P}_E^{(\tau)}(Y_b \times Y_c)}} \frac{|\Delta(Z_{ab} \cup Z_{ac} \cup Z_{bc})|}{|Y_a||Y_b||Y_c|} \cdot \mathbf{1}_{(Z_{ab} \cup Z_{ac} \cup Z_{bc}, \mathcal{H}) \text{ is useful}}.$$

The final observation is that this sum (i.e. the right-hand expression without the factor $2^{-10}\eta^2$) is precisely the probability that a uniformly random tuple $(x_1, \ldots, x_t) \in X_1 \times \cdots \times X_t$ is useful. By assumption, this probability is at least $\eta/(2\binom{t}{3}) \geqslant \eta/t^3$, and this concludes the proof of the claim (6).

In order to estimate $|\mathcal{P}_E^{(\tau+\frac{1}{2})}|$, let $\ell = |\mathcal{P}_E^{(\tau)}(Y)|$ for some vertex cylinder Y. Let $\mathcal{P}_E^{(\tau)}(Y_a \times Y_b) = \{Z_{ab}^1, \ldots, Z_{ab}^\ell\}$, and consider some Z_{ab}^μ. There are at most ℓ^2 choices of bipartite graphs in $\mathcal{P}_E^{(\tau)}(Y_a \times Y_c), \mathcal{P}_E^{(\tau)}(Y_b \times Y_c)$ for which the corresponding chain is useful. For each of these ℓ^2 choices, Lemma 3.3 partitions Z_{ab}^μ into at most $3^{\delta^{-4}}$ subgraphs, and then $\mathcal{P}_E^{(\tau+\frac{1}{2})}(Y_a \times Y_b)$ is obtained by taking the common refinement of all of these. This implies that $|\mathcal{P}_E^{(\tau+\frac{1}{2})}(Y_a \times Y_b)| \leqslant \ell \cdot 2^{\ell^2 3^{\delta^{-4}}} \leqslant 2^{\ell^3 3^{\delta^{-4}}}$. Applying the same argument to ac and bc, we conclude that

$$|\mathcal{P}_E^{(\tau+\frac{1}{2})}(Y)| \leqslant 2^{|\mathcal{P}_E^{(\tau)}|^3 3^{\delta^{-4}}} \leqslant 2^{B(\tau)^3 3^{(t^2 B(\tau)/\eta)^{4\binom{t}{2}}}} \leqslant B(\tau+1).$$

What remains now is to additionally refine the vertex cylinder partition $\mathcal{P}_V^{(\tau)}$ in order to ensure that property (iv) holds for $\mathcal{P}^{(\tau+1)}$. Fix some $Y \in \mathcal{P}_V^{(\tau)}$. The edge parts in $\mathcal{P}_E^{(\tau+\frac{1}{2})}(Y)$ form a collection of $m \leqslant \binom{t}{2} B(\tau+1)$ bipartite graphs, each with vertex set $Y(X_i) \cup Y(X_j)$ for some $i < j$. In particular, we may and will also view them as t-partite graphs[15] on the vertex set $Y(X_1) \cup \cdots \cup Y(X_t)$, which

---

[15]It would perhaps be more natural to apply the cylinder regularity lemma for *bipartite* graphs (i.e. Lemma 3.4 with $t = 2$), regularizing each of the $\binom{t}{2}$ partitions $\mathcal{P}_E^{(\tau+\frac{1}{2})}(Y_i \times Y_j)$ one by one. The problem with this approach is that when we regularize a pair, we may destroy the regularity in previously-processed pairs. This is essentially the same reason why, if one wants to prove a regularity lemma for several graphs on the same vertex set, it is better to run the proof of the regularity lemma for all of them simultaneously, as opposed to applying the regularity lemma to each of them in turn.

contain only edges between $Y(X_i)$ and $Y(X_j)$. Let $\varepsilon = (\eta/(t^2 B(\tau + 1)))^{\binom{t}{2}}$, and let $\beta = \psi(\varepsilon)$; these will be the values of δ, α, respectively, corresponding to step $\tau + 1$ of the iteration. Applying Lemma 3.4, we find a vertex cylinder partition of Y into at most $2^{mt^2\beta^{-20}}$ parts, such that a $1 - \beta/2$ fraction of the t-tuples of vertices define a β-quasirandom cylinder within each part of $\mathcal{P}_E^{(\tau+\frac{1}{2})}(Y)$. Let $\mathcal{P}_V^{(\tau+1)}$ be the vertex cylinder partition obtained by partitioning each such $Y \in \mathcal{P}_V^{(\tau)}$. Note that

$$|\mathcal{P}_V^{(\tau+1)}| \leqslant 2^{\binom{t}{2}B(\tau+1)\cdot t^2\beta^{-20}}|\mathcal{P}_V^{(\tau)}| \leqslant A(\tau+1).$$

Let $\mathcal{P}_E^{(\tau+1)}$ be the edge partition obtained by restricting $\mathcal{P}_E^{(\tau+\frac{1}{2})}$ to the new vertex cylinders in $\mathcal{P}_V^{(\tau+1)}$, and let $\mathcal{P}^{(\tau+1)} = (\mathcal{P}_V^{(\tau+1)}, \mathcal{P}_E^{(\tau+1)})$. Note that for every $Y \in \mathcal{P}_V^{(\tau+1)}$, we have that $|\mathcal{P}_E^{(\tau+1)}(Y)| \leqslant |\mathcal{P}_E^{(\tau+\frac{1}{2})}(\widehat{Y})| \leqslant B(\tau+1)$, where $\widehat{Y}$ is the (unique) vertex cylinder in $\mathcal{P}_V^{(\tau)}$ containing Y. As $\mathcal{P}^{(\tau+1)}$ is a refinement of $(\mathcal{P}_V^{(\tau)}, \mathcal{P}_E^{(\tau+\frac{1}{2})})$, we have by Lemma 3.2(d) that

$$q(\mathcal{P}^{(\tau+1)}) \geqslant q(\mathcal{P}^{(\tau)}) + 2^{-10}\eta^3/t^3 \geqslant 2^{-10}\eta^3(\tau+1)/t^3,$$

by applying property (iii) to $\mathcal{P}^{(\tau)}$. This proves that properties (i), (ii), and (iii) hold for $\mathcal{P}^{(\tau+1)}$.

Moreover, by our choice of $\mathcal{P}_V^{(\tau+1)}$, we know that for every $Y \in \mathcal{P}_V^{(\tau+1)}$, if we sample $(y_1, \ldots, y_t) \in Y(X_1) \times \cdots \times Y(X_t)$ uniformly at random, then with probability at least $1 - \beta/2$, we have that $Z := \mathcal{P}_E^{(\tau+1)}[y_1, \ldots, y_t]$ is β-quasirandom. Let us call the tuple $(y_1, \ldots, y_t)$ *sparse for* (i,j) if $d(Z_{ij}) < \varepsilon^{1/\binom{t}{2}}$, or equivalently if the number of edges in Z_{ij} is less than $\varepsilon^{1/\binom{t}{2}}|Y(X_i)||Y(X_j)|$. As there are at most $B(\tau+1)$ choices for Z_{ij}, the total number of edges contributed by the sparse tuples is less than $\varepsilon^{1/\binom{t}{2}}B(\tau+1)|Y(X_i)||Y(X_j)| = (\eta/t^2)|Y(X_i)||Y(X_j)|$.

We now recall that the graphs Z_{ij} partition the complete bipartite graph $Y(X_i) \times Y(X_j)$. Since at most an (η/t^2)-fraction of those edges come from tuples that are sparse for (i,j), we conclude that the probability that $(y_1, \ldots, y_t)$ is sparse for (i,j) is at most η/t^2. However, if $\delta(Z) < \varepsilon$ then $d(Z_{ij}) < \varepsilon^{1/\binom{t}{2}}$ for some i, j, and hence the probability that $\delta(Z) < \varepsilon$ is at most $\binom{t}{2}(\eta/t^2)$. In total, the probability of violating property (iv) is at most $\beta/2 + \binom{t}{2}(\eta/t^2) \leqslant \eta/2$. Recall that this holds for any fixed Y, and for a uniformly random sample of $(y_1, \ldots, y_t) \in Y(X_1) \times \cdots \times Y(X_t)$; in particular, this implies that the same holds for uniformly random $(x_1, \ldots, x_t) \in X_1 \times \cdots \times X_t$, demonstrating that property (iv) holds for step $\tau + 1$. This completes the proof of the inductive step.

To obtain the desired final partition, we recall by Lemma 3.2(c) that $q(\mathcal{P}^{(\tau)}) \leqslant \binom{t}{3}$ for all τ. Therefore, by property (iii), this process must terminate at some step $\tau \leqslant \binom{t}{3}/(2^{-10}\eta^3/t^3) \leqslant 10^4 t^6 \eta^{-3}$. At that step, by property (v), we must have found a partition $\mathcal{P} = (\mathcal{P}_V, \mathcal{P}_E)$ such that $\mathcal{H}$ is (η, ψ)-quasirandom with respect to $\mathcal{P}$. Moreover, by (iv), we have that a $(1-\eta)$-fraction of the t-tuples define a chain with product density at least δ, where $\delta = (\eta/(t^2 B(10^4 t^6 \eta^{-3})))^{\binom{t}{2}}$ depends only on t, η, and ψ. Finally, from (4), we have that

$$|\mathcal{P}_V| \leqslant A(10^4 t^6/\eta^3) \leqslant \mathrm{twr}(10^5 t^6/\eta^3, \psi(\eta/t^2)^{-20})$$

and

$$|\mathcal{P}_E| \leqslant B(10^4 t^3/\eta^3) \leqslant \mathrm{twr}(10^5 t^6/\eta^3, (t^2/\eta)^{10}). \qquad \square$$

4 Proof of Theorem 1.4

In this section we prove Theorem 2.9, following the proof sketch discussed in Section 1.3. Our main tools will be Theorem 3.5 from the previous section, as well as the induced counting lemma, which we now discuss.

The main property we need about 3-graphs of bounded VC_2 dimension is that in every quasirandom chain, the relative density is very close to 0 or 1. This follows from the induced counting lemma (Lemma 4.1 below), which informally states that if a tripartite 3-graph $\mathcal{H}$ is sufficiently quasirandom, and the relative density of hyperedges is bounded away from 0 and 1, then it contains any fixed tripartite 3-graph as an induced subhypergraph. For simplicity, we only state the result for finding one induced copy, but it is easy to see that one could similarly get a counting result. It is well-known that such induced statements follow from general counting lemmas (see e.g. [69, Remark 2.6.3]), and indeed, our result follows directly from Gowers's hypergraph counting lemma [35, Corollary 5.3]. Essentially the same statement, without explicit constants, is proved in [62, Corollary 2.16]; we include a proof in order to fully track the quantitative dependencies.

Lemma 4.1. *Let $\mathcal{V}$ be a tripartite 3-graph with r vertices. Let G be a tripartite graph on parts U_1, U_2, U_3, and let $\mathcal{H}_0$ be a tripartite 3-graph supported on $\Delta(G)$.*
Let $\delta, \gamma \in (0, \frac{1}{100}]$, and define $\eta = 10^{-16} \gamma^{10r^3}$ and $\psi(x) = 10^{-20} \gamma^{10r^3} x^{16r^2}$. Suppose that $\delta(G) \geqslant \delta$, that

$$\gamma \leqslant d(\mathcal{H}_0 \mid G) \leqslant 1 - \gamma,$$

and that $\mathcal{H}_0[U; G]$ is (η, ψ)-quasirandom. If $\min\{|U_1|, |U_2|, |U_3|\} \geqslant \delta^{-r^2} \gamma^{-r^3}$, then $\mathcal{V}$ is an induced subhypergraph of $\mathcal{H}$.

As the proof of Lemma 4.1 uses standard tools, we defer it to Appendix A.2.

Our next lemma is an immediate corollary of Lemma 4.1, and states that a sufficiently quasirandom 3-graph of bounded VC_2 dimension must have relative density close to 0 or 1. Results of this type are quite well-known (see e.g. [65, Proposition 4.2] or [62, Proposition 3.4]), but we include a proof in order to explicitly keep track of the quantitative dependencies.

Recall from the introduction that the tripartite 3-graph $\mathcal{V}_d$ is defined as follows. Its vertex set is $A \sqcup B \sqcup C$, where $A = B = [d]$ and $C = 2^{[d]^2}$, and its hyperedges are all triples $(a, b, S) \in A \times B \times C$ with $(a, b) \in S$.

