# eScholarship
## Combinatorial Theory

**Title**

Resilience for tight Hamiltonicity

**Permalink**

https://escholarship.org/uc/item/2x43q0st

**Journal**

Combinatorial Theory, 4(1)

**ISSN**

2766-1334

**Authors**

Allen, Peter
Parczyk, Olaf
Pfenninger, Vincent

**Publication Date**

2024

**DOI**

10.5070/C64163846

**Supplemental Material**

https://escholarship.org/uc/item/2x43q0st#supplemental

**Copyright Information**

Peer reviewed

# RESILIENCE FOR TIGHT HAMILTONICITY

Peter Allen[*1], Olaf Parczyk[†2], and Vincent Pfenninger[‡3]

[1]*Department of Mathematics, London School of Economics, Houghton Street, London, WC2A 2AE, U.K.*
*p.d.allen@lse.ac.uk*
[2]*Department of Mathematics and Computer Science, FU Berlin, Arnimallee 3, 14195 Berlin, Germany*
*parczyk@mi.fu-berlin.de*
[3]*Institute of Discrete Mathematics, Graz University of Technology, Steyrergasse 30, 8010 Graz, Austria*
*pfenninger@math.tugraz.at*

**Abstract**. We prove that random hypergraphs are asymptotically almost surely resiliently Hamiltonian. Specifically, for any $\gamma > 0$ and $k \geqslant 3$, we show that asymptotically almost surely, every subgraph of the binomial random $k$-uniform hypergraph $G^{(k)}\big(n, n^{\gamma-1}\big)$ in which all $(k-1)$-sets are contained in at least $\big(\frac{1}{2} + 2\gamma\big)pn$ edges has a tight Hamilton cycle. This is a cyclic ordering of the $n$ vertices such that each consecutive $k$ vertices forms an edge.

**Keywords.** Random graphs, hypergraphs, tight Hamilton cycles, resilience

**Mathematics Subject Classifications.** 05C80, 05C35

## 1. Introduction

The study of Hamilton cycles in graphs is one of the oldest topics in graph theory. In extremal graph theory, Dirac [Dir52] in 1952 proved the sharp result that an $n$-vertex graph with minimum degree at least $\frac{n}{2}$ contains a Hamilton cycle. In random graph theory, Pósa [Pos76] and Korshunov [Kor76, Kor77] independently showed in the 1970s that Hamilton cycles first appear in the random graph $G(n, p)$ — that is, the $n$-vertex graph where edges are present independently with probability $p$ — at a threshold $p = \Theta\big(\frac{\log n}{n}\big)$. Komlós and Szemerédi [KS83] showed that the sharp threshold for Hamiltonicity coincides with that for minimum degree 2, and Bollobás [Bol84] strengthened this by showing a hitting time version: if edges are added one by one, the edge which causes minimum degree 2 will asymptotically almost surely[1] also cause Hamiltonicity.

---

[1]Asymptotically almost surely (a.a.s.) is with probability tending to 1 as $n$ tends to infinity.

Combining these areas, Sudakov and Vu [SV08] introduced the term *resilience* (though the same concept appears earlier in work of Alon, Capalbo, Kohayakawa, Rödl, Ruciński and Szemerédi [ACK⁺00]). They proved that for each $\gamma > 0$, the random graph $\Gamma = G(n,p)$ is a.a.s. $\left(\frac{1}{2} + \gamma\right)$-resiliently Hamiltonian whenever $p = \omega(n^{-1} \log^4 n)$; that is, every subgraph of $\Gamma$ with minimum degree at least $\left(\frac{1}{2} + \gamma\right)pn$ has a Hamilton cycle. This result is sharp in the minimum degree, for the same reason as Dirac's theorem, but the probability can be improved. This was done over a succession of papers: Lee and Sudakov [LS12] showed that $p$ can be reduced to the threshold $\Omega(n^{-1} \log n)$ and very recently Montgomery [Mon19] and, independently, Nenadov, Steger, and Trujić [NST19] showed the hitting time version of this result (for which one needs to be a little more careful with edge deletion: it is permitted to delete only a $\left(\frac{1}{2} - \gamma\right)$-fraction of the edges at any given vertex).

Hamilton cycles in hypergraphs have only much more recently been attacked. There are several natural notions of paths and cycles in hypergraphs: the one that will concern us here is that of tight paths and cycles in $k$-uniform hypergraphs. That is, we work with hypergraphs in which all edges have uniformity $k$. We say that a given linear ordering of some vertices is a *tight path* if each consecutive $k$-set of vertices forms an edge; a given cyclic ordering of some vertices with the same condition forms a tight cycle. The $k = 2$ case of this definition reduces to the usual paths and cycles in graphs. For brevity, in what follows we write *$k$-graph* for $k$-uniform hypergraph.

In terms of extremal results, there are again several reasonable questions — one should place some form of 'minimum degree' condition for tight Hamilton cycles, but this can take the form of insisting that every $j$-set of vertices is in sufficiently many edges, where one can choose $j$ between $1$ and $k-1$. This leads to several significantly different problems (and even more if one considers other notions of cycle). We refer the reader to the comprehensive survey of Kühn and Osthus [KO14] for details, and focus on the version of minimum degree we want to work with. This is the case $j = k-1$, sometimes called *codegree*. Here, the Hamiltonicity problem is resolved. Rödl, Ruciński and Szemerédi [RRS06, RRS08], first for 3-uniform and then for general uniformity, showed that if $n$ is sufficiently large, any $n$-vertex $k$-graph with minimum codegree at least $\left(\frac{1}{2} + \gamma\right)n$ (i.e. every $(k-1)$-set is in at least that many edges) contains a tight Hamilton cycle. For 3-graphs, they [RRS11] were also able to give the exact result for sufficiently large $n$ (finding exactly what should replace the error term $\gamma n$).

In random hypergraphs, Dudek and Frieze [DF11, DF13] found for several different notions of 'cycle' the threshold for Hamiltonicity in the binomial random hypergraph $G^{(k)}(n,p)$, that is the $n$-vertex $k$-graph in which $k$-sets are edges independently with probability $p$. In particular, in [DF13] they showed by the second moment method that for $k = 3$ the threshold is $\omega(n^{-1})$, and for $k \geqslant 4$ the sharp threshold is at $en^{-1}$. Narayanan and Schacht [NS20] strengthened these results, in particular showing that $en^{-1}$ is also the sharp threshold for $k = 3$.

Combining these (and answering a question of Frieze [Fri23]), we prove the following corresponding codegree resilience statement.

**Theorem 1.1.** *Given any $\gamma > 0$ and $k \geqslant 3$, if $p \geqslant n^{-1+\gamma}$, we show that $\Gamma = G^{(k)}(n,p)$ a.a.s. satisfies the following. Let $G$ be any $n$-vertex subgraph of $\Gamma$ such that $\delta_{k-1}(G) \geqslant \left(\frac{1}{2} + 2\gamma\right)pn$. Then $G$ contains a tight Hamilton cycle.*

Observe that this theorem is sharp in the minimum degree requirement, but it is presumably not sharp in the probability. More precisely, when $p = \Omega(\log n/n)$ then a.a.s. in $\Gamma$ there is an $n$-vertex subgraph $G$ such that $\delta_{k-1}(G) \geqslant (1/2 - \gamma)pn$ and $G$ does not contain a tight Hamilton cycle. When $p = o(\log n/n)$, there a.a.s. are $(k-1)$-tuples in $\Gamma$ that are not contained in any edges and, therefore, no $G$ as required by the theorem exists. For this regime the resilience condition needs to be adjusted, perhaps as explained above for the hitting time results in graphs from [Mon19, NST19]. We certainly need $p \geqslant 2en^{-1}$ for any statement of this kind to be true, otherwise randomly deleting half of the edges from $\Gamma$ would a.a.s. destroy the tight Hamiltonicity.

This is the first resilience statement for tight Hamilton cycles in sparse random hypergraphs to the best of our knowledge; however for Berge cycles, Clemens, Ehrenmüller and Person [CEP20] proved a resilience statement which is both tight in the minimum degree and has only a poly-logarithmic gap in the probability. For perfect matchings Ferber and Hirschfeld [FH20] showed that the same codegree resilience as in Theorem 1.1 holds with $p = \Omega(\log n/n)$, which is significantly above the threshold for the appearance of perfect matchings, but optimal for the same reasons as discussed above. More generally, Ferber and Kwan [FK22] studied the transference of results for perfect matchings in dense hypergraphs into resilience statements in random hypergraphs.

It would be interesting to investigate this transference for other types of Hamilton cycles and other degree conditions. For example, in the case of 3-graphs Reiher, Rödl, Ruciński, Schacht, and Szemerédi [RRR$^+$19] show that any $n$-vertex 3-graph with minimum vertex degree $(\frac{5}{9} + \gamma)\binom{n}{2}$ contains a tight Hamilton cycle. Can this be extended to a resilience statement in random 3-graphs? More precisely, can the condition $\delta_2(G) \geqslant (\frac{1}{2} + \gamma)pn$ in Theorem 1.1 for $k = 3$ be replaced by $\delta_1(G) \geqslant (\frac{5}{9} + \gamma)p\binom{n}{2}$? The bound on the minimum degree would again be sharp.

## 1.1. Ideas of the proof, and outline of the paper

Our proof strategy for Theorem 1.1 uses the *reservoir method*, which was previously used in a similar way in [ABKP15] and [AKPP20] to give polynomial-time algorithms that find tight Hamilton cycles in $\Gamma$ itself for broadly similar values of $p$. Very briefly, the reservoir method is as follows.

In a first step, we identify a *reservoir set* $R$, which contains a small (but bounded away from 0) fraction of the vertices of $G$. We construct a *reservoir path* $P_{\mathrm{res}}$, which is a tight path that contains all the vertices of $R$ and in addition for any subset $R'$ of $R$, there is a tight path with the same ends as $P_{\mathrm{res}}$ whose vertex set is $V(P_{\mathrm{res}}) \setminus R'$.

In a second step, we extend $P_{\mathrm{res}}$ to an almost-spanning tight path $P_{\mathrm{almost}}$. In the final step we re-use some vertices of $R$ to extend $P_{\mathrm{almost}}$ further to a structure which is 'almost' a tight Hamilton cycle, except that some vertices $R'$ of $R$ are used twice. Finally we apply the reservoir property of $P_{\mathrm{res}}$ to obtain the desired tight Hamilton cycle.

In [AKPP20], in the random hypergraph, there are two main tools needed to put this plan into action. First, for any given ordered $(k-1)$-tuple $\mathbf{x}$ of vertices and set $S$ of 'unused' vertices which is not too small, there will be lots of ways to start a tight path from $\mathbf{x}$ and continuing with vertices of $S$. Second, for any given pair of ordered $(k-1)$-tuples $\mathbf{x}$ and $\mathbf{y}$, and any given set $S$

of unused vertices which is not too small, it is possible to find a tight path from $\mathbf{x}$ to $\mathbf{y}$ in $S$.[2]

Neither of these statements is true in the resilience setting. Instead, we make use of hypergraph regularity to help us. In the following section 2 we state our main tools, and prove some of them. We first introduce spike paths, which we need to construct our reservoir structure (much as in [AKPP20]).

We give the notational setup for hypergraph regularity, and state a sparse, strengthened version of the Strong Hypergraph Regularity Lemma, Lemma 2.4, which may be of independent interest. We show that the output of this Regularity Lemma is, for $k$-graphs with our minimum degree condition, a structure which is robustly *tightly linked*: this is a version of connectivity appropriate for tight paths.

We show that the random hypergraph has certain nice properties: in particular, once one removes a small fraction of $(k-1)$-tuples, for any remaining $(k-1)$-tuple $\mathbf{x}$ and set $S$ which is reasonably small (it cannot contain more than $n/2$ vertices) there are lots of ways to start constructing a tight path from $\mathbf{x}$ *avoiding* $S$ (Lemma 2.12), and if we do so for a sufficiently large (but independent of $n$) number of steps, we reach a positive fraction of all $(k-1)$-tuples. This statement (Lemma 2.13) is one of the key points in our proof: most of the time, we can expand in a few steps from any given $(k-1)$-tuple to a positive density of $(k-1)$-tuples (and a similar statement holds for spike paths).

Using Lemma 2.13, regularity and tight linkedness, we can prove a Connecting Lemma (Lemma 2.19) which states that for any reasonably small set $S$ and most pairs $\mathbf{x}$ and $\mathbf{y}$ of $(k-1)$-tuples, there is a short tight path from $\mathbf{x}$ to $\mathbf{y}$ which avoids $S$.

These tools are enough to prove a Reservoir Lemma 2.23, which (much as in [AKPP20]) constructs $P_{\mathrm{res}}$ mentioned above. However again at this point difficulties arise. In the random hypergraph of [AKPP20], the vertices outside $P_{\mathrm{res}}$ have no particular structure. In our setting, $P_{\mathrm{res}}$ interacts in some rather unpredictable way with the existing structure provided by the Regularity Lemma. To deal with this, we use LP-duality in Lemma 2.21 to find a fractional matching which will tell us how many vertices we should use in each part of our regular partition in order to obtain $P_{\mathrm{almost}}$. We also at this point run into the difficulty that we can only guarantee expansion from the minimum degree when we are avoiding less than $n/2$ vertices, yet $P_{\mathrm{almost}}$ is supposed to cover almost all of the vertices; it is here that we need the 'strengthened' property of our Regularity Lemma.

In Section 3, we give the proof of Theorem 1.1, assuming the so far unproved lemmas.

In Section 4 we prove the Connection Lemma, Lemma 2.19, and also Lemma 2.20 which shows how we can use the strengthened regularity to continue extending a tight path even when most vertices have been used.

In Section 5 we prove the Reservoir Lemma, Lemma 2.23.

Finally, we defer the proof of our Regularity Lemma, Lemma 2.4, together with various more-or-less standard facts about dense hypergraph regularity, to Appendix A. Although some of these results are new and Lemma 2.4 may well be useful in future, the ideas needed to prove them are not new.

---

[2]To be accurate, these statements will be true for all the sets $S$ that actually appear in the proof, by a careful revealing argument; they are not true for every $S$.

# 2. Tools

## 2.1. Spike paths

To build our reservoir structure we need spike paths, which are the following variant of a tight path that changes orientation every $(k-1)$ steps. We will only consider spike paths with a number of vertices divisible by $k-1$.

**Definition 2.1** (Spike path). In an $k$-uniform hypergraph, a spike path with $t$ spikes consists of a sequence of $t$ pairwise disjoint $(k-1)$-tuples $\mathbf{a}_1, \ldots, \mathbf{a}_t$, where $\mathbf{a}_i = (a_{i,1}, \ldots, a_{i,k-1})$ for all $i$, with the property, that the edges $\{a_{i,k-j}, \ldots, a_{i,1}, a_{i+1,1}, \ldots, a_{i+1,j}\}$ are present for all $i = 1, \ldots, t-1$ and $j = 1, \ldots, k-1$. We call $\mathbf{a}_i$ the $i$th spike.

## 2.2. Notation

A $k$-*complex* is a hypergraph $H$ all of whose edges have size at most $k$, which is down-closed, i.e. if $e \in E(H)$ and $e' \subseteq e$ then $e' \in E(H)$. The *layers* of a $k$-complex are, for each $0 \leqslant i \leqslant k$, the $i$-uniform hypergraph $H^{(i)}$ on the same vertex set, where $E(H^{(i)}) = \{e \in E(H) : |e| = i\}$.

A $k$-*multicomplex* is, informally, a $k$-complex in which multiple edges of any size between $2$ and $k$ are permitted, together with a map *boundary* $\partial$ identifying the $(i-1)$-edges which support a given $i$-edge. Formally, a $k$-multicomplex $H$ consists of a *vertex set* $V(H)$, together with a set of *edges* $E(H)$, a *vertices* map vertices : $E \to \mathcal{P}(V)$ such that vertices$(e)$ is a set of size between $0$ and $k$ for each $e \in E(H)$, and a *boundary* map $\partial : E \setminus \{\varnothing\} \to \mathcal{P}(E)$ such that $\partial e$ contains exactly one edge whose vertices are vertices$(e) \setminus \{v\}$ for each $v \in$ vertices$(e)$, and no other edges. We further insist on the following consistency condition: if $2 \leqslant i \leqslant k$, and $S$ is a set of $i$ edges each with $i-1$ vertices, such that $\left| \bigcup_{f \in S} \partial f \right| > \binom{i}{i-2}$, then there are no edges $e \in H$ such that $\partial e = S$. We say that the *uniformity* of an edge $e$ is $|$vertices$(e)|$, and we may write that $e$ is an edge *on the set* vertices$(e)$, or that $e$ is a $|$vertices$(e)|$-*edge*. We will also say, given a set $S$ consisting of $i$ edges of uniformity $(i-1)$, that $e$ is *supported* on $S$ if $\partial e = S$.

Note that the boundary of a 1-edge is necessarily $\{\varnothing\}$, and that 'down-closure' is forced by the condition of the boundary map. To better understand the consistency condition, consider the following. If $e$ is an edge of $H$ with at least two vertices, and $x$ and $y$ are distinct vertices of $e$, let $e_x$ and $e_y$ be the edges in $\partial e$ whose vertices do not contain respectively $x$ and $y$. There is an edge $e_{xy}$ in $\partial e_x$, and an edge $e_{yx}$ in $\partial e_y$, on vertices$(e) \setminus \{x, y\}$. The consistency condition is equivalent to insisting that for any $e$, $x$ and $y$ we have $e_{yx} = e_{xy}$.

Observe that a $k$-complex is a $k$-multicomplex, where the vertices of each edge are simply its members as a set, and the boundary map is the usual boundary $\partial e = \{e \setminus \{v\} : v \in e\}$ (which is in this case the only possible boundary map for the given vertices map). However in general, for a given ground set, edge set and vertices map, there may be several different boundary maps which fit the definition of $k$-multicomplex; these return different multicomplexes. The idea here is that we will need to think of a given edge (say with vertices $\{1, 2, 3\}$) as containing specific edges with vertices $\{1, 2\}$, $\{1, 3\}$ and $\{2, 3\}$, and the map $\partial$ tells us which edges these are. We should stress that it is possible to have a $k$-multicomplex in which there are two different edges which have the same boundary and vertices, and indeed the multicomplexes we consider in this

paper will have this property for edges of uniformity two and above (though for us a 1-edge will always be the unique 1-edge on a given vertex).

Given a vector $\mathbf{d} = (d_2, \dots, d_k)$ where $1/d_i \in \mathbb{N}$ for each $i$, we say that a $k$-multicomplex $H$ is $\mathbf{d}$-*equitable* if there is exactly one 1-edge on each vertex, and furthermore for any $2 \leqslant i \leqslant k$ and $i$-set $X$ of vertices the following holds. Whenever $S$ is a collection of $i$ edges of uniformity $i - 1$ in $H$, one on the vertices $X \setminus \{x\}$ for each $x \in X$, if the union $\bigcup_{f \in S} \partial f$ has exactly $\binom{i}{i-2}$ edges then the number of $i$-edges in $H$ supported on $S$ is exactly $1/d_i$. We refer to $\mathbf{d}$ as the *density vector* of the multicomplex.

Finally, we need a notion of connectedness for multicomplexes.

**Definition 2.2** (tight link, tightly linked)**.** Given a $k$-multicomplex $\mathcal{R}$, and two $(k-1)$-edges $u, v$ of $\mathcal{R}$, let $\mathbf{u}$ be $u$ together with an ordering $(u_1, \dots, u_{k-1})$ of its vertices, and similarly let $\mathbf{v}$ be $v$ together with an ordering $(v_1, \dots, v_{k-1})$ of its vertices. A *tight link from $\mathbf{u}$ to $\mathbf{v}$ in $\mathcal{R}$* is the following collection of (not necessarily distinct) vertices and edges of $\mathcal{R}$.

For each $1 \leqslant j \leqslant k-1$, there is a vertex $w_j$. There are $k$-edges $e_{1,u}$ and $e_{1,v}$ of $\mathcal{R}$, where $e_{1,u}$ is on vertices $\{u_1, \dots, u_{k-1}, w_1\}$ and $u \in \partial e_{1,u}$, and $e_{1,v}$ is on vertices $\{v_1, \dots, v_{k-1}, w_1\}$ and $v \in \partial e_{1,v}$. For each $2 \leqslant j \leqslant k - 1$, there are $k$-edges $e_{j,u}$ and $e_{j,v}$ of $\mathcal{R}$, where $e_{j,u}$ is on vertices $\{u_j, \dots, u_{k-1}, w_1, \dots, w_j\}$ and $\partial e_{j-1,u} \cap \partial e_{j,u} \neq \varnothing$, and $e_{j,v}$ is on vertices $\{v_j, \dots, v_{k-1}, w_1, \dots, w_j\}$ and $\partial e_{j-1,v} \cap \partial e_{j,v} \neq \varnothing$. Finally $\partial e_{k-1,u} \cap \partial e_{k-1,v} \neq \varnothing$.

We say that a $k$-multicomplex $\mathcal{R}$ is *tightly linked* if for any two $(k-1)$-edges in $\mathcal{R}$, and any orderings of their vertices, $\mathbf{u}$ and $\mathbf{v}$, there is a tight link from $\mathbf{u}$ to $\mathbf{v}$ in $\mathcal{R}$.

The precise sequence of vertices and edges is not critical (it is simply a particular structure we can easily construct). However it will be convenient to note that the $k$-edges of a tight link are in fact a spike path with three spikes. Note that there is $\ell \in \mathbb{N}$ and a permutation $\varrho$ on $[k-1]$ such that for any $\mathbf{u}$ and $\mathbf{v}$, if there is a tight link from $\mathbf{u}$ to $\mathbf{v}$ then there is a homomorphism from the $\ell$-vertex tight path to $\mathcal{R}$, using only the $k$-edges of the tight link, where the first $k-1$ vertices of the tight path are sent to $\mathbf{u}$ in order and the last $k-1$ vertices to the vertices of $\mathbf{v}$ in the order $\varrho$.

### 2.3. Sparse hypergraph regularity

We need a strengthened version of the Strong Hypergraph Regularity Lemma for sparse hypergraphs. The Strong Hypergraph Regularity Lemma was first proved by Rödl and Skokan [RS04] and Gowers [Gow07]; we use a version due to Rödl and Schacht [RS07], from which we deduce a strengthened version by a standard method. We then use a weak sparse regularity lemma of Conlon, Fox and Zhao [CFZ15] to transfer this strengthened version to a sparse version, following [ADS19].

In order to state our regularity lemma, we need quite a few definitions. These are either standard definitions for the dense ($p = 1$) case, or the natural sparse versions of the same, as taken from [ABCM17].

Let $\mathcal{P}$ partition a vertex set $V$ into parts $V_1, \dots, V_s$. We say that a subset $S \subseteq V$ is $\mathcal{P}$-*partite* if $|S \cap V_i| \leqslant 1$ for every $i \in [s]$ and the *index* of a $\mathcal{P}$-partite set $S \subseteq V$ is $i(S) := \{i \in [s] : |S \cap V_i| = 1\}$. For any $A \subseteq [s]$ we write $V_A$ for $\bigcup_{i \in A} V_i$. Similarly, we say that a hypergraph $H$ is $\mathcal{P}$-*partite* if all of its edges are $\mathcal{P}$-partite. In this case we refer to

the parts of $\mathcal{P}$ as the *vertex classes* of $H$. Moreover, we say that a hypergraph $H$ is *s-partite* if there is some partition $\mathcal{P}$ of $V(H)$ into $s$ parts for which $H$ is $\mathcal{P}$-partite.

Let $i \geqslant 2$, let $H_i$ be any $i$-partite $i$-graph, and let $H_{i-1}$ be any $i$-partite $(i-1)$-graph, on a common vertex set $V$ partitioned into $i$ common vertex classes. We denote by $K_i(H_{i-1})$ the $i$-partite $i$-graph on $V$ whose edges are all $i$-sets in $V$ which are supported on $H_{i-1}$ (i.e. induce a copy of the complete $(i-1)$-graph $K_i^{i-1}$ on $i$ vertices in $H_{i-1}$). Given $p \in (0,1]$, the *p-density of $H_i$ with respect to $H_{i-1}$* is then defined to be

$$d_p(H_i|H_{i-1}) := \frac{|K_i(H_{i-1}) \cap H_i|}{p|K_i(H_{i-1})|}$$

if $|K_i(H_{i-1})| > 0$. For convenience we take $d_p(H_i|H_{i-1}) := 0$ if $|K_i(H_{i-1})| = 0$, and we assume $H_1$ is the complete 1-graph on $V$, whose edge set is $V$. So $d_p(H_i|H_{i-1})$ is the proportion of copies of $K_i^{i-1}$ in $H_{i-1}$ which are also edges of $H_i$, scaled by $p$. When $H_{i-1}$ is clear from the context, we simply refer to $d_p(H_i|H_{i-1})$ as the *relative p-density of $H_i$*. We say that $H_i$ is $(d_i, \varepsilon, p)$-*regular with respect to $H_{i-1}$* if we have $d_p(H_i|H'_{i-1}) = d_i \pm \varepsilon$ for every subgraph $H'_{i-1}$ of $H_{i-1}$ such that $|K_i(H'_{i-1})| > \varepsilon|K_i(H_{i-1})|$. Given an $i$-graph $G$ whose vertex set contains that of $H_{i-1}$, we say that $G$ is $(d_i, \varepsilon, p)$-*regular with respect to $H_{i-1}$* if the $i$-partite subgraph of $G$ induced by the vertex classes of $H_{i-1}$ is $(d_i, \varepsilon, p)$-regular with respect to $H_{i-1}$. Finally, we say $G$ is $(\varepsilon, p)$-*regular with respect to $H_{i-1}$* if there exists $d_i$ such that $G$ is $(d_i, \varepsilon, p)$-regular with respect to $H_{i-1}$. Similarly as before, when $H_{i-1}$ is clear from the context, we refer to the relative density of this $i$-partite subgraph of $G$ with respect to $H_{i-1}$ as the *relative p-density of $G$*.

Now let $H$ be an $s$-partite $k$-complex on vertex classes $V_1, \ldots, V_s$, where $s \geqslant k \geqslant 3$. Recall that, since $H$ is a complex, if $e \in H$ and $e' \subseteq e$ then $e' \in H$. So if $e \in H^{(i)}$ for some $2 \leqslant i \leqslant k$, then the vertices of $e$ induce a copy of $K_i^{i-1}$ in $H^{(i-1)}$. We say that $H$ is $(d_k, \ldots, d_2, \varepsilon_k, \varepsilon, p)$-*regular* if

(a) for any $2 \leqslant i \leqslant k-1$ and any $A \in \binom{[s]}{i}$, the induced subgraph $H^{(i)}[V_A]$ is $(d_i, \varepsilon, 1)$-regular with respect to $H^{(i-1)}[V_A]$, and

(b) for any $A \in \binom{[s]}{k}$, the induced subgraph $H^{(k)}[V_A]$ is $(d_k, \varepsilon_k, p)$-regular with respect to $H^{(k-1)}[V_A]$.

So each constant $d_i$ approximates the relative density of each subgraph $H^{(i)}[V_A]$ for $A \in \binom{[s]}{i}$. For a $(k-1)$-tuple $\mathbf{d} = (d_k, \ldots, d_2)$ we write $(\mathbf{d}, \varepsilon_k, \varepsilon, p)$-regular to mean $(d_k, \ldots, d_2, \varepsilon_k, \varepsilon, p)$-regular.

The definition of a $(\mathbf{d}, \varepsilon_k, \varepsilon, p)$-regular complex $H$ is the 'right' generalisation of an $\varepsilon$-regular pair $(X, Y)$ in dense graphs to sparse hypergraphs. The Szemerédi Regularity Lemma states that there is a partition of the vertices of any graph into boundedly many parts such that most pairs of parts are regular; now our aim is to define a generalisation of 'partition' in order to say that we can partition any $k$-uniform hypergraph $G$ such that most $k$-sets lie in regular complexes. As one can guess from the phrasing, the $k$-layer of each complex will consist of (all) edges of $G$ supported by the complex. The lower layers will be in the 'partition', and we now set up the notation to define this.

Fix $k \geqslant 3$, and let $\mathcal{P}$ partition a vertex set $V$ into parts $V_1, \ldots, V_t$. For any $A \subseteq [t]$, we denote by $\mathrm{Cross}_A(\mathcal{P})$ the collection of $\mathcal{P}$-partite subsets $S \subseteq V$ of index $i(S) = A$. Likewise, we denote

by $\mathrm{Cross}_j(\mathcal{P})$ the union of $\mathrm{Cross}_A(\mathcal{P})$ for each $A \in \binom{[t]}{j}$, so $\mathrm{Cross}_j(\mathcal{P})$ contains all $\mathcal{P}$-partite subsets $S \subseteq V$ of size $j$. When $\mathcal{P}$ is clear from the context, we write simply $\mathrm{Cross}_A$ and $\mathrm{Cross}_j$. For each $2 \leqslant j \leqslant k-1$ and $A \in \binom{[t]}{j}$ let $\mathcal{P}_A$ be a partition of $\mathrm{Cross}_A$. For consistency of notation we also define the trivial partitions $\mathcal{P}_{\{s\}} := \{V_s\}$ for $s \in [t]$ and $\mathcal{P}_\varnothing := \{\varnothing\}$. Let $\mathcal{P}^*$ consist of the partitions $\mathcal{P}_A$ for each $A \in \binom{[t]}{j}$ and each $0 \leqslant j \leqslant k-1$. We say that $\mathcal{P}^*$ is a $(k-1)$-*family of partitions on* $V$ if whenever $S, T \in \mathrm{Cross}_A$ lie in the same part of $\mathcal{P}_A$ and $B \subseteq A$, then $S \cap \bigcup_{j \in B} V_j$ and $T \cap \bigcup_{j \in B} V_j$ lie in the same part of $\mathcal{P}_B$. In other words, given $A \in \binom{[t]}{j}$, if we specify one part of each $\mathcal{P}_B$ with $B \in \binom{A}{j-1}$, then we obtain a subset of $\mathrm{Cross}_A$ consisting of all $S \in \mathrm{Cross}_A$ whose $(j-1)$-subsets are in the specified parts. We say that this subset of $\mathrm{Cross}_A$ is the subset *supported* by the specified parts of $\mathcal{P}_B$. In general, we say that a $j$-set $e$ is *supported* by a collection $S$, with $|S| = j$, of $(j-1)$-graphs if exactly one $(j-1)$-subset of $e$ is in each member of $S$, and we say a set of $j$-edges $E$ is supported by $S$ if each edge of $E$ is supported by $S$.

Thus the partitions $\mathcal{P}_B$ give a natural partition of $\mathrm{Cross}_A$, and we are saying that $\mathcal{P}_A$ must refine it.

We refer to the parts of each member of $\mathcal{P}^*$ as *cells*. Also, we refer to $\mathcal{P}$ as the *ground partition* of $\mathcal{P}^*$, and the parts of $\mathcal{P}$ (i.e. the vertex classes $V_i$) as the *clusters* of $\mathcal{P}^*$. For each $0 \leqslant j \leqslant k-1$ let $\mathcal{P}^{(j)}$ denote the partition of $\mathrm{Cross}_j$ formed by the parts (which we call $j$-*cells*) of each of the partitions $\mathcal{P}_A$ with $A \in \binom{[t]}{j}$ (so in particular $\mathcal{P}^{(1)} = \mathcal{P}$).

Observe that a $(k-1)$-family of partitions $\mathcal{P}^*$ naturally form the edges of a $k$-multicomplex, whose vertex set is the (set of parts of the) ground partition, whose edges of uniformity $j \leqslant k-1$ are the $j$-cells, with the vertices map identifying the $j$ parts of the ground partition which contain a given $j$-cell, and where the boundary operator $\partial e$ identifies the $(|e|-1)$-cells supporting $e$. So far we have described a $(k-1)$-multicomplex; we extend this to a $k$-complex by adding, for each set $S$ of $k$ edges of uniformity $k-1$ which can be a boundary (i.e. which is such that $\left| \bigcup_{f \in S} \partial f \right| = \binom{k}{k-2}$ ) one edge of uniformity $k$ whose boundary is $S$. When we refer to the *multicomplex of the family of partitions* $\mathcal{P}^*$ we mean this multicomplex. Note that we have defined the word 'support' both in terms of multicomplexes and in terms of a family of partitions: but these definitions are consistent, i.e. that a given $j$-cell is supported by some $(j-1)$-cells means the same thing whether one reads 'support' in terms of the family of partitions or its multicomplex.

For any $0 \leqslant j \leqslant k-1$, any $A \in \binom{[t]}{j}$ and any $Q' \in \mathrm{Cross}_A$, let $C_{Q'}$ denote the cell of $\mathcal{P}_A$ which contains $Q'$. Then the fact that $\mathcal{P}^*$ is a family of partitions implies that for any $Q \in \mathrm{Cross}_k$ the union $\mathcal{J}(Q) := \bigcup_{Q' \subsetneq Q} C_{Q'}$ of cells containing subsets of $Q$ is a $k$-partite $(k-1)$-complex. We say that the $(k-1)$-family of partitions $\mathcal{P}^*$ is $(t_0, t_1, \varepsilon)$-*equitable* if

(a) $\mathcal{P}$ partitions $V$ into $t$ clusters of equal size, where $t_0 \leqslant t \leqslant t_1$,

(b) for each $2 \leqslant j \leqslant k-1$, $\mathcal{P}^{(j)}$ partitions $\mathrm{Cross}_j$ into at most $t_1$ cells,

(c) there exists $\mathbf{d} = (d_{k-1}, \dots, d_2)$ such that for each $2 \leqslant j \leqslant k-1$ we have $d_j \geqslant 1/t_1$ and $1/d_j \in \mathbb{N}$, and for every $Q \in \mathrm{Cross}_k$ the $k$-partite $(k-1)$-complex $\mathcal{J}(Q)$ is $(\mathbf{d}, \varepsilon, \varepsilon, 1)$-regular.

Note that conditions (a) and (c) imply that $\mathcal{J}(Q)$ is a $(1, t_1, \varepsilon)$-equitable $(k-1)$-complex (with the same density vector $\mathbf{d}$) for any $Q \in \mathrm{Cross}_k$.

Next, for any $\mathcal{P}$-partite set $Q$ with $2 \leqslant |Q| \leqslant k$, define $\hat{P}(Q; \mathcal{P}^*)$ to be the $|Q|$-partite $(|Q|-1)$-graph on $V_{i(Q)}$ with edge set $\bigcup_{Q' \in \binom{Q}{|Q|-1}} C_{Q'}$. We refer to $\hat{P}(Q; \mathcal{P}^*)$ as a $|Q|$-*polyad*; when the family of partitions $\mathcal{P}^*$ is clear from the context, we write simply $\hat{P}(Q)$ rather than $\hat{P}(Q; \mathcal{P}^*)$. Note that the condition for $\mathcal{P}^*$ to be a $(k-1)$-family of partitions can then be rephrased as saying that if $2 \leqslant |Q| \leqslant k-1$ then the cell $C_Q$ is supported on $\hat{P}(Q)$, and in the multicomplex corresponding to $\mathcal{P}^*$ we have edges corresponding to the cells of each uniformity from 1 to $k-1$ inclusive, together with edges corresponding to the $k$-polyads supported by $\mathcal{P}^*$. As shown in [ABCM17, Claim 32], if $\mathcal{P}^*$ is $(t_0, t_1, \varepsilon)$-equitable for sufficiently small $\varepsilon$, then for any $2 \leqslant j \leqslant k-1$ and any $Q \in \mathrm{Cross}_j$ the number of $j$-cells of $\mathcal{P}^*$ supported on $\hat{P}(Q)$ is precisely equal to $1/d_j$. More specifically, if $\left(d_j^{-1} - 1\right)\left(d_j + \varepsilon\right) < 1$, and $\left(d_j^{-1} + 1\right)\left(d_j - \varepsilon\right) > 1$, then by definition necessarily there are exactly $d_j^{-1}$ cells supported; it suffices to choose $\varepsilon \ll d_j^2$ to ensure these two inequalities. In other words, the multicomplex corresponding to $\mathcal{P}^*$ is $\mathbf{d}$-equitable.

Now let $G$ be a $k$-graph on $V$, and let $\mathcal{P}^*$ be a $(k-1)$-family of partitions on $V$. Let $Q \in \mathrm{Cross}_k$, so the polyad $\hat{P}(Q)$ is a $k$-partite $(k-1)$-graph. We say that $G$ is $(\varepsilon_k, p)$-*regular with respect to* $\mathcal{P}^*$ if there are at most $\varepsilon_k \binom{|V|}{k}$ sets $Q \in \mathrm{Cross}_k$ for which $G$ is not $(\varepsilon_k, p)$-regular with respect to the polyad $\hat{P}(Q)$. That is, at most an $\varepsilon_k$-proportion of subsets of $V$ of size $k$ yield polyads with respect to which $G$ is not regular (though some subsets of $V$ of size $k$ do not yield any polyad due to not being members of $\mathrm{Cross}_k$).

At this point we have the setup to state the Strong Hypergraph Regularity Lemma, which says that for any $k$-uniform hypergraph $G$ there is a $(k-1)$-family of partitions $\mathcal{P}^*$, which is $(t_0, t_1, \varepsilon)$-equitable for some $t_1$ independent of $|V(G)|$, such that $G$ is regular with respect to $\mathcal{P}^*$. However for this paper we need a stronger version, which is not standard (the dense graph version, called the Strengthened Regularity Lemma, is due to Alon, Fischer, Krivelevich and Szegedy [AFKS00], and it is folklore that the hypergraph version we now state should exist). To that end, given two families of partitions $\mathcal{P}^*$ and $\mathcal{Q}^*$ on the same vertex set, we say that $\mathcal{P}^*$ *refines* $\mathcal{Q}^*$ if every cell of $\mathcal{P}^*$ is a subset of some cell of $\mathcal{Q}^*$.

**Definition 2.3.** Given a $k$-uniform hypergraph $G$, we call a pair of families of partitions $(\mathcal{P}_c^*, \mathcal{P}_f^*)$ on $V(G)$ a $(t_0, t_1, t_2, \varepsilon_k, \varepsilon, f_k, f, p)$-*strengthened pair* for $G$ if the following are true.

(S1)  $\mathcal{P}_f^*$ refines $\mathcal{P}_c^*$.

(S2)  $\mathcal{P}_c^*$ is $(t_0, t_1, \varepsilon)$-equitable.

