Thresholds in discrete structures

Dissertation

zur Erlangung des Doktorgrades an der
Fakulät für Mathematik, Informatik und Naturwissenschaften
der Universität Hamburg
Fachbereich Mathematik

vorgelegt von
Bjarne Schülke

Hamburg
Oktober 20, 2021

Erstgutachter:	Prof. Dr. Mathias Schacht (Cobetreuer)
Zweitgutachter:	Prof. Dr. Tomasz Łuczak
Drittgutachter:	Prof. Dr. Vojtěch Rödl
Datum der Disputation:	16.12.2021

Contents

1. Introduction

Combinatorics is often very broadly defined as the study of finite or discrete structures. While the earliest combinatorial investigations seem to go back much further [117], often Euler's solution to the "Seven bridges of Königsberg" problem is mentioned as an early example of graph theory, a central part of combinatorics. This work presents several results in extremal combinatorics, in which one studies thresholds and "extremes" in the behaviour of discrete structures like graphs and hypergraphs. Classic results in this area include Turán's theorem, which provides the threshold edge density of graphs above which the existence of a clique as a subgraph is guaranteed, and Dirac's theorem, which determines the threshold minimum degree of graphs above which the existence of a Hamiltonian cycle is guaranteed.

Extremal combinatorics has progressed significantly in the recent decades. Many powerful tools, like Szemerédi's regularity lemma, its extension to hypergraphs, the container method, and the absorption method, have been developed. Due to this technological improvement, progress has been made on several major open problems and some have been solved. Simultaneously, connections to other fields of mathematics, such as number theory and probability theory, have opened up. In combinatorics, it often happens that the methods used to prove a new result are even more important than the result itself since they may be applied to different problems and offer some deeper insight into the problem. In this introduction, we shall therefore not only introduce the topics and results which are proved in the subsequent chapters of the thesis but also give sketches of the arguments used to obtain the results.

We will assume a basic knowledge of definitions and results as they can for instance be found in Diestel's standard book on graph theory [26]. Important objects investigated in this work but not extensively covered in [26] are hypergraphs. A *hypergraph* $H = (V, E)$ consists of a *vertex set* V and an *edge set* $E \subseteq \mathscr{P}(V) = \{e : e \subseteq V\}$ and similarly as for graphs, we may abbreviate the notation of an edge $\{v_1, \ldots, v_k\}$ to $v_1 \ldots v_k$. For a positive integer k, we call H a *k-uniform* hypergraph (or a *k-graph*) if $E \subseteq V^{(k)} = \{e \subseteq V : |e| = k\}$.[1] Note that graphs are 2-uniform hypergraphs.

Below, we give a very rough overview over the topics covered in this thesis. Afterwards, we provide slightly more detailed introductions to each of the topics. In order to not let this introduction become too lengthy and technical, we will refrain from defining every term we use here formally within this section if the exact definition is not very important for the introduction. The full introduction to each result, with a broader background and the necessary setup, will be given at the beginning of each chapter.

The first part of this thesis (Chapters 2-4) is devoted to several problems about Hamiltonian cycles in hypergraphs. For the following reasons, these serve as good examples of extremal combinatorics. Firstly, the results are typical threshold results, determining above which value of some parameter, such as the minimum degree, a certain behaviour, like the existence of a Hamiltonian cycle, is guaranteed. Secondly, "typical" combinatorial counting and construction techniques are used throughout the works, first and foremost the absorption method. Furthermore, we apply results on quasirandomness including a (weak[2]) hypergraph version of the regularity lemma and we make use of the probabilistic method at several points. Both quasirandom and probabilistic arguments have become important tools in extremal combinatorics.

Afterwards, in Chapter 5, we look at another problem involving spanning substructures. Here however, not only the methods but also the objects considered are random and we mix the search for spanning substructures with Ramsey theory. Namely, we will try to cover edge-coloured random graphs $G(n, p)$ with as few monochromatic trees as possible. The main threshold under investigation here is the value for p (at least the right order of magnitude) above which with high probability any colouring of $G(n, p)$ allows a covering with few monochromatic trees.[3]

The results in Chapter 6 are not directly "threshold results" yet they are strongly connected with extremal combinatorics. This connection is Sidorenko's conjecture, which basically states that, given a bipartite graph H, the number of copies of H in a graph G is minimised when G is quasirandom. In other words, it makes an assertion about the extremal behaviour of the "homomorphism density" of H. To approach this conjecture as well as for its own sake, one can investigate for which graphs H we can define a norm

[1] Often also the notation $2^V = \mathscr{P}(V)$ and $\binom{V}{k} = V^{(k)}$ is used.