Lemma 4.2. *Let $d \geqslant 1$ be an integer, and let $r = 2(d+1) + 2^{(d+1)^2}$. Let G be a tripartite graph on parts U_1, U_2, U_3, and let $\mathcal{H}_0$ be a tripartite 3-graph supported on $\Delta(G)$.*
Let $\delta, \gamma \in (0, \frac{1}{100}]$, and define $\eta = 10^{-16} \gamma^{10r^3}$ and $\psi(x) = 10^{-20} \gamma^{10r^3} x^{16r^2}$. Suppose that $\delta(G) \geqslant \delta$, that $\mathcal{H}_0[U; G]$ is (η, ψ)-quasirandom, and that $\min\{|U_1|, |U_2|, |U_3|\} \geqslant \delta^{-r^2} \gamma^{-r^3}$.
If $\mathcal{H}_0$ has VC_2 dimension at most d, then

$$d(\mathcal{H}_0 \mid G) \in [0, \gamma] \cup [1 - \gamma, 1].$$

Proof. Suppose for contradiction that $\gamma < d(\mathcal{H}_0 \mid G) < 1 - \gamma$. Then we are precisely in the setting of Lemma 4.1, which implies that the r-vertex tripartite 3-graph $\mathcal{V}_{d+1}$ is an induced subhypergraph of $\mathcal{H}_0$. That is, the VC_2 dimension of $\mathcal{H}_0$ is at least $d+1$, a contradiction. $\square$

Combining Theorem 2.13 and Lemma 4.2, we obtain the following result, which is a version of Theorem 1.4 for cylinder partitions.

Lemma 4.3. *Let $d \geqslant 1, t \geqslant 3$ be integers and let $\gamma \in (0, \frac{1}{100}]$ be a parameter. Let $\mathcal{H}$ be a sufficiently large 3-graph with VC_2 dimension at most d, and let $X_1 \sqcup \cdots \sqcup X_t$ be an arbitrary equitable partition of $V(\mathcal{H})$. There is a cylinder chain partition $\mathcal{P} = (\mathcal{P}_V, \mathcal{P}_E)$ of $\mathcal{H}$ such that the following holds. If $(x_1, \ldots, x_t) \in X_1 \times \cdots \times X_t$ is sampled uniformly at random, then with probability at least $1 - \gamma^2$, we have that*

$$d(\mathcal{H}[\mathcal{P}_V[x_1, \ldots, x_t]; \mathcal{P}_E[x_1, \ldots, x_t]; \{i, j, k\}]) \in [0, \gamma] \cup [1 - \gamma, 1]$$

for all $1 \leqslant i < j < k \leqslant t$. Moreover,

$$|\mathcal{P}_V| \leqslant \mathrm{twr}\left(\left(\frac{t}{\gamma}\right)^C\right) \qquad and \qquad |\mathcal{P}_E| \leqslant \mathrm{twr}\left(\left(\frac{t}{\gamma}\right)^C\right),$$

where $C > 0$ is a constant depending only on d.

Note that in this statement, we are simply ignoring all edges of $\mathcal{H}$ which do not transverse the partition $X_1 \sqcup \cdots \sqcup X_t$.

Proof of Lemma 4.3. Let $r = 2(d+1) + 2^{(d+1)^2}$, let $\eta = 10^{-16}\gamma^{10r^3}$ and $\psi(x) = 10^{-20}\gamma^{10r^3}x^{16r^2}$. We apply Theorem 3.5 to $\mathcal{H}$ to obtain a cylinder chain partition $\mathcal{P}$ such that $\mathcal{H}$ is (η, ψ)-quasirandom relative to $\mathcal{P}$. We also obtain some constant $\delta > 0$ depending only on t, η and ψ such that a $(1-\eta)$-fraction of the tuples lie in a chain with product density at least δ. Additionally, the claimed bounds on $|\mathcal{P}_V|$ and $|\mathcal{P}_E|$ follow immediately from Theorem 3.5.

Now, as long as $|V(\mathcal{H})|$ is sufficiently large, we conclude that at least a $(1 - 3\eta)$-fraction of the tuples $(x_1, \ldots, x_t)$ lie in a cylinder that is (η, ψ)-quasirandom, has product density at least δ, and all t of its vertex parts have size at least $\delta^{-r^2}\gamma^{-r^3}$. We now apply Lemma 4.2 to any triple of parts in this cylinder. The only thing to note is that this tripartite hypergraph has no tripartitely induced copy of $\mathcal{V}_{d+1}$, as such a copy is also tripartitely induced in $\mathcal{H}$ (recall that for tripartite inducedness, we only care about edges going between all three parts, and thus the additional edges $\mathcal{H}$ may have are irrelevant). We may thus apply Lemma 4.2. We conclude that the relative density in this chain is at most γ or at least $1 - \gamma$. The claimed result now follows since $3\eta \leqslant \gamma^2$. $\square$

With Lemma 4.3 in hand, our goal is now to convert this cylinder chain partition into a genuine chain partition, while preserving the property that almost all chains have relative density close to 0 or 1. We do this following the outline presented in Section 1.3.3, and the remainder of the proof is dedicated to this step.

Our next lemma, which is a simple consequence of Markov's inequality, states that when we partition a quasirandom chain with relative density very close to 0 or 1, most subchains also have edge density close to 0 or 1.

Lemma 4.4. *Let G be a tripartite graph on parts A, B, C, and let $\mathcal{H}_0$ be a tripartite 3-graph supported on $\Delta(G)$. Let $\gamma > 0$ be a parameter.*

Suppose that A, B, C are partitioned into $A_1, \ldots, A_K, B_1, \ldots, B_K, C_1, \ldots, C_K$, respectively. Suppose too that each bipartite graph $G[A_i \times B_j], G[A_i \times C_j], G[B_i \times C_j]$ is further partitioned into K subgraphs.

Suppose that, among all triangles in G, a triangle (x, y, z) is sampled uniformly at random. If $d(\mathcal{H}_0 \mid G) < \gamma$, then with probability at least $1 - \sqrt{\gamma}$, the unique subchain $G_\mu[A_i \times B_j] \cup G_\nu[A_i \times C_k] \cup G_\xi[B_j \times C_k]$ that contains (x, y, z) satisfies

$$d(\mathcal{H}_0 \mid G_\mu[A_i \times B_j] \cup G_\nu[A_i \times C_k] \cup G_\xi[B_j \times C_k]) < \sqrt{\gamma}.$$

Similarly, if $d(\mathcal{H}_0 \mid G) > 1 - \gamma$, then with probability at least $1 - \sqrt{\gamma}$, we have

$$d(\mathcal{H}_0 \mid G_\mu[A_i \times B_j] \cup G_\nu[A_i \times C_k] \cup G_\xi[B_j \times C_k]) > 1 - \sqrt{\gamma}.$$

Proof. Every triangle in G is contained in precisely one subgraph of the form $(A_i \cup B_j \cup C_k, G_\mu \cup G_\nu \cup G_\xi)$. Therefore, these partitions of the vertex and edge sets induce a corresponding partition of $\Delta(G)$ into a collection of subsets, say $\Delta(G) = \Delta_1 \sqcup \cdots \sqcup \Delta_M$. The uniform distribution on $\Delta(G)$ is the same as first picking an index $\ell \in [M]$ at random, with probability proportional to $|\Delta_\ell|$, and then picking a uniformly random element of Δ_ℓ. Denote by d_ℓ the relative density $d(\mathcal{H}_0 \mid G_\mu[A_i \times B_j] \cup G_\nu[A_i \times C_k] \cup G_\xi[B_j \times C_k])$, where $A_i, B_j, C_k, G_\mu, G_\nu, G_\xi$ are the subchain defining Δ_ℓ. We then have that $d(\mathcal{H}_0 \mid G)$ is equal to the weighted average of d_ℓ, where the weight of d_ℓ is proportional to $|\Delta_\ell|$; more precisely,

$$d(\mathcal{H}_0 \mid G) = \sum_{\ell=1}^{M} \frac{|\Delta_\ell|}{|\Delta(G)|} d_\ell.$$

Said differently, in the random experiment above, the expected value of d_ℓ is exactly $d(\mathcal{H}_0 \mid G)$, which we assumed is less than γ. By Markov's inequality, we conclude that

$$\Pr(d_\ell \geqslant \sqrt{\gamma}) \leqslant \frac{\mathbb{E}[d_\ell]}{\sqrt{\gamma}} < \frac{\gamma}{\sqrt{\gamma}} = \sqrt{\gamma}.$$

This is precisely the first claimed statement.

For the second statement, we simply replace $\mathcal{H}_0$ by its relative complement (namely the hypergraph whose hyperedges are precisely those triangles in G which do not form a hyperedge of $\mathcal{H}_0$), and the result follows from the first statement. $\square$

We will now begin working with chain partitions of hypergraphs, which are true chain partitions rather than partitions into cylinders. To minimize confusion, we use the letter $\mathcal{Q}$ for such objects, as in the introduction.

Definition 4.5. Let $\mathcal{P} = (\mathcal{P}_V, \mathcal{P}_E)$ be a cylinder chain partition of a t-partite hypergraph $\mathcal{H}$ on vertex set $X_1 \sqcup \cdots \sqcup X_t$. The *Venn diagram partition* of $\mathcal{P}$ is the chain partition $\mathcal{Q} = \mathcal{Q}(\mathcal{P})$ defined as follows. First, two vertices $x_i, x_i' \in X_i$ are in the same part of $\mathcal{Q}_V$ if and only if, for all $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_t) \in X_1 \times \cdots \times X_{i-1} \times X_{i+1} \times \cdots \times X_t$, and for all $Y \in \mathcal{P}_V$, we have

$$(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_t) \in Y \iff (x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_t) \in Y.$$

That is, x_i, x_i' are in the same part of $\mathcal{Q}_V$ if and only if they participate in exactly the same cylinders in $\mathcal{P}_V$.

Now, given parts $Y_i \subseteq X_i, Y_j \subseteq X_j$ of $\mathcal{Q}_V$, we define the edge partition $\mathcal{Q}_E(Y_i \times Y_j)$ as follows. Two edges $(x_i, x_j), (x_i', x_j') \in Y_i \times Y_j$ lie in the same part of $\mathcal{Q}_E(Y_i \times Y_j)$ if and only if they participate in all of the same edge parts of $\mathcal{P}_E(Y)$, over all vertex cylinders $Y \in \mathcal{P}_V$ which contain Y_i, Y_j.

Note that, if $\mathcal{Q}$ is the Venn diagram partition of $\mathcal{P}$, then $|\mathcal{Q}_V| \leqslant t 2^{|\mathcal{P}_V|}$, since, for each part X_i, there are at most $2^{|\mathcal{P}_V|}$ possible intersection patterns of cylinders in $\mathcal{P}_V$ with X_i. Similarly, $|\mathcal{Q}_E(Y_i \times Y_j)| \leqslant 2^{|\mathcal{P}_E||\mathcal{P}_V|}$ for every $Y_i, Y_j \in \mathcal{Q}_V$, since there are at most $|\mathcal{P}_V|$ options for a cylinder $Y \in \mathcal{P}_V$ containing Y_i, Y_j, and at most $|\mathcal{P}_E|$ bipartite graphs in $\mathcal{P}_E(Y)$, and thus at most $2^{|\mathcal{P}_E||\mathcal{P}_V|}$ possible intersection patterns with $Y_i \times Y_j$.

Our next lemma shows that by taking the Venn diagram partition of the cylinder chain partition given by Lemma 4.3, we can obtain a true chain partition such that almost all triples have relative density close to 0 or 1.

Lemma 4.6. *Let $d \geqslant 1, t \geqslant 3$ be integers and let $\zeta \in (0, \frac{1}{10}]$ be a parameter. Let $\mathcal{H}$ be a sufficiently large 3-graph with VC_2 dimension at most d, and let $X_1 \sqcup \cdots \sqcup X_t$ be an arbitrary equitable partition of $V(\mathcal{H})$. There is a chain partition $\mathcal{Q} = (\mathcal{Q}_V, \mathcal{Q}_E)$ of $\mathcal{H}$ such that the following holds for all $1 \leqslant i < j < k \leqslant t$. If $(x_i, x_j, x_k) \in X_i \times X_j \times X_k$ is sampled uniformly at random, then with probability at least $1 - 3\zeta$, we have that*

$$d(\mathcal{H}[\mathcal{Q}; x_i, x_j, x_k]) \in [0, \zeta] \cup [1 - \zeta, 1].$$

Moreover,

$$|\mathcal{Q}_V| \leqslant \operatorname{twr}\left(\left(\frac{t}{\zeta}\right)^C\right) \qquad \text{and} \qquad |\mathcal{Q}_E(Y \times Y')| \leqslant \operatorname{twr}\left(\left(\frac{t}{\zeta}\right)^C\right)$$

for every $Y, Y' \in \mathcal{Q}_V$, where $C > 0$ is a constant depending only on d.