(S3)  $G$ is $(\varepsilon_k, p)$-regular with respect to $\mathcal{P}_c^*$.

(S4)  $\mathcal{P}_f^*$ is $(t_0, t_2, f)$-equitable.

(S5)  $G$ is $(f_k, p)$-regular with respect to $\mathcal{P}_f^*$.

(S6)  For all but at most $\varepsilon_k^2 \binom{|V(G)|}{k}$ elements $Q$ of $\mathrm{Cross}_k(\mathcal{P}_c)$, we have $d_p\big(G\big|\hat{\mathcal{P}}(Q, \mathcal{P}_c^*)\big) = d_p\big(G\big|\hat{\mathcal{P}}(Q, \mathcal{P}_f^*)\big) \pm \varepsilon_k$.

We refer to $\mathcal{P}_c^*$ as the *coarse partition* and $\mathcal{P}_f^*$ as the *fine partition*. Slightly extending the usual definition, we say a $k$-polyad $\hat{P}(Q;\mathcal{P}_c^*)$ is *irregular* (with respect to $G$) if any one of the following three things occurs:

(i) $G$ is not $(\varepsilon_k, p)$-regular with respect to $\hat{P}(Q;\mathcal{P}_c^*)$,

(ii) for more than an $\varepsilon_k$-fraction of the $k$-sets $Q'$ supported on $\hat{P}(Q;\mathcal{P}_c^*)$, $G$ is not $(f_k, p)$-regular with respect to $\hat{P}(Q';\mathcal{P}_f^*)$, or

(iii) for more than an $\varepsilon_k$-fraction of the $k$-sets $Q'$ supported on $\hat{P}(Q;\mathcal{P}_c^*)$, $d_p\big(G\big|\hat{P}(Q';\mathcal{P}_f^*)\big) \neq d_p\big(G\big|\hat{P}(Q;\mathcal{P}_c^*)\big) \pm \varepsilon_k$.

If a polyad of $\mathcal{P}_c^*$ is not irregular, we say it is *regular*.

We will always choose $f_k$ such that $f_k \leqslant \varepsilon_k^2$, and $\varepsilon$ small enough that every $k$-polyad supports very close to the same number of $k$-edges. Under this assumption, it is straightforward to check that at most a $4\varepsilon_k$-fraction of polyads in $\mathcal{P}_c^*$ are irregular (we will prove this in Appendix A, Proposition A.3).

We need one more definition. Given any (not necessarily distinct) subsets $E_1, \ldots, E_k$ in $\binom{[n]}{k-1}$, we say a $k$-set $S \subseteq [n]$ is *rainbow* for the $E_i$ if there is an injective labelling of the $(k-1)$-subsets of $S$ with the numbers $1, \ldots, k$ such that the $(k-1)$-subset labelled $i$ is in $E_i$. We write $K_k(E_1, \ldots, E_k)$ for the set of rainbow $k$-sets in $[n]$. We say that a graph $G$ on $[n]$ is $(\eta, p)$-*upper regular* if the following holds. For any $E_1, \ldots, E_k$, we have

$$\big|E(G) \cap K_k(E_1, \ldots, E_k)\big| \leqslant p\big|K_k(E_1, \ldots, E_k)\big| + p\eta n^k .$$

Finally, we are in a position to state our strengthened sparse version of the Strong Hypergraph Regularity Lemma. Informally, what this says is that we can find $\mathcal{P}_c^*$ and $\mathcal{P}_f^*$ which are simultaneously a strengthened pair for $s$ edge-disjoint graphs, for any (fixed) regularity $\varepsilon_k$ of $\mathcal{P}_c^*$, where $\varepsilon$ and $f$ can be as small as desired depending on the number of parts in $\mathcal{P}_c^*$ and $\mathcal{P}_f^*$ respectively, and furthermore the regularity $f_k$ of $\mathcal{P}_f^*$ can depend arbitrarily on the number of parts of $\mathcal{P}_c^*$.

**Lemma 2.4** (Strengthened Sparse Strong Hypergraph Regularity Lemma). *Given integers $k \geqslant 2$ and $t_0$ and $s$, real $\varepsilon_k > 0$ and functions $\varepsilon, f_k, f : \mathbb{N} \to (0,1]$, there exists a real $\eta > 0$ and integers $T$ and $n_0$ such that the following holds for all $n \geqslant n_0$ with $T!|n$. Let $V$ be a vertex set of size $n$, suppose that $G_1, \ldots, G_s$ are $k$-uniform hypergraphs on $V$, and suppose $\mathcal{Q}^*$ is a family of partitions on $V$ which is $(1, t_0, \eta)$-equitable. Suppose furthermore that for each $1 \leqslant i \leqslant s$ there is a real $p_i \in (0,1]$ such that $G_i$ is $(\eta, p_i)$-upper regular. Then there are integers $t_1, t_2$ with $t_0 \leqslant t_1 \leqslant t_2 \leqslant T$, and families of partitions $\mathcal{P}_c^*$ and $\mathcal{P}_f^*$, both refining $\mathcal{Q}^*$, such that for each $1 \leqslant i \leqslant s$, the pair $(\mathcal{P}_c^*, \mathcal{P}_f^*)$ is a $\big(t_0, t_1, t_2, \varepsilon_k, \varepsilon(t_1), f_k(t_1), f(t_2), p_i\big)$-strengthened pair for $G_i$.*

We prove this lemma in Appendix A. Note that the case $k = 2$ will not be used here; and in this setting the 'families of partitions' are simply vertex set partitions and the functions $\varepsilon$ and $f$ play no rôle.

Given a $(t_0, t_1, t_2, \varepsilon_k, \varepsilon, f_k, f, p)$-strengthened pair $(\mathcal{P}_c^*, \mathcal{P}_f^*)$ for $G$, recall that $\mathcal{P}_c^*$ has the structure of a multicomplex. We denote by $\mathcal{R}_{\varepsilon_k}(G; \mathcal{P}_c^*, \mathcal{P}_f^*)$ the $\varepsilon_k$-*reduced multicomplex of* $G$ *with respect to* $(\mathcal{P}_c^*, \mathcal{P}_f^*)$, which is the (unique) maximal submulticomplex of $\mathcal{P}_c^*$ which has the following properties.

(RG1) Every $k$-edge of $\mathcal{R}_{\varepsilon_k}(G; \mathcal{P}_c^*, \mathcal{P}_f^*)$ is regular.

(RG2) For each $1 \leqslant i \leqslant k - 1$, each $i$-edge of $\mathcal{R}_{\varepsilon_k}(G; \mathcal{P}_c^*, \mathcal{P}_f^*)$ is in the boundary of at least

$$\left(1 - 2^{i+2}\varepsilon_k^{1/k}\right)t \prod_{j=2}^{i+1} d_j^{-\binom{i}{j-1}} \text{ if } i < k - 1, \text{ and } \left(1 - 2^{k+1}\varepsilon_k^{1/k}\right)t \prod_{j=2}^{k-1} d_j^{-\binom{k-1}{j-1}} \text{ if } i = k - 1$$

$(i + 1)$-edges of $R_{\varepsilon_k}(G; \mathcal{P}_c^*, \mathcal{P}_f^*)$.

The existence and uniqueness of the reduced multicomplex are trivial: we obtain it by simply iteratively removing from the multicomplex $\mathcal{P}_c^*$ edges which either fail one of (RG1) or (RG2), or from whose boundary we removed edges (so that they are no longer supported and cannot be in the multicomplex). It is easy, but not quite trivial, to show that most of the vertices of $\mathcal{R}_{\varepsilon_k}(G; \mathcal{P}_c^*, \mathcal{P}_f^*)$ (i.e. the parts of $\mathcal{P}_c$) are also 1-edges of $\mathcal{R}_{\varepsilon_k}(G; \mathcal{P}_c^*, \mathcal{P}_f^*)$. Now given $d > 0$, we let $\mathcal{R}_{\varepsilon_k,d}(G; \mathcal{P}_c^*, \mathcal{P}_f^*)$ be the (unique) submulticomplex of $\mathcal{R}_{\varepsilon_k}(G; \mathcal{P}_c^*, \mathcal{P}_f^*)$ obtained by removing all $k$-edges corresponding to polyads whose relative $p$-density is less than $d$. We call $\mathcal{R}_{\varepsilon_k,d}(G; \mathcal{P}_c^*, \mathcal{P}_f^*)$ the $(\varepsilon_k, d)$-*reduced multicomplex of* $G$ *with respect to* $(\mathcal{P}_c^*, \mathcal{P}_f^*)$.

In Appendix A we show the following lemma.

**Lemma 2.5.** *Given $k \in \mathbb{N}$ and $d > 0$ suppose that $t_0 \in \mathbb{N}$ is sufficiently large. Given any constants $\delta, \varepsilon_k, \nu > 0$, any function $\varepsilon : \mathbb{N} \to (0, 1]$ which tends to zero sufficiently fast, any $t_1, t_2 \in \mathbb{N}$, any $0 < f_k \leqslant \varepsilon_k^2$ and any $f > 0$, there exists $\eta > 0$ such that the following holds for any sufficiently large $n$ and any $p > 0$. Suppose $G$ is an $n$-vertex hypergraph which is $(\eta, p)$-upper regular and every $(k - 1)$-set in $V(G)$ is contained in at least $\delta pn$ edges. Suppose that $(\mathcal{P}_c^*, \mathcal{P}_f^*)$ is a $(t_0, t_1, t_2, \varepsilon_k, \varepsilon(t_1), f_k, f, p)$-strengthened pair for $G$.*

*Let $\mathcal{R} = \mathcal{R}_{\varepsilon_k,d}(G; \mathcal{P}_c^*, \mathcal{P}_f^*)$ be the $(\varepsilon_k, d)$-reduced multicomplex of $G$, and suppose that $\mathcal{P}_c^*$ has $t$ clusters and density vector $\mathbf{d} = (d_{k-1}, \ldots, d_2)$. Then $\mathcal{R}$ contains at least $\left(1 - 4\varepsilon_k^{1/k}\right)t$ 1-edges, and every $(k-1)$-edge of $\mathcal{R}$ is contained in at least*

$$\left(\delta - 2d - 2^{k+2}\varepsilon_k^{1/k}\right)t \cdot \prod_{i=2}^{k-1} d_i^{-\binom{k-1}{i-1}}$$

*$k$-edges of $\mathcal{R}$.*

*Finally, if $\delta > \frac{1}{2} + 2d + 2^{k+2}\varepsilon_k^{1/k} + \nu$, then any induced subcomplex of $\mathcal{R}$ on at least $(1 - \nu)t$ 1-edges is tightly linked.*

The next lemma, often called the Dense Counting Lemma, is a straightforward generalisation of the well-known graph Counting Lemma (in contrast to the so-called Sparse Counting Lemma, which is much harder; the difference being that in the Dense Counting Lemma the parameter $\varepsilon$ of regularity is much smaller than all the density parameters). We state the special case of counting $(k-1)$- and $k$-cliques in $(k-1)$-uniform hypergraphs. The version that we need works with parts of different sizes, but this can be easily derived from the version with parts of the same size from [KRS02, Theorem 6.5].

**Lemma 2.6** (Dense Counting Lemma). *For all integers $k \geqslant 2$ and constants $\alpha, \gamma, d_0 > 0$, there exists $\varepsilon > 0$ such that the following holds. Let $\mathbf{d} = (d_{k-1}, \ldots, d_2)$ be a vector of real numbers with $d_i \geqslant d_0$ for each $2 \leqslant i \leqslant k-1$, and let $G$ be a $k$-partite $(k-1)$-complex which is $(\mathbf{d}, \varepsilon, \varepsilon, 1)$-regular and has parts $V_1, \ldots, V_k$ of size at least $m \geqslant \alpha^{-1}\varepsilon^{-1}$. Then for $V_i' \subseteq V_i$ of size $|V_i'| \geqslant \alpha|V_i|$ for $i = 1, \ldots, k$ the number of copies of the $k$-vertex complete $(k-1)$-complex in $G[V_1', \ldots, V_k']$ is*

$$\left(1 \pm \gamma\right) \prod_{i=1}^{k} |V_i'| \prod_{i=2}^{k-1} d_i^{\binom{k}{i}},$$

*and the number of copies of the $(k-1)$-vertex complete $(k-1)$-complex in $G[V_1', \ldots, V_{k-1}']$ is*

$$\left(1 \pm \gamma\right) \prod_{i=1}^{k-1} |V_i'| \prod_{i=2}^{k-1} d_i^{\binom{k-1}{i}}.$$

Note that with $\alpha = 1$ this is the Dense Counting Lemma with parts of the same size. We give the proof for this generalisation in Appendix A. If we do not remove too many vertices from the 1-cells we still have a regular complex with slightly different parameters. We will use this to prove Lemma 2.6, but also need it in our arguments.

**Lemma 2.7** (Regular Restriction Lemma [ABCM17, Lemma 28]). *For all integers $k \geqslant 2$ and constants $\alpha, d_0 > 0$, there exists $\varepsilon > 0$ such that the following holds. Let $\mathbf{d} = (d_{k-1}, \ldots, d_2)$ be a vector of real numbers with $d_i \geqslant d_0$ for each $2 \leqslant i \leqslant k-1$, and let $G$ be a $k$-partite $(k-1)$-complex with parts $V_1, \ldots, V_k$ of size $m \geqslant \varepsilon^{-1}$ which is $(\mathbf{d}, \varepsilon, \varepsilon, 1)$-regular. Choose any $V_i' \subseteq V_i$ of size at least $\alpha m$ for $i = 1, \ldots, k$. Then the induced subcomplex $G[V_1', \ldots, V_k']$ is $(\mathbf{d}, \sqrt{\varepsilon}, 1)$-regular.*

We will also need the following two lemmas that follow from the Dense Counting Lemma and the Regular Restriction Lemma. The first allows us to control the 'degree' of most tuples within the $(k-1)$-complex. For $k = 3$ this basically says that most edges are contained in the correct number of triangles.

**Lemma 2.8** (Degree Counting Lemma). *For all integers $k \geqslant 2$ and constants $\alpha, \gamma, d_0 > 0$, there exists $\varepsilon > 0$ such that the following holds. Let $\mathbf{d} = (d_{k-1}, \ldots, d_2)$ be a vector of real numbers with $d_i \geqslant d_0$ for each $2 \leqslant i \leqslant k$, and let $G$ be a $k$-partite $(k-1)$-complex with parts $V_1, \ldots, V_k$ of size $m \geqslant \alpha^{-1}\varepsilon^{-1}$ which is $(\mathbf{d}, \varepsilon, \varepsilon, 1)$-regular. Choose any $V_i' \subseteq V_i$ of size*

*at least $\alpha m$ for $i = 1, \ldots, k$ and let $G' = G[V'_1, \ldots, V'_k]$. Then at least a $(1 - \gamma)$-fraction of the $(k - 1)$-tuples in $G[V'_1, \ldots, V'_{k-1}]$ is contained in*

$$\left(1 \pm \gamma\right)|V'_k| \prod_{i=2}^{k-1} d_i^{\binom{k-1}{i-1}}$$

*copies of the $k$-vertex complete $(k - 1)$-complex in $G'$. Furthermore, the $\gamma$-fraction of $(k - 1)$-tuples in most copies of the $k$-vertex complete $(k - 1)$-complex in $G'$ contain in total at most*

$$\tfrac{5}{2}\gamma \prod_{i=1}^{k} |V'_i| \prod_{i=2}^{k} d_i^{\binom{k}{i-1}}$$

*copies of the $k$-vertex complete $(k - 1)$-complex in $G'$. Similarly, at least a $(1 - \gamma)$-fraction of the $(k - 2)$-tuples in $G[V'_1, \ldots, V'_{k-2}]$ is contained in*

$$\left(1 \pm \gamma\right)|V'_{k-1}| \prod_{i=2}^{k-1} d_i^{\binom{k-2}{i-1}}$$

*copies of the $(k - 1)$-vertex complete $(k - 1)$-complex in $G'$ together with a vertex from $V'_{k-1}$.*

The two supplements guarantee for $k = 3$ that few edges contained in most triangles can not be responsible for a large fraction of the total number of triangles and that most vertices are contained in the right number of edges. These statements will be used repeatedly to shorten the arguments.

The second lemma looks a bit more complicated, but we only need the variant with all parts of the same size. For $k = 3$ this implies that if many vertices have high degree into two different 2-cells, then they will also support many triangles.

**Lemma 2.9** (Minimum Degree Lemma). *For all integers $k \geqslant 3$ and constants $\gamma, \delta, d_0 > 0$, there exists $\varepsilon > 0$ such that the following holds. Let $\mathbf{d} = (d_{k-1}, \ldots, d_2)$ be a vector of real numbers with $d_i \geqslant d_0$ for each $2 \leqslant i \leqslant k - 1$, and let $G$ be a $k$-partite $(k - 1)$-complex with parts of size $m \geqslant \varepsilon^{-1}$ which is $(\mathbf{d}, \varepsilon, \varepsilon, 1)$-regular. Moreover, with integers $a, b, c$ such that $a + b - c = k$, let $A$ be part of an $a$-cell, $B$ be part of a $b$-cell, and $C$ be part of a $c$-cell such that the tuples from $C$ have degree $(\delta \pm \gamma)m^{a-c} \prod_{i=2}^{k-1} d_i^{\binom{a}{i} - \binom{c}{i}}$ into $A$. Suppose that every edge of $B$ contains an edge of $C$, and that $|B| \geqslant \gamma m^b \prod_{i=2}^{k-1} d_i^{\binom{b}{i}}$. Then there are*

$$|B|(\delta \pm 2\gamma)m^{a-c} \prod_{i=2}^{k-1} d_i^{\binom{k}{i} - \binom{b}{i}}$$

*copies of the $k$-vertex complete $(k - 1)$-complex in $G$ supported by $A$ and $B$.*

All of these lemmas are broadly standard, and hence we prove them in the appendix.

## 2.4. Properties of the random hypergraph

We use the following standard versions of the Chernoff bound.

**Theorem 2.10.** *Let $X$ be a random variable with distribution $\mathrm{Bin}(n,p)$. Then for any $\varepsilon > 0$ we have*

$$\Pr(X \geqslant pn + \varepsilon n) \leqslant \exp(-D(p+\varepsilon\|p)n) \quad \text{and} \quad \Pr(X \leqslant pn - \varepsilon n) \leqslant \exp(-D(p-\varepsilon\|p)n),$$

*where $D(x\|y)$ is the Kullback–Leibler divergence. From this it follows*

$$\Pr(|X - pn| > \varepsilon pn) < 2\exp\left(-\tfrac{\varepsilon^2 pn}{3}\right) \quad \text{for any } \varepsilon \leqslant \tfrac{3}{2}$$

*and if $t \geqslant 6pn$ we have*

$$Pr(X \geqslant pn + t) < \exp(-t).$$

**Lemma 2.11.** *Given $\eta > 0$, $k \in \mathbb{N}$ there exists $C$ such that if $p \geqslant \frac{C}{n}$, then $\Gamma = G^{(k)}(n,p)$, and all its subgraphs, are a.a.s. $(\eta, p)$-upper regular.*

*Proof.* Observe that if $\Gamma = G^{(k)}(n,p)$ is $(\eta, p)$-upper-regular, then automatically all its subgraphs are also. We assume without loss of generality that $\eta < 1$, and set $C = 18k\eta^{-3}$.

Given any $E_1, \ldots, E_k \subseteq \binom{[n]}{k}$, we aim to estimate the probability that $E_1, \ldots, E_k$ witness the failure of $G^{(k)}(n,p)$ to be $(\eta, p)$-upper regular. The expected number of edges of $G^{(k)}(n,p)$ which appear on the sets $K_k(E_1, \ldots, E_k)$ is $p|K_k(E_1, \ldots, E_k)|$, and the distribution is binomial, so we may apply the Chernoff bound.

If $|K_k(E_1, \ldots, E_k)| \leqslant \tfrac{1}{6}\eta n^k$, then failure to be $(\eta, p)$-upper regular means that the number of $k$-edges appearing on $K_k(E_1, \ldots, E_k)$ is at least seven times the expected number; by the Chernoff bound the probability of this event is less than $\exp(-p\eta n^k) < \exp(-kn^{k-1})$.

If $|K_k(E_1, \ldots, E_k)| \geqslant \tfrac{1}{6}\eta n^k$, then the probability that more than $(1+\eta)p|K_k(E_1, \ldots, E_k)|$ edges appear is at most

$$\exp\left(-\frac{\eta^2 p|K_k(E_1, \ldots, E_k)|}{3}\right) \leqslant \exp\left(-\frac{\eta^2 \frac{C}{n}\eta n^k}{18}\right) = \exp\left(-kn^{k-1}\right).$$

Since there are at most $2^{\binom{n}{k-1}}$ choices for each $E_i$, by the union bound the probability that $G^{(k)}(n,p)$ is not $(\eta, p)$-upper regular is at most

$$2^{k\binom{n}{k-1}} \exp\left(-kn^{k-1}\right)$$

which tends to zero as $n$ tends to infinity. □

Given a set $S \subseteq V(\Gamma)$, we say a $(k-1)$-set $x$ is $(\varepsilon, p, 1)$-*good for* $S$ if we have

$$\left|\left\{s \in S : x \cup \{s\} \in E(\Gamma)\right\}\right| = p|S| \pm \varepsilon pn.$$

For each $\ell \geqslant 2$, we say inductively that a $(k-1)$-set $x$ is $(\varepsilon, p, \ell)$-*good for* $S$ if it is $(\varepsilon, p, \ell-1)$-good for $S$ and there are at most $\varepsilon pn$ edges of $\Gamma$ which contain $x$ and in addition contain a set which is not $(\varepsilon, p, \ell-1)$-good for $S$.

**Lemma 2.12.** *Given $\varepsilon > 0$, $k \in \mathbb{N}$ there exists $C$ such that if $p \geqslant \frac{C \log n}{n}$, then $\Gamma = G^{(k)}(n, p)$ a.a.s. has the following property. For each set $S \subseteq V(\Gamma)$, and each $1 \leqslant \ell \leqslant \frac{1}{C} \log \log n$, there are at most $o(n)$ $(k-1)$-sets in $V(\Gamma) \setminus S$ which are not $(\varepsilon, p, \ell)$-good for $S$.*

*Proof.* Given $S$, we first estimate the number of $(k-1)$-sets $x$ which are outside $S$ and not $(\varepsilon, p, 1)$-good for $S$.

If $|S| < \frac{1}{6}\varepsilon n$, then failure of a given $x$ to be $(\varepsilon, p, 1)$-good for $S$ means $x$ forms an edge with at least $7p|S|$ vertices in $S$, the probability of which is by the Chernoff bound at most $\exp(-\varepsilon pn)$, which for large enough $C$ is smaller than $n^{-k}$. If on the other hand $|S| \geqslant \frac{1}{6}\varepsilon n$, then the probability that $x$ does not form an edge with $(1 \pm \varepsilon)p|S|$ vertices of $S$ is at most $2 \exp\left(\frac{-\varepsilon^2 p|S|}{3}\right) < n^{-k}$ for large enough $C$. We see that in either case, the probability that $x$ is not $(\varepsilon, p, 1)$-good for $S$ is at most $n^{-k}$. Now if $x$ and $x'$ are two different $(k-1)$-sets outside $S$, then the events of $x$ and of $x'$ being not $(\varepsilon, p, 1)$-good for $S$ are independent, so again using the Chernoff bound we can estimate the likelihood of many sets being bad for $S$. The expected number of bad sets for $S$ is at most $n^{k-1} \cdot n^{-k} = n^{-1}$. Therefore, for any $t \geqslant 1$, we can bound the probability that there are $t$ or more bad $(k-1)$-sets for $S$ by

$$\exp\left(-D(n^{-k} + tn^{1-k}||n^{-k})n^{k-1}\right) \leqslant \exp\left(-\frac{t \log n}{2}\right).$$

In particular, taking $t = 4n/\log n$ and using the union bound, the probability that there exists a set $S$ for which more than $4n/\log n$ $(k-1)$-sets are not $(\varepsilon, p, 1)$-good is at most $2^{-n}$. Suppose that $\Gamma$ is such that this good event occurs, and in addition that every $(k-1)$-set of vertices of $\Gamma$ is contained in at most $2pn$ edges of $\Gamma$.

Let $K = 2k\varepsilon^{-1}$. Now given $S$ and $\ell \geqslant 1$, we claim that the number of $(k-1)$-sets outside $S$ which are not $(\varepsilon, p, \ell)$-good for $S$ is at most $4n \cdot K^{\ell-1}/\log n$. We prove this by induction on $\ell$; the base case $\ell = 1$ is the assumption on $\Gamma$. Suppose $\ell \geqslant 2$, and that the number of $(k-1)$-sets outside $S$ which are not $(\varepsilon, p, \ell-1)$-good for $S$ is at most $4n \cdot K^{\ell-2}/\log n$. For each $(k-1)$-set $x$ outside $S$ which is not $(\varepsilon, p, \ell-1)$-good for $S$, we assign to each $(k-1)$-set $y$ such that $x \cup y$ is an edge of $\Gamma$ one unit of badness. Observe that the total number of units of badness assigned is at most $(k-1) \cdot 2pn \cdot 4n \cdot K^{\ell-2}/\log n$. On the other hand, a set $y$ which is $(\varepsilon, p, \ell-1)$-good for $S$ can only fail to be $(\varepsilon, p, \ell)$-good for $S$ if it is assigned at least $\varepsilon pn$ units of badness. It follows that the total number of such sets is at most $2(k-1)\varepsilon^{-1} \cdot 4nK^{\ell-2}/\log n$, and so the number of $(k-1)$-sets outside $S$ which are not $(\varepsilon, p, \ell)$-good for $S$ is at most

$$2(k-1)\varepsilon^{-1} \cdot 4nK^{\ell-2}/\log n + 4n \cdot K^{\ell-2}/\log n \leqslant 4n \cdot K^{\ell-1}/\log n,$$

as desired. In particular, this formula is in $o(n)$ for all $1 \leqslant \ell \leqslant \frac{1}{C} \log \log n$ with $C$ large enough. $\qquad\square$

For a given $(k-1)$-tuple, we will find many paths starting from there. To get expansion we need to ensure that they have many different end-tuples.

**Lemma 2.13.** *For $\gamma > 0$, $k \geqslant 3$, any fixed integer $\ell > \frac{k-1}{\gamma} + k - 1$, and any $\mu > 0$ a.a.s. in $\Gamma = G^{(k)}(n, p)$ with $p = n^{-1+\gamma}$ the following holds. For any $(k-1)$-tuple $\mathbf{x}$ in $V(\Gamma)$ and a set $\mathcal{P}$ of at least $(\mu p n)^\ell$ tight paths in $\Gamma$ with $\ell + (k-1)$ vertices and rooted at $\mathbf{x}$, the number of end $(k-1)$-tuples of the paths in $\mathcal{P}$ is at least $\frac{\mu^{2\ell}}{8(2\ell)!} n^{k-1}$. Moreover, when $(k-1)|\ell$, the same holds for spike paths rooted at $\mathbf{x}$.*

To prove Lemma 2.13 we need a concentration result of Kim and Vu [KV00]. We first give some definitions and then state the result. Let $m$ be a positive integer and $H$ be a hypergraph with $|V(H)| = m$ and each edge has at most $r$ vertices. Let $p \in [0, 1]$ and let $X_i, i \in V(H)$ be independent random variables with $\mathbb{P}[X_i = 1] = p$ and $\mathbb{P}[X_i = 0] = 1 - p$. We define the random variable

$$Y_H = \sum_{f \in E(H)} \prod_{i \in f} X_i.$$

For each subset $A \subseteq V(H)$, we define the $A$-truncated subgraph $H(A)$ of $H$ to be the subgraph of $H$ with $V(H(A)) = V(H) \setminus A$ and $E(H(A)) = \{f \subseteq V(H(A)) : f \cup A \in E(H)\}$. Hence

$$Y_{H(A)} = \sum_{\substack{f \in E(H) \\ A \subseteq f}} \prod_{i \in f \setminus A} X_i.$$

Now, for $0 \leqslant i \leqslant r$, we set $\mathcal{E}_i(H) = \max_{A \subseteq V(H), |A| = i} \mathbb{E}[Y_{H(A)}]$. Note that $\mathcal{E}_0(H) = \mathbb{E}[Y_H]$. Finally, we let $\mathcal{E}(H) = \max_{0 \leqslant i \leqslant r} \mathcal{E}_i(H)$ and $\mathcal{E}'(H) = \max_{1 \leqslant i \leqslant r} \mathcal{E}_i(H)$.

**Theorem 2.14** (Kim–Vu polynomial concentration [KV00]). *In this setting we have*

$$\mathbb{P}[|Y_H - \mathbb{E}(Y_H)| > a_r (\mathcal{E}(H) \mathcal{E}'(H))^{1/2} \lambda^r)] = O(\exp(-\lambda + (r-1) \log m))$$

*for any $\lambda > 1$ and $a_r = 8^r r!^{1/2}$.*

Moreover, we will need the following definitions. Let $k \geqslant 3$, $\ell \geqslant k - 1$, and $\gamma > 0$. Roughly speaking, we define $D_\ell$ to be the $k$-graph obtained from two vertex-disjoint tight paths on $\ell + k - 1$ vertices by identifying the end $(k-1)$-tuples and let $\mathcal{D}_\ell$ be the set of hypergraphs obtained from $D_\ell$ by additionally identifying some (or none) of the not yet identified vertices from the first tight path with such vertices from the second without completely collapsing it into a tight path. More precisely, we let $U = \{u_1, \ldots, u_{\ell+k-1}\}$ and $W = \{w_1, \ldots, w_{\ell+k-1}\}$ be two sets of vertices that are disjoint except that $\mathbf{x} = (u_1, \ldots, u_{k-1}) = (w_1, \ldots, w_{k-1})$ and $\mathbf{y} = (u_{\ell+k-1}, \ldots, u_{\ell+1}) = (w_{\ell+k-1}, \ldots, w_{\ell+1})$. Then $D_\ell$ is the hypergraph with vertex set $U \cup W$ and edge set

$$\{\{u_i, \ldots, u_{i+k-1}\} : i \in [\ell]\} \cup \{\{w_i, \ldots, w_{i+k-1}\} : i \in [\ell]\}.$$

For $0 \leqslant j \leqslant \ell - (k-1)$, we denote by $\mathcal{D}_\ell^j$ the graphs obtained from $D_\ell$ by taking sets $I_1, I_2 \subseteq \{k, \ldots, \ell\}$ each of size $j$ and a bijection $\sigma : I_1 \to I_2$ and identifying $u_i$ with $w_{\sigma(i)}$ for all $i \in I_1$, where, if $j = \ell - (k-1)$, then we do not allow $\sigma$ to be the identity (since that would collapse $D_\ell$ into a tight path). We say that such a graph $F \in \mathcal{D}_\ell^j$ is rooted at $\mathbf{x}$ and call

the vertices in $\mathbf{y}$ the end-vertices of $F$. Moreover, for any edge $\{u_i, \ldots, u_{i+k-1}\}$ in $F \in \mathcal{D}_\ell^j$ as above we call $u_i$ the *first vertex* and $u_{i+k-1}$ the *last vertex* of the edge and say that the edge *starts* in $u_i$ and *ends* in $u_{i+k-1}$ (and analogously for edges contained in $W$). Finally, we let $\mathcal{D}_\ell = \bigcup_{0 \leqslant j \leqslant \ell - (k-1)} \mathcal{D}_\ell^j$.

We now prove the following lemma which we will use to prove Lemma 2.13.

**Lemma 2.15.** *For $\gamma > 0$, $k \geqslant 3$, and any fixed integer $\ell > \frac{k-1}{\gamma} + k - 1$ a.a.s. in $\Gamma = G^{(k)}(n, p)$ with $p = n^{-1+\gamma}$ the following holds. For all $(k-1)$-tuples $\mathbf{x}$ in $V(\Gamma)$, the number of copies of elements of $\mathcal{D}_\ell$ in $\Gamma$ that are rooted at $\mathbf{x}$ is at most $2p^{2\ell} n^{2\ell - (k-1)}$.*

*Proof.* Fix a $(k-1)$-tuple $\mathbf{x}$ in $V(\Gamma)$ and an integer $\ell > \frac{k-1}{\gamma} + k - 1$ and let $n$ be large enough. We call the vertices in $\mathbf{x}$ *rooted* and any other vertices *unrooted*. Let $F \in \mathcal{D}_\ell$ and consider the complete $k$-graph $K_n^{(k)}$ on $n$ vertices. We define a hypergraph $H_F$ as follows. Let $V(H_F) = E(K_n^{(k)})$ and let

$$E(H_F) = \left\{ \mathcal{F} \in \binom{E(K_n^{(k)})}{e(F)} : \mathcal{F} \text{ spans a copy of } F \text{ in } K_n^{(k)} \text{ rooted at } \mathbf{x} \right\}.$$

Note that, since $e(F) \leqslant 2\ell$, each edge in $H_F$ has size at most $2\ell$. For each $e \in V(H_F)$, let $X_e$ be the random variable for which $X_e = 1$ if $e$ is an edge of $\Gamma$ and $X_e = 0$ otherwise. Note that $\mathbb{P}[X_e = 1] = p$. It is easy to see that with these definitions $Y_{H_F}$ is the number of copies of $F$ in $\Gamma$ rooted at $\mathbf{x}$. Since $e(D_\ell) = 2\ell$, $v(D_\ell) = 2\ell$, and $k - 1$ vertices are rooted, we have $\binom{n}{2\ell - (k-1)} p^{2\ell} \leqslant \mathbb{E}[Y_{H_{D_\ell}}] \leqslant p^{2\ell} n^{2\ell - (k-1)}$, in particular $\mathbb{E}[Y_{H_{D_\ell}}] = \Theta(p^{2\ell} n^{2\ell - (k-1)})$.

**Claim 2.16.** *For $F \in \mathcal{D}_\ell \setminus \{D_\ell\}$, we have*

$$\mathbb{E}\left[Y_{H_F}\right] = o\left(\mathbb{E}\left[Y_{H_{D_\ell}}\right]\right).$$

*Proof of Claim.* We split the proof into two cases depending on the integer $j$ for which we have $F \in \mathcal{D}_\ell^j$.

First suppose that $F \in \mathcal{D}_\ell^j$ for some $j \in \{1, \ldots, \ell - 2(k-1)\}$. Note that $v(F) = 2\ell - j$. We claim that $e(F) \geqslant 2\ell - j$. This can be seen as follows. Recall that $F$ is obtained from $D_\ell$ by identifying $j$ additional vertices from the first tight path in $D_\ell$ with vertices from the second. This leaves $\ell - (k-1) - j \geqslant k - 1$ unidentified vertices in the first tight path. In addition to the $\ell$ edges in the second path, $F$ contains an edge ending in each of the unidentified vertices and one more additional edge starting with each of the last $k - 1$ unidentified vertices (these edges cannot end in an unidentified vertex, so there is no double counting). Thus

$$e(F) \geqslant \ell + \ell - (k-1) - j + (k-1) = 2\ell - j.$$

Hence, since $k - 1$ vertices are rooted,

$$\mathbb{E}[Y_{H_F}] \leqslant p^{2\ell - j} n^{2\ell - j - (k-1)} = p^{2\ell} n^{2\ell - (k-1)} (pn)^{-j} = p^{2\ell} n^{2\ell - (k-1)} n^{-j\gamma} = o\left(\mathbb{E}\left[Y_{H_{D_\ell}}\right]\right).$$

Now suppose that $F \in \mathcal{D}_\ell^j$ for some $j \in \{\ell - 2(k-1) + 1, \ldots, \ell - (k-1)\}$. As in the previous case, in addition to the $\ell$ edges in the second path, $F$ contains an edge ending in

each of the $\ell - (k-1) - j$ unidentified vertices. Thus $e(F) \geqslant 2\ell - (k-1) - j$. Hence, since $v(F) = 2\ell - j$, $k-1$ vertices are rooted, and $j > \ell - 2(k-1)$, we have

$$
\begin{aligned}
\mathbb{E}[Y_{H_F}] &\leqslant p^{2\ell-j-(k-1)}n^{2\ell-j-(k-1)} = p^{2\ell}n^{2\ell-(k-1)}(pn)^{-j}p^{-(k-1)} \\
&= p^{2\ell}n^{2\ell-(k-1)}n^{-j\gamma+(k-1)-\gamma(k-1)} < p^{2\ell}n^{2\ell-(k-1)}n^{-\gamma(\ell-2(k-1))+(k-1)-\gamma(k-1)} \\
&= p^{2\ell}n^{2\ell-(k-1)}n^{-\gamma(\ell-(k-1))+(k-1)} = o\left(\mathbb{E}\left[Y_{H_{D_\ell}}\right]\right),
\end{aligned}
$$

since $\ell > \frac{k-1}{\gamma} + k - 1$. $\qquad\qquad\square$

Combining the claim with our bound on $\mathbb{E}[Y_{H_{D_\ell}}]$ we obtain

$$
\sum_{F \in \mathcal{D}_\ell} \mathbb{E}[Y_{H_F}] \leqslant \frac{3}{2}p^{2\ell}n^{2\ell-(k-1)}. \tag{2.1}
$$

Next we show that, for all $F \in \mathcal{D}_\ell$, the random variable $Y_{H_F}$ is concentrated around its expectation.

**Claim 2.17.** *For all $F \in \mathcal{D}_\ell$, we have*

$$
\mathbb{P}\left[|Y_{H_F} - \mathbb{E}[Y_{H_F}]| > \frac{p^{2\ell}n^{2\ell-(k-1)}}{2|\mathcal{D}_\ell|}\right] = O\left(\exp(-n^{\gamma/(6\ell)})\right).
$$

*Proof of Claim.* Let $F \in \mathcal{D}_\ell$. We first show that $\mathcal{E}'(H_F) \leqslant n^{2\ell-(k-1)}p^{2\ell}n^{-\gamma}$. To that end let $1 \leqslant i \leqslant 2\ell$ and $A \subseteq V(H_F) = E(K_n^{(k)})$ with $|A| = i$. Note that it suffices to show that $\mathbb{E}[Y_{H_F(A)}] \leqslant n^{2\ell-(k-1)}p^{2\ell}n^{-\gamma}$. We let $v_A$ be the number of vertices covered by $A$, that is, $v_A = |\bigcup A|$ and we let $r_A$ be the number of rooted vertices not covered by $A$, that is, $r_A = |\mathbf{x} \setminus \bigcup A|$. Moreover, we call edges in $A$ *covered edges*, edges not in $A$ *uncovered edges*, vertices in $\bigcup A$ *covered vertices* and vertices not in $\bigcup A$ *uncovered vertices*. Note that $Y_{H_F(A)}$ is the number of copies of $F$ in $\Gamma + A$ that are rooted at $\mathbf{x}$ and contain $A$. So if $A$ is not contained in (the edge set of) any copy of $F$ rooted at $\mathbf{x}$, then $Y_{H_F(A)} = 0$ and we are done. Now assume that $A$ is contained in a copy of $F$ rooted at $\mathbf{x}$. We consider two cases.