[2] Unfortunately, this thesis does not contain a proof using the "full" hypergraph regularity lemma, we refer to [96] to see it at work.

[3] The number of trees needed could also be seen as a threshold. When we worked on this problem, the "right" number of trees that suffice had not yet been found, even for relatively large p.

via its homomorphism density (we will make this more precise below). We prove that several graphs which have been in the spotlight regarding Sidorenko's conjecture do not define a norm, which arguably makes them even more interesting for the investigation of Sidorenko's conjecture.

Lastly, we return to hypergraphs in Chapter 7. Call a hypergraph $H = (V, E)$ *hereditary* if for all e and e' with $e' \subseteq e \in E$, we have $e' \in E$.[4] In Chapter 7, the extremal behaviour of the number of edges in hereditary hypergraphs with respect to the minimum degree is studied. That is, for certain integers s, we determine the maximum real $m(s)$ such that every hereditary hypergraph with "edge density"[5] at most $m(s)$ has minimum (vertex) degree at most s.

While both the first few chapters and the last chapter are about results in hypergraphs, the flavour is quite different. In the first chapters, the hypergraphs are uniform and the uniformity is very small compared to the number of vertices. The problems and methods are relatively closely related to those in graphs (although of course the consideration of hypergraphs instead of graphs generally still brings with it major difficulties). Loosely speaking, here we consider hypergraphs as "graphs of higher dimension".

In contrast, the last chapter is set in a subarea of extremal combinatorics called extremal set theory, where the hypergraphs dealt with are not necessarily uniform and if they are uniform, often the number of vertices is only polynomial in the uniformity. The problems studied in this field are often distinct from those studied in graphs; in fact, for several of the problems, the graph versions would be rather trivial. In addition, extremal set theory encompasses its own unique set of proof techniques.

1.1 Spanning substructures in graphs and hypergraphs

The first part of this thesis concerns the question which conditions guarantee the existence of spanning substructures in a graph or hypergraph, that is, substructures containing all vertices. There has been significant progress on problems of this kind in the recent decade, much due to the absorption method introduced by Rödl, Ruciński, and Szemerédi in their paper [98] and reviewed by Szemerédi in [111]. Roughly speaking, this strategy reduces the problem of finding some spanning substructure in a graph or hypergraph to the usually

[4]Hereditary hypergraphs are also called abstract simplicial complexes or downsets.
[5]More precisely, we mean $|E|/|V|$.

much simpler problem of finding an almost spanning substructure.

1.1.1 Hamiltonian cycles

Hamiltonian cycles, that is, cycles containing every vertex of a graph, form a central theme in classic graph theory. While the problem of determining whether or not a given graph contains a Hamiltonian cycle is one of Karp's initial 21 NP-complete problems [65] and so there is probably no good characterisation of all Hamiltonian graphs, there are several structural and extremal conditions that guarantee the existence of a Hamiltonian cycle. Perhaps the best known result is Dirac's theorem [27], which states that every graph on $n \geqslant 3$ vertices with minimum degree at least $\frac{n}{2}$ contains a Hamiltonian cycle. Considering slightly imbalanced bipartite graphs shows that this result is tight.

Let us consider possible generalisations of this result to k-uniform hypergraphs. Katona and Kierstead [67] initiated the study of minimal degree conditions for Hamiltonian cycles in hypergraphs and introduced the following notation. Define a *(tight) k-uniform cycle of length ℓ* as a k-uniform hypergraph C for which there is an ordering of the vertices $V(C) = \{v_1, \ldots, v_\ell\}$ such that $E(C) = \{v_i \ldots v_{i+k-1} : i \in [\ell]\}$, where we view the indices as elements of $\mathbb{Z}/\ell\mathbb{Z}$. Since we will only consider tight cycles here, we omit the prefix "tight". Naturally, a *Hamiltonian cycle* in a k-uniform hypergraph H is a cycle in H containing all vertices of H. When looking for generalisations of Dirac's theorem to k-uniform hypergraphs, there are various minimum degrees that can be considered. For a hypergraph $H = (V, E)$ and a set $S \subseteq V$, we write $d(S) = d_H(S) = |\{e \in E : S \subseteq e\}|$ and we define the *minimum i-degree* as $\delta_i(H) = \min_{S \in V^{(i)}} d(S)$. The general problem now reads as follows.