Proof. Apply Lemma 4.3 with parameter $\gamma = \zeta^2$ to obtain a cylinder chain partition $\mathcal{P}$, and let $\mathcal{Q}$ be the Venn diagram partition of $\mathcal{P}$. Note that the claimed bounds on $|\mathcal{Q}_V|, |\mathcal{Q}_E(Y \times Y')|$ follow immediately from the corresponding bounds on $|\mathcal{P}_V|, |\mathcal{P}_E|$.

For the rest of the proof, we fix some $1 \leqslant i < j < k \leqslant t$. For a vertex cylinder $Y \in \mathcal{P}_V$, we define its *homogeneous fraction* $h(Y)$ to be the probability, for uniformly random $(y_1, \ldots, y_t) \in Y$, that $d(\mathcal{H}[\mathcal{P}_V[y_1, \ldots, y_t]; \mathcal{P}_E[y_1, \ldots, y_t; \{i, j, k\}]]) \in [0, \gamma] \cup [1 - \gamma, 1]$. Equivalently, $h(Y)$ is the fraction of triples $(y_i, y_j, y_k) \in Y_i \times Y_j \times Y_k$ which define a tripartite chain whose relative density is in $[0, \gamma] \cup [1 - \gamma, 1]$, i.e.

$$h(Y) = \sum_{\substack{Z_{ij} \in \mathcal{P}_E(Y_i \times Y_j) \\ Z_{ik} \in \mathcal{P}_E(Y_i \times Y_k) \\ Z_{jk} \in \mathcal{P}_E(Y_j \times Y_k)}} \frac{|\Delta(Z_{ij} \cup Z_{ik} \cup Z_{jk})|}{|Y_i||Y_j||Y_k|} \mathbf{1}_{d(\mathcal{H}|Z_{ij} \cup Z_{ik} \cup Z_{jk}) \in [0, \gamma] \cup [1 - \gamma, 1]}.$$

Let us say that a vertex cylinder $Y \in \mathcal{P}_V$ is *typical* if $h(Y) \geqslant 1 - \gamma$.

To justify this name, we first claim that a $(1 - \gamma)$-fraction of the tuples in $X_1 \times \cdots \times X_t$ lie in typical vertex cylinders. Indeed, Lemma 4.3 states that if we sample $(x_1, \ldots, x_t) \in X_1 \times \cdots \times X_t$ uniformly, then with probability at least $1 - \gamma^2$, we have $d(\mathcal{H}[\mathcal{P}_V[x_1, \ldots, x_t]; \mathcal{P}_E[x_1, \ldots, x_t; \{i, j, k\}]]) \in [0, \gamma] \cup [1 - \gamma, 1]$. But this probability is exactly the weighted average of $h(Y)$ over all cylinders Y, namely

$$\sum_{Y \in \mathcal{P}_V} \left(\prod_{i=1}^t \frac{|Y_i|}{|X_i|}\right) h(Y) \geqslant 1 - \gamma^2.$$

By Markov's inequality, we find that if $(x_1, \ldots, x_t) \in X_1 \times \cdots \times X_t$ is sampled uniformly at random, then with probability at least $1 - \gamma$ it lies in a typical vertex cylinder.

Now, let Y be a typical vertex cylinder. By the definition of typicality, a $(1 - \gamma)$-fraction of the triples in $Y_i \times Y_j \times Y_k$ define edge parts Z_{ij}, Z_{ik}, Z_{jk} for which the relative density is in $[0, \gamma] \cup [1 - \gamma, 1]$. Let us call such tuples $(Y_i \cup Y_j \cup Y_k, Z_{ij} \cup Z_{ik} \cup Z_{jk})$ *doubly typical*. What we have now found is that, if $(x_1, \ldots, x_t)$ is sampled uniformly at random, then it lies in a doubly typical chain with probability at least $1 - 2\gamma$.

Finally, we fix a doubly typical tuple $(Y_i \cup Y_j \cup Y_k, Z_{ij} \cup Z_{ik} \cup Z_{jk})$. By definition, $\mathcal{Q}_V$ and $\mathcal{Q}_E$ give some further partition of the vertex and edge set of this tripartite graph. By Lemma 4.4, we conclude that after this partitioning, a $(1 - \sqrt{\gamma})$-fraction of the triples of vertices in this tripartite graph still define tripartite subgraphs on which $\mathcal{H}$ has relative density in $[0, \sqrt{\gamma}] \cup [1 - \sqrt{\gamma}, 1] = [0, \zeta] \cup [1 - \zeta, 1]$.

Putting this all together, we find that if $(x_1, \ldots, x_t) \in X_1 \times \cdots \times X_t$ is sampled uniformly at random, the probability that $d(\mathcal{H}[\mathcal{Q}; x_i, x_j, x_k]) \in [0, \zeta] \cup [1 - \zeta, 1]$ is at least $1 - 2\gamma - \sqrt{\gamma} \geqslant 1 - 3\zeta$. □

The final result we need before proving Theorem 2.9 is a version of Szemerédi's regularity lemma for simultaneously regularizing multiple graphs, similar to Lemma 3.4, which is the cylinder regularity version of the same statement. This statement is well-known, e.g. [34, Theorem 7.8].

Given a chain partition $(\mathcal{Q}_V, \mathcal{Q}_E)$ and a refinement $\mathcal{Q}_V'$ of $\mathcal{Q}_V$, we denote by $\mathcal{Q}_E'$ the edge partition obtained by restricting each bipartite graph in $\mathcal{Q}_E$ to the new vertex parts in $\mathcal{Q}_V'$.

Lemma 4.7 ([34, Theorem 7.8]). *Let $\alpha > 0$, and let $\mathcal{Q} = (\mathcal{Q}_V, \mathcal{Q}_E)$ be a chain partition of some vertex set V. Let $T = |\mathcal{Q}_V|$, and let L be the maximum of $|\mathcal{Q}_E(Y \times Y')|$ over all distinct $Y, Y' \in \mathcal{Q}_V$. There is a refinement $\mathcal{Q}_V'$ of $\mathcal{Q}_V$ such that a $(1 - \alpha)$-fraction of pairs of vertices (x_1, x_2) define an α-quasirandom graph $\mathcal{Q}_E'[x_1, x_2]$. Moreover,*

$$|\mathcal{Q}_V'| \leqslant \mathrm{twr}(16\alpha^{-3}T^2 L)$$

and

$$|\mathcal{Q}_E'(Y \times Y')| \leqslant L$$

for all $Y, Y' \in \mathcal{Q}_V'$.

With these tools, we are finally ready to prove Theorem 2.9, our regularity lemma for hypergraphs of bounded VC_2 dimension.

Proof of Theorem 2.9. Apply Lemma 4.6 with $\zeta = \eta^2/16$ and $t = 3/\zeta$ to obtain a partition $\mathcal{Q}^{\mathrm{hom}}$. The choice of t guarantees that at most a ζ-fraction of the triples of vertices do not go between three different parts. Together with the conclusion of Lemma 4.6, this implies that for a $(1 - 4\zeta)$-fraction of triples of vertices, the chain containing the triple has relative density in $[0, \zeta] \cup [1 - \zeta, 1]$. Let L be the maximum of $|\mathcal{Q}_E^{\mathrm{hom}}(Y \times Y')|$ over all $Y, Y' \in \mathcal{Q}_V^{\mathrm{hom}}$, so that $L \leqslant \mathrm{twr}(\eta^{-C_1})$ by Lemma 4.6, for a constant C_1 depending only on d. Similarly, letting $T = |\mathcal{Q}_V|$, we have that $T \leqslant \mathrm{twr}(\eta^{-C_1})$ by Lemma 4.6.

We now apply Lemma 4.7 with parameter $\alpha = \psi((\eta/(9L))^3)$ to obtain a new chain partition $\mathcal{Q} = (\mathcal{Q}_V, \mathcal{Q}_E)$, obtained by refining $\mathcal{Q}_V^{\mathrm{hom}}$ and then restricting the graphs in $\mathcal{Q}_E^{\mathrm{hom}}$ to the new vertex parts. By Lemma 4.7, we find that

$$|\mathcal{Q}_V| \leqslant \mathrm{twr}(16\alpha^{-3}T^2 L) \leqslant \mathrm{twr}(\mathrm{twr}(\psi(\eta)^{-C_2}))$$

for a constant C_2, and

$$|\mathcal{Q}_E(Y \times Y')| \leqslant L \leqslant \mathrm{twr}(\eta^{-C_1})$$

for all $Y, Y' \in \mathcal{Q}_V$. By Lemma 4.7, a $(1 - \alpha)$-fraction of pairs of vertices lie in α-quasirandom bipartite graphs in $\mathcal{Q}_E$. Call such a bipartite graph *sparse* if its edge density is less than $\eta/(9L)$. At most an $(\eta/9)$-fraction of pairs of vertices (x, x') define a sparse graph $\mathcal{Q}_E[x, x']$. Therefore, if we pick a triple of vertices uniformly at random, then with probability at least $1 - 3(\eta/9 + \alpha) = 1 - \eta/3 - 3\alpha$, they lie in a chain of $\mathcal{Q}$ in which all of the three bipartite graphs are α-quasirandom

and not sparse. In particular, for every such chain, the corresponding tripartite graph G satisfies $\delta(G) \geqslant (\eta/(9L))^3$, hence G is $\psi(\delta(G))$-quasirandom by the monotonicity of ψ.

Recall that in $\mathcal{Q}_V^{\text{hom}}$, a $(1 - 4\zeta)$-fraction of triples lie in chains with relative density in $[0, \zeta] \cup [1 - \zeta, 1]$. By Lemma 4.4, we conclude that in $\mathcal{Q}_V$, a $(1 - 4\zeta - \sqrt{\zeta})$-fraction of triples lie in chains with relative density in $[0, \sqrt{\zeta}] \cup [1 - \sqrt{\zeta}, 1]$. Thus, for a random triple, the failure probability is at most

$$\frac{\eta}{3} + 3\alpha + 4\zeta + \sqrt{\zeta} \leqslant \eta$$

using our assumption that $\psi(x) \leqslant x$ and our definitions of ζ and α. This concludes the proof. $\square$

5 Proof of the lower bound

As mentioned in the introduction, Proposition 1.5 follows from the following two results of Terry. The first states, in a quantitative form, the fact that quasirandomness for 3-graphs implies weak quasirandomness. This fact is well-known and not hard to prove, see e.g. [21, 42]. However, to the best of our knowledge, [62, Lemma 5.15] is the only place in the literature where this reduction is given a quantitative form.

Proposition 5.1 (Terry [62, Lemma 5.15]). *There exists an absolute constant $C > 0$ such that the following holds for all $\eta > 0$ and all $\psi : (0, 1) \to (0, 1)$ satisfying $\psi(x) \leqslant x^C$. If a partition $\mathcal{Q} = (\mathcal{Q}_V, \mathcal{Q}_E)$ of a 3-graph is (η, ψ)-quasirandom, then the partition $\mathcal{Q}_V$ is weakly $\eta^{1/C}$-quasirandom.*

Second, we need the following result, which states that there exist 3-graphs of bounded VC_2 dimension admitting no weakly quasirandom partitions with fewer than tower-type many parts.

Proposition 5.2 (Terry [61, Theorem 5.4]). *For every sufficiently small $\eta > 0$, there are infinitely many 3-graphs $\mathcal{H}$ with the following properties. $\mathcal{H}$ has VC_2 dimension at most 1, and if $\mathcal{Q}_V$ is a weakly η-quasirandom partition of $V(\mathcal{H})$, then $|\mathcal{Q}_V| \geqslant \text{twr}(\text{poly}(1/\eta))$.*

Note that this lower bound is essentially best possible, since Theorem 2.8 produces a weakly η-quasirandom partition with $\text{twr}(\text{poly}(1/\eta))$ parts for *any* $\mathcal{H}$. Note too that by combining Propositions 5.1 and 5.2, we immediately obtain Proposition 1.5.