First suppose that the number of uncovered unrooted vertices in any copy of $F$ rooted at $\mathbf{x}$ and containing $A$ is at most $\ell + k - 2$, that is, $v(F) - v_A - r_A \leqslant \ell + k - 2$. Note that there are at least $v(F) - v_A - r_A$ uncovered edges, that is, $e(F) - i \geqslant v(F) - v_A - r_A$ as for each uncovered unrooted vertex, there is at least one uncovered edge ending in that vertex. Thus

$$
\begin{aligned}
\mathbb{E}[Y_{H_F(A)}] &\leqslant n^{v(F)-v_A-r_A}p^{e(F)-i} \leqslant (np)^{v(F)-v_A-r_A} \leqslant n^{\gamma(\ell+k-2)} = n^{\gamma\ell+\gamma(k-2)} \\
&\leqslant n^{2\gamma\ell-(k-1)-\gamma(k-1)+\gamma(k-2)} = n^{2\gamma\ell-(k-1)-\gamma} = n^{2\ell-(k-1)}p^{2\ell}n^{-\gamma},
\end{aligned}
$$

where the last inequality follows from the fact that $\ell > \frac{k-1}{\gamma} + k - 1$.

Now we consider the case where the number of uncovered unrooted vertices in any copy of $F$ rooted at $\mathbf{x}$ and containing $A$ is at least $\ell + k - 1$, that is, $v(F) - v_A - r_A \geqslant \ell + k - 1$. Consider such a copy of $F$. Recall that $F$ is obtained from two tight paths by identifying vertices. Note that in this case there are at least $k - 1$ vertices in each of the tight paths that are unrooted,

uncovered, not on the other tight path, and not part of the $k-1$ end-vertices of $F$. Let $B$ be the set of the last $k-1$ such vertices on the first tight path. We now show that $F$ has at least $v(F) - v_A - r_A + k - 1$ uncovered edges, that is, $e(F) - i \geqslant v(F) - v_A - r_A + k - 1$. Note that for each uncovered unrooted vertex there is at least one uncovered edge ending in that vertex. That gives us $v(F) - v_A - r_A$ uncovered edges. We show that for each vertex in $B$ there is at least one more uncovered edge that we have not yet counted. For each vertex in $u \in B$ consider the uncoverd edge starting in $u$. If it ends in a covered vertex, then it is one more uncovered edge that we have not yet counted. If it ends in an uncovered vertex $w$, then the corresponding edge on the other tight path is also uncovered. Thus there are two uncovered edges ending in $w$ and we have only counted one of them. Since there are $k-1$ vertices in $B$, it follows that there are at least $v(F) - v_A - r_A + k - 1$ uncovered edges in $F$. It follows that for any copy of $F$ containing $A$, we have

$$2\ell - 1 \geqslant e(F) - i \geqslant v(F) - v_A - r_A + k - 1.$$

Hence

$$\mathbb{E}[Y_{H_F(A)}] \leqslant n^{v(F)-v_A-r_A} p^{e(F)-i} \leqslant n^{e(F)-i-(k-1)} p^{e(F)-i}$$
$$\leqslant (np)^{e(F)-i} n^{-(k-1)} \leqslant (np)^{2\ell-1} n^{-(k-1)} = n^{2\ell-(k-1)} p^{2\ell} n^{-\gamma}.$$

We have shown that $\mathcal{E}'(H_F) \leqslant n^{2\ell-(k-1)} p^{2\ell} n^{-\gamma}$. Now note that, since $\mathbb{E}[Y_{H_F}] \leqslant \mathbb{E}[Y_{H_{D_\ell}}] \leqslant n^{2\ell-(k-1)} p^{2\ell}$, we have $\mathcal{E}(H_F) = \max\{\mathcal{E}'(H_F), \mathbb{E}[Y_{H_F}]\} \leqslant n^{2\ell-(k-1)} p^{2\ell}$. It follows that

$$(\mathcal{E}(H_F)\mathcal{E}'(H_F))^{1/2} \leqslant n^{2\ell-(k-1)} p^{2\ell} n^{-\gamma/2}.$$

Therefore, with

$$\lambda_F = \left(\frac{p^{2\ell} n^{2\ell-(k-1)}}{2|\mathcal{D}_\ell| a_{2\ell} (\mathcal{E}(H_F)\mathcal{E}'(H_F))^{1/2}}\right)^{1/(2\ell)} \geqslant \left(\frac{n^{\gamma/2}}{2|\mathcal{D}_\ell| a_{2\ell}}\right)^{1/(2\ell)} \geqslant n^{\gamma/(5\ell)},$$

we have, by Theorem 2.14,

$$\mathbb{P}\left[|Y_{H_F} - \mathbb{E}[Y_{H_F}]| > \frac{p^{2\ell} n^{2\ell-(k-1)}}{2|\mathcal{D}_\ell|}\right] = \mathbb{P}[|Y_{H_F} - \mathbb{E}[Y_{H_F}]| > a_{2\ell}(\mathcal{E}(H_F)\mathcal{E}'(H_F))^{1/2} \lambda_F^{2\ell}]$$
$$= O\left(\exp\left(-\lambda_F + (2\ell-1)\log\binom{n}{k}\right)\right)$$
$$= O\left(\exp(-n^{\gamma/(5\ell)} + (2\ell-1)k\log n)\right)$$
$$= O\left(\exp(-n^{\gamma/(6\ell)})\right).$$

$\square$

Now let $Z_{\mathbf{x}}$ be the number of copies of elements of $\mathcal{D}_\ell$ in $\Gamma$ rooted at $\mathbf{x}$. Note

that $Z_{\mathbf{x}} = \sum_{F \in \mathcal{D}_\ell} Y_{H_F}$. We have

$$\mathbb{P}[Z_{\mathbf{x}} > 2p^{2\ell}n^{2\ell-(k-1)}] \leqslant \mathbb{P}\left[\left|\sum_{F \in \mathcal{D}_\ell}(Y_{H_F} - \mathbb{E}[Y_{H_F}])\right| > 2p^{2\ell}n^{2\ell-(k-1)} - \sum_{F \in \mathcal{D}_\ell}\mathbb{E}[Y_{H_F}]\right]$$

$$\overset{(2.1)}{\leqslant} \mathbb{P}\left[\sum_{F \in \mathcal{D}_\ell}|Y_{H_F} - \mathbb{E}[Y_{H_F}]| > \frac{1}{2}p^{2\ell}n^{2\ell-(k-1)}\right]$$

$$\leqslant \sum_{F \in \mathcal{D}_\ell}\mathbb{P}\left[|Y_{H_F} - \mathbb{E}[Y_{H_F}]| > \frac{p^{2\ell}n^{2\ell-(k-1)}}{2|\mathcal{D}_\ell|}\right]$$

$$= O\left(|\mathcal{D}_\ell|\exp(n^{-\gamma/(6\ell)})\right),$$

where the last inequality follows from the union bound together with the fact that if the average of the values $|Y_{H_F} - \mathbb{E}[Y_{H_F}]|$ is at least $\frac{p^{2\ell}n^{2\ell-(k-1)}}{2|\mathcal{D}_\ell|}$, then at least one of the values has to be that large. Finally, the result follows by the union bound over all $(k-1)$-tuples $\mathbf{x}$ in $V(\Gamma)$. $\qquad\square$

We are now ready to prove Lemma 2.13.

*Proof of Lemma 2.13.* Let $\mathbf{x}$ be a $(k-1)$-tuple in $V(\Gamma)$ and $\mathcal{P}$ be a set of at least $(\mu p n)^\ell$ tight paths in $\Gamma$ with $\ell + (k-1)$ vertices and rooted at $\mathbf{x}$. Let $Q$ be the set of end-tuples we reach from $\mathbf{x}$ with paths in $\mathcal{P}$. For each $\mathbf{q} \in Q$, let $\mathcal{P}_{\mathbf{q}}$ be those paths in $\mathcal{P}$ that end in $\mathbf{q}$. Note that, for $\mathbf{q} \in Q$ and distinct elements $P, P' \in \mathcal{P}_{\mathbf{q}}$, we have $P \cup P' \in \mathcal{D}_\ell$. Thus for each $\mathbf{q} \in Q$, there are at least $\frac{1}{(2\ell)!}\binom{|\mathcal{P}_{\mathbf{q}}|}{2}$ copies of elements of $\mathcal{D}_\ell$ in $\Gamma$ rooted at $\mathbf{x}$ and ending in $\mathbf{q}$ (we divide by $(2\ell)!$ since there are at most $(2\ell)!$ ways the union of two paths in $\mathcal{P}$ could result in the same copy of an element of $\mathcal{D}_\ell$). Hence the number of copies of elements of $\mathcal{D}_\ell$ in $\Gamma$ rooted at $\mathbf{x}$ is at least

$$\sum_{\mathbf{q} \in Q}\frac{1}{(2\ell)!}\binom{|\mathcal{P}_{\mathbf{q}}|}{2} \geqslant \frac{|Q|}{(2\ell)!}\binom{\frac{(\mu p n)^\ell}{|Q|}}{2} \geqslant \frac{(\mu p n)^{2\ell}}{4(2\ell)!|Q|},$$

where the penultimate inequality follows by Jensen's inequality. Thus, by Lemma 2.15, we have a.a.s.

$$2p^{2\ell}n^{2\ell-(k-1)} \geqslant \frac{(\mu p n)^{2\ell}}{4(2\ell)!|Q|}$$

and thus

$$|Q| \geqslant \frac{\mu^{2\ell}}{8(2\ell)!}n^{k-1}.$$

Moreover, an analogous argument shows the result for spike paths. $\qquad\square$

Together with the definition of tuples that are $(\varepsilon, p, \ell)$-good for $S$, Lemma 2.13 implies the following.

**Corollary 2.18.** *For any $\gamma > 0$ and any $0 < \varepsilon \leqslant \frac{1}{4}\gamma$, and integers $s$, $k \geqslant 3$, and $\ell > \frac{k-1}{\gamma} + k - 1$, there exists $\nu > 0$ such that in $\Gamma = G^{(k)}(n, p)$ a.a.s. the following holds when $p = n^{-1+\gamma}$. Let $G \subseteq \Gamma$ satisfy $\delta_{k-1}(G) \geqslant \left(\frac{1}{2} + \gamma\right)pn$. Let $S, S' \subseteq V(\Gamma)$ be sets with $|S| \leqslant \frac{1}{2}n$ and $|S'| \leqslant s$. Let $\mathbf{x}$ be a $(k-1)$-tuple, which is $(\varepsilon, p, \ell)$-good for $S$. Then there are at least $\nu n^{k-1}$ different $(k-1)$-tuples $\mathbf{y}$, such that there exists a tight path in $G$ of length $\ell$ with ends $\mathbf{x}$ and $\mathbf{y}$ and no vertices of the path in $S \cup S'$ except for possibly some of the vertices in $\mathbf{x}$. Moreover, when $(k-1)|\ell$, the same holds for spike paths in $G$ of length $\ell$.*

*Proof.* We only prove the statement for tight paths as it is easy to see that the proof can be adapted for spike paths. We set $\mu = \frac{1}{4}\gamma$, and $\nu = \frac{\mu^{2\ell}}{8(2\ell)!}$. Suppose that the good event of Lemma 2.13, with input $\gamma$, $k$, and $\mu$, holds for $\Gamma = G^{(k)}(n, p)$.

Given $G$ and $\mathbf{x}$ as in the lemma statement, let $\mathbf{x} = (x_1, \dots, x_{k-1})$. We construct tight paths $x_1 \dots x_{\ell+k-1}$ rooted at $\mathbf{x}$ by choosing vertices $x_k, \dots, x_{\ell+k-1}$ one by one as follows. For each $k \leqslant i \leqslant \ell + k - 1$, we choose $x_i$ such that $x_i \notin S \cup S' \cup \{x_1, \dots, x_{i-1}\}$ and $\{x_{i-k+1}, \dots, x_i\} \in E(G)$. If $i < \ell + k - 1$, we insist in addition that $\{x_{i-k+2}, \dots, x_i\}$ is $(\varepsilon, p, \ell - (i - k + 1))$-good for $S$. Since $\mathbf{x}$ is $(\varepsilon, p, \ell)$-good for $S$, for each $k \leqslant i \leqslant \ell + k - 1$, the number of choices for each $x_i$, such that $\{x_{i-k+1}, \dots, x_i\}$ is an edge of $G$, $x_i \notin S \cup S' \cup \{x_1, \dots, x_{i-1}\}$, and $\{x_{i-k+2}, \dots, x_i\}$ is $(\varepsilon, p, \ell - (i - k + 1))$-good, is at least

$$\left(\tfrac{1}{2} + \gamma\right)pn - (p\,|S| + \varepsilon pn) - s - \ell - (k-1) - \varepsilon pn$$
$$\geqslant (\gamma - 2\varepsilon)pn - s - \ell - (k-1) \geqslant \tfrac{1}{4}\gamma pn \,.$$

Let $\mathcal{P}$ be the set of tight paths constructed in this way; then we have $|\mathcal{P}| \geqslant \left(\frac{1}{4}\gamma pn\right)^{\ell} = (\mu pn)^{\ell}$. Since the good event of Lemma 2.13 holds, the number of end $(k-1)$-tuples of these paths is at least $\frac{\mu^{2\ell}}{8(2\ell)!}n^{k-1} = \nu n^{k-1}$, as desired. $\qquad\square$

## 2.5. Connecting lemma

The next lemma will enable us to connect two $(k-1)$-tuples, which are $(\varepsilon', p, \ell)$-good for some set $S$, by a path of length at most $\ell$ avoiding $S$.

**Lemma 2.19.** *Given $k \geqslant 3$, and $\gamma > 0$, there exists an integer $\ell$ such that for any integer $s$ the following holds. For any $d, \eta > 0$, any $0 < \varepsilon' \leqslant \frac{1}{4}\gamma$, any integer $t_0$, any small enough $\nu, \varepsilon_k > 0$, any functions $\varepsilon, f, f_k : \mathbb{N} \to (0, 1]$ which tend to zero sufficiently fast, and any large enough $t_1, t_2 \in \mathbb{N}$, the following holds a.a.s. in $\Gamma = G^{(k)}(n, p)$ with $p \geqslant n^{-1+\gamma}$. Suppose $G \subseteq \Gamma$ is an $n$-vertex $k$-graph with $\delta_{k-1}(G) \geqslant \left(\frac{1}{2} + \gamma\right)pn$, that $(\mathcal{P}_c^*, \mathcal{P}_f^*)$ is a $(t_0, t_1, t_2, \varepsilon_k, \varepsilon(t_1), f_k(t_1), f(t_2), p)$-strengthened pair for $G$, and that $t$ is the number of $1$-cells in $\mathcal{P}_c^*$. Let $\mathcal{R}' \subseteq \mathcal{R} = \mathcal{R}_{\varepsilon_k, d}(G; \mathcal{P}_c^*, \mathcal{P}_f^*)$ be an induced subcomplex of the $(\varepsilon_k, d)$-reduced multicomplex of $G$ on at least $(1 - \nu)t$ $1$-edges and assume that it is tightly linked. Further, let $S \subseteq V(G)$ be such that $|S| \leqslant \frac{1}{2}n$ and it intersects all $1$-cells of $\mathcal{R}'$ in at most an $(1 - \eta)$-fraction. Then for any two $(k-1)$-tuples $\mathbf{x}$ and $\mathbf{y}$, which are $(\varepsilon', p, \ell)$-good for $S$, and any set $S'$ of size at most $s$, there exists a tight path of length $\ell$ with ends $\mathbf{x}$ and $\mathbf{y}$.*

To prove this we use the following lemma, which allows us to connect a fraction of any good $(k-1)$-cell to a fraction of an adjacent good $(k-1)$-cell, where adjacency is with respect to regular polyads.

**Lemma 2.20.** *Given $k \geqslant 3$ and $\gamma > 0$, there exists an integer $\ell$ such that for any integer $s$ the following holds. For any $d, \eta, \nu > 0$, any $t_0 \in \mathbb{N}$, any small enough $\varepsilon_k > 0$, any functions $\varepsilon, f : \mathbb{N} \to (0, 1]$ which tend to zero sufficiently fast, any integers $t_2 \geqslant t_1 \geqslant t_0$, and any small enough $f_k > 0$, the following holds a.a.s. in $\Gamma = G^{(k)}(n, p)$ with $p \geqslant n^{-1+\gamma}$. Suppose $G \subseteq \Gamma$ is an $n$-vertex $k$-graph, that $(\mathcal{P}_c^*, \mathcal{P}_f^*)$ is a $(t_0, t_1, t_2, \varepsilon_k, \varepsilon(t_1), f_k, f(t_2), p)$-strengthened pair for $G$, let $H := \hat{P}(Q; \mathcal{P}_c^*)$ be a regular polyad in the reduced complex $\mathcal{R}_{\varepsilon_k, d}(G; \mathcal{P}_c^*, \mathcal{P}_f^*)$, $V_1, \dots, V_k$ its underlying $1$-cells, and $S \subseteq V(G)$ is a set intersecting each of these in at most an $(1 - \eta)$-fraction. Further, let $E_1$ and $E_k$ be the two $(k-1)$-cells of $H$ missing $V_1$ and $V_k$, respectively.*

*Let the tuples in $E_1$ and $E_k$ be ordered according to $V_1, \dots, V_k$. Then there is $\overline{E}_k \subseteq E_k$ with $|\overline{E}_k| \geqslant (1 - \nu)|E_k|$, such that for any $\mathbf{x} \in \overline{E}_k$ and any set $S'$ of at most $s$ vertices there is a tight path from $\mathbf{x}$ to $\mathbf{y}$ of length $\ell$ with internal vertices not in $S \cup S'$ for a $(1 - \nu)$-fraction of the tuples $\mathbf{y} \in E_1$.*

With this lemma and Corollary 2.18 it is straightforward to prove Lemma 2.19. We will prove both, Lemma 2.19 and 2.20, in Section 4.

## 2.6. Fractional matchings

While the clusters of a regular partition are all the same size, and are still about the same size after we remove the reservoir set, the reservoir path may intersect the clusters in very different amounts. When we extend the reservoir path to an almost-spanning path, this means we need to use different numbers of vertices in the different clusters. To guide the construction of the almost-spanning path, the following lemma returns a fractional matching in the cluster graph such that the total weight on each cluster is at most the fraction of vertices still to use in that cluster, and the total weight of the fractional matching is very close to $\frac{1}{k}$ times the fraction of vertices in total still to use.

**Lemma 2.21.** *Let $H$ be an $m$-vertex $k$-complex, and let $w : V(H) \to [0, 1]$ be a weight function. Given $\varepsilon > 0$, suppose that $H$ has at least $(1 - \varepsilon)m$ edges of size $1$, and that for each $1 \leqslant i \leqslant k - 2$, each $i$-edge of $H$ is contained in at least $(1 - \varepsilon)m$ edges of size $i + 1$. Finally suppose that each $(k - 1)$-edge of $H$ is contained in at least $\left(\frac{1}{2} + \gamma\right)m$ edges of size $k$, and suppose $\sum_{v \in V(H)} w(v) \geqslant (1 - \gamma)m$. Then there is a weight function $w^* : E\left(H^{(k)}\right) \to [0, 1]$ such that for each $v \in V(H)$ we have $\sum_{e \ni v} w^*(e) \leqslant w(v)$ and $\sum_{e \in E\left(H^{(k)}\right)} w^*(e) \geqslant \left(\sum_{v \in V(H)} w(v) - \varepsilon m\right) \cdot \frac{1}{k}$.*

*Proof.* Consider the linear program

$$\text{maximise} \sum_{e \in E(H^{(k)})} w^*(e) \text{ subject to } \sum_{e \ni v} w^*(e) \leqslant w(v) \text{ for each } v \in V(H) \text{ and } w^*(e) \geqslant 0 \,.$$

The dual program has variables $y : V(H) \to [0, 1]$ such that for each $e \in E\left(H^{(k)}\right)$ we have $\sum_{v \in e} y(v) \geqslant 1$, where we minimise $\sum_{v \in V(H)} y(v)w(v)$. Suppose that $y$ is a feasible solution to the dual program.

We order $V(H)$ according to decreasing $y$. We find a $k$-edge of $H$ as follows. We take the last $v_1$ such that $\{v_1\}$ is a 1-edge of $H$. Then for each $2 \leqslant i \leqslant k$ in succession, we choose the last vertex $v_i$ such that $\{v_1, \ldots, v_i\}$ is an $i$-edge of $H$.

For each $1 \leqslant i \leqslant k-1$, since by construction $\{v_1, \ldots, v_{i-1}\}$ is an $(i-1)$-edge of $H$, there are at most $\varepsilon m$ choices of $v_i$ which do not give an $i$-edge of $H$, and in particular $v_i$ will be at or after position $(1-\varepsilon)m$ in the order. Finally since $\{v_1, \ldots, v_{k-1}\}$ is a $(k-1)$-edge of $H$, necessarily $v_k$ will be at position at or after $\left(\frac{1}{2} + \gamma\right)m$ in the order.

Suppose that the vertex $v$ of $H$ at position $(1-\varepsilon)m$ in the order satisfies $y(v) = a$, and let $y(v_k) = b$. Then we have $(k-1)a + b \geqslant \sum_{i=1}^{k} y(v_i) \geqslant 1$, where the second inequality is since $y$ is feasible for the dual program. On the other hand, let $\alpha$ denote the sum of $w(u)$ over vertices $u$ equal to or earlier in the order than $v_k$, and let $\beta$ denote the sum of $w(u)$ over vertices $u$ after $v_k$ but not after $v$ (where $v$ is at position $(1-\varepsilon)m$ in the order). Then we have $\sum_{v \in V(H)} w(v)y(v) \geqslant \alpha b + \beta a$.

We view this as an optimisation problem: given $0 \leqslant a \leqslant b \leqslant 1$ such that $(k-1)a + b \geqslant 1$, minimise $\alpha b + \beta a$. Trivially we can assume the minimum occurs for $(k-1)a + b = 1$, and since $k \geqslant 2$ and $\alpha > \beta$, the unique minimum occurs when $a = b = \frac{1}{k}$.

Thus we have $\sum_{v \in V(H)} w(v)y(v) \geqslant (\alpha + \beta) \cdot \frac{1}{k}$ for any feasible solution $y$ to the dual program, so the value of the dual program is at least $(\alpha + \beta) \cdot \frac{1}{k}$. By the Duality Theorem for linear programming, the value of the primal program is the same. Finally since $w(v) \in [0, 1]$ we have $\sum_{v \in V(H)} w(v) \leqslant \alpha + \beta + \varepsilon m$, and the lemma follows. $\qquad \square$

## 2.7. Reservoir path

**Definition 2.22** (Reservoir path). A *reservoir path* $P_{\mathrm{res}}$ with a *reservoir set* $R \subsetneq V(P_{\mathrm{res}})$ is a $k$-uniform hypergraph with two $(k-1)$-tuples $\mathbf{v}$ and $\mathbf{w}$, such that for *any* $R' \subseteq R$, $P_{\mathrm{res}}$ contains a tight path with the vertex set $V(P_{\mathrm{res}}) \setminus R'$ and end-tuples $\mathbf{v}$ and $\mathbf{w}$.

**Lemma 2.23** (Reservoir Lemma). *Given $k \geqslant 3$, $\gamma > 0$, and $\ell' \in \mathbb{N}$, there exist an integer $c$, such that for $0 < \varepsilon' \leqslant \frac{1}{4}\gamma$, $0 < d \leqslant \frac{1}{8}\gamma$, large enough $t_0$, small enough $\nu, \varepsilon_k > 0$, any functions $\varepsilon, f_k, f : \mathbb{N} \to (0, 1]$ which tend to zero sufficiently fast and any integers $t_2 \geqslant t_1 \geqslant t_0$ the following holds a.a.s. in $\Gamma = G^{(k)}(n, p)$ with $p \geqslant n^{-1+\gamma}$. Suppose $G \subseteq \Gamma$ with $\delta_{k-1}(G) \geqslant (\frac{1}{2} + \gamma)pn$, that $(\mathcal{P}_c^*, \mathcal{P}_f^*)$ is a $\left(t_0, t_1, t_2, \varepsilon_k, \varepsilon(t_1), f_k(t_1), f(t_2), p\right)$-strengthened pair for $G$, and that $t$ is the number of 1-cells in $\mathcal{P}_c^*$. Let $\mathcal{R} = \mathcal{R}_{\varepsilon_k, d}(G; \mathcal{P}_c^*, \mathcal{P}_f^*)$ be the $(\varepsilon_k, d)$-reduced multicomplex of $G$ and let $S$ be the union of the 1-cells that are not in $\mathcal{R}$. Then given $R \subseteq V(G)$ with $|R| \leqslant \nu n$ there exists a reservoir path $P_{\mathrm{res}}$ in $G$ with reservoir set $R$ and ends $\mathbf{v}$ and $\mathbf{w}$, such that $\mathbf{v}, \mathbf{w}$ are $(\varepsilon', p, \ell')$-good for $S \cup V(P_{\mathrm{res}})$ and $|V(P_{\mathrm{res}})| \leqslant c|R|$.*

We prove Lemma 2.23 in Section 5.

# 3. Proof of Theorem 1.1

*Proof.* Given $\gamma > 0$ and $k \geqslant 3$, let $\ell_{\mathrm{L2.20}} \geqslant \frac{k-1}{\gamma} + k$ be returned by Lemma 2.20 for input $k$ and $\gamma$. Similarly, let $\ell_{\mathrm{L2.19}}$ be given by Lemma 2.19 with input $k$ and $\frac{1}{2}\gamma$. Let $\nu_{\mathrm{C2.18}}$ be returned by Corollary 2.18 for input $\gamma$, $\varepsilon = \frac{1}{4}\gamma$, $s = k$, $k$, and $\ell_{\mathrm{L2.20}}$ and let $\nu_{\mathrm{L2.20}} = \frac{1}{4}\nu_{\mathrm{C2.18}}$. Let $c$ be the integer

returned by Lemma 2.23 for input $\gamma$, $k$ and $\ell_{\text{L2.20}}$ and then let $\frac{1}{8}\gamma \geqslant d > 0$. Let $t_0 \geqslant k! \nu_{\text{C2.18}}^{-2}$ be sufficiently large for Lemma 2.23 with input as above, $\varepsilon' = \frac{1}{4}\gamma$ and $d$ and $s = 3\ell_{\text{L2.19}} + 3k$, and for Lemma 2.5 with input $k$ and $d$.

We then choose $\nu_{\text{res}} < \frac{\gamma}{8c}$ such that $2\nu_{\text{res}}$ is sufficiently small for Lemma 2.23 with the given input and $\nu_{\text{L2.19}} > 0$ is small enough for Lemma 2.19 with input as above and $\eta = \frac{1}{2}$, $\varepsilon' = \frac{1}{8}\gamma$. We let $\eta_{\text{L2.20}} = 10^{-6}\ell_{\text{L2.19}}^{-1}\nu_{\text{res}}\nu_{\text{L2.19}}$. Next we choose $\varepsilon_k \leqslant 10^{-6}k^{-k}\nu_{\text{C2.18}}^k\nu_{\text{res}}^k\eta_{\text{L2.20}}^k$ small enough for Lemma 2.20 with input as above and $s = \ell_{\text{L2.20}}$, $d$, $\eta_{\text{L2.20}}$, $\nu_{\text{L2.20}}$, and $\varepsilon' = \frac{1}{8}\gamma$, for Lemma 2.23 with input as above, and also such that $2\nu_{\text{res}}^{-k}\varepsilon_k$ is small enough (as $\varepsilon_k$) for Lemma 2.19 with input as above.

We choose functions $\varepsilon, f_k, f : \mathbb{N} \to (0,1]$ such that $\sqrt{\varepsilon}$, $2\nu_{\text{res}}^{-k}f_k$ and $\sqrt{f}$ are all smaller than $\varepsilon_k$, small enough (as $\varepsilon$, $f_k$, and $f$, respectively) for each of Lemmas 2.19, 2.20 and 2.23 with the above inputs, and that for each $t$, both $\varepsilon(t)$ and $f(t)$ are small enough for Lemma 2.7 with input $k$, $\alpha = \frac{1}{2}\nu_{\text{res}}$ and $d_0 = \frac{1}{2t}$. Let $\eta_{\text{L2.4}}$ and $T_{\text{L2.4}}$ be returned by Lemma 2.4 for input $k$, $t_0$, $s = 1$, $\varepsilon_k$, $\varepsilon$, $f_k$, $f$.

Given $n$, let $p \geqslant n^{-1+\gamma}$. Let $\tilde{L}$ be a set of at most $T_{\text{L2.4}}! - 1$ vertices in $[n]$ such that $n - |\tilde{L}|$ is divisible by $T_{\text{L2.4}}!$. Suppose that $\tilde{\Gamma} = G^{(k)}(n,p)$ and its induced subgraph $\Gamma = \tilde{\Gamma} - \tilde{L}$ are in the good events of Corollary 2.18, Lemmas 2.20 and 2.23 with inputs as above and Lemma 2.12 with input $\frac{1}{4}\gamma\nu_{\text{res}}$ and $k$. Suppose that $\Gamma$ and all its subgraphs are $(\eta_{\text{L2.4}}, p)$-upper regular, which by Lemma 2.11 holds a.a.s. In addition, suppose that $\Gamma$ satisfies the following: if $R$ is a set of vertices chosen independently with probability $\nu_{\text{res}}$ from $V(\Gamma)$, then a.a.s. $\Gamma[R]$ is in the good event of Lemma 2.19 with input as above. Note that this last event occurs a.a.s. for the following reason: if we first choose $R$ randomly then expose the edges of $\Gamma$, a.a.s. we obtain a set $R$ of size $\left(1\pm\frac{1}{2}\right)\nu_{\text{res}}n$, and given this Lemma 2.19 states that a.a.s. $\Gamma[R]$ will be in the good event. Thus the probability of obtaining a pair $(R, \Gamma)$ such that $\Gamma[R]$ is not in the good event of Lemma 2.19 is $o(1)$, and it follows that, for any $\iota > 0$, the probability of choosing $\Gamma$ such that $\bigl(\, \Gamma[R]$ has probability at least $\iota$ of not being in the good event of Lemma 2.19 $\bigr)$, is $o(1)$.

Given $\tilde{G} \subseteq \Gamma$ with $\delta_{k-1}(\tilde{G}) \geqslant \left(\frac{1}{2} + 2\gamma\right)pn$, we remove $\tilde{L}$ to obtain an induced subgraph $G$ with $T_{\text{L2.4}}!|v(G)$. Observe that $\delta_{k-1}(G) \geqslant \left(\frac{1}{2} + \gamma\right)pn$. We apply Lemma 2.4 to $G$, with input as above, to obtain a $\left(t_0, t_1, t_2, \varepsilon_k, \varepsilon(t_1), f_k(t_1), f(t_2), p\right)$-strengthened pair $(\mathcal{P}_c^*, \mathcal{P}_f^*)$, where $t_0 \leqslant t_1 \leqslant t_2 \leqslant T_{\text{L2.4}}$. Let $t$ be the number of clusters of $\mathcal{P}_c$; by definition we have $t_0 \leqslant t \leqslant t_1$. Applying Lemma 2.5, we see that the $(\varepsilon_k, d)$-reduced multicomplex $\mathcal{R}$ of $G$, with respect to this strengthened pair, has at least $\left(1 - 4\varepsilon_k^{1/k}\right)t$ 1-edges, every $(k-1)$-edge of $\mathcal{R}$ is contained in at least

$$\left(\tfrac{1}{2} + \gamma - 2d - 2^{k+2}\varepsilon_k^{1/k}\right)t\prod_{i=2}^{k-1}d_i^{-\binom{k-1}{i-1}} \geqslant \left(\tfrac{1}{2} + \tfrac{1}{2}\gamma\right)t\prod_{i=2}^{k-1}d_i^{-\binom{k-1}{i-1}}$$

$k$-edges, and every induced subcomplex of $\mathcal{R}$ on at least $\left(1 - \gamma + 2d + 2^{k+2}\varepsilon_k^{1/k}\right)t < \left(1 - \tfrac{1}{2}\gamma\right)t$ vertices is tightly linked, where both inequalities use that $d$ and $\varepsilon_k$ are chosen small enough.

We choose a subset $R$ of $[n]$ by selecting vertices uniformly at random with probability $\nu_{\text{res}}$. A.a.s. we have $|R| = \left(1 \pm o(1)\right)\nu_{\text{res}}n$. By Chernoff's inequality and the union bound, a.a.s. for each $V$ which is a part of either $\mathcal{P}_c$ or $\mathcal{P}_f$, we have $|V \cap R| = \left(1 \pm o(1)\right)\nu_{\text{res}}|V|$. Furthermore, for each $S$ which is the neighbourhood in $\tilde{G}$ or in $\Gamma$ of some $(k-1)$-set of vertices, we

have $|S \cap R| = \big(1 \pm o(1)\big)\nu_{\mathrm{res}}|S|$ (recall that any such set $S$ has size at least $\frac{1}{2}pn \geqslant n^{\gamma/2}$). Finally, by our assumption on $\Gamma$, we have a.a.s. that $\Gamma[R]$ is in the good event of Lemma 2.19 with input as above. Suppose that $R$ is such that all of these likely events occur.

We apply Lemma 2.23, with inputs as above, to find a reservoir path $P_{\mathrm{res}}$ in $G$ with reservoir set $R$ whose ends are $\mathbf{v}_{\mathrm{res}}$ and $\mathbf{w}_{\mathrm{res}}$, such that $\mathbf{v}_{\mathrm{res}}$ and $\mathbf{w}_{\mathrm{res}}$ are both $(\frac{1}{4}\gamma, p, \ell_{\mathrm{L2.20}})$-good for $S \cup V(P_{\mathrm{res}})$, where $S$ is the union of all 1-cells not in $\mathcal{R}$, and such that $\big|V(P_{\mathrm{res}})\big| \leqslant c|R| \leqslant \frac{1}{8}\gamma n$.

We now aim to extend $P_{\mathrm{res}}$, from its end $\mathbf{w}_{\mathrm{res}}$, to a path $P_{\mathrm{almost}}$ covering almost all vertices of $G$. To begin with, let $\mathcal{R}'$ denote the complex on $V(\mathcal{R})$ obtained by letting $e'$ be an edge of $\mathcal{R}'$ whenever there is an edge $e$ of $\mathcal{R}$ such that $\mathrm{vertices}(e) = e'$. Thus the 1-edges of $\mathcal{R}$ and $\mathcal{R}'$ are identical, and it follows inductively from the definition of an $(\varepsilon_k, d)$-reduced multicomplex that for each $1 \leqslant i \leqslant k-2$, each $i$-edge of $\mathcal{R}'$ is contained in at least $\big(1 - 2^{i+2}\varepsilon_k^{1/k}\big)t$ $(i+1)$-edges, and each $(k-1)$-edge of $\mathcal{R}'$ is contained in at least $\big(\frac{1}{2} + \frac{1}{2}\gamma\big)t$ $k$-edges. We define a weight function $\omega$ on $V(\mathcal{R}')$ as follows. Given a cluster $V_i \in V(\mathcal{R}')$, if $|V_i \setminus V(P_{\mathrm{res}})| < 2\eta_{\mathrm{L2.20}}\frac{n}{t}$, we set $\omega(V_i) = 0$. Otherwise, we set

$$\omega(V_i) = \frac{|V_i \setminus V(P_{\mathrm{res}})| - 2\eta_{\mathrm{L2.20}}\frac{n}{t}}{(1 - \nu_{\mathrm{res}})\frac{n}{t}}\,.$$

Note that since $|V_i \setminus V(P_{\mathrm{res}})| \leqslant |V_i \setminus R| \leqslant \big(1 + o(1)\big)(1 - \nu_{\mathrm{res}})\frac{n}{t}$, this weight function takes values in $[0,1]$. Furthermore, we have

$$\sum_{V_i \in V(\mathcal{R}')} \omega(V_i) = \frac{n - |V(P_{\mathrm{res}})| - 2\eta_{\mathrm{L2.20}}n}{(1 - \nu_{\mathrm{res}})\frac{n}{t}} > \big(1 - \tfrac{1}{2}\gamma\big)t\,.$$

This is the required setup to apply Lemma 2.21, with input $2^{k+2}\varepsilon_k^{1/k}$ and $\frac{1}{2}\gamma$. The result is a weight function $\omega^* : E(\mathcal{R}'^{(k)}) \to [0,1]$ such that for each $V_i \in V(\mathcal{R}')$ we have $\sum_{e \ni V_i} \omega^*(e) \leqslant \omega(V_i)$, and

$$\sum_{e \in \mathcal{R}'^{(k)}} \omega^*(e) \geqslant \tfrac{1}{k}\Big( \sum_{V_i \in V(\mathcal{R}')} \omega(V_i) - 2^{k+2}\varepsilon_k^{1/k}t \Big) > \tfrac{1}{k} \cdot \frac{n - |V(P_{\mathrm{res}})| - 2\eta_{\mathrm{L2.20}}n - 2^{k+3}\varepsilon_k^{1/k}n}{(1 - \nu_{\mathrm{res}})\frac{n}{t}}$$

$$> \tfrac{1}{k} \cdot \frac{n - |V(P_{\mathrm{res}})| - 3\eta_{\mathrm{L2.20}}n}{(1 - \nu_{\mathrm{res}})\frac{n}{t}}\,.$$

Recall that $\mathcal{R}$ is tightly linked, even if an arbitrary set of $\frac{1}{2}\gamma t$ vertices is removed. If a cluster of $\mathcal{P}_c$ has $\omega$-weight zero, then it contains at least $\frac{n}{2t}$ vertices of $P_{\mathrm{res}}$, so there are at most $\frac{2c\nu_{\mathrm{res}}n \cdot 2t}{n} = 4c\nu_{\mathrm{res}}t \leqslant \frac{1}{2}\gamma t$ clusters with $\omega$-weight zero. In particular, the submulticomplex of $\mathcal{R}$ induced by removing clusters of $\omega$-weight zero is tightly linked.