Problem 1.1.1. *For $k \in \mathbb{N}$ and $i \in [k-1]$, determine the infimal $d^k_{k-i} \in [0, 1]$ such that every (large) k-uniform hypergraph H with $\delta_{k-i}(H) \geqslant (d^k_{k-i} + o(1))\binom{n}{i}$ contains a Hamiltonian cycle.*

Note that lower bounds on $\delta_j(H)$ become more restrictive, that is, carry more information, when j increases. Thus, it is not surprising that Problem 1.1.1 was first solved for $i = 1$. In [99], Rödl, Ruciński, and Szemerédi generalised Dirac's result to hypergraphs by proving that k-uniform hypergraphs H with $\delta_{k-1}(H) \geqslant (\frac{1}{2} + o(1))n$ contain a Hamiltonian cycle, which is asymptotically tight. The next step according to the "monotonicity" mentioned above, namely to determine the minimum $(k-2)$-degree guaranteeing a Hamiltonian cycle in k-uniform hypergraphs (the case $i = 2$ of Problem 1.1.1) required several new tricks.

First, Reiher, Rödl, Ruciński, Schacht, and Szemerédi [95] solved the 3-uniform case, after which Polcyn, Reiher, Rödl, Ruciński, Schacht, and the author [92] proved the respective result for 4-uniform hypergraphs. Recently, the general case was solved independently by Lang and Sanhueza-Matamala [77] and by Polcyn, Reiher, Rödl, and myself [93], that is, we proved the following theorem.

Theorem 1.1.2. *For all integers $k \geqslant 3$ and all $\alpha > 0$, there exists an integer n_0 such that every k-uniform hypergraph H on $n \geqslant n_0$ vertices with $\delta_{k-2}(H) \geqslant (\frac{5}{9} + \alpha)\binom{n}{2}$ contains a Hamiltonian cycle.*

Chapter 2 contains the detailed discussion and proof of Theorem 1.1.2. There, we also provide an example showing that this result is asymptotically optimal.

For $i \geqslant 3$, the optimal bounds are not yet known. However, the solutions of the cases $i = 1$ and $i = 2$ of Problem 1.1.1, in particular their proofs, and lower bound constructions due to Han and Zhao [57] make the author cautiously believe that, in fact, d^k_{k-i} is independent of k.

Conjecture 1.1.3. *For every integer $i \geqslant 1$, there exists $d_i \in [0,1]$ such that for all integers $k \geqslant i+1$ and all $\alpha > 0$, there is an integer n_0 such that every k-uniform hypergraph H on $n \geqslant n_0$ vertices with $\delta_{k-i}(H) \geqslant (d_i + \alpha)\binom{n}{i}$ contains a Hamiltonian cycle.*

An even stronger conjecture by Lang and Sanhueza-Matamala [77] states that those d_i exist and that they match the lower bounds due to Han and Zhao [57].

Although Dirac's theorem is arguably the most famous result on Hamiltonian cycles, there are strengthenings of it, in which several vertices are allowed to have smaller degrees than $n/2$. Let $G = ([n], E)$ be a graph on $n \geqslant 3$ vertices and let $d(1) \leqslant \cdots \leqslant d(n)$ be its degree sequence. Pósa [94] proved that if $d(i) \geqslant i+1$ for all $i < (n-1)/2$ and if furthermore $d(\lceil n/2 \rceil) \geqslant \lceil n/2 \rceil$ when n is odd, then G contains a Hamiltonian cycle.

The strongest result of this kind is due to Chvátal [18]. For an integer $n \geqslant 3$, we say that an integer sequence $a_1 \leqslant \cdots \leqslant a_n$ is *Hamiltonian* if every graph $G = ([n], E)$ whose degree sequence $d(1) \leqslant \cdots \leqslant d(n)$ satisfies $a_i \leqslant d(i)$, for all $i \in [n]$, contains a Hamiltonian cycle. Chvátal characterised all Hamiltonian sequences by showing that for $n \geqslant 3$, an integer sequence $0 \leqslant a_1 \leqslant \cdots \leqslant a_n < n$ is Hamiltonian if and only if for all $i < \frac{n}{2}$, we have: $a_i \leqslant i \Rightarrow a_{n-i} \geqslant n - i$.