For completeness, we briefly sketch the proof of Proposition 5.2, in order to explain where the tower-type bound comes from. The construction used in the proof is the following simple operation, which turns a bipartite graph into a tripartite 3-graph.

Definition 5.3. Let G be a bipartite graph with parts A, B, and let n be an integer. The *n-fold cone hypergraph* of G, denoted $n \otimes G$, is the 3-graph with parts A, B, C, where $|C| = n$, and with edge set

$$E(n \otimes G) = \{(a, b, c) : ab \in E(G), c \in C\}.$$

In other words, $n \otimes G$ is obtained from G by adding n new vertices, each of which has link precisely equal to G. The basic observation about this construction is that $n \otimes G$ has bounded VC_2 dimension, regardless of the choice of G.

Lemma 5.4. *For every bipartite graph G and every integer n, the 3-graph $n \otimes G$ has VC_2 dimension at most 1.*

Proof. It suffices to prove that $n \otimes G$ has no shattered $K_{2,2}$. Let the parts of $n \otimes G$ be A, B, C, as in Definition 5.3. As $n \otimes G$ is tripartite, any shattered $K_{2,2}$ must have its two parts contained in two of the parts A, B, C.

Consider first some $a, a' \in A, b, b' \in B$. They clearly cannot form a shattered $K_{2,2}$, since all $c \in C$ have the same link when restricted to these four vertices, whereas in a shattered $K_{2,2}$ we would need to see 16 different links on these four vertices.

Next consider some $a, a' \in A, c, c' \in C$. We claim that for every $b \in B$, if $ac \in N(b)$, then $ac' \in N(b)$ as well. Note that this immediately implies that that a, a', c, c' do not form a shattered $K_{2,2}$, as for example we can never see the subgraph $\{(a, c)\} \subseteq \{a, a'\} \times \{c, c'\}$ in the link of some $b \in B$. To prove the claim, suppose that $ac \in N(b)$ for some $b \in B$. By definition of $n \otimes G$, this implies that $ab \in E(G)$, hence that $abc' \in E(n \otimes G)$, and therefore that $ac' \in N(b)$ as well, as claimed.

The final case, of a shattered $K_{2,2}$ among $B \times C$, is ruled out identically to the previous case, by simply interchanging the roles of A and B. $\square$

With this construction in hand, it is straightforward to describe the proof of Proposition 5.2. We begin by letting G be Gowers's example of a graph admitting no η-quasirandom partition with fewer than $\text{twr}(\text{poly}(1/\eta))$ parts. We may moreover assume that G is bipartite[16] with both parts of size n. Now, we let $\mathcal{H} = n \otimes G$. By Lemma 5.4, we know that $\mathcal{H}$ has bounded VC_2 dimension.

To complete the proof, it suffices to show that $\mathcal{H}$ has no weakly η-quasirandom partition into fewer than $\text{twr}(\text{poly}(1/\eta))$ parts. While proving this is somewhat long and tedious (see [61, Lemma 5.3] for details), the fact that such a statement is true should hopefully not be surprising: the 3-graph $n \otimes G$ contains "no extra information" beyond the graph G, hence a weakly quasirandom partition of the former should immediately yield a quasirandom partition of the latter. Once one formalizes this intuitive argument, the proof of Proposition 5.2 is complete.

6 Concluding remarks and open problems

6.1 Other applications of cylinder regularity for hypergraphs

The major technical tool introduced in this paper is the cylinder regularity lemma for hypergraphs, Theorem 2.13, and we expect this result to have many other applications. We now briefly discuss two such applications, as well as some open problems related to them.

The first of these applications is a hypergraph extension of the quasirandom subset lemma of Conlon and Fox [18]. For a graph G and a set $U \subseteq V(G)$, we say that U is *ε-quasirandom* if the pair (U, U) is ε-quasirandom. More formally, U is ε-quasirandom if the pair (U, U) is ε-quasirandom in the graph $G \times K_2$ (the *bipartite double cover* of G), whose vertex set is two copies of $V(G)$, and whose edges are all pairs crossing the two parts that correspond to an edge of G. The quasirandom subset lemma then states that every graph contains a linear-sized quasirandom subset.

Lemma 6.1 ([18, Lemma 5.2]). *For every $\varepsilon > 0$, there exists $\delta > 0$ such that for every graph G, there is a set $U \subseteq V(G)$ with $|U| \geqslant \delta|V(G)|$ such that U is ε-quasirandom.*

Lemma 6.1 is extremely useful (see e.g. [19, 20] and [68, Chapter 6] for several applications). This statement can be deduced from Szemerédi's regularity lemma, but the remarkable insight of

---

[16]Gowers's [33] construction does not explicitly provide a balanced bipartite graph G, but it is easy to see that his construction can be modified to yield this, and this is also explicitly the case in the construction of Fox–Lovász [24].

Conlon and Fox is that one can find an alternative proof which yields much stronger quantitative bounds; Conlon and Fox proved that $1/\delta$ can be taken to be merely double-exponential in $1/\text{poly}(\varepsilon)$. The proof of Lemma 6.1 with such a bound is quite short: it follows from an application of the cylinder regularity lemma and Ramsey's theorem (although see [68, Lemma 6.1.4] for a direct energy-increment proof).

One of the many immediate consequences of Lemma 6.1 is a famous result of Rödl [49], which we now discuss. Recall that a graph G is said to be *induced-F-free* if it does not have an induced subgraph isomorphic to F. Rödl proved that if G is induced-F-free, it has a linear-sized vertex subset whose edge density is close to 0 or 1. More precisely, there is a set $U \subseteq V(G)$ with $d_G(U) \in [0, \varepsilon] \cup [1 - \varepsilon, 1]$, where $|U| \geqslant \delta|V(G)|$ and $\delta > 0$ depends only on F and ε. Rödl's theorem turns out to be closely connected to other fundamental questions in graph theory such as the Erdős–Hajnal conjecture, see e.g. [10] for details.

There are a number of ways known for proving Rödl's theorem, but one of the simplest is as a direct consequence of Lemma 6.1: by the induced counting lemma, a quasirandom set in an induced-F-free cannot have edge density bounded away from 0 or 1. See e.g. [68, Theorem 6.3.3] for a detailed proof via this method.

As Rödl himself observed [49], the obvious extension of this statement to hypergraphs is false. Indeed, the well-known *random tournament* construction—the 3-graph whose hyperedges are the cyclic triangles in a random tournament—is induced-$K_4^{(3)}$-free, but all of its linear-sized subsets have edge density $\frac{1}{4} + o(1)$. In fact, the largest subsets of this 3-graph with density close to 0 or 1 have logarithmic order, and a result of Amir, Shapira, and Tyomkyn [5] shows that this is the smallest possible in an induced-F-free 3-graph.

The same random tournament construction is one of the classical examples demonstrating the difficulty in defining a successful notion of hypergraph regularity. Indeed, the same construction demonstrates that there is no direct hypergraph analogue of Lemma 6.1, in the sense that every linear-sized vertex subset of this random tournament hypergraph spans a 3-graph which is not quasirandom. As such, one can learn the lessons of hypergraph regularity, and decide that in order to sensibly extend Rödl's theorem to hypergraphs, one should not only restrict the vertex set of $\mathcal{H}$, but also the set of pairs. Indeed, we prove the following theorem.

Theorem 6.2. *Let F be a 3-graph, let $\varepsilon > 0$, and let $\psi : (0, 1) \to (0, 1)$ be a polynomial function. There exists $\delta > 0$ such that the following holds. If $\mathcal{H}$ is an induced-F-free 3-graph, then there is a subset $U \subseteq V(\mathcal{H})$ and a graph G with vertex set U such that $|U| \geqslant \delta|V(\mathcal{H})|$, $e(G) \geqslant \delta|U|^2$, G is $\psi(\delta)$-quasirandom, and*

$$d(\mathcal{H}[U, G] \mid G) \in [0, \varepsilon] \cup [1 - \varepsilon, 1].$$

Moreover, we may take $1/\delta \leqslant \text{twr}(2^{2^{\text{poly}(1/\varepsilon)}})$.

Note that it is natural to insist that G is highly quasirandom, as otherwise the result may be vacuously true (e.g. if G is a complete bipartite graph, which has positive density but no triangles). Moreover, this can be viewed as a direct extension of the condition in Rödl's theorem, since every set of vertices is "highly quasirandom in uniformity 1".

We prove Theorem 6.2 via the same proof technique discussed above, and as such, we begin by proving an analogue of Lemma 6.1 for 3-graphs. Given a hypergraph $\mathcal{H}$, a vertex subset $U \subseteq V(\mathcal{H})$, and a graph G with vertex set U, let $\mathcal{H}[U, G]$ denote the subhypergraph of $\mathcal{H}$ obtained by only keeping the vertices in U, and only keeping the hyperedges supported on $\Delta(G)$. Let us say that $\mathcal{H}[U, G]$ is (η, ψ)-quasirandom if the following tripartite chain is (η, ψ)-quasirandom: its vertex set

consists of three copies of U, its tripartite graph consists of three copies of the bipartite double cover of G, and its hyperedges are all triples across the three parts corresponding to hyperedges in $\mathcal{H}$. With this terminology, we can state the following hypergraph extension of Lemma 6.1.

Theorem 6.3. *For every $\eta > 0$ and every polynomial function $\psi : (0,1) \to (0,1)$, there exists $\delta > 0$ such that the following holds for every 3-graph $\mathcal{H}$. There is a subset $U \subseteq V(\mathcal{H})$ and a graph G with vertex set U such that $|U| \geqslant \delta|V(\mathcal{H})|$, $e(G) \geqslant \delta|U|^2$, and $\mathcal{H}[U,G]$ is (η,ψ)-quasirandom.*
Moreover, we may take $1/\delta \leqslant \mathrm{twr}(2^{2^{\mathrm{poly}(1/\eta)}})$.

In other words, we may restrict $V(\mathcal{H})$ to a constant-density subset, and restrict the complete graph on U to a constant-density subgraph, such that the hypergraph $\mathcal{H}[U,G]$ supported on these substructures is quasirandom. Note that, as above, we must include the graph G for such a statement to be true: the random tournament construction witnesses that no such statement can hold if we are not allowed to restrict both the set of vertices and the set of pairs. The proof of Theorem 6.3 closely follows that of Lemma 6.1, with the main difference being an application of Theorem 2.13 rather than the cylinder regularity lemma for graphs. Note that, as in the graph case, one can easily deduce a statement like Theorem 6.3 from the hypergraph regularity lemma, but such a proof would yield wowzer-type bounds; the main upshot of Theorem 6.3 is the quantitative improvement over this.

Proof sketch of Theorem 6.3. Let t be a large parameter to be chosen later, and arbitrarily partition $V(\mathcal{H})$ into t equal-sized parts. Apply Theorem 2.13 to obtain an (η^2, ψ^2)-quasirandom cylinder chain partition of this t-partite 3-graph. In fact, we will not need the whole partition, and we may simply restrict to a single quasirandom cylinder (Y, Z). Let us assume for simplicity that the vertex cylinder is equitable, i.e. that $|Y_1| = \cdots = |Y_t|$.

By randomly subsampling, we may assume that all bipartite graphs appearing in Z have the same edge density δ; it is well-known (and follows straightforwardly[17] from Azuma's inequality) that random subsampling preserves quasirandomness, up to some small losses. We now define an edge-coloring of the complete t-vertex 3-graph with $1/\eta^2$ colors, as follows. The triple $\{i, j, k\}$ receives color ℓ if $d(\mathcal{H}[Y; Z; \{i,j,k\}]) \in [(\ell-1)\eta^2, \ell\eta^2]$ (if a triple receives two colors, we break ties arbitrarily). By Ramsey's theorem for hypergraphs, we may pass to a subset of s parts such that all triples among these parts receive the same color.

By renaming the indices, this means we have found disjoint vertex subsets $Y_1, \ldots, Y_s$ as well as bipartite graphs $Z_{ij} \subseteq Y_i \times Y_j$, all of the same edge density, such that for all $1 \leqslant i < j < k \leqslant s$, all the chains $\mathcal{H}[Y; Z; \{i,j,k\}]$ are (η^2, ψ^2)-quasirandom, and have the same relative density up to an additive error of η^2. At this point, we let $U = Y_1 \cup \cdots \cup Y_s$, and let $E(G) = \bigcup_{i \leqslant j} E(Z_{ij})$, where we define Z_{ii} to be a random graph with vertex set Y_i and edge density δ for all i. In particular, with high probability as $|V(\mathcal{H})| \to \infty$, each graph Z_{ii} is $\psi(\delta)^2$-quasirandom. Thus G is the union of graphs, each of which has edge density δ and is $\psi(\delta)^2$-quasirandom, which implies that G itself has edge density δ and is $\psi(\delta)$-quasirandom.