We next construct a path $P_{\mathrm{almost}}$ extending $P_{\mathrm{res}}$ from $\mathbf{w}_{\mathrm{res}}$ as follows. Recall that $\mathbf{w}_{\mathrm{res}}$ is $\big(\frac{1}{4}\gamma, p, \ell_{\mathrm{L2.20}}\big)$-good for $S \cup V(P_{\mathrm{res}})$. To begin with, we use Corollary 2.18 to obtain a collection of $(k-1)$-tuples, of size at least $\nu_{\mathrm{C2.18}}n^{k-1}$, each of which is the end-tuple of a path of length $\ell_{\mathrm{L2.20}}$ starting at $\mathbf{w}_{\mathrm{res}}$ whose vertices, other than those in $\mathbf{w}_{\mathrm{res}}$, are disjoint from $V(P_{\mathrm{res}})$. Note that all these tuples are by construction outside $V(P_{\mathrm{res}})$ and so also outside $R$. By definition of a strengthened pair and $(\varepsilon_k, d)$-reduced multicomplex, and choice of $t_0$ and $\varepsilon_k$, at least

half of these end-tuples are contained in $(k-1)$-cells of $\mathcal{P}_c^*$ which are in $\mathcal{R}$. In particular, by averaging there is a $(k-1)$-cell of $\mathcal{R}$, with clusters in a given order, $f_0$, such that at least a $\frac{1}{2}\nu_{\text{C2.18}}$-fraction of these end-tuples are in $f_0$ in the given order. Let $P_0 = P_{\text{res}}$, and let $Q_0$ denote the set of $(k-1)$-tuples in $f_0$ which are ends of paths of length $\ell_{\text{L2.20}}$ starting from $\mathbf{w}_0 := \mathbf{w}_{\text{res}}$ whose vertices outside $\mathbf{w}_0$ are disjoint from $P_0$.

We order arbitrarily the $k$-edges of $\mathcal{R}'$ with positive $\omega^*$-weight, and for each $j$, let $g_j$ be a $k$-edge of $\mathcal{R}$ whose vertices are the same as the $j$th $k$-edge of $\mathcal{R}'$; we let $\omega^*(g_j)$ be given by $\omega^*$ at the $j$th edge of $\mathcal{R}'$. We now create a sequence $e_1, \ldots$ of $k$-edges of $\mathcal{R}$ as follows. To begin with, we choose a tight link in $\mathcal{R}$ from $f_0$ to a $(k-1)$-tuple in $g_1$ using only clusters of positive weight, and we let the first edges $e_1, \ldots$ be the edges of a homomorphic copy of a minimum length tight path following this tight link. We then repeat $g_1$ in the sequence

$$\left\lceil \frac{k(1-\nu_{\text{res}})\frac{n}{t} \cdot \omega^*(g_1)}{\ell_{\text{L2.20}}} \right\rceil$$

times, follow a tight link to $g_2$, and so on. When we follow a tight link, we always do so such that the edges $e_1, \ldots$ form a homomorphic copy of a tight path in $\mathcal{R}$, using only vertices whose weight according to $\omega$ is positive; this is possible since the vertices of each $g_j$ have weight at least $\omega^*(g_j) > 0$, and since the positive-weight induced submulticomplex of $\mathcal{R}$ is tightly linked. Note that the number of repetitions of $g_1$ fixes the ordered $(k-1)$-cell in the boundary of $g_1$ from which we follow a tight link to $g_2$, and so on.

Since $\mathcal{R}$ is a bounded size multicomplex — it contains in total at most $t_1^k \cdot t_1^{\binom{k}{2}} \ldots t_1^{\binom{k}{k-1}} \leqslant t_1^{2^k}$ edges of size $k$ — the total number of edges $e_i$ used in following tight links is at most $4k^3 \cdot t_1^{2^{k+1}}$.

We now use the following procedure repeatedly for $i \geqslant 1$. We are given $P_{i-1}$ which is a path from $\mathbf{v}_{\text{res}}$ to $\mathbf{w}_{i-1}$, an ordered $(k-1)$-cell $f_{i-1}$ of $\mathcal{R}$ (which is contained in $e_{i-1}$ and also in $e_i$), and a set $Q_{i-1}$ of $(k-1)$-tuples in $f_{i-1}$ which are ends of paths of length $\ell_{\text{L2.20}}$ from $\mathbf{w}_{i-1}$ whose vertices outside $\mathbf{w}_{i-1}$ are disjoint from $P_{i-1}$. We suppose $Q_{i-1}$ contains at least a $2\nu_{\text{L2.20}}$-fraction of the $(k-1)$-tuples in $f_{i-1}$. We let $f_i$ be the $(k-1)$-cell in the boundary of $e_i$ on the last $k-1$ clusters of $e_i$, with the order inherited from $e_i$.

By Lemma 2.20, with input as above, and $S = V(P_{i-1})$, and choice of $\nu_{\text{L2.20}}$ there is a tuple $\mathbf{w}_i$ in $Q_{i-1}$ such that the following holds. Let $P_i$ denote the extension of $P_{i-1}$ to $\mathbf{w}_i$ by adding a path of length $\ell_{\text{L2.20}}$ witnessing $\mathbf{w}_i \in Q_{i-1}$; let $S'$ be the vertices $V(P_i) \setminus V(P_{i-1})$. There is a set $Q_i$ of $(1-\nu_{\text{L2.20}})$-fraction of the tuples of $f_i$, each of which is the end of a path of length $\ell_{\text{L2.20}}$ from $\mathbf{w}_i$, whose vertices outside $\mathbf{w}_i$ are disjoint from $V(P_{i-1})$ and from $S'$. Note that this is the setup required to iterate the application of Lemma 2.20, provided that we ensure that at no stage does $S = V(P_{i-1})$ intersect any cluster of $e_i$ in more than a $(1-\eta_{\text{L2.20}})$-fraction. This is guaranteed for the following reason. Given a cluster $V_j$ of $\mathcal{P}_c$, if $\omega(V_j) = 0$ then $V_j$ is not a vertex of any $e_i$. If on the other hand $\omega(V_j) > 0$, then the total number of vertices used in $V_j$ is at most

$$\ell_{\text{L2.20}} \cdot 4k^3 \cdot t_1^{2^{k+1}} + 2\ell_{\text{L2.20}} \cdot \binom{t_1}{k-1} + \frac{\ell_{\text{L2.20}}}{k} \cdot \frac{k(1-\nu_{\text{res}})\frac{n}{t}}{\ell_{\text{L2.20}}} \cdot \sum_{g_i \ni V_j} \omega^*(g_i) \, ,$$

where the first term counts vertices used in following tight links, the second accounts for the rounding up in the weighting at each edge $g_i$ containing $V_j$ and the (at most) one vertex per $g_i$

extra since the tight path may use one more vertex in some clusters than others (since $\frac{\ell_{\text{L2.20}}}{k}$ may not be an integer). Note that these first two terms are bounded from above by a constant. Since $\sum_{g_i \ni V_j} \omega^*(g_i) \leqslant \omega(V_j)$, we see that the number of vertices used in $V_j$ is at most

$$O(1) + \frac{\ell_{\text{L2.20}}}{k} \cdot \frac{k(1 - \nu_{\text{res}})\frac{n}{t}}{\ell_{\text{L2.20}}} \cdot \omega(V_j) = O(1) + |V_j \setminus V(P_{\text{res}})| - 2\eta_{\text{L2.20}}\frac{n}{t} \leqslant \left|V_j \setminus V(P_{\text{res}})\right| - \eta_{\text{L2.20}}\frac{n}{t},$$

and in particular at all times at least $\eta_{\text{L2.20}}\frac{n}{t}$ vertices remain in $V_j$. We let $P_{\text{almost}}$ denote the final tight path from $\mathbf{v}_{\text{res}}$ to $\mathbf{w}_{\text{alm}}$ obtained by this procedure.

Observe that, just counting repetitions of the $g_i$, the total number of vertices $\left|V(P_{\text{almost}}) \setminus V(P_{\text{res}})\right|$ is at least

$$\ell_{\text{L2.20}} \cdot \frac{k(1-\nu_{\text{res}})\frac{n}{t}}{\ell_{\text{L2.20}}} \cdot \sum_{e \in \mathcal{R}'^{(k)}} \omega^*(e) > k(1 - \nu_{\text{res}})\frac{n}{t} \cdot \frac{1}{k} \cdot \frac{n - |V(P_{\text{res}})| - 3\eta_{\text{L2.20}}n}{(1 - \nu_{\text{res}})\frac{n}{t}}$$

$$= n - |V(P_{\text{res}})| - 3\eta_{\text{L2.20}}n.$$

It follows that $n - \left|V(P_{\text{almost}})\right| \leqslant 3\eta_{\text{L2.20}}n$. Let $L = \left(V(G) \setminus V(P_{\text{almost}})\right) \cup \tilde{L}$. Recall that $\tilde{L}$ is the set of at most $T_{\text{L2.4}}! - 1$ vertices we removed from $\tilde{G}$ in order to guarantee the required divisibility condition.

Our final task is to extend $P_{\text{almost}}$, re-using some vertices of $R$, to cover the vertices of $L$ and connect the ends. Critically, observe that $|L|$ is much smaller than $|R|$, and that by assumption on $\Gamma$ and $R$, the good event of Lemma 2.19 holds for $\Gamma[R]$, for the input given at the start of the proof. Recall that $G[R]$ has minimum codegree at least $\left(\frac{1}{2} + \frac{1}{2}\gamma\right)p|R|$. Let $\mathcal{P}_{cr}^*$ and $\mathcal{P}_{fr}^*$ denote the families of partitions obtained from $\mathcal{P}_c^*$ and $\mathcal{P}_f^*$ respectively by reducing each cell to only those elements contained in $R$. By Lemma 2.7 and choice of $\varepsilon_k$, $\varepsilon$, $f_k$ and $f$, $(\mathcal{P}_{cr}^*, \mathcal{P}_{fr}^*)$ is a $(t_0, t_1, t_2, 2\nu_{\text{res}}^{-k}\varepsilon_k, \sqrt{\varepsilon(t_1)}, 2\nu_{\text{res}}^{-k}f_k, \sqrt{f(t_2)}, p)$-strengthened pair for $G[R]$. Let $\mathcal{R}_r$ denote the multicomplex obtained from $\mathcal{R}$ by replacing the cells of $\mathcal{P}_c^*$ with those of $\mathcal{P}_{cr}^*$. Note that $\mathcal{R}_r$ is still the $(\varepsilon_k, d)$-reduced multicomplex for this strengthened pair, so it is contained in the $(2\nu_{\text{res}}^{-k}\varepsilon_k, d)$-reduced multicomplex.

Let $S_{-1} = \varnothing$. We now construct for $i = 0, 1, \dots$ two disjoint tight paths $P_{v,i}$ and $P_{w,i}$ and $S_i = V(P_{v,i}) \cup V(P_{w,i})$, where one end of $P_{v,i}$ is $\overleftarrow{\mathbf{v}_{\text{res}}}$ and the other, $\mathbf{v}_i$, is $(\frac{1}{4}\gamma\nu_{\text{res}}, p, \ell_{\text{L2.19}})$-good for $S_{i-1}$, and $P_{v,i}$ contains $i$ vertices of $L$ and all other vertices, except those of $\mathbf{v}_{\text{res}}$, are in $R$. Similarly one end of $P_{w,i}$ is $\mathbf{w}_{\text{alm}}$ and the other, $\mathbf{w}_i$, is $(\frac{1}{4}\gamma\nu_{\text{res}}, p, \ell_{\text{L2.19}})$-good for $S_{i-1}$, and $P_{w,i}$ contains $i$ vertices of $L$, not in $P_{v,i}$, and all other vertices, except those of $\mathbf{w}_{\text{alm}}$, are in $R$. We do this as follows. To begin with, we find a tight path $P_{v,0}$ of length $k - 1$, one of whose end tuples is $\overleftarrow{\mathbf{v}_{\text{res}}}$ and the other of which, $\mathbf{v}_0$, is contained in $R$. Recall that every $(k-1)$-set in $V(G)$ contains at least $\left(\frac{1}{2} + \frac{1}{2}\gamma\right)p|R|$ edges of size $k$ with the extra vertex in $R$, so in particular we can greedily build the required path of length $k - 1$. We construct $P_{w,0}$ from $\mathbf{w}_{\text{alm}}$ to $\mathbf{w}_0$ similarly. Observe that, by definition, both $\mathbf{v}_0$ and $\mathbf{w}_0$ are $(\frac{1}{4}\gamma\nu_{\text{res}}, p, \ell_{\text{L2.19}})$-good for $S_{-1}$.

Now suppose $i \geqslant 1$ and that we have constructed tight paths $P_{v,i-1}$ and $P_{w,i-1}$ as above, whose ends $\mathbf{v}_{i-1}$ and $\mathbf{w}_{i-1}$ are both $(\frac{1}{4}\gamma\nu_{\text{res}}, p, \ell_{\text{L2.19}})$-good for $S_{i-2}$, and we have $|S_{i-1}| \leqslant 4(i - 1)\ell_{\text{L2.19}}$. We first extend $P_{v,i-1}$ to a path $P_{v,i}$ as follows. We choose any $u \in L \setminus S_{i-1}$ and vertices $v_1, \dots, v_{k-2}$ from $R \setminus S_{i-1}$ such that the tuple $(u, v_1, \dots, v_{k-2})$

is $(\frac{1}{4}\gamma\nu_{\mathrm{res}}, p, \ell_{\mathrm{L2.19}} + k - 1)$-good for $S_{i-1}$. This step always succeeds, as $o(n)$ of these tuples are not $(\frac{1}{4}\gamma\nu_{\mathrm{res}}, p, \ell_{\mathrm{L2.19}} + k - 1)$-good for $S_{i-1}$, by the good event of Lemma 2.12 assumed above. Then we can easily choose additional vertices $v_{k-1}, u_1, \ldots, u_{k-1}$ from $R \setminus S_{i-1}$ such that for $j = 1, \ldots, k$ there is a $k$-edge $\{u_j, \ldots, u_{k-1}, u, v_1, \ldots, v_{j-1}\}$ and the tuples $\mathbf{v}_i = (v_1, \ldots, v_{k-1})$ and $\mathbf{u} = (u_1, \ldots, u_{k-1})$ are $(\frac{1}{4}\gamma\nu_{\mathrm{res}}, p, \ell_{\mathrm{L2.19}})$-good for $S_{i-1}$. For example, there are at least $(\frac{1}{2} + \frac{1}{4}\gamma)p|R|$ edges $\{u, v_1, \ldots, v_{k-1}\}$ in $G$ with $v_{k-1} \in R$, of which at most $p|S| + \frac{1}{4}\gamma\nu_{\mathrm{res}}pn \leqslant \frac{1}{4}p|R|$ have $v_{k-1} \in S_{i-1}$ and at most $\frac{1}{4}\gamma\nu_{\mathrm{res}}pn$ are such that $(v_1, \ldots, v_{k-1})$ is not $(\frac{1}{4}\gamma\nu_{\mathrm{res}}, p, \ell_{\mathrm{L2.19}} + k - 2)$-good for $S_{i-1}$.

Next, with Lemma 2.19, we connect $\mathbf{v}_{i-1}$ to $\mathbf{u}$ with a tight path of length $\ell_{\mathrm{L2.19}}$ and internal vertices not in $S_{i-1}$. Note that here we added the set $S'$ containing the vertices $V(P_{v,i-1}) \setminus V(P_{v,i-2})$, $V(P_{w,i-1}) \setminus V(P_{w,i-2})$, and $\{u, v_1, \ldots, v_{k-1}\}$ and that $|S'| \leqslant 2\ell_{\mathrm{L2.19}} + 2k$. To see that the conditions of Lemma 2.19 are satisfied, recall that $|S_{i-1}| \leqslant 4(i-1)\ell_{\mathrm{L2.19}} \leqslant 6\eta_{\mathrm{L2.20}}\ell_{\mathrm{L2.19}}n$. By the choice of $\eta_{\mathrm{L2.20}}$, this is at most $\frac{1}{4}|R|$ and there can bet at most $\nu_{\mathrm{L2.19}}t$ 1-cells in $\mathcal{R}_r$ which intersect $S_{i-1}$ in at least a $\frac{1}{2}$-fraction. We then let $P_{v,i}$ be the path obtained by concatenating $P_{v,i-1}$, the path from $\mathbf{v}_{i-1}$ to $\mathbf{u}$, and the path from $\mathbf{u}$ via $u$ to $\mathbf{v}_i$.

If there remain uncovered vertices in $L$, we repeat the same procedure to extend $P_{w,i-1}$ to $P_{w,i}$, where in the last step we also add the vertices from $V(P_{v,i}) \setminus V(P_{v,i-1})$ to $S'$ and get $|S'| \leqslant 3\ell_{\mathrm{L2.19}} + 3k$. Note that afterwards with $S_i = V(P_{v,i}) \cup V(P_{w,i})$ we have $|S_i| \leqslant 4i\ell_{\mathrm{L2.19}}$ and all conditions of $P_{v,i}$ and $P_{w,i}$ needed for the next iterations are satisfied. We stop this procedure as soon as all vertices of $L$ are used; we let $P_v$ denote the final $P_{v,i}$ with end tuple $\mathbf{v} = \mathbf{v}_i$, and $P_w$ denote either $P_{w,i}$ or $P_{w,i-1}$ (depending on whether $|L|$ is even or odd, respectively) with end tuple $\mathbf{w}$ either $\mathbf{w}_i$ or $\mathbf{w}_{i-1}$, respectively. Finally we make a last use of Lemma 2.19 to find a tight path in $R$ whose interior vertices are disjoint from $V(P_v) \cup V(P_w)$, and whose ends are $\overleftarrow{\mathbf{v}}$ and $\mathbf{w}$. This is possible for the same reasons as above. Concatenating these three tight paths, we obtain a tight path $P_{\mathrm{cover}}$ whose end tuples are $\mathbf{w}_{\mathrm{alm}}$ and $\mathbf{v}_{\mathrm{res}}$, such that $L \subseteq V(P_{\mathrm{cover}})$, and such that all interior vertices of $P_{\mathrm{cover}}$ are contained in $L \cup R$.

Let $R'$ denote the set of vertices $V(P_{\mathrm{cover}}) \cap R$. By the reservoir property of $P_{\mathrm{res}}$, there is a tight path $P_{\mathrm{res}}^*$ whose end tuples are identical to $P_{\mathrm{res}}$ and whose vertex set is $V(P_{\mathrm{res}}) \setminus R'$. We replace $P_{\mathrm{res}}$ with $P_{\mathrm{res}}^*$ in $P_{\mathrm{almost}}$ to obtain a tight path $P_{\mathrm{almost}}^*$ whose end tuples are identical to those of $P_{\mathrm{almost}}$ and whose vertex set is $V(P_{\mathrm{almost}}) \setminus R' = V(\tilde{G}) \setminus (L \cup R')$. Concatenating $P_{\mathrm{almost}}^*$ and $P_{\mathrm{cover}}$, we obtain the desired tight Hamilton cycle in $\tilde{G}$. $\qquad\square$

## 4. Connecting within the partition

In this section we prove Lemma 2.19 and 2.20. For the first, the strategy is to expand from the tuples $\mathbf{x}$ and $\mathbf{y}$ using Corollary 2.18 and then connecting two of the many ends that we found with Lemma 2.20 by following a tight link given by Lemma 2.5.

*Proof of Lemma 2.19.* Let $k \geqslant 3$ and $\gamma > 0$. Further let $\ell_{\mathrm{C2.18}}$ be the smallest integer exceeding $\frac{k-1}{\gamma} + k - 1$, and let $\ell_{\mathrm{L2.20}}$ be given by Lemma 2.20 on input $k$ and $\gamma$. Let $h$ be the shortest length of a tight path which admits a homomorphism to the edges of a tight link with first $k - 1$ vertices going to the start $(k - 1)$-tuple of the tight link in order and last $k - 1$ vertices going to the end $(k - 1)$-tuple of the tight link; let $\varrho$ be such that they are in the order $\varrho$. We

let $\ell := h\ell_{\text{L2.20}} + 2\ell_{\text{C2.18}} + 2$. Further let $s$ be any integer, and set $s_{\text{C2.18}} = s + \ell_{\text{L2.20}}$ and $s_{\text{L2.20}} = \ell$. Then let $d, \eta > 0$, $0 < \varepsilon' \leqslant \frac{1}{2}\gamma$, and $\nu_{\text{C2.18}}$ be given by Corollary 2.18 on input with $\gamma$, $\varepsilon'$, $s_{\text{C2.18}}$, $k$, and $\ell_{\text{C2.18}}$. Next, let $\eta_{\text{L2.20}}, \nu_{\text{L2.20}}, \nu > 0$ and an integer $t_0$ be such that

$$\eta_{\text{L2.20}} < \tfrac{1}{2}\eta, \quad 6\nu_{\text{L2.20}} < \eta_{\text{L2.20}}^{k-1}, \quad (k-1)!\nu_{\text{L2.20}} \leqslant \tfrac{1}{2}\nu_{\text{C2.18}}, \quad \nu \leqslant \tfrac{1}{4}\nu_{\text{C2.18}}, \quad \text{and} \quad \tfrac{k}{t_0} \leqslant \tfrac{1}{8}\nu_{\text{C2.18}}.$$

Then let $\varepsilon_k > 0$, functions $\varepsilon, f : \mathbb{N} \mapsto (0, 1]$), integers $t_1, t_2$, and $f_k > 0$ be such that Lemma 2.20 is applicable with input as above and $s_{\text{L2.20}}, d, \eta_{\text{L2.20}}, \nu_{\text{L2.20}}$, and $t_0$. We additionally require that $\varepsilon(t_1)$ is small enough for Lemma 2.6 with input $k$, $\alpha = \eta_{\text{L2.20}}$, $\gamma_c = \frac{1}{2}$, and $d_0 = \frac{1}{t_0}$.

Given $n$, let $p \geqslant n^{-1+\gamma}$. Suppose that $\Gamma = G^{(k)}(n, p)$ is in the good events of Corollary 2.18 and Lemma 2.20. Let $G \subseteq \Gamma$ be an $n$-vertex $k$-graph with $\delta_{k-1}(G) \geqslant (\frac{1}{2} + \gamma)pn$. Further, suppose that $(\mathcal{P}_c^*, \mathcal{P}_f^*)$ is a $(t_0, t_1, t_2, \varepsilon_k, \varepsilon(t_1), f_k, f(t_2), p)$-strengthened pair for $G$, let $\mathcal{R} = \mathcal{R}_{\varepsilon_k, d}(G; \mathcal{P}_c^*, \mathcal{P}_f^*)$ be the $(\varepsilon_k, d)$-reduced multicomplex of $G$, and let $t$ be the number of 1-cells. Then let $\mathcal{R}' \subseteq \mathcal{R}$ be an induced subcomplex of $\mathcal{R}$ on at least $(1 - \nu)t$ 1-edges, assume that it is tightly linked, and fix the coarse density vector $\mathbf{d} = (d_{k-1}, \ldots, d_2)$ with $d_i \geqslant t_1^{-1}$ for $i = 1, \ldots, k-1$. Further, let $S \subseteq V(G)$ with $|S| \leqslant \frac{1}{2}n$ be such that it intersects every 1-cell of $\mathcal{R}'$ in at most an $(1 - \eta)$-fraction. Next, let $\mathbf{x} = (x_1, \ldots, x_{k-1})$ and $\mathbf{y} = (y_1, \ldots, y_{k-1})$ be two $(k-1)$-tuples, which are $(\varepsilon', p, \ell_{\text{C2.18}})$-good for $S$ and also fix a set of vertices $S'$ of size at most $s$.

From now on we will solely work in $\mathcal{R}'$. To avoid clashes when constructing the paths we arbitrarily split the vertices in $V(G) \setminus (S \cup S')$ into two sets $T_{\mathbf{x}}$ and $T_{\mathbf{y}}$ such that each of them intersects every 1-cell in the same number of vertices and, in particular, more than an $\eta_{\text{L2.20}}$-fraction. We then define $S_{\mathbf{x}} := S \cup T_{\mathbf{y}}$ and $S_{\mathbf{y}} := S \cup T_{\mathbf{x}}$. For any coarse $(k-1)$-cell $E$ we get by Lemma 2.6 that $|E| \leqslant \frac{3}{2}\left(\frac{n}{t}\right)^{k-1} \prod_{i=2}^{k-1} d_i^{\binom{k-1}{i}}$. Similarly, applying Lemma 2.6 to the 1-cells restricted to $T_{\mathbf{x}}$ and $T_{\mathbf{y}}$ respectively, we get that $|E \setminus S_{\mathbf{x}}^{k-1}|, |E \setminus S_{\mathbf{y}}^{k-1}| \geqslant \frac{1}{2}\left(\eta_{\text{L2.20}}\frac{n}{t}\right)^{k-1} \prod_{i=2}^{k-1} d_i^{\binom{k-1}{i}}$. Therefore, we have

$$\frac{|E \setminus S_{\mathbf{x}}^{k-1}|}{|E|}, \frac{|E \setminus S_{\mathbf{y}}^{k-1}|}{|E|} \geqslant \frac{1}{3}\eta_{\text{L2.20}}^{k-1} > 2\nu_{\text{L2.20}} \quad \text{for all } (k-1)\text{-cells } E. \tag{4.1}$$

We apply Corollary 2.18 with $S$, $S'' = S' \cup \{y_1, \ldots, y_{k-1}\}$, and $\mathbf{x}$ to obtain a set $X$ of $\nu_{\text{C2.18}}n^{k-1}$ different $(k-1)$-tuples $\mathbf{x}'$ that are reachable from $\mathbf{x}$ by a path of length $\ell_{\text{C2.18}}$ with no vertices in $S \cup S''$ but possibly some of $\mathbf{x}$. We now want to estimate how many of the tuples from $X$ we cannot use. By assumption, in $\mathcal{R}'$ there are at most $\nu t$ 1-cells missing, which accumulate to at most $\nu t \frac{n}{t}n^{k-2} \leqslant \frac{1}{4}\nu_{\text{C2.18}}n^{k-1}$ tuples. Also, there can be at most $(k-1)\frac{n}{t}n^{k-2} \leqslant \frac{1}{8}\nu_{\text{C2.18}}n^{k-1}$ tuples which are not crossing with respect to the partition. Therefore, we have at least $\frac{1}{2}\nu_{\text{C2.18}}n^{k-1} \geqslant (k-1)!\nu_{\text{L2.20}}n^{k-1}$ usable tuples in $X$ and there exists a coarse $(k-1)$-cell $E_0$, such that at least a $\nu_{\text{L2.20}}$-fraction of this cell is contained in $X$ and they all have the same ordering.

As $\mathcal{R}'$ is tightly linked, we can fix a $k$-edge $H_{\mathbf{x}} := \hat{P}(Q_{\mathbf{x}}, \mathcal{P}_c^*)$ in $\mathcal{R}'$ that contains $E_0$. By Lemma 2.20 applied with $S_{\mathbf{x}}$ to $H_{\mathbf{x}}$ and another $(k-1)$-cell $E_1$ from $H_{\mathbf{x}}$, using (4.1), there exists a $(k-1)$-tuple $\mathbf{x}_0 \in (X \cap E_0)$ such that for a $(1 - \nu_{\text{L2.20}})$-fraction of the tuples $\mathbf{x}' \in E_1$ there exists a tight path of length $\ell_{\text{L2.20}}$ from $\mathbf{x}_0$ to $\mathbf{x}'$ with internal vertices not in $S_{\mathbf{x}} \cup S''$ for any set $S''$ of at most $s_{\text{L2.20}}$ vertices.

We repeat the above for $\mathbf{y}$ with $S_\mathbf{y}$ and $S''$ the union of $S'$ with the set of all vertices on the path from $\mathbf{x}$ to $\mathbf{x}_0$, where $|S''| \leqslant s + \ell_{\text{L2.20}} = s_{\text{C2.18}}$. From this we obtain $Y$, $E_{h+1}$, $E_h$, $H_\mathbf{y}$, and $\mathbf{x}_{h+1} \in Y \cap E_{h+1}$ with the analogous properties as above. As the 1-cells of $E_0$ are incident to at most $(k-1)\frac{n}{t}n^{k-2} \leqslant \frac{1}{8}\nu_{\text{C2.18}}n^{k-1}$ $(k-1)$-tuples we can assume that $E_0$ and $E_{h+1}$ use a disjoint set of 1-cells and, therefore, the paths starting in $\mathbf{x}_0$ and $\mathbf{x}_{h+1}$ do not overlap.

We will now use that, by assumption, $\mathcal{R}'$ is tightly linked to connect $\mathbf{x}_0$ to $\mathbf{x}_{h+1}$. We apply tight linkedness with $\mathbf{u}$ being the clusters of $E_1$ in the order induced by the order of $\mathbf{x}_0$ on $H_\mathbf{x}$, and $\varrho^{-1}(\mathbf{v})$ defined similarly on $H_\mathbf{y}$. By definition of $h$ and $\varrho$, there exists a sequence of $h$ edges $E_1, \ldots, E_h$ that give a homomorphism of a tight path, from $\mathbf{u}$ to $\mathbf{v}$ in that order, and the edges $E_i$ and $E_{i+1}$ are contained in a $k$-edge of $\mathcal{R}'$ for $i = 1, \ldots, h-1$.

We will connect $\mathbf{x}_0$ to $\mathbf{x}_{h+1}$ by following $E_1, \ldots, E_h$. Assume that for some $i = 0, \ldots, h-2$ we have a tuple $\mathbf{x}_i$ in $E_i$, tight paths from $\mathbf{x}$ to $\mathbf{x}_i$ and from $\mathbf{y}$ to $\mathbf{x}_{h+1}$ with no vertices in $S \cup S'$ but possibly some of $\mathbf{x}$ and $\mathbf{y}$, and denote the set of vertices of these paths by $S''$. Further assume that for at least a $(1-\nu_{\text{L2.20}})$-fraction of the $(k-1)$-tuples $\mathbf{x}'$ from $E_{i+1}$ there exists a tight path of length at most $\ell_{\text{L2.20}}$ from $\mathbf{x}_i$ to $\mathbf{x}'$ with no internal vertices in $S_\mathbf{x} \cup S''$. Then we apply Lemma 2.20 with $S_\mathbf{x} \cup S''$ to the $k$-edge of $\mathcal{R}'$ containing $E_{i+1}$ and $E_{i+2}$ to obtain with (4.1) that there exists a $(k-1)$-tuple $\mathbf{x}_{i+1}$ from $E_{i+1} \setminus (S_\mathbf{x} \cup S'')^{k-1}$ with a path from $\mathbf{x}_i$ to $\mathbf{x}_{i+1}$ of length $\ell_{\text{L2.20}}$ with no internal vertices in $S_\mathbf{x} \cup S''$, such that for at least a $(1-\nu_{\text{L2.20}})$-fraction of the $(k-1)$-tuples $\mathbf{x}'$ from $E_{i+2}$ there exists a tight path of length at most $\ell_{\text{L2.20}}$ from $\mathbf{x}_{i+1}$ to $\mathbf{x}'$ with no internal vertices in $S_\mathbf{x} \cup S''$ for any set $S''$ of at most $s_{\text{L2.20}}$ vertices. This implies the condition above for $i+1$ and, therefore, we can advance to the next step.

For the final step let $S''$ be the vertices on the tight paths from $\mathbf{x}$ to $\mathbf{x}_{h-1}$ and from $\mathbf{y}$ to $\mathbf{x}_{h+1}$ that we now have and note that $|S''| \leqslant s_{\text{L2.20}}$. Observe, that for a $(1-2\nu_{\text{L2.20}})$-fraction of the $(k-1)$-tuples $\mathbf{x}'$ from $E_{h-1}$ there exist a tight path from $\mathbf{x}_{h-1}$ and $\mathbf{x}_{h+1}$ to $\mathbf{x}'$ avoiding $S_\mathbf{x} \cup S''$ and $S_\mathbf{y} \cup S''$ respectively. Then, by (4.1), there exist an $\mathbf{x}_h$ in $E_{h-1} \setminus (S \cup S' \cup S'')^{k-1}$ and two paths from $\mathbf{x}_{h-1}$ to $\mathbf{x}_h$ and from $\mathbf{x}_{h+1}$ to $\mathbf{x}_h$ of length $\ell_{\text{L2.20}}$ that only overlap in $\mathbf{x}_h$. This finishes the path from $\mathbf{x}_0$ to $\mathbf{x}_{h+1}$ and, therefore, we have a path from $\mathbf{x}$ to $\mathbf{y}$, of length $\ell$ with no internal vertices from $S \cup S'$. $\qquad\qquad\square$

The proof of Lemma 2.20 is fairly long and intricate. Before explaining it, however, let us sketch an easier version. Suppose that $k = 2$ (i.e. we are dealing with graphs, not hypergraphs) and rather than having two clusters which are adjacent in the reduced graph, we have a path of $\ell + 1$ clusters $V_1, \ldots, V_\ell, V_{\ell+1}$ in the reduced graph. We want to show that for most vertices $x \in V_1$, there is a path from $x$ to $y$ for most $y \in V_{\ell+1}$. To begin with, we look at the fine parts within $V_\ell$. We discard those fine parts which do not form $(f_2, \frac{1}{2}d, p)$-regular pairs with most fine parts in $V_{\ell+1}$; by definition of the reduced graph, there are few such, and we let $X_\ell$ be the remaining subset of $V_\ell$. Next, for each $i = \ell-1, \ldots, 1$ we discard from $V_i$ those vertices with fewer than $(d - \varepsilon_2)p|X_{i+1}|$ neighbours in $X_{i+1}$ to obtain $X_i$. Again, by regularity we discard few vertices at each step, so $X_1$ is most of $V_1$. Now if we choose any $x \in X_1$, we claim there is a path from $x$ to $y$ for most $y \in V_{\ell+1}$.

To see this, note that there are many paths which start at $x$ and go out to $X_\ell$: we can construct these paths greedily starting from $x$, and we have at least $\frac{1}{2}dp|V_i|$ choices in each $X_i$. By Lemma 2.13 and choice of $\ell$, there are linearly many different endvertices of these paths in $X_\ell$. We call this the *coarse expansion*. However the number of these endvertices will be

much *smaller* than $\varepsilon_2 |X_\ell|$, so we cannot use the coarse regularity to say anything about the set of endvertices. This is where we need the fine partition: we can ensure the fine regularity constant $f_2$ is so small that the number of endvertices is much larger than $f_2 |X_\ell|$. By averaging, there is a fine part $Z$ contained in $X_\ell$ which contains a set $R_0$ of endvertices, where $|R_0| \geqslant f_2 |Z|$. Now $Z$ forms a $(f_2, \frac{1}{2} d, p)$-regular pair with most fine parts in $V_{\ell+1}$. For any such fine part $Z'$, by $(f_2, \frac{1}{2} d, p)$-regularity, the set $\overline{R_1}$ of vertices in $Z'$ which we cannot reach, i.e. which do not send an edge to $R_0$, is of size at most $f_2 |Z'|$. In other words, we have found, for most fine parts $Z'$ in $V_{\ell+1}$, a path from $x$ to most vertices of $Z'$; that is the desired paths to most vertices of $V_{\ell+1}$. We call this second step the *fine expansion*.

It is fairly easy to see that this strategy still works with sets $S$ and $S'$ to avoid. It is also not very hard to modify it to work with one regular pair rather than a path of regular pairs: we split off a small fraction of each cluster to use for the coarse expansion (and we do not reuse this part for the fine expansion). What is not, however, so easy is to make this argument work for $k$-graphs for $k \geqslant 3$. The coarse expansion step works much as described above, but the fine expansion requires more care. If we are given $k = 3$ and a regular polyad on parts $(X, Y, Z)$, and a significant fraction of the $XY$ 2-cell are marked as end-tuples of tight paths from some given $\mathbf{x}$, then we cannot necessarily conclude that almost all pairs in the $YZ$ 2-cell are end-tuples of tight paths from $\mathbf{x}$. We can only conclude this for those pairs whose vertex in $Y$ is also in many marked pairs. However this does then imply that most vertices of $Z$ are in $YZ$ pairs which form an edge with a marked pair, and taking another step, using another regular polyad $(Y, Z, W)$, we can finally argue that most $ZW$ pairs are end-tuples of tight paths from $\mathbf{x}$; so the $ZW$ pairs play the same role as $Z'$ in the argument sketched above. For higher uniformity, we generalise this argument; in uniformity $k$, we need $k - 1$ steps.

Before we prove Lemma 2.20, we give the following lemma, which deals with the fine expansion mentioned above. We will also reuse it in proving Lemma 2.23.

**Lemma 4.1.** *Given $k \geqslant 3$ and $\delta$, $d_k > 0$, for all sufficiently small $f'_k > 0$ we have: given $d_0 > 0$, for all sufficiently small $f' > 0$ and all sufficiently large $m$ the following holds.*

*Given a set $V$ of vertices, suppose that we have a ground partition $\mathcal{P} = \{X_0, \ldots, X_{2k-3}\}$ with $|X_i| = m$ for each $i$, and for each $2 \leqslant i \leqslant k-1$ a $\mathcal{P}$-partite $i$-graph $G_i$ on $V$ such that for each $Y \subseteq \{0, \ldots, 2k-3\}$ the graph $G_i\big[\prod_{y \in Y} X_y\big]$ is $(d_i, f', 1)$-regular with respect to $G_{i-1}$ (where we assume $E(G_1) = V$). Furthermore suppose that we have a $\mathcal{P}$-partite $k$-graph $G_k$, such that $G_k[X_j, \ldots, X_{j+k-1}]$ is $(d_k, f'_k, p)$-regular with respect to $G_{k-1}$ for each $0 \leqslant j \leqslant k-2$. Suppose that $d_i \geqslant d_0$ for each $2 \leqslant i \leqslant k-1$. Suppose that for each $2 \leqslant i \leqslant k$, all the edges of $G_i$ are supported by $G_{i-1}$.*

*Suppose that we are given a set $R_0 \subseteq G_{k-1}[X_0, \ldots, X_{k-2}]$ of size at least*

$$\delta m^{k-1} \prod_{\ell=2}^{k-1} d_\ell^{\binom{k-1}{\ell}}.$$

*Let $R_{k-1} \subseteq G_{k-1}[X_{k-1}, \ldots, X_{2k-3}]$ be those $(k-1)$-edges which are the end-tuples of some tight path in $G_k$ with one vertex in each of $X_0, \ldots, X_{2k-3}$ and whose start $(k-1)$-tuple is in $R_0$.*

*Then we have $|R_{k-1}| \geqslant (1-\delta) m^{k-1} \prod_{\ell=2}^{k-1} d_\ell^{\binom{k-1}{\ell}}$.*

*Proof.* Given $k$ and $\delta, d_k > 0$, we set $\eta = 4^{-10k}\delta$, and we set $\gamma = \frac{1}{1000k^2}\eta^{2k^2}$. Suppose $0 < f'_k < \frac{1}{100k}\gamma^3 d_k$, and given $d_0$, let $f' > 0$ be sufficiently small for all the below applications of Lemmas 2.6, 2.8 and 2.9 with input $k$, $\alpha = 1$, $\gamma$, $\delta$, and $d_0$ as required.