Building on work from my master thesis, I showed a Pósa-type strengthening of the Dirac-type result for 3-uniform hypergraphs that Rödl, Ruciński, and Szemerédi proved in [98]. Call a (symmetric) matrix $(d_{ij})_{ij}$ *Hamiltonian* if every 3-uniform hypergraph $H = ([n], E)$

with $d(i,j) = d(\{i,j\}) \geqslant d_{ij}$, for all $\{i,j\} \in [n]^{(2)}$, contains a Hamiltonian cycle. It would be very desirable to have a result for 3-uniform hypergraphs similar to the one by Chvátal for degree sequences in graphs, that is, a characterisation of all Hamiltonian matrices. For the graph case, Pósa's result was a step towards the characterisation by Chvátal. The following theorem, which we prove in Chapter 3, can be seen as a 3-uniform (asymptotic) analogue of the theorem by Pósa and a step towards a full characterisation of Hamiltonian matrices.

Theorem 1.1.4. *For $\alpha > 0$, there exists $n_0 \in \mathbb{N}$ such that for all $n \in \mathbb{N}$ with $n \geqslant n_0$, the following holds. If $H = ([n], E)$ is a 3-uniform hypergraph with $d(i,j) \geqslant \min\left(i, j, \frac{n}{2}\right) + \alpha n$ for all $ij \in [n]^{(2)}$, then H contains a Hamiltonian cycle.*

In Chapter 3, we also see that this result is tight in a certain sense.

A variation of the foregoing line of results concerns decompositions into Hamiltonian cycles. A decomposition of a k-uniform hypergraph H into Hamiltonian cycles is a collection of edge-disjoint Hamiltonian cycles in H such that the union of their edges is $E(H)$. One of the earliest results regarding decompositions of graphs is Walecki's theorem from the 1890s, which states that a complete graph on an odd number of vertices has a decomposition into Hamiltonian cycles. An interesting follow-up of Problem 1.1.1 is how many edge-disjoint Hamiltonian cycles can be guaranteed in k-uniform hypergraphs H satisfying the minimum degree condition $\delta_{k-i}(H) \geqslant (d^k_{k-i} + o(1))\binom{n}{i}$ (for some $i \in [k-1]$) and, in particular, whether such H can in fact be decomposed into Hamiltonian cycles if it is degree-regular. For graphs, it was shown by Csaba, Kühn, Lo, Osthus, and Treglown [24] that a decomposition is indeed possible but for $k \geqslant 3$, a proper decomposition even of complete k-uniform hypergraphs into Hamiltonian cycles is not yet known to exist. Previous results about approximate decompositions assumed strong quasirandomness properties [5, 30] or did not include tight Hamiltonian cycles [37]. In recent work, Joos, Kühn, and the author [64] proved an approximate decomposition result for regular k-uniform hypergraphs $H = (V, E)$ with $\delta_{k-1}(H) \geqslant (1/2 + o(1))|V|$.

In fact, a much stronger result is shown. For $\eta, \varrho, r > 0$, we say that a k-uniform hypergraph $H = (V, E)$ is η-intersecting if for any two sets $e, f \in V^{(k-1)}$, we have that their neighbourhoods intersect in at least $\eta|V|$ vertices, i.e., $|\{v \in V : e \cup \{v\}, f \cup \{v\} \in E\}| \geqslant \eta|V|$ and we call H ϱ-almost r-regular if for every vertex $v \in V$, we have $d(v) = (1 \pm \varrho)r$. Note that if H satisfies $\delta_{k-1}(H) \geqslant (1/2 + \eta)|V|$, then it is 2η-intersecting and if H allows a decomposition into r/k Hamiltonian cycles, then H is (0-almost) r-regular.

Our main result guarantees an approximate decomposition of η-intersecting ϱ-almost regular k-uniform hypergraphs not only into Hamiltonian cycles but into any cycle factors of not too small girth.[6] Previously, it was not even known whether a single such cycle factor is guaranteed by $\delta_{k-1}(H) \geqslant (1/2 + o(1))|V|$.

Theorem 1.1.5. *For all integers $k \geqslant 2$ and all $\eta, \varepsilon > 0$, there exist integers L and n_0, and $\varrho > 0$ such that every η-intersecting ϱ-almost r-regular k-uniform hypergraph H on $n \geqslant n_0$ vertices contains edge-disjoint copies of any given cycle factors $\mathcal{C}_1, \ldots, \mathcal{C}_{r'}$, where $r' \leqslant (1-\varepsilon)r/k$, whose girths are at least L.*

For the sake of clarity, let us also state the following result, which follows from Theorem 1.1.5. Let $\mathrm{reg}_k(H)$ be the largest integer r divisible by k such that H contains a spanning r-regular subhypergraph and note that H can have at most $\mathrm{reg}_k(H)/k$ edge-disjoint Hamiltonian cycles.