Finally, we claim that if $s \geqslant 10/\eta^2$, then $\mathcal{H}[U, G]$ is (η, ψ)-quasirandom. Indeed, if s is so large, then at least a $(1 - \eta^2)$-fraction of triples in U^3 lie in three distinct parts Y_i, Y_j, Y_k. As all chains $\mathcal{H}[Y; Z; \{i,j,k\}]$ are (η^2, ψ^2)-quasirandom and have the same density up to an additive error of η^2,

---

[17]For this proof, it is very convenient to use the definition of quasirandomness in Definition 2.1, rather than other equivalent notions like ε-regularity, for which naive applications of Azuma's inequality and the union bound do not give sufficiently strong estimates.

it is not hard to conclude that $\mathcal{H}[U, G]$ is indeed (η, ψ)-quasirandom (see [18, Lemma 5.5] for the analogous argument in uniformity 2).

This concludes the proof sketch; all that remains is to understand its quantitative aspects. We set $s = 10/\eta^2$. We then need t to be at least the $(1/\eta^2)$-color hypergraph Ramsey number of s, so we may take $t = 2^{2^{100\eta^{-4}}}$. Finally, Theorem 2.13 then implies that both δ and $|Y_i|/|V(\mathcal{H})|$ can be lower-bounded as $1/\operatorname{twr}(2^{2^{\operatorname{poly}(1/\eta)}})$, as claimed. $\qquad\square$

It is now straightforward to deduce our generalization of Rödl's theorem, Theorem 6.2, from Theorem 6.3.

Proof sketch of Theorem 6.2. Let η be polynomially small with respect to ε. By potentially replacing ψ with a smaller polynomial function, we may assume that the induced counting lemma (Lemma 4.1) applies, that is, that whenever $\mathcal{H}[U, G]$ is (η, ψ)-quasirandom with relative density in $[\varepsilon, 1 - \varepsilon]$, then $\mathcal{H}[U, G]$ contains an induced copy of F. In fact, by applying the induced counting lemma for chains [35, Corollary 5.3], we may assume that in this situation, we can find an induced copy of F so that all pairs of vertices in F span an edge of G.

We now apply Theorem 6.3 to $\mathcal{H}$ to obtain U and G as above. We are done if $d(\mathcal{H}[U, G] \mid G) \in [0, \varepsilon] \cup [1 - \varepsilon, 1]$. If this does not happen, then we can find an induced copy of F in $\mathcal{H}[U, G]$ as above. In particular, the fact that all pairs of vertices in F span an edge of G implies that this copy is also induced in $\mathcal{H}$, a contradiction. $\qquad\square$

The proofs above give tower-type bounds in Theorems 6.2 and 6.3, and we think it would be extremely interesting to determine whether these bounds are close to best possible.

Question 6.4. *Can one prove Theorem 6.3 or Theorem 6.2 with bounds of constant tower height?*

The analogous questions in uniformity 2 are also very interesting: in Lemma 6.1, it remains a major open problem to determine whether the double-exponential dependence can be improved to single exponential (which would be best possible). In Rödl's theorem, quasipolynomial bounds are known [29], and improving these bounds to polynomial is now known to be equivalent to proving the Erdős–Hajnal conjecture [10].

In order to answer Question 6.4, it is natural to try to improve the bounds in the cylinder regularity lemma. However, we believe that this is impossible, and expect the bound in Theorem 2.13 to be essentially best possible.

Conjecture 6.5. *There exists a tripartite 3-graph $\mathcal{H}$ for which every (η, ψ)-quasirandom cylinder partition has at least $\operatorname{twr}(\psi(\eta)^{-C})$ parts, for some constant C.*

It may be possible to prove Conjecture 6.5 by mimicking the technique of Moshkovitz and Shapira [47]. Alternately, it may be possible to deduce Conjecture 6.5 as a black-box reduction to the Moshkovitz–Shapira theorem: it may be possible to prove the full hypergraph regularity lemma by iterative applications of Theorem 2.13, in which case sub-tower bounds for Theorem 2.13 would yield sub-wowzer bounds for the 3-graph regularity lemma, contradicting the lower bound of Moshkovitz–Shapira. In uniformity 2, it is well-known (see [51, Section 2.2] or [18, Section 5.4]) that iterative applications of weak regularity lemmas such as those of Duke–Lefmann–Rödl [22] or Frieze–Kannan [31] can be used to prove Szemerédi's regularity lemma, hence we expect a similar technique to exist in higher uniformities as well.

6.2 Further open problems

Another natural problem left open by our results is to close the gap between Proposition 1.5 and Theorem 2.9, i.e. to determine whether the true behavior for 3-graphs with bounded VC_2 dimension is of tower type or of double-tower type. We conjecture that the lower bound is closer to the truth.

Conjecture 6.6. *The bound in Theorem 2.9 can be improved to $|\mathcal{Q}_V| \leqslant \mathrm{twr}(\psi(\eta)^{-C})$.*

We recall that throughout this paper, we have worked with Gowers's notion of hypergraph regularity, because this notion is known to support a counting lemma with polynomial bounds (concretely, we required that in Lemmas 4.1 and 4.2, η and ψ depend polynomially on γ, and that ψ itself is a polynomial). As far as we are aware, all known proofs of the counting lemma for the regularity notion of Rödl et al. use, as an intermediate step of the process, an application of the hypergraph regularity lemma. As such, the counting lemma requires terrible bounds. However, it is not clear if this is the truth, or merely an artifact of the known proofs.

Question 6.7. *Can one prove a counting lemma with polynomial bounds for other notions of hypergraph regularity?*

Finally, we remark that Theorem 1.4 gives a rough structural characterization of all 3-graphs with bounded VC_2 dimension: in some sense, the only way to obtain such a 3-graph is to start with some graph, and include all of its triangles as hyperedges. More precisely, Theorem 1.4 implies that if $\mathcal{H}$ has bounded VC_2 dimension, then it is constructed in such a way from a bounded number of graphs, apart from a small amount of "noise"; this is precisely the meaning of the fact that $\mathcal{H}$ has a regularity partition in which almost all triples have relative density close to 0 or 1.

From the perspective of such a structural characterization, it is no longer necessary to demand that the graphs arising in the partition are themselves highly quasirandom (recall that in the setting of hypergraph regularity, and specifically the counting lemma, such an assumption is absolutely crucial). To make this notion precise, let us assume for simplicity that $\mathcal{H}$ is tripartite with parts X, Y, Z, each of size n. We then define a ε-*graph partition* of $\mathcal{H}$ to be partitions $X \times Y = E_1 \sqcup \cdots \sqcup E_M, X \times Z = F_1 \sqcup \cdots \sqcup F_M, Y \times Z = G_1 \sqcup \cdots \sqcup G_M$, as well as a function $f : [M]^3 \to \{0, 1\}$, with the property that for all but εn^3 choices of $(x, y, z) \in X \times Y \times Z$, we have

$$xyz \in E(\mathcal{H}) \Longleftrightarrow f(i, j, k) = 1, \qquad \text{where } xy \in E_i, xz \in F_j, yz \in G_k.$$

Theorem 1.4 implies that if $\mathcal{H}$ has bounded VC_2 dimension, then it admits an ε-graph partition with M bounded by a tower-type function. Note that although the results of Terry [59, 63] imply polynomial bounds on the number of edge parts in a regularity decomposition, they only give wowzer-type bounds on the size of an ε-graph partition, because they still require a wowzer-type number of vertex parts, and hence a wowzer-type number of graphs in such a partition. However, we conjecture that the bound can be improved to polynomial.

Conjecture 6.8. *If $\mathcal{H}$ has bounded VC_2 dimension, then it has an ε-graph partition with $M = \mathrm{poly}(1/\varepsilon)$ many graphs.*

Note that, if true, Conjecture 6.8 would imply Conjecture 6.6, as we could simply apply Szemerédi's regularity lemma to all the graphs in such an ε-graph partition, while maintaining the homogeneity property by Lemma 4.4.

References

[1] N. Alon, E. Fischer, M. Krivelevich, and M. Szegedy, Efficient testing of large graphs, *Combinatorica* **20** (2000), 451–476.

[2] N. Alon, E. Fischer, and I. Newman, Efficient testing of bipartite graphs for forbidden induced subgraphs, *SIAM J. Comput.* **37** (2007), 959–976.

[3] N. Alon, J. Fox, and Y. Zhao, Efficient arithmetic regularity and removal lemmas for induced bipartite patterns, *Discrete Anal.* (2019), Paper No. 3, 14pp.

[4] N. Alon, R. Livni, M. Malliaris, and S. Moran, Private PAC learning implies finite Littlestone dimension, in *STOC'19—Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, ACM, New York, 2019, 852–860.

[5] M. Amir, A. Shapira, and M. Tyomkyn, Two Erdős–Hajnal-type theorems in hypergraphs, *J. Combin. Theory Ser. B* **146** (2021), 417–438.

[6] J. Balogh, A. Bernshteyn, M. Delcourt, A. Ferber, and H. T. Pham, Sunflowers in set systems with small VC-dimension, 2024. Preprint available at arXiv:2408.04165.

[7] Ö. Beyarslan, Random hypergraphs in pseudofinite fields, *J. Inst. Math. Jussieu* **9** (2010), 29–47.

[8] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth, Learnability and the Vapnik-Chervonenkis dimension, *J. Assoc. Comput. Mach.* **36** (1989), 929–965.

[9] N. Brukhim, D. Carmon, I. Dinur, S. Moran, and A. Yehudayoff, A characterization of multiclass learnability, in *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science—FOCS 2022*, IEEE Computer Soc., Los Alamitos, CA, 2022, 943–955.

[10] M. Bucić, J. Fox, and H. T. Pham, Equivalence between Erdős–Hajnal and polynomial Rödl and Nikiforov conjectures, 2024. Preprint available at arXiv:2403.08303.

[11] A. Chernikov and N. Hempel, On n-dependent groups and fields II, *Forum Math. Sigma* **9** (2021), Paper No. e38, 51.

[12] A. Chernikov, D. Palacin, and K. Takeuchi, On n-dependence, *Notre Dame J. Form. Log.* **60** (2019), 195–214.

[13] A. Chernikov and S. Starchenko, Definable regularity lemmas for NIP hypergraphs, *Q. J. Math.* **72** (2021), 1401–1433.

[14] A. Chernikov and H. Towsner, Hypergraph regularity and higher arity VC-dimension, 2020. Preprint available at arXiv:2010.00726.

[15] F. R. K. Chung, Regularity lemmas for hypergraphs and quasi-randomness, *Random Structures Algorithms* **2** (1991), 241–252.

[16] F. R. K. Chung, R. L. Graham, and R. M. Wilson, Quasi-random graphs, *Combinatorica* **9** (1989), 345–362.

[17] G. Conant, A. Pillay, and C. Terry, Structure and regularity for subsets of groups with finite VC-dimension, *J. Eur. Math. Soc. (JEMS)* **24** (2022), 583–621.

[18] D. Conlon and J. Fox, Bounds for graph regularity and removal lemmas, *Geom. Funct. Anal.* **22** (2012), 1191–1256.

[19] D. Conlon, J. Fox, and B. Sudakov, Hereditary quasirandomness without regularity, *Math. Proc. Cambridge Philos. Soc.* **164** (2018), 385–399.

[20] D. Conlon, J. Fox, and Y. Wigderson, Ramsey numbers of books and quasirandomness, *Combinatorica* **42** (2022), 309–363.

[21] D. Conlon, H. Hàn, Y. Person, and M. Schacht, Weak quasi-randomness for uniform hypergraphs, *Random Structures Algorithms* **40** (2012), 1–38.

[22] R. A. Duke, H. Lefmann, and V. Rödl, A fast approximation algorithm for computing the frequencies of subgraphs in a given graph, *SIAM J. Comput.* **24** (1995), 598–620.

[23] J. Fox and R. Li, On edge-ordered Ramsey numbers, *Random Structures Algorithms* **57** (2020), 1174–1204.

[24] J. Fox and L. M. Lovász, A tight lower bound for Szemerédi's regularity lemma, *Combinatorica* **37** (2017), 911–951.