For each $1 \leqslant j \leqslant k - 1$, we let $R_j$ be those $(k-1)$-edges which are the end-tuples of some tight path in $G_k$ with one vertex in each of $X_j, \ldots, X_{k+j-2}$ and whose start $(k-1)$-tuple is in $R_0$. For each $1 \leqslant j \leqslant k - 1$, let $\overline{R_j} = E\big(G_{k-1}[X_j, \ldots, X_{j+k-2}]\big) \setminus R_j$. That is, $\overline{R_j}$ is the part of the $j$th $(k-1)$-cell which we *cannot* reach from $R_0$.

By definition, for each $1 \leqslant j \leqslant k-1$, there is no edge of $G_k$ which contains both a $(k-1)$-set in $R_{j-1}$ and one in $\overline{R_j}$. Since $G_k[X_{j-1}, \ldots, X_{j+k-2}]$ is $(d_k, f'_k, p)$-regular with respect to $G_{k-1}$, we conclude that the number of copies of $K_k^{(k-1)}$ in $G_{k-1}$ which contain both an edge of $R_{j-1}$ and one of $\overline{R_j}$ is smaller than an $f'_k$-fraction of all the copies of $K_k^{(k-1)}$ in $G_{k-1}[X_{j-1}, \ldots, X_{j+k-2}]$, i.e. it is at most

$$2f'_k m^k \prod_{i=2}^{k-1} d_i^{\binom{k}{i}} . \tag{4.2}$$

The remainder of the proof of this lemma will consist of repeatedly using this fact, together with counting in $G_{k-1}$, to argue that $R_{k-1}$ is necessarily large. We will not need to use $G_k$ (or $(d_k, f'_k, p)$-regularity) again. We do this by the following induction.

Let $U_{0,0} = \{\varnothing\}$ (which we think of as the 0-edge in a complex).

**Claim 4.2.** *There exist sets with the following properties. For each $1 \leqslant j \leqslant k - 1$, $U_{j,j}$ is a subgraph of $G_j[X_{k-1}, \ldots, X_{k+j-2}]$, with*

$$|U_{j,j}| = \big(1 - j\eta\big)m^j \prod_{\ell=2}^{j} d_\ell^{\binom{j}{\ell}} \pm 1 .$$

*For each $0 \leqslant j < k-1$ and each $j < i \leqslant k-1$, $U_{j,i}$ is a subgraph of $G_i[X_{k+j-i-1}, \ldots, X_{k+j-2}]$. For each $0 \leqslant j \leqslant k - 2$ and each $j \leqslant i \leqslant k - 2$, each edge of $U_{j,i}$ is contained in*

$$\eta^{j+1} m \prod_{\ell=2}^{i+1} d_\ell^{\binom{i}{\ell-1}} \pm 1$$

*edges of $U_{j,i+1}$.*

*Each edge of $U_{j,k-1}$ is an edge of $R_j$ for each $0 \leqslant j \leqslant k - 1$.*

*Proof of Claim.* We begin with the base case $j = 0$. Let $U'_{0,k-1} = R_0$. For each $k - 2 \geqslant i \geqslant 1$ successively, we let $U'_{0,i}$ contain all the $i$-edges with one vertex in each of $X_{k-1-i}, \ldots, X_{k-2}$ which lie in

$$\text{between} \quad \eta m \prod_{\ell=2}^{i+1} d_\ell^{\binom{i}{\ell-1}} \quad \text{and} \quad 2m \prod_{\ell=2}^{i+1} d_\ell^{\binom{i}{\ell-1}}$$

edges of $U'_{0,i+1}$. We now check $|U'_{0,i}|$ is sufficiently large. By the third part of Lemma 2.8, there are at most $\gamma m^i \prod_{\ell=2}^{i} d_\ell^{\binom{i}{\ell}}$ elements of $G_i$ that violate the upper bound, which by the second part of Lemma 2.8 are in total contained in at most $3\gamma m^{i+1} \prod_{\ell=2}^{i+1} d_\ell^{\binom{i+1}{\ell}}$ edges of $U'_{0,i+1}$. By definition

and by Lemma 2.6, the total number of edges of $U'_{0,i+1}$ containing edges of $G_i$ violating the lower bound is at most $(1 + \gamma)\eta m^{i+1} \prod_{\ell=2}^{i+1} d_\ell^{\binom{i+1}{\ell}}$, and therefore there are at least

$$|U'_{0,i+1}| - 2\eta m^{i+1} \prod_{\ell=2}^{i+1} d_\ell^{\binom{i+1}{\ell}}$$

edges of $U'_{0,i+1}$ which contain edges of $U'_{0,i}$. We check inductively that this is for each $i$ at least $\frac{1}{2}|U'_{0,i+1}|$, and hence obtain

$$|U'_{0,i}| \geqslant \frac{\frac{1}{2}|U'_{0,i+1}|}{2m \prod_{\ell=2}^{i+1} d_\ell^{\binom{i}{\ell-1}}} \geqslant 4^{i+1-k}\delta m^i \prod_{\ell=2}^{i} d_\ell^{\binom{i}{\ell}}.$$

By choice of $\eta$, in particular we have $|U'_{0,1}| > \eta m$.

Now we let $U_{0,1} \subseteq U'_{0,1}$ be some set of size $\eta m \pm 1$, and for each $i \geqslant 1$ successively and each edge $e$ of $U_{0,i}$, we put into $U_{0,i+1}$ some

$$\eta m \prod_{\ell=2}^{i+1} d_\ell^{\binom{i}{\ell-1}} \pm 1$$

edges in $U'_{0,i+1}$ which contain $e$. These sets witness that Claim 4.2 holds for $j = 0$.

Now suppose $1 \leqslant j \leqslant k - 1$. Given sets $U_{j-1,j-1}, \ldots, U_{j-1,k-1}$ as in the claim statement for $j - 1$, let $C$ consist of those edges in $U_{j-1,k-2}$ which are contained in at least

$$\gamma m \prod_{\ell=2}^{k-1} d_\ell^{\binom{k-2}{\ell-1}}$$

members of $\overline{R_j}$. Let $A = R_{j-1}$ and let $B$ consist of those members of $\overline{R_j}$ which contain an edge of $C$. By Lemma 2.9 (with input $\gamma/2$) and choice of $f'_k$, if $|C| \geqslant \gamma m^{k-2} \prod_{\ell=2}^{k-2} d_\ell^{\binom{k-2}{\ell}}$, then the number of copies of $K_k^{(k-1)}$ containing both an edge of $R_{j-1}$ and of $\overline{R_j}$ is in contradiction to (4.2). We conclude

$$|C| < \gamma m^{k-2} \prod_{\ell=2}^{k-2} d_\ell^{\binom{k-2}{\ell}}.$$

By definition, we have

$$\left|U_{j-1,k-2}\right| = \left(\prod_{i=j-1}^{k-3} \left(\eta^j m \prod_{\ell=2}^{i+1} d_\ell^{\binom{i}{\ell-1}} \pm 1\right)\right) \cdot \left(1 - (j-1)\eta\right) m^{j-1} \prod_{\ell=2}^{j} d_\ell^{\binom{j}{\ell}},$$

and hence

$$\left|U_{j-1,k-2} \setminus C\right| \geqslant (1 - (j-1)\eta)(1 - 2\gamma\eta^{-k^2})\eta^{(k-j-1)j} m^{k-2} \prod_{\ell=2}^{k-2} d_\ell^{\binom{k-2}{\ell}}.$$

We let $U'_{j,k-1}$ be the set of all edges in $R_j$ which contain an edge of $U_{j-1,k-2}$. Suppose that a given edge $e \in U_{j-1,k-2} \setminus C$ is contained in at least $(1-\gamma)m \prod_{\ell=2}^{k-1} d_\ell^{\binom{k-2}{\ell-1}}$ edges of $G_{k-1}$ together with a vertex of $X_{k+j-2}$. Then by definition of $C$, at least $(1-2\gamma)m \prod_{\ell=2}^{k-1} d_\ell^{\binom{k-2}{\ell-1}}$ of these are edges of $R_j$. By the third part of Lemma 2.8 there are at most $\gamma m^{k-2} \prod_{\ell=2}^{k-2} d_\ell^{\binom{k-2}{\ell}}$ edges in $U_{j-1,k-2}$ that are contained in less than $(1-\gamma)m \prod_{\ell=2}^{k-1} d_\ell^{\binom{k-2}{\ell-1}}$ edges of $G_{k-1}$ and we conclude

$$
\begin{aligned}
\left|U'_{j,k-1}\right| &\geqslant (1-2\gamma)m \prod_{\ell=2}^{k-1} d_\ell^{\binom{k-2}{\ell-1}} \cdot (1-(j-1)\eta)(1-3\gamma\eta^{-k^2})\eta^{(k-j-1)j}m^{k-2} \prod_{\ell=2}^{k-2} d_\ell^{\binom{k-2}{\ell}} \\
&\geqslant (1-5\gamma\eta^{-k^2})(1-(j-1)\eta)\eta^{(k-j-1)j}m^{k-1} \prod_{\ell=2}^{k-1} d_\ell^{\binom{k-1}{\ell}} .
\end{aligned}
\tag{4.3}
$$

If $j = k-1$, we set $U_{k-1,k-1} = U'_{k-1,k-1}$ and by choice of $\gamma$ we are done, so we now suppose $1 \leqslant j \leqslant k-2$.

We next let for each $i = k-2, \ldots, j$ in succession the set $U'_{j,i}$ consist of all those edges of $G_i$ with one vertex in each of $X_{k+j-i-1}, \ldots, X_{k+j-2}$ which contain at least

$$
\eta^{j+1}m \prod_{\ell=2}^{i+1} d_\ell^{\binom{i}{\ell-1}}
$$

edges of $U'_{j,i+1}$. Finally we let $U_{j,j} = U'_{j,j}$, and for each $j+1 \leqslant i \leqslant k-1$, we create $U_{j,i}$ by, for each $e \in U_{j,i-1}$, putting in $\eta^{j+1}m \prod_{\ell=2}^{i+1} d_\ell^{\binom{i}{\ell-1}} \pm 1$ edges of $U'_{j,i}$ which contain $e$.

We claim that these sets witness the $j$ case of Claim 4.2. To show this, we need to show $|U_{j,j}|$ is sufficiently large, and for that purpose we establish bounds on $|U'_{j,i}|$ for each $k-2 \geqslant i \geqslant j$ in succession.

Given $i$, consider the set of edges $Y_i$ in $G_i[X_{k+j-i-1}, \ldots, X_{k+j-2}]$ which contain an element of $U_{j-1,i-1}$. By the third part of Lemma 2.8 and because $U_{j-1,i-1}$ is sufficiently large, we have

$$
|Y_i| = (1 \pm \gamma)|U_{j-1,i-1}|m \prod_{\ell=2}^{i} d_\ell^{\binom{i-1}{\ell-1}} = (1 \pm 2\gamma)\eta^{(i-j)j}m^i \prod_{\ell=2}^{i} d_\ell^{\binom{i}{\ell}} .
$$

By Lemma 2.9 with input $\gamma/2$, for any subset $Y'$ of $Y_i$ with $|Y'| \geqslant \frac{1}{2}|Y_i|$, the number of edges of $G_{i+1}$ which contain an edge of $Y'$ and an edge of $U_{j-1,i}$ is

$$
|Y'|(\eta^j \pm \gamma)m \prod_{\ell=2}^{i+1} d_\ell^{\binom{i}{\ell-1}} .
$$

Observe that by construction every edge of $U'_{j,i+1}$ contains an edge of $Y_i$ and an edge of $U_{j-1,i}$. Let $Y'$ contain all the edges of $U'_{j,i}$, and if necessary additional edges to match the lower bound

of Lemma 2.9. Then putting the bound of Lemma 2.9 together with the definition of $U'_{j,i}$, we have

$$
\begin{aligned}
|U'_{j,i+1}| &\leqslant |U'_{j,i}|(\eta^j + \gamma)m \prod_{\ell=2}^{i+1} d_\ell^{\binom{i}{\ell-1}} + |Y_i \setminus U'_{j,i}|\eta^{j+1} m \prod_{\ell=2}^{i+1} d_\ell^{\binom{i}{\ell-1}} \\
&\leqslant \left( \left(1 + \gamma\eta^{-j}\right)\frac{|U'_{j,i}|}{|Y_i|} + \eta\left(1 - \frac{|U'_{j,i}|}{|Y_i|}\right)\right)\eta^j m |Y_i| \prod_{\ell=2}^{i+1} d_\ell^{\binom{i}{\ell-1}} \\
&\leqslant \left( 1 + \gamma\eta^{-j} - (1-\eta)\left(1 - \frac{|U'_{j,i}|}{|Y_i|}\right)\right)\eta^{(i+1-j)j} m^{i+1} \prod_{\ell=2}^{i+1} d_\ell^{\binom{i+1}{\ell}} .
\end{aligned}
$$

Comparing the last line, in the case $i = k - 2$, with (4.3), we see

$$
(1 - 5\gamma\eta^{-k^2})\left(1 - (j-1)\eta\right) \leqslant \left(1 + \gamma\eta^{-j} - (1-\eta)\left(1 - \frac{|U'_{j,k-2}|}{|Y_{k-2}|}\right)\right)
$$

and hence

$$
\begin{aligned}
|U'_{j,k-2}| &\geqslant |Y_i|\left(1 - (j-1)\eta - 2j\eta^2 - 9\gamma\eta^{-k^2}\right) \\
&\geqslant \left(1 - (j-1)\eta - 2j\eta^2 - 10\gamma\eta^{-k^2}\right)\eta^{(k-2-j)j} m^{k-2} \prod_{\ell=2}^{k-2} d_\ell^{\binom{k-2}{\ell}} .
\end{aligned}
$$

Repeating the same argument for $i = k - 3, k - 4, \ldots, j$ in succession, we end up with

$$
|U'_{j,j}| \geqslant \left(1 - (j-1)\eta - 2(k-1-j)j\eta^2 - 5 \cdot 2^{k-1-j}\gamma\eta^{-k^2}\right)m^j \prod_{\ell=2}^{j} d_\ell^{\binom{j}{\ell}} ,
$$

which by choice of $\eta$ and $\gamma$ is as required. □

The lemma follows from the case $j = k - 1$ of Claim 4.2 directly, by choice of $\eta$. □

The proof of Lemma 2.20 proceeds as follows. We start by setting up the parameters and the graph together with the properties that we assume. Afterwards, we choose the fine $(k-1)$-cells $\mathcal{C}$ (Properties (FR1) and (FR2)) that we would like to use later and show that these make up most of the coarse $(k-1)$-cell (Claim 4.3). Then we select a small fraction $\overline{\mathcal{C}}$ of these 1-cells, which gives a sufficiently large vertex set $\overline{V}$ for the coarse expansion. Having this setup, we can use the coarse partition and $\overline{V}$ to expand (Claim 4.4). After this we reached a significant fraction of a coarse $(k-1)$-cell, carefully chosen to also give significant fraction of a well-behaved fine $(k-1)$-cell. We then use Lemma 4.1 to argue that from a significant fraction of this one fine cell, we can reach almost all of almost all fine $(k-1)$-cells, as required.

*Proof of Lemma 2.20.* **Setting the parameters.** For $k \geqslant 3$ and $\gamma > 0$, let $\ell$ be the smallest integer exceeding $\frac{k-1}{\gamma} + 2k$, such that $\ell \equiv 1 \pmod{k}$ and let $s \geqslant 3k$. Let $d, \eta, \nu > 0$, where

w.l.o.g. we can assume $d, \eta, \nu \leqslant 1$, and $t_0$ be an integer. We let $0 < \gamma_c \leqslant \frac{\nu}{10^\ell} t_0^{-2k2^k}$ and $\varepsilon_k > 0$ such that

$$(2k)^{3k} \varepsilon_k^{1/2k} \leqslant \frac{\nu^k \eta}{10^\ell}. \tag{4.4}$$

Next, we let $\alpha_c = \frac{\nu}{10k} \varepsilon_k^{1/k}$ and $\varepsilon : \mathbb{N} \mapsto (0,1]$ tend to zero sufficiently fast, such that for any $t_1 \geqslant t_0$ we have that $\varepsilon := \varepsilon(t_1) < 2^{-2k2^k} t_1^{-1} \varepsilon_k$ is small enough for Lemma 2.6 on input $k$, $\alpha_c$, $d_0 = t_0^{-1}$, and $\gamma_c$ and Lemma 2.8 on input $k$, $\alpha_c$, $\gamma_c$, $d_0 = t_0^{-1}$.

We then define

$$\mu = \frac{d \varepsilon_k^{1/k} \nu}{t_1 20k^2} t_1^{-2^k}, \quad \delta = \frac{\mu^{2\ell}}{10(2\ell)!} \frac{20k^2 t_0}{\varepsilon_k^{1/2k} \nu} t_1^{-2^k}, \tag{4.5}$$

and note that $\delta \leqslant \nu$. We let

$$0 < \gamma_f \leqslant \frac{\delta}{10^k} \frac{\nu}{3^k} t_1^{-2k2^k} \tag{4.6}$$

and

$$0 < f_k \leqslant \frac{\delta^3 \nu}{30^{3k}}. \tag{4.7}$$

Suppose $f_k$ is small enough for Lemma 4.1 on input $k$, $\frac{1}{2}\delta$, $\frac{1}{2}d$.

Next, we let $f : \mathbb{N} \mapsto (0,1]$ tend to zero sufficiently fast, such that for $t_2 \geqslant t_1$ we have that $f := f(t_2) < \varepsilon t_2^{-1}$ is small enough for Lemma 2.6 on input $k$, $\alpha = 1$, $d_0 = t_1^{-1}$, and $\gamma_f$ and Lemma 2.7 on input $k$, $\alpha = \varepsilon_k^{1/2k}$, and $d_0 = t_1^{-1}$. Additionally assume that $\sqrt{f}$ is small enough for Lemma 4.1 with input as above, $f_k$, and $d_0 = \frac{1}{2t_2}$. For convenience we summarise the relative order of the parameters (besides $\gamma, k, \ell, s$) in a simplified form

$$f \ll f_k, \gamma_f, t_2^{-1} \ll \delta, \varepsilon \ll \varepsilon_k, \gamma_c, t_1^{-1} \ll t_0^{-1}, d, \eta, \nu, \varepsilon'.$$

Given $n$ we let $p \geqslant n^{-1+\gamma}$ and $\Gamma = G^{(k)}(n, p)$. We assume that $\Gamma$ is in the good event of Lemma 2.13 with $\gamma, k, \ell_0 = \ell - 3k + 4$, and $\mu$. We are given $G \subseteq \Gamma$ and the reduced complex $\mathcal{R}_{\varepsilon_k, d}(G; \mathcal{P}_c^*, \mathcal{P}_f^*)$ of a $(t_0, t_1, t_2, \varepsilon_k, \varepsilon^2, f_k, f^2, p)$-strengthened pair $(\mathcal{P}_c^*, \mathcal{P}_f^*)$ for $G$. Fix the density vectors $\mathbf{d} = (d_{k-1}, \ldots, d_2)$ and $\mathbf{d_f} = (d_{k-1}^f, \ldots, d_2^f)$ with $d_i \geqslant t_1^{-1}$ and $d_i^f \geqslant t_2^{-1}$ for $i = 1, \ldots, k$ and let $t$ and $t_f$ be the number of coarse and fine 1-cells respectively.

Take a $k$-set $Q_0$ such that $H := \hat{P}(Q_0; \mathcal{P}_c^*)$ is in $\mathcal{R}_{\varepsilon_k, d}(G; \mathcal{P}_c^*, \mathcal{P}_f^*)$. For convenience, we equip the complex $H$ with a $k$-level consisting of all $k$-sets supported by the $(k-1)$-edges of $H$. By definition, the relative $p$-density of $G$ with respect to $H$ is at least $d$. Note that every coarse 1-cell contains exactly $t_f/t$ fine 1-cells.

We let $V_1, \ldots, V_k$ be the 1-cells of $H$ and $E_1, \ldots, E_k$ be the $(k-1)$-cells of $H$, where $E_i \subseteq \prod_{j \neq i} V_j$ for $i = 1, \ldots, k$. Further, let $S \subseteq V(G)$ be such that $|S \cap V_i| \leqslant (1-\eta)n/t$. By the good event of Lemma 2.12 assumed above there are $o(n)$ $(k-1)$-sets in $V(\Gamma)$ outside of $S$ that are not $(f, p, \ell)$-good for $S$. Before we can expand using the coarse partition we need to ensure that we use edges that behave well with respect to the fine cells.

**Preparing the fine cells.** The polyad $H$ corresponds to a $(k-1)$-complex $\mathcal{H}(G;H)$ in the multicomplex of the family of partitions $\mathcal{P}_c^*$. We denote by $\mathcal{H} = \mathcal{H}(G;H,\mathcal{P}_f^*)$ the multicomplex of the family of partitions $\mathcal{P}_f^*$ restricted to $\mathcal{H}(G;H)$, i.e. we keep all the edges of the multicomplex of $\mathcal{P}_f^*$ that correspond to cells that are contained in cells that correspond to edges of $\mathcal{H}(G;H)$. For $2 \leqslant j \leqslant k-1$ and any $j$-set $Q$ supported by $\mathcal{H}$, the number of fine $j$-cells supported by the polyad $\hat{P}(Q;\mathcal{P}_c^*)$ is

$$\frac{d_j \pm \varepsilon}{d_j^f \pm f} = \left(1 \pm \frac{2\varepsilon}{d_j}\right)\frac{d_j}{d_j^f}, \tag{4.8}$$

where the equality follows as $f < \varepsilon t_2^{-1}$. This follows by simple double counting of the $j$-sets supported by $\hat{P}(Q;\mathcal{P}_c^*)$ that are contained in $\mathcal{H}$. This allows us for each $1 \leqslant i \leqslant k-1$ and any $i$-edge of $\mathcal{H}$ to control the number of $(i+1)$-edges it is contained in. But we have to carefully select the edges of $\mathcal{H}$, which we are using.

For this, we denote by $\mathcal{H}_{\varepsilon_k} = \mathcal{H}_{\varepsilon_k}(G;H,\mathcal{P}_f^*)$ the $\varepsilon_k$-reduced multicomplex of $G$ with respect to $(H, \mathcal{P}^f)$, which is the (unique) maximal subcomplex of $\mathcal{H}$ which has the following properties:

(FR1) For every $k$-edge $\hat{P}(Q;\mathcal{P}_f^*)$ of $\mathcal{H}_{\varepsilon_k}(G;H,\mathcal{P}_f^*)$ we have that $G$ is $(f_k, p)$-regular with respect to $\hat{P}(Q;\mathcal{P}_f^*)$ and $d_p(G|\hat{P}(Q;\mathcal{P}_f^*)) = d_p(G|H) \pm \varepsilon_k$.

(FR2) For each $1 \leqslant i \leqslant k-1$, each $i$-edge of $\mathcal{H}_{\varepsilon_k}(G;H,\mathcal{P}_f^*)$ is in the boundary of at least

$$\left(1 - k2^{i+4}\varepsilon_k^{1/k}\right)\frac{t_f}{t}\prod_{j=2}^{i+1}\left(\frac{d_j}{d_j^f}\right)^{\binom{i}{j-1}} \qquad \text{if } i < k-1,$$

$$\text{and} \quad \left(1 - k2^{i+4}\varepsilon_k^{1/k}\right)\frac{t_f}{t}\prod_{j=2}^{k-1}\left(\frac{d_j}{d_j^f}\right)^{\binom{i}{j-1}} \qquad \text{if } i = k-1$$

$(i+1)$-edges of $\mathcal{H}_{\varepsilon_k}(G;H,\mathcal{P}_f^*)$ with respect to any other 1-cell of $H$.

We will show below that by this construction we keep sufficiently many 1-edges and $(k-1)$-edges.

**Claim 4.3.** *There are at most $6\varepsilon_k^{1/k}\frac{t_f}{t}$ 1-edges removed from $\mathcal{H}$ to get $\mathcal{H}_{\varepsilon_k}$. Furthermore, for any $i = 1, \ldots, k$, there are at least*

$$(1 - k\,k!\varepsilon_k^{1/k})\left(\frac{t_f}{t}\right)^{k-1}\prod_{j=2}^{k-1}\left(\frac{d_j}{d_j^f}\right)^{\binom{k-1}{j}}$$

$(k-1)$-*edges in $\mathcal{H}_{\varepsilon_k}$ that are also contained in $E_i$.*

*Proof of Claim 4.3.* Since $H$ is a regular polyad, for at most $3\varepsilon_k(\frac{n}{t})^k\prod_{j=2}^k d_j^{\binom{k}{j}}$ of the $k$-sets $Q'$ supported on $H$, the fine polyad $\hat{P}(Q';\mathcal{P}_f^*)$ fails (FR1). As any $k$-edge in $\mathcal{H}$ supports at

least $\frac{3}{4}\left(\frac{n}{t_f}\right)^k \prod_{j=2}^{k}(d_j^f)^{\binom{k}{j}}$ of these $Q'$, we have at most

$$4\varepsilon_k \left(\frac{t_f}{t}\right)^k \prod_{j=2}^{k}\left(\frac{d_j}{d_j^f}\right)^{\binom{k}{j}} \tag{4.9}$$

$k$-edges in $\mathcal{H}$ failing (FR1). We mark all these $k$-edges as *bad*. Then, for each $i = k-1, \ldots, 1$, in succession, we mark as *bad* all $i$-edges which are contained in the boundary of at least

$$6\varepsilon_k^{1/k}(k-i)\frac{t_f}{t}\prod_{j=2}^{i+1}\left(\frac{d_j}{d_j^f}\right)^{\binom{i}{j-1}}$$

bad $(i+1)$-edges.

Now consider the following construction. We begin by taking any bad 1-edge, then any bad 2-edge containing it, and so on until we obtain a bad $k$-edge together with an order on its vertices. Clearly we obtain any given bad $k$-edge in at most $k!$ ways by following this process (since a $k$-edge together with an order determines the chosen edges in the process). If there are more than $6\varepsilon_k^{1/k}\frac{t_f}{t}$ bad 1-edges, it follows that the number of bad $k$-edges is at least

$$\frac{1}{k!}\prod_{i=0}^{k-1}\left(5\varepsilon_k^{1/k}(k-i)\frac{t_f}{t}\prod_{j=2}^{i+1}\left(\frac{d_j}{d_j^f}\right)^{\binom{i}{j-1}}\right) = 5^k\varepsilon_k\left(\frac{t_f}{t}\right)^k\prod_{i=2}^{k}\left(\frac{d_i}{d_i^f}\right)^{\binom{k}{i}},$$

which is a contradiction.

We claim that any edge of $\mathcal{P}_f^*$ which is not in $\mathcal{H}_{\varepsilon_k}(G; H, \mathcal{P}_f^*)$ is either bad or contains a bad edge. To see this, consider the process of obtaining $\mathcal{H}_{\varepsilon_k}(G; H, \mathcal{P}_f^*)$ by successively removing edges which either fail one of (FR2) or (FR1), or which contain a removed edge. Suppose for a contradiction that at some stage in this process we remove an edge which is neither bad nor contains a bad edge; let $e$ be the first such edge removed. Observe that $e$ cannot have been removed for failing (FR1), since edges which fail this condition are bad. Furthermore $e$ cannot have been removed for being unsupported, because all edges previously removed either were bad or contain bad edges, and by assumption $e$ contains no bad edges. So $e$ was removed for failing (FR2). In other words, we have $|e| \leqslant k-1$ and $e$ is neither bad nor contains a bad edge, but nevertheless there are many $(|e|+1)$-edges containing $e$ which either are bad or contain a bad edge.

Suppose that $f$ is a bad edge such that $|f \setminus e| = 1$. If $|f| = 1$, then there are at most $6\varepsilon_k^{1/k}\frac{t_f}{t}$ choices of $f$, each of which, by (4.8), is contained in at most

$$\prod_{i=2}^{|e|+1}\left(\left(1+\frac{2\varepsilon}{d_i}\right)\frac{d_i}{d_i^f}\right)^{\binom{|e|+1}{i}-\binom{|e|}{i}} \leqslant 2\prod_{i=2}^{|e|+1}\left(\frac{d_i}{d_i^f}\right)^{\binom{|e|}{i-1}} \tag{4.10}$$

edges of $\mathcal{P}_f^*$ of uniformity $|e|+1$ which contain $e$. If $|f| > 1$, then $f \cap e$ is non-empty and not a bad edge. There are at most $\binom{|e|}{\ell}$ choices of $f \cap e$ with $\ell$ elements, each of which by definition

is contained in less than

$$6\varepsilon_k^{1/k}(k-\ell)\frac{t_f}{t}\prod_{i=2}^{\ell+1}\left(\frac{d_i}{d_i^f}\right)^{\binom{\ell}{i-1}}$$

bad edges. Thus there are at most

$$\binom{|e|}{\ell}\cdot 6\varepsilon_k^{1/k}(k-\ell)\frac{t_f}{t}\prod_{i=2}^{\ell+1}\left(\frac{d_i}{d_i^f}\right)^{\binom{\ell}{i-1}}$$

choices of $f$, each of which is contained in at most

$$2\prod_{i=2}^{|e|+1}\left(\frac{d_i}{d_i^f}\right)^{\binom{|e|}{i-1}-\binom{\ell}{i-1}}$$

edges of $\mathcal{P}_f^*$ of uniformity $|e|+1$ which contain $e$.

Summing up, the number of $(|e|+1)$-edges containing $e$ which are either bad or contain a bad edge is at most

$$6\varepsilon_k^{1/k}\frac{t_f}{t}\cdot 2\prod_{i=2}^{|e|+1}\left(\frac{d_i}{d_i^f}\right)^{\binom{|e|}{i-1}}$$

$$+\sum_{\ell=1}^{|e|}\binom{|e|}{\ell}\cdot 6\varepsilon_k^{1/k}(k-\ell)\frac{t_f}{t}\left(\prod_{i=2}^{\ell+1}\left(\frac{d_i}{d_i^f}\right)^{\binom{\ell}{i-1}}\right)\cdot 2\prod_{i=2}^{|e|+1}\left(\frac{d_i}{d_i^f}\right)^{\binom{|e|}{i-1}-\binom{\ell}{i-1}}$$

$$\leqslant k16\varepsilon_k^{1/k}\frac{t_f}{t}\left(\sum_{\ell=0}^{|e|}\binom{|e|}{\ell}\right)\prod_{i=2}^{|e|+1}\left(\frac{d_i}{d_i^f}\right)^{\binom{|e|}{i-1}}=k2^{|e|+4}\varepsilon_k^{1/k}\frac{t_f}{t}\prod_{i=2}^{|e|+1}\left(\frac{d_i}{d_i^f}\right)^{\binom{|e|}{i-1}}.$$

This last equation simply states that $e$ does *not* fail (FR2), which is our desired contradiction. In particular, every 1-edge which is not bad is in $\mathcal{H}_{\varepsilon_k}(G;H,\mathcal{P}_f^*)$.

With this at hand we want to estimate the number of $(k-1)$-edges of $\mathcal{H}_{\varepsilon_k}$ that are not in $E_1$ (the same argument applies to any other). A bad $(k-1)$-edge is in the boundary of at least

$$6\varepsilon_k^{1/k}\frac{t_f}{t}\prod_{j=2}^{k-1}\left(\frac{d_j}{d_j^f}\right)^{\binom{k-1}{j-1}}$$

bad $k$-edges and a $k$-edge supports exactly $k$ different $(k-1)$-edges. With the bound on the number of bad $k$-edges in (4.9) it then follows that there are at most

$$k\varepsilon_k^{(k-1)/k}\left(\frac{t_f}{t}\right)^{k-1}\prod_{j=2}^{k-1}\left(\frac{d_j}{d_j^f}\right)^{\binom{k-1}{j}}$$

bad $(k-1)$-edges.

More generally, for $i = k - 2, \ldots, 1$, a bad $i$-edge is in the boundary of at least

$$6\varepsilon_k^{1/k}(k-i)\frac{t_f}{t}\prod_{j=2}^{i+1}\left(\frac{d_j}{d_j^f}\right)^{\binom{i}{j-1}}$$

bad $(i+1)$-edges and an $(i+1)$-edge supports exactly $i+1$ different $i$-edges. It follows that there are at most

$$k\underline{^{k-i}}\varepsilon_k^{i/k}\left(\frac{t_f}{t}\right)^i\prod_{j=2}^{i+1}\left(\frac{d_j}{d_j^f}\right)^{\binom{i}{j}}$$

bad $i$-edges. In particular, there are at most $k!\varepsilon_k^{1/k}\frac{t_f}{t}$ 1-edges removed from $\mathcal{H}$ to get $\mathcal{H}_{\varepsilon_k}$.

Then, with (4.8), for $i = 1, \ldots, k - 2$, there are at most

$$2 \cdot k\underline{^{k-i}}\varepsilon_k^{i/k}\left(\frac{t_f}{t}\right)^i\prod_{j=2}^{i+1}\left(\frac{d_j}{d_j^f}\right)^{\binom{i}{j}} \cdot \left(\frac{t_f}{t}\right)^{k-1-i}\prod_{j=2}^{k-1}\left(\frac{d_j}{d_j^f}\right)^{\binom{k-1}{j}-\binom{i}{j}}$$

$$\leqslant 2 \cdot k!\varepsilon_k^{i/k}\left(\frac{t_f}{t}\right)^{k-1}\prod_{j=2}^{k-1}\left(\frac{d_j}{d_j^f}\right)^{\binom{k-1}{j}}$$

fine $(k-1)$-edges in $\mathcal{H}$ supported by a bad $i$-edge, but no bad $i'$-edge for $1 \leqslant i' \leqslant i-1$. Combining this with the number of $(k-1)$-edges of $\mathcal{H}$ that are in $E_1$, which we can derive from (4.8) and the bound on $\varepsilon$, we get that there are at least

$$\left(1 - \varepsilon_k - 2\sum_{i=1}^{k-1}k!\varepsilon_k^{i/k}\right)\left(\frac{t_f}{t}\right)^{k-1}\prod_{j=2}^{k-1}\left(\frac{d_j}{d_j^f}\right)^{\binom{k-1}{j}} \geqslant \left(1 - kk!\varepsilon_k^{1/k}\right)\left(\frac{t_f}{t}\right)^{k-1}\prod_{j=2}^{k-1}\left(\frac{d_j}{d_j^f}\right)^{\binom{k-1}{j}}$$

$(k-1)$-edges in $\mathcal{H}_{\varepsilon_k}$ that are also contained in $E_1$.                                                    □

**Partitioning the vertex sets.** For $i = 1, \ldots, k$ we let $\mathcal{C}_i$ be the family of fine 1-cells in $V_i$ that correspond to 1-edges of $\mathcal{H}_{\varepsilon_k}(G; H, P_f^*)$. For building the paths we have to avoid $S$. To avoid clash of vertices during the expansion in the coarse partition, we set aside a small fraction of the fine cells that do not overlap with $S$ too much. For $i = 1, \ldots, k$ let $\overline{\mathcal{C}_i} \subseteq \mathcal{C}_i$ be selections of fine 1-cells $C$ with $|C \setminus S| \geqslant \varepsilon_k^{1/2k}\frac{n}{t_f}$ such that

$$\sum_{C \in \overline{\mathcal{C}_i}}|C| = \frac{\nu}{16k^2}|V_i| \pm \frac{t|V_i|}{t_f}.$$

This is possible, because by the condition on $S$ and $\varepsilon_k$ in (4.4) we have $|V_i \setminus S| \geqslant \eta\frac{n}{t} \geqslant 2\varepsilon_k^{1/2k}\frac{n}{t}$ and there can be at most $\frac{t_f}{t}\varepsilon_k^{1/2k}\frac{n}{t_f} = \varepsilon_k^{1/2k}\frac{n}{t}$ vertices of $V_i \setminus S$ in 1-cells $C \in \mathcal{C}_i$ with $|C \setminus S| < \varepsilon_k^{1/2k}\frac{n}{t_f}$. Then for $i = 1, \ldots, k$ we define $\overline{V}_i = \cup_{C \in \overline{\mathcal{C}_i}}C \setminus S$ and get

$$|\overline{V}_i| \geqslant \varepsilon_k^{1/2k}\frac{\nu}{16k^2}|V_i| - \frac{t|V_i|}{t_f} \geqslant \frac{\varepsilon_k^{1/2k}\nu}{20k^2}\frac{n}{t}. \tag{4.11}$$

We will use $\overline{\mathcal{C}_i}$ for the expansion and then in the end use $\mathcal{C}_i$ to reach a $(1-\nu)$-fraction of $E_k$.

We now want to argue that the fine 1-cells from the $\mathcal{C}_i$ are enough to reach a large fraction of any $E_i$. For this we let $\mathcal{H}' = \mathcal{H}'_{\varepsilon_k}(G; H, \mathcal{P}^*_f, \mathcal{C}_1, \ldots, \mathcal{C}_k)$ be the sub-multicomplex of $\mathcal{H}_{\varepsilon_k}$ induced by $\mathcal{C}_j$, $j = 1, \ldots, k$. For any $i = 1, \ldots, k$, by Claim 4.3, there are at most

$$2kk!\varepsilon_k^{1/k}\left(\frac{n}{t}\right)^{k-1}\prod_{j=2}^{k-1}d_j^{\binom{k-1}{j}} \leqslant \nu\frac{\varepsilon_k\nu^k}{10^\ell 20^k k^{2k}2}\left(\frac{n}{t}\right)^{k-1}\prod_{j=2}^{k-1}d_j^{\binom{k-1}{j}} \tag{4.12}$$

$(k-1)$-tuples in $E_i$ that are not contained in a $(k-1)$-cell of $\mathcal{H}_{\varepsilon_k}(G; H, \mathcal{P}^*_f)$, where the inequality follows from the choice of $\varepsilon_k$ in (4.4). Therefore, for $i = 1, \ldots, k$ at least $(1 - \frac{\nu}{4})|E_i|$ tuples from $E_i$ are contained in fine $(k-1)$-cells of $\mathcal{H}'$. This justifies that it will be sufficient to restrict the fine expansion to $\mathcal{H}'$.

For the coarse expansion, similarly to $\mathcal{H}'$, let $\overline{\mathcal{H}} = \overline{\mathcal{H}_{\varepsilon_k}}(G; H, \mathcal{P}^*_f, \overline{\mathcal{C}_1}, \ldots, \overline{\mathcal{C}_k})$ be the sub-multicomplex of $\mathcal{H}_{\varepsilon_k}$ induced by $\overline{\mathcal{C}_j}$, $j = 1, \ldots, k$. Then, let $\overline{E}_1, \ldots, \overline{E}_k$ be the $(k-1)$-cells of $H$ restricted to the 1-cells $\overline{V}_i$, where $\overline{E}_i \subseteq \prod_{j \neq i}\overline{V}_j$ for $i = 1, \ldots, k$. Then, for $i = 1, \ldots, k$, we have with Lemma 2.6

$$|\overline{E}_i| \geqslant \frac{1}{2}\prod_{j\neq i}|\overline{V}_i|\prod_{j=2}^{k-1}d_j^{\binom{k-1}{j}} \geqslant \frac{\varepsilon_k\nu^k}{20^k k^{2k}2}\left(\frac{n}{t}\right)^{k-1}\prod_{j=2}^{k-1}(d_j)^{\binom{k-1}{j}},$$

which with (4.12) gives that at least

$$(1 - \nu 10^{-\ell})|\overline{E}_i| \text{ tuples from } \overline{E}_i \text{ are contained in fine } (k-1)\text{-cells of } \overline{\mathcal{H}}. \tag{4.13}$$

This will be essential for the coarse expansion.

**Preparing for coarse expansion.** We will construct paths starting in a $(1-\nu)$-fraction of the tuples from $E_k$. In order to do this, we need to know that most of these tuples have high degree into $\overline{C}_k$, and that the tuples we then reach have high degree into $\overline{C}_1$, and so on. The following claim allows us to get this.

**Claim 4.4.** *Let* $\nu' \in (0,1)$ *with* $\nu' \geqslant \frac{1}{4}\varepsilon_k^{1/k}$ *and* $\nu' \geqslant 3\gamma_c\prod_{i=2}^{k-1}d_i^{-\binom{k}{i}}$. *Let* $U_i \subseteq V_i$ *for* $i = 1, \ldots, k$ *be subsets of size at least* $\frac{\varepsilon_k^{1/2k}\nu}{20k^2}\frac{n}{t}$ *and for* $i = 1, \ldots, k$ *let* $F_i$ *be the* $(k-1)$*-edges of* $\mathcal{H}'$ *in* $\prod_{j\neq i}U_j$. *Then for any* $i_1 \neq i_2$ *the number of* $(k-1)$*-tuples from* $F_{i_1}$ *which are contained in less than*

$$\frac{p}{2}|U_{i_1}|d\prod_{j=2}^{k-1}d_j^{\binom{k-1}{j-1}} \tag{4.14}$$

*edges of* $G$ *that are supported by* $\mathcal{H}_{\varepsilon_k}$ *and any* $(1-\nu')$*-fraction* $\hat{F}_{i_2}$ *of* $F_{i_2}$ *is at most a* $10\nu'$*-fraction of* $F_{i_1}$.

*Proof of Claim 4.4.* W.l.o.g. let $i_1 = 1$ and $i_2 = 2$. First we note that analogous to (4.13) with (4.12) we get that $F_1$ is at least a $(1 - \frac{\nu'}{10})$-fraction of all $(k-1)$-tuples of $\mathcal{H}$ supported

by the respective 1-cells. Now fix any $(1 - \nu')$-fraction $\hat{F}_2$ of $F_2$. Next let $\hat{F}_1 \subseteq F_1$ be those $(k-1)$-tuples in $F_1$ which are contained in less than

$$\frac{p}{2}|U_1|d\prod_{i=2}^{k-1} d_i^{\binom{k-1}{i-1}}$$

edges of $G$ that are supported by $\mathcal{H}$ and $\hat{F}_2$ and assume for a contradiction that

$$|\hat{F}_1| \geqslant 10\nu'|F_1| \geqslant \frac{89}{10}\nu' \prod_{i=2}^{k} |U_i| \prod_{i=2}^{k-1} d_i^{\binom{k-1}{i}},$$

where the latter inequality holds by Lemma 2.6 and as $10(1 - \gamma_f)(1 - \frac{\nu'}{10}) \geqslant \frac{89}{10}$. By the first part of Lemma 2.8 at least $(1 - \frac{1}{100})|\hat{F}_1|$ of these are contained in at least $(1 - \frac{1}{100})|U_1| \prod_{i=2}^{k-1} d_i^{\binom{k-1}{i-1}}$ $k$-vertex complete $(k-1)$-graphs in $\mathcal{H}$ using vertices from $U_1$.

On the other hand, with Lemma 2.6, there are at most

$$\nu'|F_2| \leqslant \left(1 + \frac{1}{100}\right)\nu' \prod_{i\neq 2} |U_i| \prod_{i=2}^{k-1} d_i^{\binom{k-1}{i}}$$

$(k-1)$-tuples in $F_2 \setminus \hat{F}_2$. By the first part of Lemma 2.8 all but $2\gamma_c|F_2|$ of these are contained in at most $(1 + \frac{1}{100})|U_2| \prod_{i=2}^{k-1} d_i^{\binom{k-1}{i-1}}$ $k$-vertex complete $(k-1)$-graphs. By Claim 4.3, the first part of Lemma 2.8, (4.4), and the bound on $\gamma_c$ there are at most

$$4k^2 6^k k! \varepsilon_k^{1/k} \left(\frac{n}{t}\right)^k \prod_{i=2}^{k-1} d_i^{\binom{k}{i}} + \gamma_c k^2 6^k k! \varepsilon_k^{1/k} \left(\frac{n}{t}\right)^k \prod_{i=2}^{k-1} d_i^{\binom{k-1}{i}} \leqslant \frac{\nu'}{2} \prod_{i=1}^{k} |U_i| \prod_{i=2}^{k-1} d_i^{\binom{k}{i}}$$

$k$-vertex complete $(k-1)$-graphs supported in $\mathcal{H}$ but not in $\mathcal{H}_{\varepsilon_k}$. So with $2\gamma_c(1 + \gamma_c) \leqslant \frac{\nu'}{2} \prod_{i=2}^{k-1} d_i^{\binom{k}{i}}$ this gives us at least

$$\left(1 - \frac{1}{100}\right)^2 |\hat{F}_1||U_1| \prod_{i=2}^{k-1} d_i^{\binom{k-1}{i-1}} - \left(1 + \left(1 + \frac{1}{100}\right)^2\right)\nu' \prod_{i=1}^{k} |U_i| \prod_{i=2}^{k-1} d_i^{\binom{k}{i}}$$

$$\geqslant \frac{3}{4}|\hat{F}_1||U_1| \prod_{i=2}^{k-1} d_i^{\binom{k-1}{i-1}}$$

$k$-vertex complete $(k-1)$-graphs in $\mathcal{H}_{\varepsilon_k}$ that are also in $\hat{F}_2$.

By definition of $\hat{F}_1$, they support at most

$$|\hat{F}_1| \cdot \frac{p}{2}|U_1|d\prod_{i=2}^{k-1} d_i^{\binom{k-1}{i-1}}$$

edges of $G$. This gives a $p$-density of at most $\frac{d}{2}(\frac{3}{4})^{-1} = \frac{2}{3}d$ and, therefore, by $(\varepsilon_k, 1)$-regularity of $H$ (recall that the relative $p$-density of $G$ with respect to $H$ is at least $d$), we have

$$\frac{3}{4}|\hat{F}_1||U_1|\prod_{i=2}^{k-1}d_i^{\binom{k-1}{i-1}} \leqslant \varepsilon_k\left(\frac{n}{t}\right)^k\prod_{i=2}^{k-1}d_i^{\binom{k}{i}},$$

which with (4.4) and $\nu' \geqslant \frac{1}{4}\varepsilon_k^{1/k}$ implies that

$$|\hat{F}_1| < 2\varepsilon_k\left(\frac{20k^2}{\varepsilon_k^{1/2k}\nu}\right)^k\prod_{i=2}^k|U_i|\prod_{i=2}^{k-1}d_i^{\binom{k-1}{i}} \leqslant 2\varepsilon_k^{1/k}\prod_{i=2}^k|U_i|\prod_{i=2}^{k-1}d_i^{\binom{k-1}{i}} \leqslant 8\nu'\prod_{i=2}^k|U_i|\prod_{i=2}^{k-1}d_i^{\binom{k-1}{i}}$$

and we have the desired contradiction. □

We let $\ell_0 = \ell - k + 1$. As discussed, there are two stages to building our paths. Given a $(k-1)$-tuple $\mathbf{x}$, we first look at all the ways to add $\ell_0$ vertices, all contained in the $\overline{V}_i$, to get tight paths. We then use the fine partition to complete these by adding $k-1$ vertices to get to a $(1-\frac{\nu}{4})$-fraction of most fine $(k-1)$-cells in $E_1$, which then gives a $(1-\nu)$-fraction of $E_1$.

We define the sequence $r_i$ for $i = 0, \ldots, \ell$ with $r_0 = k$ and $r_{i+1} \equiv r_i + 1 \pmod{k}$ for $i = 0, \ldots, \ell-1$. As $\ell \equiv 1 \pmod{k}$, we have $r_\ell = 1$ and $r_{\ell_0} = 2$.

Let $F_{\ell_0} \subseteq \overline{E}_2$ denote those $(k-1)$-edges which are contained in $(k-1)$-edges of $\overline{\mathcal{H}}$. That is, every $(k-1)$-edge of $F_{\ell_0}$ has vertices in $\overline{V}_1, \overline{V}_3, \ldots, \overline{V}_k$ and is in a fine $(k-1)$-cell which (and all of whose supporting $j$-cells) satisfies (FR2).

Let $\nu_0 = \nu$, for $i = 1, \ldots, \ell_0-1$ let $\nu_i = \nu_{i-1}/10$, and observe that with (4.4) and the bounds on $\nu$ and $\gamma_c$ we have that $\nu_i$ satisfies the requirements of Claim 4.4 for $i = 1, \ldots, \ell_0 - 1$. We apply Claim 4.4 to obtain that a $(1-\nu_{\ell_0+2k-4})$-fraction $F_{\ell_0-1}$ of the edges of $\overline{E}_{r_{\ell_0-1}}$ have high degree as in (4.14) into $\overline{V}_{r_{\ell_0+2k-4}}$ with respect to $F_{\ell_0}$. Repeating this for $i = \ell_0 - 2, \ldots, 2$ we get that a $(1-\nu_i)$-fraction $F_i$ of the edges of $\overline{E}_{r_i}$ have high degree as in (4.14) into $\overline{V}_{r_i}$ with respect to $F_{i+1}$. Finally, repeating the same procedure except replacing $\overline{E}_k$ with $E_k$, we arrive at a $(1-\nu)$-fraction $F_0$ of the tuples from $E_k$. This set $F_0$ is the set in the lemma statement from which we can construct paths, and we have verified that it is sufficiently large for the lemma statement. We now need to justify that we can indeed construct paths from any tuple in $F_0$ of length $\ell$ to most of $E_1$ avoiding any given small $S'$.

**Coarse expansion.** We fix any tuple $\mathbf{x} \in F_0$ and any set $S'$ of size at most $s$. We let $P$ be the set of tight paths starting at $\mathbf{x} = (x_1, \ldots, x_{k-1})$ which can be constructed as follows.

For each $1 \leqslant j \leqslant \ell_0$ in succession, we pick a vertex $x_{k+j-1}$ such that $\{x_j, \ldots, x_{k+j-1}\}$ is an edge of $G$, and $\{x_{j+1}, \ldots, x_{k+j-1}\}$ is an edge of $F_j$. In addition, we insist that $x_{k+j-1}$ is not in $S'$, nor equal to $x_i$ for any $i < k + j - 1$.

Because of our choice of $F_{j-1}$ to satisfy (4.14), at any given step $j$, the number of choices for $x_{k+j-1}$ is at least

$$\frac{p}{2}|\overline{V}_{r_j}|d\prod_{i=2}^{k-1}d_i^{\binom{k-1}{i-1}} - s - k - \ell_0 \geqslant \frac{p}{4}|\overline{V}_{r_j}|d\prod_{i=2}^{k-1}d_i^{\binom{k-1}{i-1}}.$$

It follows that

$$|P| \geqslant \prod_{j=1}^{\ell_0} \left( \frac{p}{4} |\overline{V}_{r_j}| d \prod_{i=2}^{k-1} d_i^{\binom{k-1}{i-1}} \right) \geqslant \left( \frac{1}{4} pd \frac{\varepsilon_k^{1/k} \nu n}{20 k^2 t} \prod_{i=2}^{k-1} d_i^{\binom{k-1}{i-1}} \right)^{\ell_0} \geqslant (\mu p n)^{\ell_0}$$

paths starting in $\mathbf{x}$ and ending with tuples in $F_{\ell_0} \subseteq \overline{E}_2$, where the last inequality follows from (4.5). By the good event of Lemma 2.13, since $\ell_0 \geqslant \frac{k-1}{\gamma}$ we get that for all $\mathbf{x}$ there are at least

$$\frac{\mu^{2\ell_0}}{8(2\ell_0)!} n^{k-1} \geqslant \delta \prod_{j \in \{1,3,\dots,k\}} |\overline{V}_j| \prod_{i=2}^{k-2} d_i^{\binom{k-1}{i}}$$

different end-tuples from the end-tuples above, where the lower bound follows from (4.5).

**Expansion in the fine partition.** We pick a $(k-1)$-cell $\hat{C}_0$ of $\overline{\mathcal{H}}$ such that at least $\delta|\hat{C}_0|$ of the $(k-1)$-edges in $\hat{C}_0$ are end-tuples of paths in $P$, which is possible by averaging. Let $R'_0$ denote the subset of $\hat{C}_0$ which are end $(k-1)$-tuples of paths in $P$.

We now consider all the $(k-1)$-cells $\hat{C}_{k-1}$ in $\mathcal{H}'$, whose 1-cells are contained in $\mathcal{C}_2 \setminus \overline{\mathcal{C}}_2, \dots, \mathcal{C}_k \setminus \overline{\mathcal{C}}_k$ respectively, which we can obtain by the following procedure. For each $1 \leqslant j \leqslant k-1$ in succession, we pick a $(k-1)$-cell $\hat{C}_j$ in $\mathcal{H}'$ such that $\hat{C}_{j-1}$ and $\hat{C}_j$ are in the boundary of some $k$-edge of $\mathcal{H}'$ and the new 1-cell is in $\mathcal{C}_{j+1} \setminus \overline{\mathcal{C}}_{j+1}$.

For any such choice of $\hat{C}_{k-1}$, we do the following. Let $m = \varepsilon_k^{1/2k} \frac{n}{t_f}$, choose $X_0, \dots, X_{k-2}$ subsets of the 1-edges of $\hat{C}_0$, each of size $m$, which contain a maximum number of the sets $R'_0$, and let the contained sets be $R_0$. By averaging, we have

$$|R_0| \geqslant \tfrac{1}{2} \delta m^{k-1} \prod_{i=2}^{k-1} d_i^{\binom{k-1}{i}}.$$

We choose any $X_{k-1}, \dots, X_{2k-3}$ subsets of the 1-edges of $\hat{C}_{k-1}$ each of size $m$. To complete the setup for Lemma 4.1, we put for each $2 \leqslant i \leqslant k-1$ an $i$-graph consisting of the restrictions of the $i$-cells in the union of the $\hat{C}_j$ to the $X_j$. We obtain the required $(\mathbf{d}_f, \sqrt{f}, 1)$-regularity from Lemma 2.7; and we let $G_k$ of Lemma 4.1 be the supported subgraph of $G$, which as previously observed is $\left( \frac{d}{2}, f_k, p \right)$-regular with respect to each of the required polyads. Thus Lemma 4.1 which returns a set $R_{k-1} \subseteq \hat{C}_{k-1}[X_{k-1}, \dots, X_{2k-3}]$ of size at least $(1-\delta)\big|\hat{C}_{k-1}[X_{k-1}, \dots, X_{2k-3}]\big|$ of $(k-1)$-tuples which, together with some tuple of $R_0$, make a tight path in $G$. Apart from the at most $k n^{k-1}$ of these tuples which share a vertex with $\mathbf{x}$, these are end-tuples of tight paths of length $\ell$ from $\mathbf{x}$ whose internal vertices avoid $S \cup S'$. By averaging, we conclude that at least $(1-2\delta)\big|\hat{C}_{k-1}\big|$ of the edges of $\hat{C}_{k-1}$ are ends of tight paths of length $\ell$ from $\mathbf{x}$ whose internal vertices avoid $S \cup S'$.

**Claim 4.5.** *There are at least*

$$\left(1 - \tfrac{\nu}{2k}\right) \left(\tfrac{t_f}{t}\right)^{k-1} \prod_{i=2}^{k-1} \left( \frac{d_i}{d_i^f} \right)^{\binom{k-1}{i}}$$

*valid choices of $\hat{C}_{k-1}$.*

Assuming this claim, by Lemma 2.6 we conclude that there are at least

$$\left(1 - \tfrac{\nu}{2k}\right)\left(\tfrac{t_f}{t}\right)^{k-1} \prod_{i=2}^{k-1} \left(\tfrac{d_i}{d_i^f}\right)^{\binom{k-1}{i}}(1 - 2\delta)(1 - \gamma_f)\left(\tfrac{n}{t_f}\right)^{k-1} \prod_{i=2}^{k-1}(d_i^f)^{\binom{k-1}{i}} \geqslant \left(1 - \tfrac{\nu}{2}\right)\left(\tfrac{n}{t}\right)^{k-1} \prod_{i=2}^{k-1} d_i^{\binom{k-1}{i}}$$

$(k - 1)$-tuples in $E_1$ which are ends of tight paths from $\mathbf{x}$ whose internal vertices avoid $S \cup S'$, as required.

*Proof of Claim 4.5.* Recall that for each $i \in [k]$ we have

$$\sum_{C \in \overline{\mathcal{C}}_i} |C| \leqslant \tfrac{\nu}{16k^2}|V_i| + t|V_i|/t_f \,.$$

Thus the number of fine 1-cells in $\overline{\mathcal{C}}_i$ is at most $\tfrac{\nu}{8k^2} \cdot \tfrac{t_f}{t}$ for each $i$, and so the number of $(k-1)$-cells of $\mathcal{H}$ which cannot be $\hat{C}_{k-1}$ due to being supported by a 1-cell in some $\overline{\mathcal{C}}_i$ is at most

$$2k\tfrac{\nu}{8k^2}\left(\tfrac{t_f}{t}\right)^{k-1} \prod_{i=2}^{k-1} \left(\tfrac{d_i}{d_i^f}\right)^{\binom{k-1}{i}} \,.$$

We consider constructing the $\hat{C}_j$ in order $j = 1, \ldots, k - 1$, ignoring the restriction of not using 1-cells in the $\overline{\mathcal{C}}_i$, and at each step $j$ keep track of the number of ways to construct $\hat{C}_j$ which will lead to different choices for $\hat{C}_{k-1}$.

Given $j \geqslant 1$, suppose we have fixed $\hat{C}_{j-1}$. This means that we have already fixed the $(j - 1)$-cell supporting $\hat{C}_{k-1}$ on $V_2, \ldots, V_j$, and when we choose $\hat{C}_j$ we will fix the $i$-cells supporting $\hat{C}_{k-1}$ for each $1 \leqslant i \leqslant j$ that are on the vertex set $V_{j+1}$ and some $i - 1$ of $V_2, \ldots, V_j$. Suppose that we have a particular choice $X$ of all these $i$-cells with $1 \leqslant i \leqslant j$ that are on the vertex set $V_{j+1}$ and some $i - 1$ of $V_2, \ldots, V_j$, which is consistent with the choice of $\hat{C}_{j-1}$ (i.e. each $i$-cell in $X$ has boundary whose $(i-1)$-cell not on $V_{j+1}$ is in the support of $\hat{C}_{j-1}$). Consider the number of $k$-polyads in $\mathcal{H}$ on $V_1, \ldots, V_k$ which are consistent with both $\hat{C}_{j-1}$ and $X$. The number of these is at most

$$2\prod_{i=2}^{k-1} \left(\tfrac{d_i}{d_i^f}\right)^{\binom{k-1}{i-1}-\binom{j-1}{i-1}} \,.$$

Suppose that more than

$$k2^{k+4}\varepsilon_k^{1/k}\left(\tfrac{t_f}{t}\right)\prod_{i=2}^{k-1} \left(\tfrac{d_i}{d_i^f}\right)^{\binom{j-1}{i-1}}$$

choices of $X$ are not consistent with any $k$-edge of $\mathcal{H}'$ whose boundary contains $\hat{C}_{j-1}$. Then in particular $\hat{C}_{j-1}$ is contained in less than

$$\left(1 - k2^{k+4}\varepsilon_k^{1/k}\right)\left(\tfrac{t_f}{t}\right)\prod_{i=2}^{k-1} \left(\tfrac{d_i}{d_i^f}\right)^{\binom{k-1}{i-1}}$$

$k$-edges of $\mathcal{H}'$, which is a contradiction to (FR2).

We conclude that the number of ways to construct $\hat{C}_{k-1}$, ignoring the restriction to avoid the $\overline{\mathcal{C}}_i$, is at least

$$\prod_{j=1}^{k-1}\left(\left(1-k2^{k+5}\varepsilon_k^{1/k}\right)\left(\tfrac{t_f}{t}\right)\prod_{i=2}^{k-1}\left(\tfrac{d_i}{d_i^f}\right)^{\binom{j-1}{i-1}}\right)\geqslant\left(1-k^22^{k+5}\varepsilon_k^{1/k}\right)\left(\tfrac{t_f}{t}\right)^{k-1}\prod_{i=2}^{k-1}\left(\tfrac{d_i}{d_i^f}\right)^{\binom{k-1}{i}}.$$

Finally the number of valid choices of $\hat{C}_{k-1}$ is at least

$$\left(1-\tfrac{\nu}{4k}-k^22^{k+5}\varepsilon_k^{1/k}\right)\left(\tfrac{t_f}{t}\right)^{k-1}\prod_{i=2}^{k-1}\left(\tfrac{d_i}{d_i^f}\right)^{\binom{k-1}{i}}\geqslant\left(1-\tfrac{\nu}{2k}\right)\left(\tfrac{t_f}{t}\right)^{k-1}\prod_{i=2}^{k-1}\left(\tfrac{d_i}{d_i^f}\right)^{\binom{k-1}{i}},$$

as required.                                                                                                                              □

□

## 5. Proof of reservoir lemma

In this section we prove Lemma 2.23. Before giving the details we outline the strategy of the proof. We fix $G$ and let $R\subseteq V(G)$ be a set of size $r=|R|\leqslant\nu n$. For every $u\in R$ we need a reservoir path $P_u$ with reservoir set $\{u\}$ on a constant number of vertices with end-tuples $\mathbf{v}_u$ and $\mathbf{w}_u$; this is a tight path with end-tuples $\mathbf{v}_u$ and $\mathbf{w}_u$ and vertex set $V(P_u)$ such that there also exists a tight path with the same end-tuples and vertex set $V(P_u)\setminus\{u\}$.

To build $P_{\mathrm{res}}$ we begin with an arbitrary $(k-1)$-tuple $\mathbf{v}=\mathbf{w}$, which is a reservoir path with an empty reservoir set, and call this $P_0$. Assume we have built a reservoir path $P_{i-1}$ with reservoir set $R'\subseteq R$ of size $i-1$ and end-tuples $\mathbf{v}_{i-1}$ and $\mathbf{w}_{i-1}$ such that $V(P_{i-1})$ does not intersect $R\setminus R'$. Then, for some $u\in R\setminus R'$, we construct a reservoir path $P_u$ with end-tuples $\mathbf{v}_u$ and $\mathbf{w}_u$ that is disjoint from $P_{i-1}$. If $i$ is odd we connect $\mathbf{w}_{i-1}$ to $\mathbf{v}_u$ by a tight path (using Lemma 2.19) and let $\mathbf{w}_i=\mathbf{w}_u$ and $\mathbf{v}_i=\mathbf{v}$. If $i-1$ is even we connect $\mathbf{v}_{i-1}$ to $\mathbf{w}_u$ by a tight path and let $\mathbf{v}_i=\mathbf{v}_u$ and $\mathbf{w}_i=\mathbf{w}$. In both cases we obtain a reservoir path $P_i$ with reservoir set $R'\cup\{u\}$, end-tuples $\mathbf{v}_i$ and $\mathbf{w}_i$, and continue. By alternating between the endpoints we ensure that the end-tuples are always $(\varepsilon',p,\ell')$-good for $V(P'_{\mathrm{res}})$.

Finally, let us sketch how we construct $P_u$, a picture of which (for $k=5$) is in Figure 5.1. We begin by finding a $(2k-1)$-vertex tight path with $u$ its central vertex; this gives the spikes $\mathbf{u}$ and $\mathbf{x}_1$ in the figure. We look at all the ways to fill in the upper and lower spike paths in the figure. Using Lemma 2.13, we see that from each we can get to a positive density of end-tuples. In particular, we can get to a positive density of each of two vertex-disjoint coarse $(k-1)$-cells in the regular partition, and two applications of Lemma 4.1 gives us the tuple $\mathbf{v}$ connecting the paths, completing the spikes. We then use Lemma 2.19 repeatedly to create the paths between pairs of spikes. The only point where we need to be a bit careful is to ensure that when creating the upper and lower spike paths, and when connecting them, we do not reuse vertices; for this purpose we randomly split the vertex set into three parts and use one for each of the upper spike path, the lower spike path, and the connection.

*Proof of Lemma 2.23.* Let $\gamma > 0$, $k \geqslant 3$ and $\ell'$ be integers. Then let $\ell$ be the smallest multiple of $k-1$ which is both larger than $\ell'$ and sufficiently large for Lemma 2.19 with input $k$ and $\gamma$. Let $\delta = 10^{-4\ell}/(32(2\ell)!k!)$.

Set $c = \frac{\ell}{k-1}\big(k-1+\ell\big) + 2k-1+\ell$, and set $s = 4c$. Let $0 < \varepsilon' \leqslant \frac{1}{4}\gamma$ and $0 < d \leqslant \frac{1}{8}\gamma$ be given.

Next, we let $t_0$ be large enough for Lemma 2.5 with input $k$, $\gamma$, and $d$ and for Lemma 2.19 with input as above, $d$, $\eta = \frac{1}{2}$, and $\varepsilon'$. Now let $0 < \nu_{\mathrm{L2.19}} < \frac{1}{4}\gamma$ and $0 < \varepsilon_k < (\gamma 2^{-k-4})^k$ be small enough for Lemma 2.19 with input as above, and for $\sqrt{\varepsilon_k}$ to play the role of $f'_k$ in Lemma 4.1 with input $k, \delta, d$. Let $\nu = \frac{1}{10c}\nu_{\mathrm{L2.19}}$. In addition let $\varepsilon, f, f_k : \mathbb{N} \to (0,1]$ tend to zero sufficiently fast for Lemma 2.19 and such that $\varepsilon(t)$ is small enough for Lemma 2.7 with input $k$, $\alpha = \frac{1}{100}$ and density $d_0 = \frac{1}{t}$, and such that for any $d_0$, if $1/t < d_0$ then $\sqrt{\varepsilon(t)}$ is small enough to play the role of $f'$ in Lemma 4.1, and be such that Lemma 2.5 is applicable with inputs as above and minimum degree $\frac{1}{2} + \gamma$, $\varepsilon_k$, $10c\nu$. Finally let $\eta_{\mathrm{L2.5}}$ be given by Lemma 2.5.

Given $n$, let $p \geqslant n^{-1+\gamma}$. Suppose that $\Gamma = G^{(k)}(n,p)$ is in the good events of Lemma 2.19 with input as above and Lemma 2.12 with input $\frac{1}{2}\varepsilon'$ and $k$, and Lemma 2.13 with input $\gamma$, $k$, both $\ell$ and $\ell + 1 - k$, and $\mu = \frac{1}{100}$, with Lemma 2.11, that $\Gamma$ and all of its subgraphs are $(\eta_{\mathrm{L2.5}}, p)$-upper regular. Suppose $G \subseteq \Gamma$ with $\delta_{k-1}(G) \geqslant (\frac{1}{2}+\gamma)pn$, that $(\mathcal{P}^*_c, \mathcal{P}^*_f)$ is a $\big(t_0, t_1, t_2, \varepsilon_k, \varepsilon(t_1), f_k(t_1), f(t_2), p\big)$-strengthened pair for $G$, and that $t$ is the number of 1-cells in $\mathcal{P}^*_c$. Let $\mathcal{R} = \mathcal{R}_{\varepsilon_k, d}(G; \mathcal{P}^*_c, \mathcal{P}^*_f)$ be the $(\varepsilon_k, d)$-reduced multicomplex of $G$ and $S$ be the union of the vertices that are not contained in 1-cells of $\mathcal{R}$. Then, by Lemma 2.5, $\mathcal{R}$ contains at least $(1 - 4\varepsilon_k^{1/k})t$ 1-edges and every induced subcomplex on at least $(1 - 10c\nu)t$ 1-cells is tightly linked. We let $S$ be the set of vertices of $V(G)$ that are not in 1-edges of $\mathcal{R}$ and note $|S| \leqslant 4\varepsilon_k^{1/k}n \leqslant \frac{1}{4}n$. Finally let $R \subseteq V(G)$ with $|R| \leqslant \nu n$.

Our goal is to construct a reservoir path $P_{\mathrm{res}}$ in $G$ with reservoir set $R$ and ends $\mathbf{v}$ and $\mathbf{w}$, such that $|V(P_{\mathrm{res}})| \leqslant c|R|$. Moreover, we require that $\mathbf{v}$ and $\mathbf{w}$ are $(\varepsilon', p, \ell')$-good for $V(P_{\mathrm{res}}) \cup S$. For the construction of the path $P_{\mathrm{res}}$ we proceed as outlined above. We start with an arbitrary $(k-1)$-tuple outside $R$ that is $(\frac{1}{2}\varepsilon', p, \ell)$-good for $S \cup R$ (which exists by the good event of Lemma 2.12), denote it by $\mathbf{v}_0$, and set $\mathbf{w}_0 = \overleftarrow{\mathbf{v}}_0$. Let this tuple be $P_0$, and let $S_{-1} := S \cup R$.

Suppose that for some $i \geqslant 1$ we have constructed a reservoir path $P_{i-1}$ whose ends $\mathbf{v}_{i-1}$ and $\mathbf{w}_{i-1}$ are both $(\frac{1}{2}\varepsilon', p, \ell)$-good for $S_{i-2}$, with $|V(P_{i-1})| \leqslant c(i-1)$, whose reservoir set is some $i$ vertices of $R$ and which does not intersect $R$ outside its reservoir set. Let $S_{i-1} := S_{i-2} \cup V(P_{i-1})$.

We first choose any $u \in R \setminus S_{i-1}$ and construct a reservoir path $P_u$ with reservoir set $\{u\}$ (see Figure 5.1) disjoint from $S_{i-1}$.

Let $S^*$ be the set of vertices that are contained in $(k-1)$-tuples that are not $(\frac{1}{2}\varepsilon', p, \ell)$-good for $S_{i-1}$. Note that we have $|S^*| = o(n)$ by the good event of Lemma 2.12 assumed above. Let $S' := S_{i-1} \cup S^*$. We will construct $P_u$ disjoint from $S' \setminus \{u\}$. This automatically means that any $(k-1)$-tuple of vertices in $P_u$ is $(\frac{1}{2}\varepsilon', p, \ell)$-good for $S_{i-1}$.

We choose vertices $u_1, \ldots, u_{k-2}$ from $V(G) \setminus (S^* \cup S_{i-1})$ such that the tuple $(u, u_1, \ldots, u_{k-2})$ is $(\frac{1}{2}\varepsilon', p, \ell_s + k - 1)$-good for $S'$. This is possible since (however we choose $u_1, \ldots, u_{k-3}$) when we come to choose $u_{k-2}$ we have at least $\frac{1}{2}n$ vertices to choose from, and by the good event of Lemma 2.12 at most $o(n)$ of these can give a $(k-1)$-tuple which is not $(\frac{1}{2}\varepsilon', p, \ell_s + k - 1)$-good for $S'$.

By the minimum degree of $G$, the $(k-1)$-tuple $(u, u_1, \ldots, u_{k-2})$ is contained in at least $\frac{1}{2}pn$
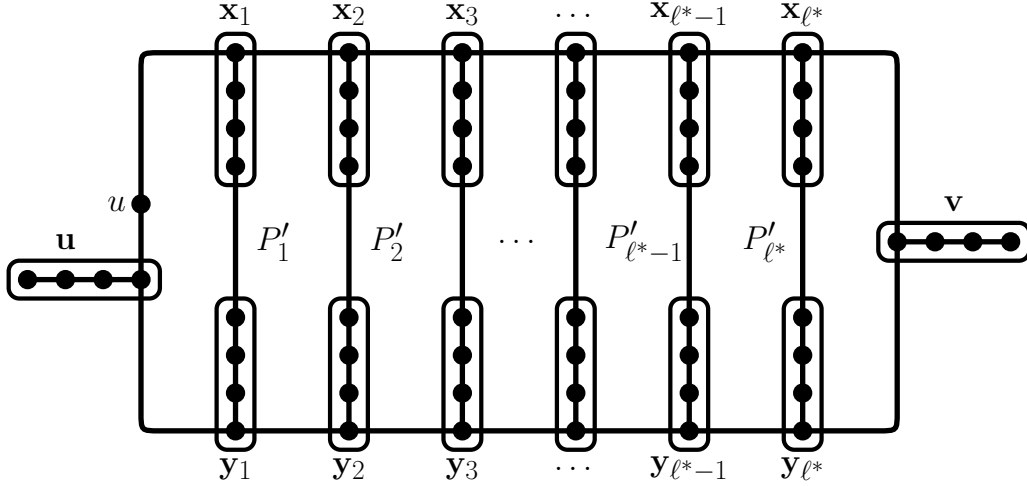
Figure 5.1: Reservoir structure in the case $k = 5$ with $\ell^* = \ell/(k-1)$ for one vertex $u$ with two tight paths that both have end-tuples $\mathbf{u}$ and $\mathbf{v}$, where one is using all vertices and the other all but $u$.

edges of $G$. Since $(u, u_1, \ldots, u_{k-2})$ is $(\frac{1}{2}\varepsilon', p, \ell_s + k - 1)$-good for $S'$, at most $p|S'| + \frac{1}{2}\varepsilon'pn$ of these edges contain a vertex of $S'$ and at most another $\frac{1}{2}\varepsilon'pn$ of these edges contain a $(k-1)$-set which is not $(\frac{1}{2}\varepsilon', p, \ell_s + k - 2)$-good for $S'$. Pick two of these vertices not in $S'$ that are in good tuples, and let them be $u_{k-1}$ and $x_{1,1}$.

We now, for each $2 \leqslant j \leqslant k-1$ in succession, choose a vertex $x_{1,j}$ such that $\{u_{k-j-1}, \ldots, u_1,$ $u, x_{1,1}, \ldots, x_{1,j}\}$ is an edge of $G$, such that $x_{1,j}$ is not in $S' \cup \{u_1, \ldots, u_k\}$, and such that if $j \leqslant k-2$ then $\{u_{k-j-2}, \ldots, u_1, u, x_{1,1}, \ldots, x_{1,j}\}$ is $(\frac{1}{2}\varepsilon', p, \ell + k - j - 1)$-good for $S'$, while if $j = k-1$ then $\{x_{1,1}, \ldots, x_{1,j}\}$ is $(\frac{1}{2}\varepsilon', p, \ell + k - j - 1)$-good for $S'$. This is possible for each $j$ by the same argument as above.

At this point, we have constructed the top left part of Figure 5.1, consisting of the spikes $\mathbf{u} = (u_1, \ldots, u_{k-1})$ and $\mathbf{x}_1 = (x_{1,1}, \ldots, x_{1,k-1})$. Our next goal is to construct the remaining spikes.

To begin with, consider constructing spike paths starting from $\mathbf{x}_1$ of type 1, with $\ell + 1 - k$ vertices. We can do this greedily, adding at each step $j$ one more vertex and one more edge. We insist on choosing our new vertex outside $S'$, and we insist on all the $(k-1)$-subsets of the $j$th edge being $(\frac{1}{2}, p, \ell - j)$-good for $S'$. This is possible for each $j \leqslant \ell + 1 - k$ by the same argument as above, and critically at each step we have a choice of at least $\frac{1}{10}pn$ vertices that satisfy these conditions. The same statement is true for spike paths of length $\ell$ starting from $\mathbf{u}$.

To avoid clashes between the spike paths we construct, let $Z_1, Z_2, Z_3$ be subsets of $V(G)$ obtained by putting each vertex of $G$ independently into one of the three sets with equal probability $\frac{1}{3}$. By the Chernoff bound and the union bound, with probability at least $1 - n^k \exp\left(-\frac{1}{1600}pn\right)$, the following holds. When we construct spike paths from $\mathbf{x}$ satisfying the above conditions and in addition with all new vertices in $Z_1$ greedily, at each step we have at least $\frac{1}{100}pn$ choices. Similarly, when we construct spike paths from $\mathbf{u}$ satisfying the above conditions and in addition with all new vertices in $Z_2$ greedily, at each step we have

at least $\frac{1}{100}pn$ choices. Finally, for each cluster $V' \in \mathcal{P}_c$ such that $|V' \setminus S'| > \varepsilon'|V'|$ and each $1 \leqslant j \leqslant 3$, we have $|(V' \setminus S') \cap Z_j| \geqslant \frac{1}{4}|V' \setminus S'|$. Suppose that our choice of $Z_1, Z_2, Z_3$ is such that all these statements hold.