[25] J. Fox, J. Pach, A. Sheffer, A. Suk, and J. Zahl, A semi-algebraic version of Zarankiewicz's problem, *J. Eur. Math. Soc. (JEMS)* **19** (2017), 1785–1810.

[26] J. Fox, J. Pach, and A. Suk, A polynomial regularity lemma for semialgebraic hypergraphs and its applications in geometry and property testing, *SIAM J. Comput.* **45** (2016), 2199–2223.

[27] J. Fox, J. Pach, and A. Suk, Erdős-Hajnal conjecture for graphs with bounded VC-dimension, *Discrete Comput. Geom.* **61** (2019), 809–829.

[28] J. Fox, J. Pach, and A. Suk, Bounded VC-dimension implies the Schur-Erdős conjecture, *Combinatorica* **41** (2021), 803–813.

[29] J. Fox and B. Sudakov, Induced Ramsey-type theorems, *Adv. Math.* **219** (2008), 1771–1800.

[30] P. Frankl and V. Rödl, Extremal problems on set systems, *Random Structures Algorithms* **20** (2002), 131–164.

[31] A. Frieze and R. Kannan, Quick approximation to matrices and applications, *Combinatorica* **19** (1999), 175–220.

[32] L. Gishboliner, A. Shapira, and Y. Wigderson, Is it easy to regularize a hypergraph with easy links?, 2025. Preprint available at arXiv:2506.15582.

[33] W. T. Gowers, Lower bounds of tower type for Szemerédi's uniformity lemma, *Geom. Funct. Anal.* **7** (1997), 322–337.

[34] W. T. Gowers, Quasirandomness, counting and regularity for 3-uniform hypergraphs, *Combin. Probab. Comput.* **15** (2006), 143–184.

[35] W. T. Gowers, Hypergraph regularity and the multidimensional Szemerédi theorem, *Ann. of Math. (2)* **166** (2007), 897–946.

[36] B. Green, A Szemerédi-type regularity lemma in abelian groups, with applications, *Geom. Funct. Anal.* **15** (2005), 340–376.

[37] D. Haussler, Sphere packing numbers for subsets of the Boolean n-cube with bounded Vapnik-Chervonenkis dimension, *J. Combin. Theory Ser. A* **69** (1995), 217–232.

[38] N. Hempel, On n-dependent groups and fields, *MLQ Math. Log. Q.* **62** (2016), 215–224.

[39] E. Hrushovski, Y. Peterzil, and A. Pillay, Groups, measures, and the NIP, *J. Amer. Math. Soc.* **21** (2008), 563–596.

[40] X. Huang, H. Liu, M. Rong, and Z. Xu, Interpolating chromatic and homomorphism thresholds, 2025. Preprint available at arXiv:2502.09576.

[41] O. Janzer and C. Pohoata, On the Zarankiewicz problem for graphs with bounded VC-dimension, *Combinatorica* **44** (2024), 839–848.

[42] Y. Kohayakawa, B. Nagle, V. Rödl, and M. Schacht, Weak hypergraph regularity and linear hypergraphs, *J. Combin. Theory Ser. B* **100** (2010), 151–160.

[43] L. Lovász and B. Szegedy, Regularity partitions and the topology of graphons, in *An irregular mind*, *Bolyai Soc. Math. Stud.*, vol. 21, János Bolyai Math. Soc., Budapest, 2010, 415–446.

[44] T. Łuczak and S. Thomassé, Coloring dense graphs via VC-dimension, 2010. Preprint available at arXiv:1007.1670.

[45] M. Malliaris and S. Shelah, Regularity lemmas for stable graphs, *Trans. Amer. Math. Soc.* **366** (2014), 1551–1585.

[46] J. Matoušek, *Lectures on discrete geometry*, *Graduate Texts in Mathematics*, vol. 212, Springer-Verlag, New York, 2002.

[47] G. Moshkovitz and A. Shapira, A tight bound for hypergraph regularity, *Geom. Funct. Anal.* **29** (2019), 1531–1578.

[48] B. Nagle, V. Rödl, and M. Schacht, Equivalent regular partitions of three-uniform hypergraphs, *Random Structures Algorithms* **65** (2024), 703–718.

[49] V. Rödl, On universality of graphs with uniformly distributed edges, *Discrete Math.* **59** (1986), 125–134.

[50] V. Rödl, B. Nagle, J. Skokan, M. Schacht, and Y. Kohayakawa, The hypergraph regularity method and its applications, *Proc. Natl. Acad. Sci. USA* **102** (2005), 8109–8113.

[51] V. Rödl and M. Schacht, Regularity lemmas for graphs, in *Fete of combinatorics and computer science*, *Bolyai Soc. Math. Stud.*, vol. 20, János Bolyai Math. Soc., Budapest, 2010, 287–325.

[52] N. Sauer, On the density of families of sets, *J. Combin. Theory Ser. A* **13** (1972), 145–147.

[53] S. Shelah, A combinatorial problem; stability and order for models and theories in infinitary languages, *Pacific J. Math.* **41** (1972), 247–261.

[54] S. Shelah, Strongly dependent theories, *Israel J. Math.* **204** (2014), 1–83.

[55] S. Shelah, Definable groups for dependent and 2-dependent theories, *Sarajevo J. Math.* **13** (2017), 3–25.

[56] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, *Acta Arith.* **27** (1975), 199–245.

[57] E. Szemerédi, Regular partitions of graphs, in *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*, Colloq. Internat. CNRS, vol. 260, CNRS, Paris, 1978, 399–401.

[58] C. Terry, VC_ℓ-dimension and the jump to the fastest speed of a hereditary $\mathcal{L}$-property, *Proc. Amer. Math. Soc.* **146** (2018), 3111–3126.

[59] C. Terry, An improved bound for regular decompositions of 3-uniform hypergraphs of bounded VC$_2$-dimension, *Model Theory* **2** (2023), 325–356.

[60] C. Terry, Growth of regular partitions 1: Improved bounds for small slicewise VC-dimension, 2024. Preprint available at arXiv:2404.01274.

[61] C. Terry, Growth of regular partitions 2: Weak regularity, 2024. Preprint available at arXiv:2404.01293.

[62] C. Terry, Growth of regular partitions 3: Strong regularity and the vertex partition, 2024. Preprint available at arXiv:2404.02024.

[63] C. Terry, Growth of regular partitions 4: Strong regularity and the pairs partition, 2024. Preprint available at arXiv:2404.02030.

[64] C. Terry and J. Wolf, Stable arithmetic regularity in the finite field model, *Bull. Lond. Math. Soc.* **51** (2019), 70–88.

[65] C. Terry and J. Wolf, Irregular triads in 3-uniform hypergraphs, *Mem. Amer. Math. Soc.* (2025), to appear. Preprint available at arXiv:2111.01737.

[66] L. G. Valiant, A theory of the learnable, *Comm. ACM* **27** (1984), 1134–1142.

[67] V. N. Vapnik and A. J. Červonenkis, The uniform convergence of frequencies of the appearance of events to their probabilities, *Teor. Verojatnost. i Primenen.* **16** (1971), 264–279.

[68] Y. Wigderson, Ramsey theory—lecture notes, 2024. Available online at `https://n.ethz.ch/~ywigderson/math/static/RamseyTheory2024LectureNotes.pdf`.

[69] Y. Zhao, *Graph theory and additive combinatorics—exploring structure and randomness*, Cambridge University Press, Cambridge, 2023.

A Proofs of technical lemmas

A.1 Proof of Lemma 3.2

Proof of Lemma 3.2. For Lemma 3.2(a), we first note that the definition (2) of $q(\mathcal{P}_E(G))$ is a sum of non-negative numbers, hence is $q(\mathcal{P}_E(G))$ is certainly non-negative. For the upper bound, we have that $d(\mathcal{H} \mid Z_{ij} \cup Z_{ik} \cup Z_{jk}) \leqslant 1$ for all Z_{ij}, Z_{ik}, Z_{jk}, and therefore

$$
\begin{aligned}
q(\mathcal{P}_E(G)) &= \sum_{\substack{Z_{ij} \in \mathcal{P}_E(G; Y_i \times Y_j) \\ Z_{ik} \in \mathcal{P}_E(G; Y_i \times Y_k) \\ Z_{jk} \in \mathcal{P}_E(G; Y_j \times Y_k)}} \frac{|\Delta(Z_{ij} \cup Z_{ik} \cup Z_{jk})|}{|\Delta(G)|} d(\mathcal{H} \mid Z_{ij} \cup Z_{ik} \cup Z_{jk})^2 \\
&\leqslant \sum_{\substack{Z_{ij} \in \mathcal{P}_E(G; Y_i \times Y_j) \\ Z_{ik} \in \mathcal{P}_E(G; Y_i \times Y_k) \\ Z_{jk} \in \mathcal{P}_E(G; Y_j \times Y_k)}} \frac{|\Delta(Z_{ij} \cup Z_{ik} \cup Z_{jk})|}{|\Delta(G)|} \\
&= 1,
\end{aligned}
$$

as every triangle in G appears in exactly one tripartite subgraph $Z_{ij} \cup Z_{ik} \cup Z_{jk}$. This proves Lemma 3.2(a)

For Lemma 3.2(c), we argue similarly. First, the lower bound $q(\mathcal{P}) \geqslant 0$ is again immediate as $q(\mathcal{P})$ is defined as a sum of non-negative numbers. For the upper bound, we have

$$
\begin{aligned}
q(\mathcal{P}) &= \sum_{Y \in \mathcal{P}_V} \left(\prod_{i=1}^{t} \frac{|Y_i|}{|X_i|} \right) \sum_{1 \leqslant i < j < k \leqslant t} q(\mathcal{P}_E(K(Y_i, Y_j, Y_k))) \\
&\leqslant \sum_{Y \in \mathcal{P}_V} \left(\prod_{i=1}^{t} \frac{|Y_i|}{|X_i|} \right) \sum_{1 \leqslant i < j < k \leqslant t} 1 \\
&= \binom{t}{3} \sum_{Y \in \mathcal{P}_V} \left(\prod_{i=1}^{t} \frac{|Y_i|}{|X_i|} \right) \\
&= \binom{t}{3},
\end{aligned}
$$

where the first inequality uses Lemma 3.2(a) applied to the complete tripartite graph $K(Y_i, Y_j, Y_k)$, and the final equality uses the fact that $\mathcal{P}_V$ is a vertex cylinder partition of $X_1 \times \cdots \times X_t$.

It remains to prove Lemmas 3.2(b) and 3.2(d), namely the monotonicity of the mean-squared density with respect to refinements. Both of these are consequences of the Cauchy–Schwarz inequality, as follows.