By Lemma 2.13, there is a set $X$ of $(k-1)$-tuples which are end-tuples of spike-paths starting from $\mathbf{x}_1$ of type 1, with $\ell + 1 - k$ vertices other than those in $\mathbf{x}_1$, none of whose vertices are in $S'$ and all of whose vertices outside $\mathbf{x}_1$ are in $Z_1$, such that $|X| \geqslant \frac{10^{-4\ell-4+4k}}{8(2\ell+2-2k)!}n^{k-1} \geqslant \delta n^{k-1}$. Similarly, there is a set $U$ of $(k-1)$-tuples which are end-tuples of spike-paths starting from $\mathbf{u}$ of type 1, with $\ell$ vertices, none of whose vertices are in $S'$ and all of whose vertices outside $\mathbf{u}$ are in $Z_2$, such that $|U| \geqslant \frac{10^{-4\ell}}{8(2\ell)!}n^{k-1} \geqslant \delta n^{k-1}$.

We now aim to find two disjoint $(k-1)$-cells $C_x$ and $C_u$ of $\mathcal{R}$ with the following properties. At least $2k!\delta|C_x|$ of the $(k-1)$-tuples in $X$ are orderings of edges of $C_x$; and at least $2k!\delta|C_u|$ of the $(k-1)$-tuples in $U$ are orderings of edges of $C_u$; and for each cluster $V'$ supporting either $C_u$ or $C_x$, we have $|V' \setminus S'| \geqslant \frac{1}{2}|V'|$.

Observe that since $|S'| \leqslant 2c|R|$ and $|R| \leqslant \nu n$ we have $|S'| \leqslant 2c\nu n$, and consequently there are at most $5c\nu t - k$ clusters $V'$ of $\mathcal{R}$ with $|V' \setminus S'| < \frac{1}{2}|V'|$. The total number of $k$-sets intersecting $S'$ or 1-cells with too many elements in $S'$ is at most $7c\nu n^k$, while by choice of $\nu$ the number of $k$-sets which are not supported by the $(k-1)$-cells of $\mathcal{R}$ is at most $\nu n^k$. By choice of $\delta$, and since $|X| \geqslant 4\delta\binom{n}{k-1}k!$, at least $3\delta\binom{n}{k-1}k!$ of the $(k-1)$-tuples in $X$ are on $(k-1)$-sets not intersecting $S'$, and by averaging the desired $C_x$ exists. A similar calculation, this time removing additionally elements of $U$ which lie in the clusters of $C_x$, gives $C_u$. Let $X''$ denote the tuples in $X$ which are orderings of edges of $C_x$. Choose an ordering $(V_{k-1}, \ldots, V_1)$ of the clusters of $C_x$ which is consistent with most tuples of $X''$, and let $X'$ be the consistently ordered tuples of $X''$. Thus a spike path with end in $X'$ has its final vertex in $V_1$. We obtain $|X'| \geqslant 2\delta|C_x|$. Similarly, we choose an ordering $(V_{2k-1}, \ldots, V_{3k-3})$ of the clusters of $C_u$ and let $U'$ be the consistently ordered tuples of $U$ in $C_u$, obtaining $|U'| \geqslant 2\delta|C_u|$.

Now by construction of $\mathcal{R}$ there is a tight link between $C_x$ and $C_u$ with the given orderings that does not use any cluster $V'$ with $|V' \setminus S'| < \frac{1}{2}|V'|$, because $|S'| < 2c\nu n$ and hence less than $5c\nu t$ clusters are more than half covered by $S'$. Let the clusters witnessing this tight link be $V_k, \ldots, V_{2k-2}$, and let $C'$ be the $(k-1)$-cell in the tight link whose clusters are $V_k, \ldots, V_{2k-2}$. Thus for each $1 \leqslant j \leqslant k-1$ the set $\{V_j, \ldots, V_{j+k-1}\}$ and $\{V_{2k-1-j}, \ldots, V_k, V_{2k-1}, \ldots, V_{2k-2+j}\}$ are $k$-edges of $\mathcal{R}$.

We now choose subsets $V_1', \ldots, V_{3k-3}'$ of $V_1, \ldots, V_{3k-3}$ respectively, each of size $\frac{n}{10t}$ and disjoint from $S'$, with $V_1', \ldots, V_{k-1}'$ in $Z_1$, with $V_k', \ldots, V_{2k-2}'$ in $Z_3$, and with $V_{2k-1}', \ldots, V_{3k-3}'$ in $Z_2$, and such that $|X' \cap C_x[V_1', \ldots, V_{k-1}']| \geqslant 2\delta|C_x[V_1', \ldots, V_{k-1}']|$ and $|U' \cap C_u[V_{2k-1}', \ldots, V_{3k-3}']| \geqslant 2\delta|C_u[V_{2k-1}', \ldots, V_{3k-3}']|$. Note that it is possible to find sets of the given sizes by choice of the $V_i$ and definition of the $Z_j$; while the condition about $X'$ and about $U'$ is satisfied on average and hence a choice exists. By Lemma 2.7 all the $j$-cells with $2 \leqslant j \leqslant k-1$ between these chosen sets are $(d_j, \sqrt{\varepsilon}, 1)$-regular, and by definition of regularity it follows that the supported $k$-polyads corresponding to $k$-edges of $\mathcal{R}$ are $(d, \sqrt{\varepsilon_k}, p)$-regular.

By Lemma 4.1, with $R_0 = X' \cap C_x[V_1', \ldots, V_{k-1}']$, there is a set $R \subseteq C'$ with $|R| \geqslant (1-\delta)|C'|$ such that for each $(v_1, \ldots, v_{k-1}) \in R$ there is a tight path from a tuple of $C_x[V_1, \ldots, V_{k-1}]$ to $(v_1, \ldots, v_{k-1})$. Applying Lemma 4.1 again, with $R_0 := R$, there is a set $R' \subseteq C_u$ with $|R'| \geqslant (1-\delta)|C_u|$ such that for each $(y_{\ell/(k-1),1}, \ldots, y_{\ell/(k-1),k-1}) \in R'$ there is a tight path from $R$ to

$(y_{\ell/(k-1),1}, \ldots, y_{\ell/(k-1),k-1})$. We can choose $\mathbf{y}_{\ell/(k-1)} := (y_{\ell/(k-1),1}, \ldots, y_{\ell/(k-1),k-1}) \in U' \cap R'$, and obtain the corresponding $\mathbf{v} := (v_1, \ldots, v_{k-1}) \in R$ which in turn gives us $\mathbf{x}_{\ell/(k-1)} := (x_{\ell/(k-1),1}, \ldots, x_{\ell/(k-1),k-1}) \in X'$. This structure is the three right-hand-most spikes of Figure 5.1. By definition of $X'$ and $U'$, and since $Z_1, Z_2, Z_3$ are by construction disjoint, there exist vertex-disjoint spike-paths completing the spikes of Figure 5.1, none of whose vertices are in $S'$. That is, we find $(k-1)$-tuples $\mathbf{x}_2, \ldots, \mathbf{x}_{\ell/(k-1)-1}$ and $\mathbf{y}_1, \ldots, \mathbf{y}_{\ell/(k-1)-1}$ such that $\mathbf{x}_1, \ldots, \mathbf{x}_{\ell/(k-1)}$ form a spike path, and $\mathbf{y}_1, \ldots, \mathbf{y}_{\ell/(k-1)}$ form a spike path.

Next, for $j = 1, \ldots, \ell/(k-1)$, we use Lemma 2.19 to connect the tuple $\mathbf{x}_j$ to the tuple $\overleftarrow{\mathbf{y}}_j$ by a tight path $P'_j$ of length $\ell$ with internal vertices not in $S \cup S_{i-1}$ and any of the previously chosen vertices. Finally, we denote the whole structure by $P_u$ and note $|V(P_u)| = \frac{\ell}{k-1}(k-1+\ell)+2k-1$. Observe (see Figure 5.1) that there are two tight paths with end-tuples $\mathbf{u}$ and $\mathbf{v}$; one with vertex set $V(P_u)$ and one with vertex set $V(P_u) \setminus \{u\}$, i.e. $P_u$ is a reservoir path with reservoir set $\{u\}$.

To finish the step, we use Lemma 2.19 to connect $\mathbf{u}$ to one of the ends of $P_{i-1}$, either $\mathbf{u}_{i-1}$ if $i$ is odd, or $\mathbf{w}_{i-1}$ if $i$ is even. Repeating until $i = |R|$ proves the lemma.                              $\square$

## A. Regularity lemmas and properties

We first prove Lemma 2.4. This we do in two steps: first, we prove the special case that $p_i = 1$ for each $1 \leqslant i \leqslant s$ and the $G_i$ are edge-disjoint (the *dense disjoint case*), and then we deduce from this special case the general case. Note that in the dense case the assumption of upper regularity is trivially satisfied with $p_i = \eta = 1$.

To prove the dense disjoint case of Lemma 2.4, we use a standard approach, borrowed from [AFKS00]. That is, we begin with the input family of partitions $\mathcal{Q}^*$, and iteratively apply the Strong Hypergraph Regularity Lemma of Rödl and Schacht [RS07], obtaining a collection of families of partitions $\mathcal{P}_1^*, \mathcal{P}_2^*$ and so on, where $\mathcal{P}_1^*$ refines $\mathcal{Q}^*$ and for each $j \geqslant 2$ the family $\mathcal{P}_j^*$ refines $\mathcal{P}_{j-1}^*$. We choose parameters for these applications of the Regularity Lemma such that for each $j \geqslant 1$ the pair $(\mathcal{P}_j^*, \mathcal{P}_{j+1}^*)$ satisfies the regularity properties of being a strengthened pair for each $G_i$. Then for each $j \geqslant 1$, one of the following two things occurs. First, the density property of being a strengthened pair holds for each $G_i$. Second, there is some $G_i$ for which the density property does not hold. We will define an energy $\mathcal{E}_j$ of the family of partitions $\mathcal{P}_j^*$, and see that in the second case $\mathcal{E}_{j+1}$ is significantly larger than $\mathcal{E}_j$. Since we will see $\mathcal{E}_j$ is bounded above by $s$, we conclude that the first case must occur for some bounded $j$, and the lemma follows. We now quote the Strong Hypergraph Regularity Lemma from [RS07], and give the details.

**Lemma A.1** ([RS07, Lemma 23]). *Let $k \geqslant 2$ be a fixed integer. For all positive integers $t_0$ and $s$, and all $\varepsilon_k > 0$ and functions $\varepsilon : \mathbb{N} \to (0,1]$, there are integers $t_1$ and $n_0$ such that the following holds for all $n \geqslant n_0$ which are divisible by $t_1!$. Let $V$ be a vertex set of size $n$, let $G_1, \ldots, G_s$ be edge-disjoint $k$-uniform hypergraphs on $V$, and suppose $\mathcal{Q}^*$ is a $(1, t_0, \eta)$-equitable family of partitions on $V$. Then there exists a family of partitions $\mathcal{P}^*$ refining $\mathcal{Q}^*$ such that $\mathcal{P}^*$ is $\big(t_0, t_1, \varepsilon(t_1)\big)$-equitable, and for each $1 \leqslant i \leqslant s$, the hypergraph $G_i$ is $(\varepsilon_k, 1)$-regular with respect to $\mathcal{P}^*$.*

Note that in [RS07] this lemma is stated for $k \geqslant 3$; the case $k = 2$ is the (much older) Szemerédi Regularity Lemma, in which the 'families of partitions' are simply vertex partitions

and the function $\varepsilon$ plays no rôle.

*Proof of Lemma 2.4, dense disjoint case.* Given $k \geqslant 2$, $q$, $t_0$, $s$ integers and $\varepsilon_k > 0$, and functions $f, f_k, \varepsilon : \mathbb{N} \to (0, 1]$, we define sequences of numbers $t_1, t_2, \ldots$ and $n_1, n_2, \ldots$ as follows. For each $j \geqslant 1$, let $t_j$ and $n_j$ be returned by Lemma A.1 with input $k, t_{j-1}, s, \min\big(\varepsilon_k, f_k(t_0)\big)$ and the function $\min(\varepsilon, f)$.

Let Lemma 2.4 return the parameters $\eta = 1$, $T = t_{s\varepsilon_k^{-4}+2}$, and $n_0 = \max(n_1, \ldots, n_{s\varepsilon_k^{-4}+2})$. Given an initial family of partitions $\mathcal{Q}^*$ which is $(1, t_0, \eta)$-equitable and edge-disjoint $k$-uniform hypergraphs $G_1, \ldots, G_s$, we proceed as follows.

We apply Lemma A.1, with input $k, t_0, s, \min\big(\varepsilon_k, f(t_0)\big)$ and $\varepsilon$, with input family of partitions $\mathcal{Q}^*$, to the hypergraphs $G_1, \ldots, G_s$. Let $\mathcal{P}_1^*$ be the returned family of partitions.

We now, for each $j \geqslant 1$ successively such that the conditions are met, apply Lemma A.1, with input $k, t_j, s, \min\big(\varepsilon_k, f_k(t_j)\big)$ and $\min(\varepsilon, f)$, with input family of partitions $\mathcal{P}_j^*$, to the hypergraphs $G_1, \ldots, G_s$. Let $\mathcal{P}_{j+1}^*$ be the returned family of partitions.

For each $j \geqslant 1$ such that $\mathcal{P}_j^*$ exists, we define

$$\mathcal{E}_j := \binom{n}{k}^{-1} \sum_{i=1}^{s} \sum_{Q \in \mathrm{Cross}_k(\mathcal{P}_j)} d_1\big(G_i\big|\hat{P}(Q, \mathcal{P}_j^*)\big)^2 .$$

Observe that if $j \geqslant 1$ and $\mathcal{P}_{j+1}^*$ exists, writing temporarily $D_{j,i}(Q) := d_1\big(G_i\big|\hat{P}(Q, \mathcal{P}_j^*)\big)$ and $D_{j+1,i}(Q) := d_1\big(G_i\big|\hat{P}(Q, \mathcal{P}_{j+1}^*)\big)$, we have

$$\mathcal{E}_{j+1} - \mathcal{E}_j \geqslant \binom{n}{k}^{-1} \sum_{i=1}^{s} \sum_{Q \in \mathrm{Cross}_k(\mathcal{P}_j)} D_{j+1,i}(Q)^2 - D_{j,i}(Q)^2$$

$$= \binom{n}{k}^{-1} \sum_{i=1}^{s} \sum_{Q \in \mathrm{Cross}_k(\mathcal{P}_j)} \big(D_{j,i}(Q) - (D_{j,i}(Q) - D_{j+1,i}(Q))\big)^2 - D_{j,i}(Q)^2$$

$$= \binom{n}{k}^{-1} \sum_{i=1}^{s} \sum_{Q \in \mathrm{Cross}_k(\mathcal{P}_j)} -2D_{j,i}(Q)\big(D_{j,i}(Q) - D_{j+1,i}(Q)\big) + \big(D_{j,i}(Q) - D_{j+1,i}(Q)\big)^2$$

$$= \binom{n}{k}^{-1} \sum_{i=1}^{s} \sum_{Q \in \mathrm{Cross}_k(\mathcal{P}_j)} \big(D_{j,i}(Q) - D_{j+1,i}(Q)\big)^2 ,$$

where the inequality comes from the fact that some $Q$ may be in $\mathrm{Cross}_k(\mathcal{P}_{j+1})$ but not in $\mathrm{Cross}_k(\mathcal{P}_j)$, and the final equality is since (by definition of density) the sum of $D_{j,i}(Q) - D_{j+1,i}(Q)$ over any polyad of $\mathcal{P}_j^*$ is zero, and $-2D_{j,i}(Q)$ is constant on any such polyad. From this we observe that $\mathcal{E}_{j+1} - \mathcal{E}_j$ is always nonnegative, and furthermore if there is some $1 \leqslant i \leqslant s$ and some $\varepsilon_k^2 \binom{n}{k}$ choices of $Q$ such that $D_{j+1,i}(Q) \neq D_{j,i}(Q) \pm \varepsilon_k$, then $\mathcal{E}_{j+1} - \mathcal{E}_j \geqslant \varepsilon_k^4$.

If for some $1 \leqslant j \leqslant s\varepsilon_k^{-4}+1$ the pair $(\mathcal{P}_j^*, \mathcal{P}_{j+1}^*)$ is a $(t_0, t_j, t_{j+1}, \varepsilon_k, \varepsilon(t_j), f_k(t_j), f(t_{j+1}), 1)$-strengthened pair for $G_i$ for each $1 \leqslant i \leqslant s$, then by choice of $T$ we are done. It follows that for each $1 \leqslant j \leqslant s\varepsilon_k^{-4} + 1$ the pair $(\mathcal{P}_j^*, \mathcal{P}_{j+1}^*)$ is not such a strengthened pair. By construction,

the conditions (S1)–(S5) in the definition of a strengthened pair are satisfied, and we conclude that (S6) fails, i.e. for some $1 \leqslant i \leqslant s$ there are $\varepsilon_k^2 \binom{n}{k}$ elements $Q \in \mathrm{Cross}_k(\mathcal{P}_j)$ such that

$$d_1\big(G_i\big|\hat{P}(Q, \mathcal{P}_j^*)\big) \neq d_1\big(G_i\big|\hat{P}(Q, \mathcal{P}_{j+1}^*)\big) \pm \varepsilon_k \,.$$

and so by our observation above $\mathcal{E}_{j+1} - \mathcal{E}_j \geqslant \varepsilon_k^4$.

Summing over $j$, we obtain $\mathcal{E}_{s\varepsilon_k^{-4}+1} - \mathcal{E}_1 \geqslant (s\varepsilon_k^{-4}+1)\varepsilon_k^4 > s$, which by definition of $\mathcal{E}_{s\varepsilon_k^{-4}+1}$ is not possible, completing the proof. $\qquad\square$

Next, we deduce the general case of Lemma 2.4, using the Weak Regularity Lemma of Conlon, Fox and Zhao [CFZ15]. This follows the approach in [ADS19]. Specifically, for each $G_i$ on vertex set $V$ we create a dense model $G_i''$ on $V$ by first using the Weak Regularity Lemma, which returns a dense model $G_i'$ of $G_i$ with weighted edges, and then randomly picking a $k$-edge into $G_i''$ with probability proportional to the weight of that edge in $G_i'$. The conclusion of the Weak Regularity Lemma, together with a simple application of the Chernoff bound, tell us that density and regularity properties with respect to $G_i''$ of any family of partitions $\mathcal{P}^*$ on $V$ with sufficiently large parts carry over (with a small loss of parameters) to $G_i$. In particular, we can apply the dense disjoint case of Lemma 2.4 to the $G_i''$ and the resulting strengthened pair is also a strengthened pair for the $G_i$. We now quote the Weak Hypergraph Regularity Lemma of [CFZ15] from [ADS19] (where a simplified statement which is all we need is given) and give the details. We need a couple of definitions.

Let $V$ be a vertex set and let $k \geqslant 2$. Let $g, h$ be two functions from $\binom{V}{k}$ to $\mathbb{R}_{\geqslant 0}$, which we think of as weighted hypergraphs. Given any collection $F_1, \ldots, F_k$ of $(k-1)$-uniform hypergraphs on $V$, let $S$ be the collection of $k$-sets in $V$ whose $(k-1)$-subsets can be labelled using each label from 1 to $i$ exactly once, such that the label $i$ subset is in $F_i$ (we say the edges of $S$ are *rainbow* for the $F_i$). If for any choice of the $F_i$ we have

$$\left| \sum_{e \in S} (g(e) - h(e)) \right| \leqslant \gamma |V|^k,$$

then $(g, h)$ is a $\gamma$-*discrepancy pair*. In addition, given $\eta > 0$, if for any choice of the $F_i$ we have

$$\sum_{e \in S} \big(g(e) - 1\big) \leqslant \eta |V|^k,$$

then we say $g$ is upper $\eta$-regular. Note that this definition is not quite the same as the previously defined $(\eta, p)$-upper regular; however if we have an $(\eta, p)$-upper regular hypergraph $G$, and we define a function $g$ by setting $g(e) = p^{-1}$ for edges $e$ of $G$, and $g(e) = 0$ otherwise, then $g$ is upper $\eta$-regular.

**Theorem A.2** (Theorem 19 from [ADS19], simplified from Theorem 2.16 of [CFZ15]). *For any $k \geqslant 2$ and $\gamma > 0$ and $g : \binom{V}{k} \to \mathbb{R}_{\geqslant 0}$ which is upper $\eta$-regular with $\eta \leqslant 2^{-80k/\gamma^2}$, there exists $\tilde{g} : \binom{V}{k} \to [0, 1]$ such that $(g, \tilde{g})$ form a $\gamma$-discrepancy pair.*

The following proof is very similar to the proof of [ADS19, Lemma 23, general case] and we copy it from there, making the appropriate modifications, for completeness.

*Proof of Lemma 2.4, general case.* Given $k \geqslant 2$, $q$, $t_0$, $s$ integers and $\varepsilon_k > 0$, and functions $f, f_k, \varepsilon : \mathbb{N} \to (0, 1]$, let $T$ and $n_0$ be returned by the dense disjoint case of Lemma 2.4 for input as above but with $\frac{1}{2s}\varepsilon_k$ and $\frac{1}{2s}f$ replacing $\varepsilon_k$ and $f$. Without loss of generality, we can assume $\varepsilon, f, f_k$ are decreasing functions.

We let $\gamma = \frac{1}{100s^4}\varepsilon_k^2\varepsilon(T)^2f(T)^2f_k(T)^2T^{-2^k}$, and $\eta = 2^{-160k/\gamma^2}$, and let Lemma 2.4 return $\eta$, $T$ and $n_0$.

Given an initial family of partitions $\mathcal{Q}^*$ which is $(1, t_0, \eta)$-equitable and $k$-uniform hypergraphs $G_1, \ldots, G_s$, where $G_i$ is $(\eta, p_i)$-upper regular for each $i$, we proceed as follows. First, for each $1 \leqslant i \leqslant s$, let $g_i : \binom{[n]}{k} \to \mathbb{R}_{\geqslant 0}$ be defined by $g_i(e) = p_i^{-1}$ if $e \in G_i$, and $g_i(e) = 0$ otherwise. Observe that each $g_i$ is upper $\eta$-regular.

Applying Theorem A.2, with input $\gamma$ separately to each $g_i$, we obtain functions $\tilde{g}_i : \binom{[n]}{k} \to [0, 1]$, such that $(g_i, \tilde{g}_i)$ is a $\gamma$-discrepancy pair for each $i$. It follows that $(s^{-1}g_i, s^{-1}\tilde{g}_i)$ is also a $\gamma$-discrepancy pair for each $i$.

We now create pairwise disjoint unweighted $k$-graphs $G_i''$ by, for each $e \in \binom{[n]}{k}$ independently, choosing to put $e$ into either exactly one of the $G_i''$, or into none of them, choosing to put $e$ in $G_i''$ with probability $s^{-1}\tilde{g}_i(e)$. Since $0 \leqslant \tilde{g}_i(e) \leqslant 1$ for each $i$, we have $0 \leqslant \sum_{i \in [s]} \tilde{g}_i(e) \leqslant s$, so that the distribution we just described is as required a probability distribution. By construction, the $G_i''$ are edge-disjoint.

We claim that a.a.s. $(s^{-1}\tilde{g}_i, G_i'')$ is a $\gamma$-discrepancy pair for each $i$ (where we temporarily abuse notation by equating $G_i''$ and the characteristic function of its edges). Indeed, suppose $i$ and unweighted $(k-1)$-graphs $F_1, \ldots, F_k$ on $[n]$ are fixed before the sampling of the $G_i''$. The expected number of edges of $G_i''$ which are rainbow for $F_1, \ldots, F_k$ is exactly equal to the sum of $g_i'(e)$ over $e$ rainbow for $F_1, \ldots, F_k$. By the Chernoff bound, the probability of an additive error of $\gamma n^k$ is $o(2^{-kn^{k-1}})$. In other words, a given $F_1, \ldots, F_k$ and $i$ witness the failure of our claim with probability $o(2^{-kn^{k-1}})$. Taking the union bound over the at most $s2^{kn^{k-1}}$ choices of $F_1, \ldots, F_k$ and $i$, our claim fails with probability $o(1)$ as desired.

We now apply the dense disjoint case of Lemma 2.4 to the $G_i''$, with inputs as above. We obtain integers $t_1, t_2 \leqslant T$ and families of partitions $\mathcal{P}_c^*$ and $\mathcal{P}_f^*$, both refining $\mathcal{Q}^*$, such that for each $1 \leqslant i \leqslant s$ the pair $(\mathcal{P}_c^*, \mathcal{P}_f^*)$ is a $(t_0, t_1, t_2, \frac{1}{2s}\varepsilon_k, \frac{1}{2s}\varepsilon(t_1), f_k(t_1), f(t_2), 1)$-strengthened pair for $G_i''$. We claim that this is the required strengthened pair for each $G_i$.

To see this, we need to check (S3), (S5) and (S6) hold. Expanding out the definition of regularity with respect to a family of partitions, each of these three statements boils down to a collection of claims of the following form. Given a $k$-polyad $\hat{P}$ (of either the coarse or fine partition) and a set of $(k-1)$-graphs $F_1, \ldots, F_k$ which are subgraphs of $\hat{P}$, let $S$ be the set of $k$-edges rainbow for $F_1, \ldots, F_k$. If $|S|$ is not too small, then $\frac{|E(G_i) \cap S|}{p_i|S|}$ is close to the relative density $d_{p_i}(G_i|\hat{P})$. Note that for each such condition, the corresponding statement for $G_i''$, namely that $\frac{|E(G_i'') \cap S|}{|S|}$ is close to $d_1(G_i''|\hat{P})$, is guaranteed by definition of a strengthened pair.

It is therefore enough to show the following: for any $(k-1)$-graphs $F_1, \ldots, F_k$, letting $S$ be the set of $k$-edges rainbow for $F_1, \ldots, F_k$, we have

$$\left|E(G_i) \cap S\right| = p_i s\left|E(G_i'') \cap S\right| \pm 2p_i s\gamma n^k.$$

Note that this statement also implies $d_{p_i}(G_i|\hat{P})$ is close to $sd_1(G_i''|\hat{P})$ by taking the $F_j$ to be the

$(k-1)$-partite subgraphs of $\hat{P}$, and the error term $2p_i s\gamma n^k$ is genuinely small compared to the main term because of the requirement that $S$ is not too small in the definition of 'regular' and by choice of $\gamma$.

This last equation is immediate since $(s^{-1}g_i, s^{-1}\tilde{g}_i)$ and $(s^{-1}\tilde{g}_i, G_i'')$ are both $\gamma$-discrepancy pairs. $\qquad\qquad\square$

In this second proof we could allow each $G_i$ to be a weighted $k$-uniform hypergraph (i.e. a function from $\binom{[n]}{k}$ to $\mathbb{R}_{\geqslant 0}$) without any change; thus Lemma 2.4 applies in the weighted setting of [ADS19]. We will not need this strengthening here, but note it for future reference.

We now prove, as promised, that a strengthened partition contains few irregular polyads.

**Proposition A.3.** *Given $k \in \mathbb{N}$, suppose that $t_0 \in \mathbb{N}$ is sufficiently large. Given any constants $\delta, d, \varepsilon_k > 0$, any function $\varepsilon : \mathbb{N} \to (0,1]$ which tends to zero sufficiently fast, any $t_1, t_2 \in \mathbb{N}$, any $0 < f_k \leqslant \varepsilon_k^2$ and any $f > 0$, there exists $\eta > 0$ such that the following holds for any sufficiently large $n$ and any $p > 0$. Suppose $G$ is an $n$-vertex hypergraph. Suppose that $(\mathcal{P}_c^*, \mathcal{P}_f^*)$ is a $(t_0, t_1, t_2, \varepsilon_k, \varepsilon, f_k, f, p)$-strengthened pair for $G$. Suppose that $\mathcal{P}_c^*$ has $t$ clusters and density vector $\mathbf{d} = (d_{k-1}, \ldots, d_2)$.*

*Then the number of irregular polyads of $\mathcal{P}_c^*$ for $G$ is at most*

$$4\varepsilon_k \binom{t}{k} \prod_{i=2}^{k-1} d_i^{-\binom{k}{i}}.$$

*Proof.* We require $t_0$ to be sufficiently large that $3\left(1 - \frac{1}{100}\right)^{-1} \frac{t_0^k}{k!} \leqslant 4\binom{t_0}{k}$ (and so the same holds for each $t \geqslant t_0$). We also require $\varepsilon(t_1)$ to be small enough, and $n/t_1$ large enough, for Lemma 2.6, with $\alpha = 1$, $d_0 = t_1^{-1}$, and $\gamma = \frac{1}{100}$.

Recall that a polyad $\hat{P}(Q; \mathcal{P}_c^*)$ is irregular with respect to $G$ if either $G$ is not $(\varepsilon_k, p)$-regular with respect to $\hat{P}(Q; \mathcal{P}_c^*)$, or for more than an $\varepsilon_k$-fraction of the $k$-sets $Q'$ supported on $\hat{P}(Q; \mathcal{P}_c^*)$, $G$ is not $(f_k, p)$-regular with respect to $\hat{P}(Q'; \mathcal{P}_f^*)$, or for more than an $\varepsilon_k$-fraction of the $k$-sets $Q'$ supported on $\hat{P}(Q; \mathcal{P}_c^*)$, we have $d_p\big(G\big|\hat{P}(Q'; \mathcal{P}_f^*)\big) \neq d_p\big(G\big|\hat{P}(Q; \mathcal{P}_c^*)\big) \pm \varepsilon_k$.

By definition and (S3), there are at most $\varepsilon_k\binom{n}{k}$ sets $Q$ in $\mathrm{Cross}_k(\mathcal{P}_c)$ such that $G$ is not $(\varepsilon_k, p)$-regular with respect to $\hat{P}(Q; \mathcal{P}_c^*)$. Since by Lemma 2.6 for each $Q \in \mathrm{Cross}_k(\mathcal{P}_c)$ the number of $k$-sets supported by $\hat{P}(Q; \mathcal{P}_c^*)$ is at least

$$(1 - \tfrac{1}{100})\left(\tfrac{n}{t}\right)^k \prod_{i=2}^{k-1} d_i^{\binom{k}{i}},$$

it follows that there are at most

$$\varepsilon_k \binom{n}{k}(1 - \tfrac{1}{100})^{-1}\left(\tfrac{t}{n}\right)^k \prod_{i=2}^{k-1} d_i^{-\binom{k}{i}} \leqslant (1 - \tfrac{1}{100})^{-1}\varepsilon_k \frac{t^k}{k!} \prod_{i=2}^{k-1} d_i^{-\binom{k}{i}}$$

$k$-polyads in $\mathcal{P}_c^*$ with respect to which $G$ is not $(\varepsilon_k, p)$-regular.

Suppose that $\mathbf{d}^f = (d^f_{k-1}, \ldots, d^f_2)$ is the density vector of $\mathcal{P}^*_f$, and $t_f$ is the number of its clusters. By definition and (S5), there are at most $f_k \binom{n}{k}$ sets in $\mathrm{Cross}_k(\mathcal{P}_f)$ such that $G$ is not $(f_k, p)$-regular with respect to $\hat{P}(Q; \mathcal{P}^*_f)$, and by (S6), there are at most $\varepsilon^2_k \binom{n}{k}$ sets $Q$ in $\mathrm{Cross}_k(\mathcal{P}_c)$ such that $d_p\big(G\big|\mathcal{P}^*_c(Q)\big) \neq d_p\big(G\big|\mathcal{P}^*_f(Q)\big) \pm \varepsilon_k$. By choice of $f_k$, there are in total at most $2\varepsilon^2_k \binom{n}{k}$ sets $Q$ in $\mathrm{Cross}_k(\mathcal{P}_f)$ which fail either condition, and so at most $2\varepsilon^2_k \binom{n}{k}$ sets $Q$ in $\mathrm{Cross}_k(\mathcal{P}_c)$ which fail either condition. Now a polyad $\hat{P}(Q; \mathcal{P}^*_c)$ which is irregular, but with respect to which $G$ is $(\varepsilon_k, p)$-regular, by Lemma 2.6 supports at least

$$\varepsilon_k(1 - \tfrac{1}{100})\big(\tfrac{n}{t}\big)^k \prod_{i=2}^{k-1} d_i^{\binom{k}{i}}$$

$k$-sets which fail one of these two conditions. It follows that the number of polyads $\hat{P}(Q; \mathcal{P}^*_c)$ which are irregular, but with respect to which $G$ is $(\varepsilon_k, p)$-regular, is at most

$$2\varepsilon^2_k \binom{n}{k} \cdot \varepsilon_k^{-1}(1 - \tfrac{1}{100})^{-1}\big(\tfrac{t}{n}\big)^k \prod_{i=2}^{k-1} d_i^{-\binom{k}{i}} \leqslant 2\varepsilon_k(1 - \tfrac{1}{100})^{-1}\tfrac{t^k}{k!} \prod_{i=2}^{k-1} d_i^{-\binom{k}{i}} .$$

Putting these together, we see that the total number of irregular $k$-polyads in $\mathcal{P}^*_c$ is at most

$$3\varepsilon_k(1 - \tfrac{1}{100})^{-1}\tfrac{t^k}{k!} \prod_{i=2}^{k-1} d_i^{-\binom{k}{i}} \leqslant 4\varepsilon_k \binom{t}{k} \prod_{i=2}^{k-1} d_i^{-\binom{k}{i}} .$$

$\square$

Building on this, we next prove Lemma 2.5.

*Proof of Lemma 2.5.* Given $k$ and $d$, we require $t_0 > \tfrac{4k}{d}$, and we then choose $0 < \gamma < \tfrac{1}{100}$ small enough so that

$$\frac{\delta - \tfrac{2k}{t_0} - (1 + \gamma)^4 \cdot 2^{k+1}\varepsilon_k^{1/k} - d}{(1 + \gamma)^4} > \delta - 2d - 2^{k+2}\varepsilon_k^{1/k} .$$

We additionally require $t_0$ to be large enough, and $\varepsilon(t_1)$ small enough, for Proposition A.3, so that $\mathcal{P}^*_c$ supports at most

$$4\varepsilon_k \binom{t}{k} \prod_{i=2}^{k-1} d_i^{-\binom{k}{i}}$$

irregular $k$-polyads, and also $\varepsilon(t_1)$ small enough, and $\tfrac{n}{t_1}$ large enough, for Lemma 2.6 with $d_0 = t_1^{-1}$ and $\gamma$. For convenience in this proof, we define $d_k = 1$.

We require $\eta$ to be sufficiently small compared to all the products of $d_i$ and $\tfrac{1}{t_1}$ which appear in the following proof: concretely, any $\eta \leqslant 100t_1^{-2k2^k}$ suffices.

Our first aim is to show that $\mathcal{R}_{\varepsilon_k}(G; \mathcal{P}_c^*, \mathcal{P}_f^*)$ contains many 1-edges. To start with, we mark the $k$-edges of the multicomplex $\mathcal{P}_c^*$ which are irregular as *bad*. We then, for each $i = k-1, \ldots, 1$ in succession, mark as bad all $i$-edges which are contained in at least

$$4\varepsilon_k^{1/k} t \prod_{j=2}^{i+1} d_j^{-\binom{i}{j-1}}$$

bad $(i+1)$-edges. Now consider the following construction. We begin by taking any bad 1-edge, then any bad 2-edge containing it, and so on until we obtain a bad $k$-edge together with an order on its vertices. Clearly we obtain any given bad $k$-edge in at most $k!$ ways by following this process (since a $k$-edge together with an order determines the chosen edges in the process). If there exist at least $4\varepsilon_k^{1/k} t$ bad 1-edges, it follows that the number of bad $k$-edges is at least

$$\frac{1}{k!} \prod_{i=0}^{k-1} \left( 4\varepsilon_k^{1/k} t \prod_{j=2}^{i+1} d_j^{-\binom{i}{j-1}} \right) \geqslant 4\varepsilon_k \frac{t^k}{k!} \prod_{i=2}^{k} d_i^{-\binom{k}{i}} > 4\varepsilon_k \binom{t}{k} \prod_{i=2}^{k} d_i^{-\binom{k}{i}},$$

which is in contradiction to Proposition A.3.

We claim that any edge of $\mathcal{P}_c^*$ which is not in $\mathcal{R}_{\varepsilon_k}(G; \mathcal{P}_c^*, \mathcal{P}_f^*)$ is either bad or contains a bad edge. To see this, consider the process of obtaining $\mathcal{R}_{\varepsilon_k}(G; \mathcal{P}_c^*, \mathcal{P}_f^*)$ by successively removing edges which either fail one of (RG1) or (RG2), or which contain a removed edge. Suppose for a contradiction that at some stage in this process we remove an edge which is neither bad nor contains a bad edge; let $e$ be the first such edge removed. Observe that $e$ cannot have been removed for failing (RG1), since edges which fail this condition are bad. Furthermore $e$ cannot have been removed for being unsupported, because all edges previously removed either were bad or contain bad edges, and by assumption $e$ contains no bad edges. So $e$ was removed for failing (RG2). In other words, we have $|e| \leqslant k - 1$ and $e$ is neither bad nor contains a bad edge, but nevertheless there are many $(|e| + 1)$-edges containing $e$ which either are bad or contain a bad edge.

Suppose that $f$ is a bad edge such that $|f \setminus e| = 1$. If $|f| = 1$, then there are at most $4\varepsilon_k^{1/k} t$ choices of $f$, each of which is contained in $\prod_{i=2}^{|e|+1} d_i^{-\binom{|e|}{i-1}}$ edges of $\mathcal{P}_c^*$ of uniformity $|e| + 1$ which contain $e$. If $|f| > 1$, then $f \cap e$ is non-empty and not a bad edge. There are at most $\binom{|e|}{\ell}$ choices of $f \cap e$ with $\ell$ elements, each of which by definition is contained in less than

$$4\varepsilon_k^{1/k} t \prod_{i=2}^{\ell+1} d_i^{-\binom{\ell}{i-1}}$$

bad edges. Thus, there are at most

$$\binom{|e|}{\ell} \cdot 4\varepsilon_k^{1/k} t \prod_{i=2}^{\ell+1} d_i^{-\binom{\ell}{i-1}}$$

choices of $f$, each of which is contained in

$$\prod_{i=2}^{|e|+1} d_i^{-\sum_{j=1}^{i-1} \binom{\ell}{i-j-1}\binom{|e|-\ell}{j}}$$

edges of $\mathcal{P}_c^*$ of uniformity $|e| + 1$ which contain $e$. This holds as $\mathcal{P}_f^*$ is $(t_0, t_1, \varepsilon)$-equitable partition.

Summing up, the number of $(|e| + 1)$-edges containing $e$ which are either bad or contain a bad edge is at most

$$4\varepsilon_k^{1/k} t \cdot \prod_{i=2}^{|e|+1} d_i^{-\binom{|e|}{i-1}} + \sum_{\ell=1}^{|e|} \binom{|e|}{\ell} \cdot 4\varepsilon_k^{1/k} t \Big( \prod_{i=2}^{\ell+1} d_i^{-\binom{\ell}{i-1}} \Big) \cdot \prod_{i=2}^{|e|+1} d_i^{-\sum_{j=1}^{i-1} \binom{\ell}{i-j-1}\binom{|e|-\ell}{j}}$$

$$= 4\varepsilon_k^{1/k} t \bigg( \sum_{\ell=0}^{|e|} \binom{|e|}{\ell} \bigg) \prod_{i=2}^{|e|+1} d_i^{-\binom{|e|}{i-1}} = 2^{|e|+2} \varepsilon_k^{1/k} t \prod_{i=2}^{|e|+1} d_i^{-\binom{|e|}{i-1}} .$$

But this last equation simply states that $e$ does *not* fail (RG2), which is our desired contradiction. In particular, every 1-edge which is not bad is in $\mathcal{R}_{\varepsilon_k}(G; \mathcal{P}_c^*, \mathcal{P}_f^*)$, and so also in $\mathcal{R}_{\varepsilon_k, d}(G; \mathcal{P}_c^*, \mathcal{P}_f^*)$. There are at least $\big(1 - 4\varepsilon_k^{1/k}\big) t$ such 1-edges.

Next, we prove that every $(k-1)$-edge of $\mathcal{R}_{\varepsilon_k, d}(G; \mathcal{P}_c^*, \mathcal{P}_f^*)$ is contained in sufficiently many $k$-edges.

By Lemma 2.6, given any $(k-1)$-edge $E$ of $\mathcal{R}_{\varepsilon_k, d}(G; \mathcal{P}_c^*, \mathcal{P}_f^*)$, we know that $E$ contains $e$ $(k-1)$-sets, where $e$ is in the range

$$(1 \pm \gamma)(\tfrac{n}{k})^{k-1} \prod_{i=2}^{k-2} d_i^{\binom{k-1}{i}}$$

$(k-1)$-sets; we need the lower bound to justify some applications of $(\eta, p)$-upper regularity. First, consider the edges of $G$ which contain a $(k-1)$-set in $E$ and another vertex from the clusters of $E$. These edges are all rainbow for $E$ together with $k-1$ copies of the complete $(k-1)$-graph on the union of the clusters of $E$, and any such edge contains at most two sets in $E$. The total number of $k$-sets which are rainbow for $E$ together with the $k-1$ complete graphs on the clusters of $E$ is at least $\frac{1}{2} e \cdot (\frac{n}{t_1} - 1) > \eta n^k$, by choice of $\eta$ and since every cluster has at least $\frac{n}{t_1}$ vertices. It is also at most $(k-1)\frac{n}{t_0} e$, since every cluster has at most $\frac{n}{t_0}$ vertices, and so by $(\eta, p)$-upper regularity of $G$, the total number of edges of $G$ which contain a $(k-1)$-set in $E$ and another vertex from the clusters of $E$ is at most $p \cdot (k-1)\frac{n}{t_0} e + p\eta n^k < kp \frac{n}{t_0} e$.

Since each $(k-1)$-set of $E$ is contained in at least $\delta pn$ edges of $G$, it follows that the total number of edges of $G$ which consist of an $(k-1)$-set of $E$ together with a vertex not in the clusters of $E$ is at least

$$\big(\delta - \tfrac{2k}{t_0}\big) epn .$$

Now these edges of $G$ are partitioned according to the $k$-polyad of $\mathcal{P}_c^*$ containing them. By Lemma 2.6, each of these $k$-polyads supports

$$\big(1 \pm \gamma\big)^3 e \tfrac{n}{t} \prod_{i=2}^{k-1} d_i^{\binom{k-1}{i-1}}$$

$k$-sets, and hence by $(\eta, p)$-upper regularity of $G$, at most

$$\big(1 + \gamma\big)^4 ep \tfrac{n}{t} \prod_{i=2}^{k-1} d_i^{\binom{k-1}{i-1}}$$

edges of $G$. Since $E$ is in $\mathcal{R}_{\varepsilon_k, d}(G; \mathcal{P}_c^*, \mathcal{P}_f^*)$, at most

$$2^{k+1} \varepsilon_k^{1/k} t \prod_{i=2}^{k-1} d_i^{-\binom{k-1}{i-1}}$$

of these polyads are not in $\mathcal{R}_{\varepsilon_k}(G; \mathcal{P}_c^*, \mathcal{P}_f^*)$, so the total number of edges of $G$ in such $k$-polyads of $\mathcal{P}_c^*$ containing $E$ is at most

$$2^{k+1} \varepsilon_k^{1/k} t \left( \prod_{i=2}^{k-1} d_i^{-\binom{k-1}{i-1}} \right) \cdot (1+\gamma)^4 e p \frac{n}{t} \prod_{i=2}^{k-1} d_i^{\binom{k-1}{i-1}} = (1+\gamma)^4 \cdot 2^{k+1} \varepsilon_k^{1/k} e p n\,.$$

By definition, the number of edges of $G$ in $k$-polyads of $\mathcal{P}_c^*$ containing $E$ with respect to which the $p$-density of $G$ is less than $d$ is at most $depn$. It follows that all the remaining edges of $G$ which contain a $k$-set of $E$ together with a vertex outside the clusters of $E$ are contained in edges of $\mathcal{R}_{\varepsilon_k, d}(G; \mathcal{P}_c^*, \mathcal{P}_f^*)$, and so there are at least

$$\frac{\left(\delta - \frac{2k}{t_0}\right) epn - (1+\gamma)^4 \cdot 2^{k+1} \varepsilon_k^{1/k} epn - depn}{(1+\gamma)^4 e p \frac{n}{t} \prod_{i=2}^{k-1} d_i^{\binom{k-1}{i-1}}}$$

$$= \frac{\delta - \frac{2k}{t_0} - (1+\gamma)^4 \cdot 2^{k+1} \varepsilon_k^{1/k} - d}{(1+\gamma)^4} t \prod_{i=2}^{k-1} d_i^{-\binom{k-1}{i-1}}$$

such edges, as desired.

Finally, we need to show that if $\delta > \frac{1}{2} + 2d + 2^{k+2} \varepsilon_k^{1/k} + \nu$ then any induced subcomplex $\mathcal{R}' \subseteq \mathcal{R}$ on at least $(1-\nu)t$ 1-edges is tightly linked. In other words, we need to show that if $\mathbf{u}$ on the vertices $(u_1, \ldots, u_{k-1})$ and $\mathbf{v}$ on the vertices $(v_1, \ldots, v_{k-1})$ are any two ordered $(k-1)$-edges of $\mathcal{R}' \subseteq \mathcal{R} = \mathcal{R}_{\varepsilon_k, d}(G; \mathcal{P}_c^*, \mathcal{P}_f^*)$, there is a tight link in $\mathcal{R}'$ from $\mathbf{u}$ to $\mathbf{v}$. This proof does not require any further regularity theory; simply the properties of $\mathcal{R}$ we already deduced.

The critical observation we need is the following. If $f$ is any $(k-1)$-edge in $\mathcal{R}'$, then it is an edge in $\mathcal{R}$ and so is in at least

$$\left(\delta - 2d - 2^{k+2} \varepsilon_k^{1/k}\right) t \prod_{i=2}^{k-1} d_i^{-\binom{k-1}{i-1}} > \left(\tfrac{1}{2} + \nu\right) t \prod_{i=2}^{k-1} d_i^{-\binom{k-1}{i-1}}$$

$k$-edges of $\mathcal{R}$. Of these, at most

$$\nu t \prod_{i=2}^{k-1} d_i^{-\binom{k-1}{i-1}}$$

use one of the 1-edges not in $\mathcal{R}'$, and so $f$ is in strictly more than

$$\tfrac{1}{2} t \prod_{i=2}^{k-1} d_i^{-\binom{k-1}{i-1}}$$

$k$-edges of $\mathcal{R}'$.

We build up the desired tight link vertex by vertex. For each $1 \leqslant j \leqslant k-1$, we will choose a 1-cell $w_j$ of $\mathcal{R}'$, disjoint from $\{u_1, \ldots, u_{k-1}, v_1, \ldots, v_{k-1}\}$. At the $j$th step, we in addition choose for each $S \subseteq [j-1]$ with $S \neq \varnothing$ an $(|S|+1)$-cell $w_{S \cup \{j\}}$ of $\mathcal{R}'$ supported by $w_S$ and the cells $w_{S' \cup \{j\}}$ for $S' \subseteq S$ with $|S'| = |S| - 1$. Finally we choose two $k$-edges $e_{j,u}$ and $e_{j,v}$ of $\mathcal{R}$, where $e_{j,u}$ has underlying 1-cells $u_j, \ldots, u_{k-1}, w_1, \ldots, w_j$ and $e_{j,v}$ has underlying 1-cells $v_j, \ldots, v_{k-1}, w_1, \ldots, w_j$ and where both $e_{j,u}$ and $e_{j,v}$ have $w_{[j]}$ in their $(k-j)$-times-iterated boundaries. We insist additionally that $\partial e_{1,u}$ contains the $(k-1)$-cell underlying $\mathbf{u}$, and $\partial e_{1,v}$ contains the $(k-1)$-cell underlying $\mathbf{v}$, and that for each $2 \leqslant j \leqslant k-1$ we have $\bigl(\partial e_{j-1,u}\bigr) \cap \bigl(\partial e_{j,u}\bigr), \bigl(\partial e_{j-1,v}\bigr) \cap \bigl(\partial e_{j,v}\bigr) \neq \varnothing$.

For $j = 1$, consider $\mathbf{u}$ and $\mathbf{v}$ as $(k-1)$-edges of $\mathcal{R}'$. We need to find a 1-cell $w_1$ of $\mathcal{R}'$ such that there are $k$-edges of $\mathcal{R}'$ whose boundaries contain $\mathbf{u}$ and $\mathbf{v}$ respectively, and both of which use the 1-edge $w_1$. If no such edges exist, then for one of $\mathbf{u}$ and $\mathbf{v}$ (without loss of generality, suppose it is $\mathbf{u}$) there are at most $\frac{1}{2}t$ 1-edges which support edges of $\mathcal{R}'$ using $\mathbf{u}$. Thus $\mathbf{u}$ is in at most

$$\tfrac{1}{2}t \prod_{i=2}^{k-1} d_i^{-\binom{k-1}{i-1}}$$

$k$-edges of $\mathcal{R}'$, which is a contradiction since $\mathbf{u}$ is an edge of $\mathcal{R}'$.

Given $2 \leqslant j \leqslant k-1$, suppose $w_S$ has been constructed for each $S \subseteq [j-1]$ with $S \neq \varnothing$. Consider the collection $W^*$ of $j$-edges in $\mathcal{R}'$ whose boundary contains $w_{[j-1]}$. Any $k$-edge of $\mathcal{R}'$ whose $(k+1-j)$-times-iterated boundary contains $w_{[j-1]}$ has $(k-j)$-times-iterated boundary containing some $w^* \in W^*$. In particular, consider the $(k-1)$-edges $e_u^*$ in $\partial e_{u,j-1}$ which does not use $u_{j-1}$, and $e_v^*$ in $\partial e_{v,j-1}$ which does not use $v_{j-1}$. Both these edges have $w_{[j-1]}$ in their $(k-j)$-times iterated boundary. What we want is some $w^* \in W^*$ such that there are $k$-edges $e_{u,j}$ whose boundary contains $e_u^*$, and $e_{v,j}$ whose boundary contains $e_u^*$, both of whose $(k-j)$-times-iterated boundaries contain $w^*$. Assuming such a $w^*$ exists, we can then let $w_{[j]} = w^*$ and for each $S \subsetneq [j-1]$ with $S \neq \varnothing$, we let $w_{S \cup \{j\}}$ be the unique $(|S|+1)$-edge on the vertices $\{w_i : i \in S \cup \{j\}\}$ that is in the $(j-|S|-1)$-iterated boundary of $w_{[j]}$.

Suppose for a contradiction that no such $w^*$ exists. Then each $w^* \in W^*$ can be assigned to at most one of $u$ and $v$, according to whether it is in an edge with $e_u^*$ or with $e_v^*$. We have

$$|W^*| \leqslant t \prod_{i=2}^{j} d_i^{-\binom{j-1}{i-1}}$$

and hence for one of $u$ and $v$, the number of elements of $W^*$ assigned to it is at most

$$\tfrac{1}{2}t \prod_{i=2}^{j} d_i^{-\binom{j-1}{i-1}}.$$

Suppose without loss of generality this is $u$. For any given $w^* \in W^*$, the number of $k$-edges of $\mathcal{R}$ whose boundary contains $e_u^*$ and whose $(k-j)$-times iterated boundary contains $w^*$ is at most

$$\prod_{i=2}^{k-1} d_i^{-\binom{k-1}{i-1} + \binom{j-1}{i-1}}$$

where we use the convention $\binom{j-1}{i-1} = 0$ if $i > j$. Hence the number of $k$-edges of $\mathcal{R}'$ whose boundary contains $e_u^*$ is at most

$$\tfrac{1}{2}t \prod_{i=2}^{j} d_i^{-\binom{j-1}{i-1}} \cdot \prod_{i=2}^{k-1} d_i^{-\binom{k-1}{i-1}+\binom{j-1}{i-1}} = \tfrac{1}{2}t \prod_{i=2}^{k-1} d_i^{-\binom{k-1}{i-1}} .$$

But $e_u^*$ is a $(k-1)$-edge of $\mathcal{R}'$, so this is a contradiction. $\qquad\square$

To prove the remaining lemmas we first state the Dense Counting Lemma with parts of the same size for more general graphs.

**Lemma A.4** ([KRS02, Theorem 6.5]). *For all integers $k \geqslant 2$ and constants $\gamma, d_0 > 0$, there exists $\varepsilon > 0$ such that the following holds. Let $\mathbf{d} = (d_{k-1}, \ldots, d_2)$ be a vector of real numbers with $d_i \geqslant d_0$ for each $2 \leqslant i \leqslant k-1$, let $G$ be a $k$-partite $(k-1)$-complex with parts of size $m \geqslant \varepsilon^{-1}$ which is $(d_{k-1}, \ldots, d_2, \varepsilon, 1)$-regular, and let $H$ be any $k$-partite $(k-1)$-complex on at most $2k$ vertices with fixed partition classes. Then the number of copies of $H$ in $G$, with the vertices of the $i$th class of $H$ embedded into the $i$th class of $G$ for $i = 1, \ldots, k$, is*

$$\left(1 \pm \gamma\right) m^{v(H)} \prod_{i=2}^{k-1} d_i^{e_i(H)} ,$$

*where $e_i(H)$ is the number of edges of size $i$ in $H$.*

*Proof of Lemma 2.6.* Let $k \geqslant 2$ be an integer and $\alpha, \gamma, d_0 > 0$. Let $\varepsilon > 0$ be small enough got Lemma 2.7 on input $k$, $\alpha$, and $d_0$. Additionally assume that $\sqrt{\varepsilon}$ is small enough for Lemma A.4 on input $k$, $\gamma$, and $d_0$. Let $\mathbf{d} = (d_{k-1}, \ldots, d_2)$ be a vector of real numbers with $d_i \geqslant d_0$ for each $2 \leqslant i \leqslant k$, and let $G$ be a $k$-partite $(k-1)$-complex with parts $V_1, \ldots, V_k$ of size $m \geqslant \alpha^{-1}\varepsilon^{-1}$ which is $(\mathbf{d}, \varepsilon, 1)$-regular.

Let $V_i' \subseteq V_i$ be of size $|V_i'| \geqslant \alpha|V_i|$ for $i = 1, \ldots, k$. Our goal for he first part is to estimate the number of copies of the $k$-vertex complete $(k-1)$-complex in $G[V_1', \ldots, V_k']$. We denote this number by $E$. Note that for any $V_i'' \subseteq V_i'$ each of size $\alpha m$ the induced subcomplex $G[V_1'', \ldots, V_k'']$ is $(\mathbf{d}, \sqrt{\varepsilon}, 1)$-regular by Lemma 2.7. Therefore, by Lemma A.4, there number of copies of the $k$-vertex complete $(k-1)$-complex in $G[V_1'', \ldots, V_k'']$ is

$$(1 \pm \gamma)(\alpha m)^k \prod_{i=2}^{k-1} d_i^{\binom{k}{i}} .$$

We choose sets $V_i'' \subseteq V_i$ of size $\alpha m$ uniformly at random and note that the expected number of copies of the $k$-vertex complete $(k-1)$-complex in $G[V_1'', \ldots, V_k'']$ is $E \prod_{i=1}^{k} \frac{\alpha m}{|V_i'|}$. With the estimate from above it follows that

$$E \prod_{i=1}^{k} \frac{\alpha m}{|V_i'|} = (1 \pm \gamma)(\alpha m)^k \prod_{i=2}^{k-1} d_i^{\binom{k}{i}}$$

and rearranging this for $E$ finishes the proof of the first part. The second part follows analogously. $\qquad\square$

*Proof of Lemma 2.8.* We will prove the first statement, then deduce the second, and note that the third also follows. Let $k \geqslant 2$ be an integer and $\alpha, \gamma, d_0 > 0$. Let $0 < \varepsilon < \frac{\gamma}{4}$ be small enough for Lemma 2.6 on input $k$, $\alpha$, $\gamma_0 = \frac{1}{36}\gamma^3$, and $d_0$. Additionally, assume that $\varepsilon$ is small enough for Lemma 2.6 with the same input and $k - 1$ instead of $k$ and also small enough for the generalisation of Lemma 2.6 for counting two copies of the $k$-vertex complete $(k - 1)$-complex that overlap on the first $k - 1$ vertices[3] with the same input as above. Let $\mathbf{d} = (d_{k-1}, \ldots, d_2)$ be a vector of real numbers with $d_i \geqslant d_0$ for each $2 \leqslant i \leqslant k$, and let $G$ be a $k$-partite $(k - 1)$-complex with parts $V_1, \ldots, V_k$ of size $m \geqslant \alpha^{-1}\varepsilon^{-1}$ which is $(\mathbf{d}, \varepsilon, 1)$-regular.

Let $V_i' \subseteq V_i$ be of size $|V_i'| \geqslant \alpha|V_i|$ for $i = 1, \ldots, k$ and $G' = G[V_1', \ldots, V_k']$. For a random $(k - 1)$-tuple $e$ from $G[V_1', \ldots, V_{k-1}']$ we denote by $X_e$ the number of copies of the $k$-vertex complete $(k - 1)$-complex in $G'$ that $e$ is contained in. With Lemma 2.6 we can count the total number of copies of the $k$-vertex complete $(k - 1)$-complex in $G'$ and the number of choices for $e$ to obtain

$$\mathbb{E}[X_e] = \frac{(1 \pm \gamma_0)}{(1 \mp \gamma_0)} \prod_{i=1}^{k} |V_i'| \prod_{i=2}^{k-1} d_i^{\binom{k}{i}} \cdot \prod_{i=1}^{k-1} |V_i'|^{-1} \prod_{i=2}^{k-1} d_i^{-\binom{k-1}{i}} = (1 \pm 3\gamma_0)|V_k'| \prod_{i=2}^{k-1} d_i^{\binom{k-1}{i-1}}.$$

Similarly, we can count the number of two copies of the $k$-vertex complete $(k - 1)$-complex that overlap on the first $k - 1$ vertices to get

$$\mathbb{E}[X_e^2] = \frac{(1 \pm \gamma_0)}{(1 \mp \gamma_0)} |V_k'| \prod_{i=1}^{k} |V_i'| \prod_{i=2}^{k-1} d_i^{\binom{k}{i}+\binom{k-1}{i-1}} \cdot \prod_{i=1}^{k-2} |V_i'| \prod_{i=2}^{k-1} d_i^{-\binom{k-1}{i}} = (1 \pm 3\gamma_0)|V_k'|^2 \prod_{i=2}^{k-1} d_i^{2\binom{k-1}{i-1}}.$$

This implies

$$V[X_e] = \mathbb{E}[X_e^2] - \mathbb{E}[X_e]^2 \leqslant 9\gamma_0 |V_k'|^2 \prod_{i=2}^{k-1} d_i^{2\binom{k-1}{i-1}}$$

and, with Chebyshev's inequality, we infer

$$\mathbb{P}\left[|X_e - \mathbb{E}[X]| \geqslant \frac{\gamma}{2}|V_k'| \prod_{i=2}^{k-1} d_i^{\binom{k-1}{i-1}}\right] \leqslant \frac{9\gamma_0}{(\frac{\gamma}{2})^2} \leqslant \gamma.$$

Thus, at least a $(1 - \gamma)$-fraction of the $(k - 1)$-tuples in $G[V_1', \ldots, V_{k-1}']$ is contained in

$$\left(1 \pm \tfrac{\gamma}{2}\right)(1 \pm 3\gamma_0)|V_k'| \prod_{i=2}^{k-1} d_i^{\binom{k-1}{i-1}} = (1 \pm \gamma)|V_k'| \prod_{i=2}^{k-1} d_i^{\binom{k-1}{i-1}}$$

copies of the $k$-vertex complete $(k - 1)$-complex in $G'$.

For the counting statement, let $S$ denote the $\gamma$-fraction of edges $e$ with largest $X_e$. Suppose that $\mathbb{E}[X_e|e \in S] \geqslant 2\mathbb{E}[X_e]$, as otherwise we get

$$\sum_{e \in S} X_e \leqslant |S| \cdot \mathbb{E}[X_e|e \in S] < 2|S| \cdot \mathbb{E}[X_e] \leqslant \tfrac{5}{2}\gamma \prod_{i=1}^{k} |V_i'| \prod_{i=2}^{k} d_i^{\binom{k}{i-1}}.$$

---

[3]This follows from Lemma A.4 by exactly the same argument as in our proof of Lemma 2.6.

Then we have

$$\mathbb{E}[X_e^2] = \gamma\mathbb{E}\big[X_e^2|e \in S\big] + (1-\gamma)\mathbb{E}\big[X_e^2|e \notin S\big] \geqslant \gamma\mathbb{E}\big[X_e|e \in S\big]^2 + (1-\gamma)\mathbb{E}\big[X_e|e \notin S\big]^2$$

where the inequality is Jensen's inequality (since the second moment function is convex). Now we have $\mathbb{E}\big[X_e|e \in S\big] = \mathbb{E}\big[X_e\big] + c$ for some $c \geqslant \mathbb{E}\big[X_e\big]$, and $\mathbb{E}\big[X_e|e \notin S\big] = \mathbb{E}\big[X_e\big] - \frac{\gamma}{1-\gamma}c$. Plugging these in we get

$$\mathbb{E}[X_e^2] \geqslant (1+\gamma)\mathbb{E}[X_e]^2$$

which is in contradiction to the above bound on $V[X_e]$.

For the third statement we apply the first statement to the first $k-1$ parts $V_1, \ldots, V_{k-1}$, which is possible by the choice of $\varepsilon$ above. We find that a $(1-\gamma)$-fraction of the $(k-2)$-tuples in $G[V_1', \ldots, V_{k-2}']$ is contained in

$$\big(1 \pm \tfrac{\gamma}{2}\big)(1 \pm 3\gamma_0)|V_{k-1}'| \prod_{i=2}^{k-2} d_i^{\binom{k-2}{i-1}}$$

copies of the $(k-1)$-vertex $(k-2)$-complex in $G'$ together with a vertex from $V_{k-1}'$. Then, with $(d_{k-1}, \varepsilon, 1)$-regularity, we get that a $(1\pm\varepsilon)d_{k-1}$-fraction of these give copies of the $(k-1)$-vertex $(k-1)$-complex in $G'$. The statement follows as $(1 \pm \tfrac{\gamma}{2})(1 \pm 3\gamma_0)(1 \pm \varepsilon) = (1 \pm \gamma)$. $\qquad\square$

*Proof of Lemma 2.9.* Let $k \geqslant 3$ be an integer $\gamma, \delta, d_0 > 0$ and w.l.o.g. assume $\delta \geqslant 2\gamma$. We choose $0 < \gamma_0 < \gamma^2\delta 2^{-k}$ and $0 < \varepsilon < \frac{1}{2}\gamma_0 d_0^{10k2^{10k}}$. Let $\mathbf{d} = (d_{k-1}, \ldots, d_2)$ be a vector of real numbers with $d_i \geqslant d_0$ for each $2 \leqslant i \leqslant k-1$, and let $G$ be a $k$-partite $(k-1)$-complex with parts $V_1, \ldots, V_k$ of size $m \geqslant \varepsilon^{-1}$ which is $(\mathbf{d}, \varepsilon, 1)$-regular. Given integers $a, b, c$ such that $a + b - c = k$, let $A$ be part of an $a$-cell, $B$ be part of a $b$-cell of size $|B| \geqslant \gamma m^b \prod_{i=2}^{k-2} d_i^{\binom{b}{i}}$, and $C$ be part of a $c$-cell such that the tuples from $C$ have degree $(\delta \pm \gamma)m^{a-c} \prod_{i=2}^{k-1} d_i^{\binom{a}{i}-\binom{c}{i}}$ into $A$ and every tuple of $B$ is contained in a tuple of $C$.

We denote by $F_0$ the $k$-vertex $(k-1)$-complex obtained from the union of a complete $a$-vertex $a$-complex with a complete $b$-vertex $b$-complex identified on $c$ vertices. We fix the canonical labelling $v_1, \ldots, v_k$ of the vertices from $F_0$ such that the $v_i$ vertices of the $a$- and $b$-vertex graph constructed above correspond to the clusters $V_i$ of $G$ in $A$ and $B$, respectively. Then each tuple from $B$ is contained in $(\delta \pm \gamma)m^{a-c} \prod_{i=2}^{k-1} d_i^{\binom{a}{i}-\binom{c}{i}}$ copies of $F_0$ supported by $A$ and $B$ and, therefore, there are

$$|B|(\delta \pm \gamma)m^{a-c} \prod_{i=2}^{k-1} d_i^{\binom{a}{i}-\binom{c}{i}} = \eta m^k \prod_{i=2}^{k-1} d_i^{e_i(F_0)}$$

copies of $F_0$ in $G$ supported by $A$ and $B$, where $\eta = (\delta \pm \gamma)|B|m^{-b} \prod_{i=2}^{k-2} d_i^{-\binom{b}{i}} \geqslant \delta^2\gamma_0$.

Starting from $F_0$ and adding edges of size at most $k-1$ that are supported we obtain a sequence $F_0, \ldots, F_t$ of $k$-vertex $(k-1)$-complexes, where $F_t$ is the complete $k$-vertex $(k-1)$-complex. Note that $t = 2^k - 2 - e(F_0)$. We show by induction on $j$ that for each $0 \leqslant j \leqslant t$, we have

$$\big(1 \pm j\gamma_0\big)\eta m^k \prod_{i=2}^{k-1} d_i^{e_i(F_j)}$$

copies of $F_j$ in $G$ supported by $A$ and $B$. The $j = t$ case, and the choice of $\eta$ and as $\gamma_0 \leqslant \gamma^2 \delta / t$, proves the lemma.

For a given $1 \leqslant j \leqslant t$, let $e$ be the edge in $E(F_j) \setminus E(F_{j-1})$. Let $F'$ be the subgraph of $F_{j-1}$ induced on $V(F) \setminus e$, and denote by $X$ the $(k - |e|)$-tuples in $G$ from the clusters $V_i$ such that $v_i \notin e$ that are copies of $F'$. For each $x \in X$, let $F_{j-1}(x)$ and $F_j(x)$ be the number of copies of $F_{j-1}$ and $F_j$, respectively, in $G$ supported by $A$ and $B$ that contain $x$. Then, by assumption, we have that

$$\sum_{x \in X} F_{j-1}(x) = (1 \pm (j-1)\gamma_0)\eta m^k \prod_{i=2}^{k-1} d_i^{e_i(F_{j-1})}.$$

We let $X' \subseteq X$ be those tuples, for which $F_{j-1}(x) \geqslant \varepsilon m^{|e|}$. Then for any fixed $x \in X'$ we get from $(\mathbf{d}, \varepsilon, 1)$-regularity of $G$ that $F_j(x) = (d_{|e|} \pm \varepsilon)F_{j-1}(x)$. For any $x \in X$, we additionally have $F_j(x) \leqslant F_{j-1}(x)$.

Observe that we have

$$\sum_{x \in X \setminus X'} F_j(x) \leqslant \sum_{x \in X \setminus X'} F_{j-1}(x) \leqslant \varepsilon m^k,$$

where we use the bound $|X \setminus X'| \leqslant m^{k-|e|}$ and the definition of $X'$ for the final inequality. This gives

$$\sum_{x \in X'} F_{j-1}(x) = (1 \pm \sqrt{\varepsilon}) \sum_{x \in X} F_{j-1}(x).$$

We then have

$$\sum_{x \in X} F_j(x) = \sum_{x \in X'} F_j(x) + \sum_{x \in X \setminus X'} F_j(x) = \left( \sum_{x \in X'} (d_{|e|} \pm \varepsilon)F_{j-1}(x) \right) \pm \varepsilon m^k$$

$$= \left( (d_{|e|} \pm \varepsilon)(1 \pm \sqrt{\varepsilon}) \sum_{x \in X} F_{j-1}(x) \right) \pm \varepsilon m^k = \left( 1 \pm \tfrac{1}{2}\gamma_0 \right) d_{|e|} \sum_{x \in X} F_{j-1}(x),$$

as desired. $\qquad\square$

# References

[ABCM17] Peter Allen, Julia Böttcher, Oliver Cooley, and Richard Mycroft. Tight cycles and regular slices in dense hypergraphs. *J. Comb. Theory, Ser. A*, 149:30–100, 2017. `doi:10.1016/j.jcta.2017.01.003`.

[ABKP15] Peter Allen, Julia Böttcher, Yoshiharu Kohayakawa, and Yury Person. Tight Hamilton cycles in random hypergraphs. *Random Struct. Algorithms*, 46(3):446–465, 2015. `doi:10.1002/rsa.20519`.

[ACK⁺00] Noga Alon, Michael Capalbo, Yoshiharu Kohayakawa, Vojtěch Rödl, Andrzej Ruciński, and Endre Szemerédi. Universality and tolerance (extended abstract). In *41st Annual Symposium on Foundations of Computer Science (Redondo Beach,*

*CA, 2000)*, pages 14–21. IEEE Comput. Soc. Press, Los Alamitos, CA, 2000. `doi:10.1109/SFCS.2000.892007`.

[ADS19]     Peter Allen, Ewan Davies, and Jozef Skokan. Regularity inheritance in hypergraphs. 2019. `arXiv:1901.05955`.

[AFKS00]    Noga Alon, Eldar Fischer, Michael Krivelevich, and Mario Szegedy. Efficient testing of large graphs. *Combinatorica*, 20(4):451–476, 2000. `doi:10.1007/s004930070001`.

[AKPP20]    Peter Allen, Christoph Koch, Olaf Parczyk, and Yury Person. Finding tight Hamilton cycles in random hypergraphs faster. *Comb. Probab. Comput.*, 30(2):239–257, 2020. `doi:10.1017/S0963548320000450`.

[Bol84]     Béla Bollobás. The evolution of sparse graphs. Graph theory and combinatorics, (Cambridge 1983), Academic Press, London, 35-37, 1984.

[CEP20]     Dennis Clemens, Julia Ehrenmüller, and Yury Person. A Dirac-type theorem for Berge cycles in random hypergraphs. *Electron. J. Comb.*, 27(3):research paper p3.39, 23, 2020. `doi:10.37236/8611`.

[CFZ15]     David Conlon, Jacob Fox, and Yufei Zhao. A relative Szemerédi theorem. *Geom. Funct. Anal.*, 25(3):733–762, 2015. `doi:10.1007/s00039-015-0324-9`.

[DF11]      Andrzej Dudek and Alan Frieze. Loose Hamilton cycles in random uniform hypergraphs. *Electron. J. Comb.*, 18(1):p48, 14, 2011. `doi:10.37236/535`.

[DF13]      Andrzej Dudek and Alan Frieze. Tight Hamilton cycles in random uniform hypergraphs. *Random Struct. Algorithms*, 42(3):374–385, 2013. `doi:10.1002/rsa.20404`.

[Dir52]     G. A. Dirac. Some theorems on abstract graphs. *Proc. Lond. Math. Soc. (3)*, s3-2(1):69–81, 1952. `doi:10.1112/plms/s3-2.1.69`.

[FH20]      Asaf Ferber and Lior Hirschfeld. Co-degrees resilience for perfect matchings in random hypergraphs. *Electron. J. Comb.*, 27(1):research paper p1.40, 13, 2020. `doi:10.37236/8167`.

[FK22]      Asaf Ferber and Matthew Kwan. Dirac-type theorems in random hypergraphs. *J. Comb. Theory, Ser. B*, 155:318–357, 2022. `doi:10.1016/j.jctb.2022.02.009`.

[Fri23]     Alan Frieze. Hamilton cycles in random graphs: a bibliography. 2023. `arXiv:1901.07139`.

[Gow07]     W. T. Gowers. Hypergraph regularity and the multidimensional Szemerédi theorem. *Ann. Math. (2)*, 166(3):897–946, 2007. `doi:10.4007/annals.2007.166.897`.

[KO14]      Daniela Kühn and Deryk Osthus. Hamilton cycles in graphs and hypergraphs: an extremal perspective. In *Proceedings of the International Congress of Mathematicians (ICM 2014), Seoul, Korea, August 13–21, 2014. Vol. IV: Invited lectures*, pages 381–406. Seoul: KM Kyung Moon Sa, 2014.

[Kor76]     A. D. Korsunov. Solution of a problem of Erdős and Renyi on Hamiltonian cycles in nonoriented graphs. *Sov. Math., Dokl.*, 17:760–764, 1976.

[Kor77] A. D. Korsunov. Solution of a problem of P. Erdős and A. Renyi on Hamiltonian cycles in undirected graphs. *Metody Diskretn. Anal.*, 31:17–56, 1977.

[KRS02] Yoshiharu Kohayakawa, Vojtěch Rödl, and Jozef Skokan. Hypergraphs, quasi-randomness, and conditions for regularity. *J. Comb. Theory, Ser. A*, 97(2):307–352, 2002. `doi:10.1006/jcta.2001.3217`.

[KS83] János Komlós and Endre Szemerédi. Limit distribution for the existence of Hamiltonian cycles in a random graph. *Discrete Math.*, 43(1):55–63, 1983. `doi:10.1016/0012-365X(83)90021-3`.

[KV00] Jeong Han Kim and Van H. Vu. Concentration of multivariate polynomials and its applications. *Combinatorica*, 20(3):417–434, 2000. `doi:10.1007/s004930070014`.

[LS12] Choongbum Lee and Benny Sudakov. Dirac's theorem for random graphs. *Random Struct. Algorithms*, 41(3):293–305, 2012. `doi:10.1002/rsa.20419`.

[Mon19] Richard Montgomery. Hamiltonicity in random graphs is born resilient. *J. Comb. Theory, Ser. B*, 139:316–341, 2019. `doi:10.1016/j.jctb.2019.04.002`.

[NS20] Bhargav Narayanan and Mathias Schacht. Sharp thresholds for nonlinear Hamiltonian cycles in hypergraphs. *Random Struct. Algorithms*, 57(1):244–255, 2020. `doi:10.1002/rsa.20919`.

[NST19] Rajko Nenadov, Angelika Steger, and Miloš Trujić. Resilience of perfect matchings and Hamiltonicity in random graph processes. *Random Struct. Algorithms*, 54(4):797–819, 2019. `doi:10.1002/rsa.20827`.

[Pos76] L. Posa. Hamiltonian circuits in random graphs. *Discrete Math.*, 14(4):359–364, 1976. `doi:10.1016/0012-365X(76)90068-6`.

[RRR+19] Christian Reiher, Vojtěch Rödl, Andrzej Ruciński, Mathias Schacht, and Endre Szemerédi. Minimum vertex degree condition for tight Hamiltonian cycles in 3-uniform hypergraphs. *Proc. Lond. Math. Soc. (3)*, 119(2):409–439, 2019. `doi:10.1112/plms.12235`.

[RRS06] Vojtěch Rödl, Andrzej Ruciński, and Endre Szemerédi. A Dirac-type theorem for 3-uniform hypergraphs. *Comb. Probab. Comput.*, 15(1-2):229–251, 2006. `doi:10.1017/S0963548305007042`.

[RRS08] Vojtěch Rödl, Andrzej Ruciński, and Endre Szemerédi. An approximate Dirac-type theorem for $k$-uniform hypergraphs. *Combinatorica*, 28(2):229–260, 2008. `doi:10.1007/s00493-008-2295-z`.

[RRS11] Vojtech Rödl, Andrzej Ruciński, and Endre Szemerédi. Dirac-type conditions for Hamiltonian paths and cycles in 3-uniform hypergraphs. *Adv. Math.*, 227(3):1225–1299, 2011. `doi:10.1016/j.aim.2011.03.007`.

[RS04] Vojtěch Rödl and Jozef Skokan. Regularity lemma for $k$-uniform hypergraphs. *Random Struct. Algorithms*, 25(1):1–42, 2004. `doi:10.1002/rsa.20017`.

[RS07]   Vojtěch Rödl and Mathias Schacht. Regular partitions of hypergraphs: regularity lemmas. *Comb. Probab. Comput.*, 16(6):833–885, 2007. `doi:10.1017/S0963548307008553`.

[SV08]   Benny Sudakov and V. H. Vu. Local resilience of graphs. *Random Struct. Algorithms*, 33(4):409–433, 2008. `doi:10.1002/rsa.20235`.