Let $(G, \mathcal{H})$ be a tripartite chain with vertex set $Y_1 \cup Y_2 \cup Y_3$, and let $\mathcal{P}_E(G), \mathcal{P}'_E(G)$ be two edge partitions of it with $\mathcal{P}'_E(G)$ refining $\mathcal{P}_E(G)$. Fix some $Z_{12} \in \mathcal{P}_E(G; Y_1 \times Y_2), Z_{13} \in \mathcal{P}_E(G; Y_1 \times Y_3), Z_{23} \in \mathcal{P}_E(G; Y_2 \times Y_3)$. As $\mathcal{P}'_E(G)$ is a refinement of $\mathcal{P}_E(G)$, in $\mathcal{P}'_E(G)$ each of these bipartite graphs is further partitioned, say into $Z^1_{12}, \ldots, Z^m_{12}, Z^1_{13}, \ldots, Z^m_{13}, Z^1_{23}, \ldots, Z^m_{23}$; we may assume that the number of parts in all three cases equals m, by potentially letting some of these bipartite graphs

be empty. We now have that

$$d(\mathcal{H} \mid Z_{12} \cup Z_{13} \cup Z_{23}) = \frac{|E(\mathcal{H}) \cap \Delta(Z_{12} \cup Z_{13} \cup Z_{23})|}{|\Delta(Z_{12} \cup Z_{13} \cup Z_{23})|}$$

$$= \sum_{a,b,c=1}^{m} \frac{|E(\mathcal{H} \cap \Delta(Z_{12}^{a} \cup Z_{13}^{b} \cup Z_{23}^{c}))|}{|\Delta(Z_{12} \cup Z_{13} \cup Z_{23})|}$$

$$= \sum_{a,b,c=1}^{m} \frac{|\Delta(Z_{12}^{a} \cup Z_{13}^{b} \cup Z_{23}^{c})|}{|\Delta(Z_{12} \cup Z_{13} \cup Z_{23})|} d(\mathcal{H} \mid Z_{12}^{a} \cup Z_{13}^{b} \cup Z_{23}^{c}),$$

where the final step uses the definition of the relative density $d(\mathcal{H} \mid Z_{12}^{a} \cup Z_{13}^{b} \cup Z_{23}^{c})$. Therefore, by Cauchy–Schwarz,

$$d(\mathcal{H} \mid Z_{12} \cup Z_{13} \cup Z_{23})^2 = \left(\sum_{a,b,c=1}^{m} \frac{|\Delta(Z_{12}^{a} \cup Z_{13}^{b} \cup Z_{23}^{c})|}{|\Delta(Z_{12} \cup Z_{13} \cup Z_{23})|} d(\mathcal{H} \mid Z_{12}^{a} \cup Z_{13}^{b} \cup Z_{23}^{c}) \right)^2$$

$$\leqslant \left(\sum_{a,b,c=1}^{m} \frac{|\Delta(Z_{12}^{a} \cup Z_{13}^{b} \cup Z_{23}^{c})|}{|\Delta(Z_{12} \cup Z_{13} \cup Z_{23})|} \right) \left(\sum_{a,b,c=1}^{m} \frac{|\Delta(Z_{12}^{a} \cup Z_{13}^{b} \cup Z_{23}^{c})|}{|\Delta(Z_{12} \cup Z_{13} \cup Z_{23})|} d(\mathcal{H} \mid Z_{12}^{a} \cup Z_{13}^{b} \cup Z_{23}^{c})^2 \right)$$

$$= \sum_{a,b,c=1}^{m} \frac{|\Delta(Z_{12}^{a} \cup Z_{13}^{b} \cup Z_{23}^{c})|}{|\Delta(Z_{12} \cup Z_{13} \cup Z_{23})|} d(\mathcal{H} \mid Z_{12}^{a} \cup Z_{13}^{b} \cup Z_{23}^{c})^2,$$

where in the final step we use that every triangle in $\Delta(Z_{12} \cup Z_{13} \cup Z_{23})$ appears in exactly one subgraph $Z_{12}^{a} \cup Z_{13}^{b} \cup Z_{23}^{c}$. Plugging this inequality into the definition (2), we find that

$$q(\mathcal{P}_E(G)) = \sum_{\substack{Z_{12} \in \mathcal{P}_E(G;Y_1 \times Y_2) \\ Z_{13} \in \mathcal{P}_E(G;Y_1 \times Y_3) \\ Z_{23} \in \mathcal{P}_E(G;Y_2 \times Y_3)}} \frac{|\Delta(Z_{12} \cup Z_{13} \cup Z_{23})|}{|\Delta(G)|} d(\mathcal{H} \mid Z_{12} \cup Z_{13} \cup Z_{23})^2$$

$$\leqslant \sum_{\substack{Z_{12} \in \mathcal{P}_E(G;Y_1 \times Y_2) \\ Z_{13} \in \mathcal{P}_E(G;Y_1 \times Y_3) \\ Z_{23} \in \mathcal{P}_E(G;Y_2 \times Y_3)}} \frac{|\Delta(Z_{12} \cup Z_{13} \cup Z_{23})|}{|\Delta(G)|} \sum_{a,b,c=1}^{m} \frac{|\Delta(Z_{12}^{a} \cup Z_{13}^{b} \cup Z_{23}^{c})|}{|\Delta(Z_{12} \cup Z_{13} \cup Z_{23})|} d(\mathcal{H} \mid Z_{12}^{a} \cup Z_{13}^{b} \cup Z_{23}^{c})^2$$

$$= \sum_{\substack{Z_{12} \in \mathcal{P}_E(G;Y_1 \times Y_2) \\ Z_{13} \in \mathcal{P}_E(G;Y_1 \times Y_3) \\ Z_{23} \in \mathcal{P}_E(G;Y_2 \times Y_3)}} \sum_{a,b,c=1}^{m} \frac{|\Delta(Z_{12}^{a} \cup Z_{13}^{b} \cup Z_{23}^{c})|}{|\Delta(G)|} d(\mathcal{H} \mid Z_{12}^{a} \cup Z_{13}^{b} \cup Z_{23}^{c})^2$$

$$= \sum_{\substack{Z'_{12} \in \mathcal{P}'_E(G;Y_1 \times Y_2) \\ Z'_{13} \in \mathcal{P}'_E(G;Y_1 \times Y_3) \\ Z'_{23} \in \mathcal{P}'_E(G;Y_2 \times Y_3)}} \frac{|\Delta(Z'_{12} \cup Z'_{13} \cup Z'_{23})|}{|\Delta(G)|} d(\mathcal{H} \mid Z'_{12} \cup Z'_{13} \cup Z'_{23})^2$$

$$= q(\mathcal{P}'_E(G)),$$

where in the penultimate equality we simply note that summing over all parts in $\mathcal{P}_E(G)$, and then over all their subparts, is the same as summing over all parts in $\mathcal{P}'_E(G)$. This proves Lemma 3.2(b).

Note that in the special case that G is complete tripartite, the result just proved shows that $q(\mathcal{P}'_E(K(Y_i, Y_j, Y_k))) \geqslant q(\mathcal{P}_E(K(Y_i, Y_j, Y_k)))$ whenever $\mathcal{P}'_E(K(Y_i, Y_j, Y_k))$ is an edge partition of a vertex cylinder $K(Y_i, Y_j, Y_k)$ which refines another edge partition $\mathcal{P}_E(K(Y_i, Y_j, Y_k))$.

We now turn to Lemma 3.2(d), so we fix a t-partite 3-graph $\mathcal{H}$ as well as two cylinder chain partitions $\mathcal{P}, \mathcal{P}'$ of it, with $\mathcal{P}'$ refining $\mathcal{P}$. Lt $\widehat{\mathcal{P}}$ be the cylinder chain partition whose vertex partition is $\mathcal{P}'_V$, but whose edge partitions are simply obtained by restricting the original edge partition $\mathcal{P}_E$ to the new vertex cylinders. That is, for all $Y' \in \mathcal{P}'_V$, the edge partition $\widehat{\mathcal{P}}_E(Y'_i \times Y'_j)$ is $\mathcal{P}_E(Y_i \times Y_j)|_{Y'_i \times Y'_j}$, where Y is the unique cylinder in $\mathcal{P}_V$ containing Y'. Note that $\mathcal{P}'$ is obtained from $\widehat{\mathcal{P}}$ by only refining the edge partition. As a consequence, we find from the above that

$$
\begin{aligned}
q(\mathcal{P}') &= \sum_{Y \in \mathcal{P}'_V} \left(\prod_{\ell=1}^{t} \frac{|Y_\ell|}{|X_\ell|} \right) \sum_{1 \leqslant i < j < k \leqslant t} q(\mathcal{P}'_E(K(Y_i, Y_j, Y_k))) \\
&\geqslant \sum_{Y \in \mathcal{P}'_V} \left(\prod_{\ell=1}^{t} \frac{|Y_\ell|}{|X_\ell|} \right) \sum_{1 \leqslant i < j < k \leqslant t} q(\widehat{\mathcal{P}}_E(K(Y_i, Y_j, Y_k))) \\
&= q(\widehat{\mathcal{P}})
\end{aligned}
$$

Therefore, it suffices to prove that $q(\widehat{\mathcal{P}}) \geqslant q(\mathcal{P})$. For this, fix some cylinder $Y \in \mathcal{P}_V$. As $\mathcal{P}'_V$ refines $\mathcal{P}_V$, there are cylinders $Y^1, \ldots, Y^m \in \mathcal{P}'_V$ which partition Y. For $1 \leqslant i < j \leqslant t$ and $Z_{ij} \in \mathcal{P}_E(Y_i \times Y_j)$, let $Z^r_{ij} \in \widehat{\mathcal{P}}_E(Y^r_i \times Y^r_j)$ be the restriction of Z_{ij} to the vertex subsets $Y^r_i \cup Y^r_j$.

For the moment, we focus on $Y_1 \cup Y_2 \cup Y_3$. Note that for any $Z_{12} \in \mathcal{P}_E(Y_1 \times Y_2), Z_{13} \in \mathcal{P}_E(Y_1 \times Y_3), Z_{23} \in \mathcal{P}_E(Y_2 \times Y_3)$, we have

$$
|\Delta(Z_{12} \cup Z_{13} \cup Z_{23})| \cdot \prod_{\ell=4}^{t} |Y_\ell| = \sum_{r=1}^{m} |\Delta(Z^r_{12} \cup Z^r_{13} \cup Z^r_{23})| \cdot \prod_{\ell=4}^{t} |Y^r_\ell|,
$$

as both sides count the number of t-tuples in $Y_1 \times \cdots \times Y_t$ in which the first three vertices span a triangle in $Z_{12} \cup Z_{13} \cup Z_{23}$. Similarly,

$$
|E(\mathcal{H}) \cap \Delta(Z_{12} \cup Z_{13} \cup Z_{23})| \cdot \prod_{\ell=4}^{t} |Y_\ell| = \sum_{r=1}^{m} |E(\mathcal{H}) \cap \Delta(Z^r_{12} \cup Z^r_{13} \cup Z^r_{23})| \cdot \prod_{\ell=4}^{t} |Y^r_\ell|,
$$

as both sides count the number of t-tuples in $Y_1 \times \cdots \times Y_t$ in which the first three vertices span a triangle in $Z_{12} \cup Z_{13} \cup Z_{23}$ and form an edge of $\mathcal{H}$. Therefore, we have that

$$
\begin{aligned}
d(\mathcal{H} \mid Z_{12} \cup Z_{13} \cup Z_{23}) &= \frac{|E(\mathcal{H}) \cap \Delta(Z_{12} \cup Z_{13} \cup Z_{23})|}{|\Delta(Z_{12} \cup Z_{13} \cup Z_{23})|} \\
&= \frac{|E(\mathcal{H}) \cap \Delta(Z_{12} \cup Z_{13} \cup Z_{23})| \cdot \prod_{\ell=4}^{t} |Y_\ell|}{|\Delta(Z_{12} \cup Z_{13} \cup Z_{23})| \cdot \prod_{\ell=4}^{t} |Y_\ell|} \\
&= \sum_{r=1}^{m} \frac{|E(\mathcal{H}) \cap \Delta(Z^r_{12} \cup Z^r_{13} \cup Z^r_{23})| \cdot \prod_{\ell=4}^{t} |Y^r_\ell|}{|\Delta(Z_{12} \cup Z_{13} \cup Z_{23})| \cdot \prod_{\ell=4}^{t} |Y_\ell|} \\
&= \sum_{r=1}^{m} \frac{|\Delta(Z^r_{12} \cup Z^r_{13} \cup Z^r_{23})| \cdot \prod_{\ell=4}^{t} |Y^r_\ell|}{|\Delta(Z_{12} \cup Z_{13} \cup Z_{23})| \cdot \prod_{\ell=4}^{t} |Y_\ell|} d(\mathcal{H} \mid Z^r_{12} \cup Z^r_{13} \cup Z^r_{23}).
\end{aligned}
$$

Again by Cauchy–Schwarz, we conclude that

$$d(\mathcal{H} \mid Z_{12} \cup Z_{13} \cup Z_{23})^2 \leqslant \sum_{r=1}^{m} \frac{|\Delta(Z_{12}^r \cup Z_{13}^r \cup Z_{23}^r)| \cdot \prod_{\ell=4}^{t} |Y_\ell^r|}{|\Delta(Z_{12} \cup Z_{13} \cup Z_{23})| \cdot \prod_{\ell=4}^{t} |Y_\ell|} d(\mathcal{H} \mid Z_{12}^r \cup Z_{13}^r \cup Z_{23}^r)^2.$$

Plugging this in to the definition of $q(\mathcal{P}_E(K(Y_1, Y_2, Y_3)))$, we find that

$$q(\mathcal{P}_E(K(Y_1, Y_2, Y_3))) = \sum_{\substack{Z_{12} \in \mathcal{P}_E(Y_1 \times Y_2) \\ Z_{13} \in \mathcal{P}_E(Y_1 \times Y_3) \\ Z_{23} \in \mathcal{P}_E(Y_2 \times Y_3)}} \frac{|\Delta(Z_{12} \cup Z_{13} \cup Z_{23})|}{|Y_1||Y_2||Y_3|} d(\mathcal{H} \mid Z_{12} \cup Z_{13} \cup Z_{23})^2$$

$$\leqslant \sum_{\substack{Z_{12} \in \mathcal{P}_E(Y_1 \times Y_2) \\ Z_{13} \in \mathcal{P}_E(Y_1 \times Y_3) \\ Z_{23} \in \mathcal{P}_E(Y_2 \times Y_3)}} \sum_{r=1}^{m} \frac{|\Delta(Z_{12}^r \cup Z_{13}^r \cup Z_{23}^r)| \cdot \prod_{\ell=4}^{t} |Y_\ell^r|}{\prod_{\ell=1}^{t} |Y_\ell|} d(\mathcal{H} \mid Z_{12}^r \cup Z_{13}^r \cup Z_{23}^r)^2$$

$$= \sum_{\substack{Z_{12} \in \mathcal{P}_E(Y_1 \times Y_2) \\ Z_{13} \in \mathcal{P}_E(Y_1 \times Y_3) \\ Z_{23} \in \mathcal{P}_E(Y_2 \times Y_3)}} \sum_{r=1}^{m} \left(\prod_{\ell=1}^{t} \frac{|Y_\ell^r|}{|Y_\ell|} \right) \frac{|\Delta(Z_{12}^r \cup Z_{13}^r \cup Z_{23}^r)|}{|Y_1^r||Y_2^r||Y_3^r|} d(\mathcal{H} \mid Z_{12}^r \cup Z_{13}^r \cup Z_{23}^r)^2$$

$$= \sum_{r=1}^{m} \left(\prod_{\ell=1}^{t} \frac{|Y_\ell^r|}{|Y_\ell|} \right) \sum_{\substack{\widehat{Z_{12}} \in \widehat{\mathcal{P}_E}(Y_1^r \times Y_2^r) \\ \widehat{Z_{13}} \in \widehat{\mathcal{P}_E}(Y_1^r \times Y_3^r) \\ \widehat{Z_{23}} \in \widehat{\mathcal{P}_E}(Y_2^r \times Y_3^r)}} \frac{|\Delta(\widehat{Z_{12}} \cup \widehat{Z_{13}} \cup \widehat{Z_{23}})|}{|Y_1^r||Y_2^r||Y_3^r|} d(\mathcal{H} \mid \widehat{Z_{12}} \cup \widehat{Z_{13}} \cup \widehat{Z_{23}})^2$$

$$= \sum_{r=1}^{m} \left(\prod_{\ell=1}^{t} \frac{|Y_\ell^r|}{|Y_\ell|} \right) q(\widehat{\mathcal{P}_E}(K(Y_1^r, Y_2^r, Y_3^r))),$$

where in the penultimate equality we use the definition of $\widehat{\mathcal{P}_E}$, namely that the set of edge parts appearing in it is precisely the set of restrictions of edge parts of $\mathcal{P}_E$.

By the same argument, the above inequality holds for all indices $1 \leqslant i < j < k \leqslant t$. Summing up over all choices for i, j, k, we conclude that for any $Y \in \mathcal{P}_V$, we have

$$q(\mathcal{P}_E(Y)) = \sum_{1 \leqslant i < j < k \leqslant t} q(\mathcal{P}_E(K(Y_i, Y_j, Y_k))) \leqslant \sum_{1 \leqslant i < j < k \leqslant t} \sum_{r=1}^{m} \left(\prod_{\ell=1}^{t} \frac{|Y_\ell^r|}{|Y_\ell|} \right) q(\widehat{\mathcal{P}_E}(K(Y_i^r, Y_j^r, Y_k^r))).$$

Therefore, by the definition of $q(\mathcal{P})$, we have

$$
\begin{aligned}
q(\mathcal{P}) &= \sum_{Y \in \mathcal{P}_V} \left(\prod_{\ell=1}^{t} \frac{|Y_\ell|}{|X_\ell|} \right) q(\mathcal{P}_E(Y)) \\
&\leqslant \sum_{Y \in \mathcal{P}_V} \left(\prod_{\ell=1}^{t} \frac{|Y_\ell|}{|X_\ell|} \right) \sum_{1 \leqslant i < j < k \leqslant t} \sum_{r=1}^{m} \left(\prod_{\ell=1}^{t} \frac{|Y_\ell^r|}{|Y_\ell|} \right) q(\widehat{\mathcal{P}}_E(K(Y_i^r, Y_j^r, Y_k^r))) \\
&= \sum_{Y \in \mathcal{P}_V} \sum_{r=1}^{m} \left(\prod_{\ell=1}^{t} \frac{|Y_\ell^r|}{|X_\ell|} \right) \sum_{1 \leqslant i < j < k \leqslant t} q(\widehat{\mathcal{P}}_E(K(Y_i^r, Y_j^r, Y_k^r))) \\
&= \sum_{Y' \in \mathcal{P}_V'} \left(\prod_{\ell=1}^{t} \frac{|Y_\ell'|}{|X_\ell|} \right) \sum_{1 \leqslant i < j < k \leqslant t} q(\widehat{\mathcal{P}}_E(K(Y_i', Y_j', Y_k'))) \\
&= q(\widehat{\mathcal{P}}).
\end{aligned}
$$

As we have already proved that $q(\widehat{\mathcal{P}}) \leqslant q(\mathcal{P}')$, this concludes the proof of Lemma 3.2(d). □

A.2   Proof of Lemma 4.1

We begin by recalling the statement of [35, Corollary 5.3], specialized to the case of counting cliques in 3-uniform hypergraphs.

Theorem A.1 (Special case of [35, Corollary 5.3])**.** *Let Z be a t-partite graph on parts $X_1, \ldots, X_t$, and let $\mathcal{H}$ be an t-partite 3-graph supported on $\Delta(Z)$.*
Let $\varepsilon, \gamma \in (0, \frac{1}{100}]$, and define $\eta = \varepsilon^8 \gamma^{10t^3}$ and $\psi(x) = \varepsilon^{10} \gamma^{10t^3} x^{16t^2}$. Suppose that

$$
d(\mathcal{H}[X; Z; \{i, j, k\}]) \geqslant \gamma
$$

and that $\mathcal{H}[X; Z; \{i, j, k\}]$ is (η, ψ)-quasirandom for all $\{i, j, k\} \in \binom{[t]}{3}$.
If $x_1, \ldots, x_t \in X_1 \times \cdots \times X_t$ are picked uniformly at random, then the probability that they span a $K_t^{(3)}$ in $\mathcal{H}$—that is, the probability that $x_i x_j x_k \in E(\mathcal{H})$ for all $\{i, j, k\} \in \binom{[t]}{3}$—is at least

$$
(1 - \varepsilon) \left(\prod_{1 \leqslant i < j \leqslant t} d(Z_{ij}) \right) \left(\prod_{1 \leqslant i < j < k \leqslant t} d(\mathcal{H}[X; Z; \{i, j, k\}]) \right).
$$

As promised, Lemma 4.1 follows as a straightforward corollary of Theorem A.1.

Proof of Lemma 4.1. Let the three vertex parts of $\mathcal{V}$ be V_1, V_2, V_3, and let the vertices of $\mathcal{V}$ be $v_1, \ldots, v_t$. Let $\iota : \{v_1, \ldots, v_t\} \to \{1, 2, 3\}$ be the "index" function that associates to each v_i the unique index $\iota(i)$ such that $v_i \in V_{\iota(i)}$.

We construct a t-partite graph Z and a t-partite 3-graph $\mathcal{H}$ as follows. For every v_i, we construct a set X_i, which is a copy of $U_{\iota(i)}$. Thus, the parts $X_1, \ldots, X_t$ consist of $|V_1|$ copies of U_1, $|V_2|$ copies of U_2, and $|V_3|$ copies of U_3.

Next, we define the $\binom{t}{2}$ bipartite graphs making up Z as follows. If $\iota(v_i) \neq \iota(v_j)$, then we define Z_{ij} to be a copy of $G[U_{\iota(i)} \times U_{\iota(j)}]$ between X_i and X_j. If $\iota(i) = \iota(j)$, we define Z_{ij} to be the

complete bipartite graph between X_i and X_j. Note that in this way, we ensure that each Z_{ij} is $\psi(\delta)$-quasirandom, and that each tripartite subgraph satisfies $\delta(Z_{ij} \cup Z_{ik} \cup Z_{jk}) \geqslant \delta$.

Finally, we define each of the tripartite 3-graphs making up $\mathcal{H}$. For indices $i, j, k \in \binom{[t]}{3}$, in case $|\iota(i), \iota(j), \iota(k)| \leqslant 2$, we include in $\mathcal{H}$ all triangles in the tripartite graph $Z_{ij} \cup Z_{ik} \cup Z_{jk}$. In the remaining case, where $\iota(i), \iota(j), \iota(k)$ are all distinct, we do the following. If $\{v_i, v_j, v_k\} \in E(\mathcal{V})$, we put a copy of $\mathcal{H}_0$ between X_i, X_j, X_k. However, if $\{v_i, v_j, v_k\} \notin E(\mathcal{V})$, we put between these three parts the *relative complement* of $\mathcal{H}_0$; that is, we include as a hyperedge every triangle in $Z_{ij} \cup Z_{ik} \cup Z_{jk}$ which is not a hyperedge of $\mathcal{H}_0$.

We now claim that $\mathcal{H}[X; Z; \{i, j, k\}]$ is (η, ψ)-quasirandom with $d(\mathcal{H}[X; Z; \{i, j, k\}]) \geqslant \gamma$ for all i, j, k. Indeed, for triples i, j, k with $|\iota(i), \iota(j), \iota(k)| \leqslant 2$, this is immediate since all triangles in $Z_{ij} \cup Z_{ik} \cup Z_{jk}$ are present. For the remaining triples, this follows from the fact that we assumed $\gamma \leqslant d(\mathcal{H}_0[U; G]) \leqslant 1 - \gamma$, and that a tripartite hypergraph is η-quasirandom with respect to a tripartite graph if and only if its relative complement is η-quasirandom with respect to the same tripartite graph.

Now suppose that we pick $(x_1, \ldots, x_t) \in X_1 \times \cdots \times X_t$ uniformly at random. By Theorem A.1, applied with $\varepsilon = \frac{1}{100}$, the probability that they form a $K_t^{(3)}$ in $\mathcal{H}$ is at least

$$(1 - \varepsilon) \left(\prod_{1 \leqslant i < j \leqslant t} d(Z_{ij}) \right) \left(\prod_{1 \leqslant i < j < k \leqslant t} d(\mathcal{H}[X; Z; \{i, j, k\}]) \right) \geqslant \frac{1}{2} \delta^{\binom{t}{2}} \gamma^{\binom{t}{3}}.$$

Additionally, for a fixed i, j, the probability that x_i and x_j correspond to the same vertex of $V(\mathcal{H}_0)$ is at most $1/\min\{|U_1|, |U_2|, |U_3|\}$, which is at most $\delta^{t^2} \gamma^{t^3}$ by assumption. Hence, the probability that there is any collision among $\{x_1, \ldots, x_t\}$ is at most $t^2 \delta^{t^2} \gamma^{t^3}$. Therefore, the probability that $x_1, \ldots, x_t$ correspond to distinct vertices in $\mathcal{H}_0$ and span a $K_t^{(3)}$ in $\mathcal{H}$ is at least

$$\frac{1}{2} \delta^{\binom{t}{2}} \gamma^{\binom{t}{3}} - t^2 \delta^{t^2} \gamma^{t^3} > 0.$$

We conclude that there is a choice of $(x_1, \ldots, x_t) \in X_1 \times \cdots \times X_t$ satisfying both of these properties.

The final observation is that, by our construction of $\mathcal{H}$, these vertices $x_1, \ldots, x_t$ correspond to an induced copy of $\mathcal{V}$ in $\mathcal{H}_0$. Indeed, since both $\mathcal{V}$ and $\mathcal{H}_0$ are tripartite, all we need to check is that if x_i, x_j, x_k lie in distinct parts of $\mathcal{H}_0$, then they form a hyperedge of $\mathcal{H}_0$ if and only if v_i, v_j, v_k form a hyperedge of $\mathcal{V}$. But this is immediate from our choice of $\mathcal{H}$, since we included in $\mathcal{H}$ all edges of $\mathcal{H}_0$ when $\{v_i, v_j, v_k\} \in E(\mathcal{V})$, and only non-edges of $\mathcal{H}_0$ when $\{v_i, v_j, v_k\} \notin E(\mathcal{V})$. This concludes the proof. □