Theorem 1.1.6. *For all integers $k \geqslant 2$ and all $\varepsilon > 0$, there exists an integer n_0 such that every k-uniform hypergraph H on $n \geqslant n_0$ vertices satisfying $\delta_{k-1}(H) \geqslant (1/2 + \varepsilon)n$ contains $(1-\varepsilon)\,\mathrm{reg}_k(H)/k$ edge-disjoint Hamiltonian cycles.*

Thus, the (asymptotically) tight minimum $(k-1)$-degree condition that ensures one Hamiltonian cycle in fact already implies that H contains almost as many Hamiltonian cycles as it possibly can, given trivial (degree-regularity) reasons. In particular, this gives an approximate decomposition if H is vertex degree regular. We will prove these results in Chapter 4.

Let us mention a few words about the proofs of the results above. The general strategy is that of absorption which, in its modern form, was introduced by Rödl, Ruciński, and Szemerédi [98] and surveyed by Szemerédi [111]. Assuming our goal is to construct a Hamiltonian cycle in a k-uniform hypergraph H, then the main idea is as follows[7]. We begin by setting aside a special structure, the *absorbing path*. This is a relatively short path P_A which can absorb any small set of vertices, that is, P_A has the property that for every small set of vertices X, there is a path P'_A such that $V(P'_A) = V(P_A) \cup X$ and such that the $(k-1)$-tuples at both ends of P'_A are the same as those at the ends of P_A. Next, in the complement of P_A, we find a long path Q that contains almost all vertices. If we have proved good connectivity properties (which often is a major obstacle in the proofs), we can subsequently connect Q and P_A to a cycle and absorb the remaining vertices into P_A to

[6] As in graphs, we call a k-graph $\mathcal{C}$ a *cycle factor* (with respect to H) if $\mathcal{C}$ is a union of vertex-disjoint cycles and has the same number of vertices as H. In particular, a Hamiltonian cycle is a cycle factor. The *girth* of a cycle factor is the length of its shortest cycle.

[7] Figure 3.2.1 provides an illustration of this general strategy.

obtain a Hamiltonian cycle. We can thus divide a proof via absorption into the following steps: connecting, absorbing, and covering[8]. The crucial point that makes this approach so powerful is that for many problems, it is much easier to find a substructure containing almost all vertices than one containing all.

Before giving a slightly more detailed explanation of the individual steps of the absorption method, let us briefly sketch how it is used to prove Theorem 1.1.5. We begin by setting aside a small, randomly chosen spanning subhypergraph $F \subseteq H$ for later use. Afterwards, by using a fractional cycle decomposition and a pseudorandom matching in an auxiliary hypergraph, we (almost) decompose the other edges of H into edge-disjoint collections of vertex-disjoint paths such that each of these collections covers almost all vertices. Next, we aim to use the edges in F to complete each collection $\mathcal{P}$ of paths to a cycle factor. To this end, we essentially apply the usual steps of the absorption method to F induced on the vertices not covered by $\mathcal{P}$. By choosing F (and some paths in $\mathcal{P}$ to be "dissolved") randomly, we can guarantee enough quasirandom properties in that hypergraph to be able to perform these steps. However, we also need to take care that we do not use any edge of F during the completion of two different path collections. This can be achieved by performing each construction step probabilistically and using Freedman's inequality to ensure that F retains quasirandom properties when used edges are deleted.

Having set up the overall picture, let us take a closer look at some of the new developments that the proofs of the theorems above brought to the three main steps of the absorption method. For this discussion, let us introduce the following definition. It is often useful to consider something like a projection of a hypergraph with respect to some (small) vertex set. Given a hypergraph $H = (V, E)$ and $S \subseteq V$, we define the link of S (with respect to H) as the hypergraph $L_S = (V \setminus S, \{e \subseteq V \setminus S : e \cup S \in E\})$ and if $S = \{v\}$ or $S = \{u, v\}$, we may simply say the link of u or the link of uv and write L_u, L_{uv}, respectively. Often the vertex set of the link is not very important and so it varies in different definitions used for different problems.

Connecting

Usually, we need to connect several so called absorbers (see below), to the absorbing path P_A and several long paths to the almost covering path Q. Moreover, as mentioned

[8]Note that we will "connect" at several points and the actual "absorption" happens at a later stage than what we usually refer to when speaking about "absorbing" as a step. More accurately, we should perhaps say the steps are "proof that connecting is possible", "construction of the absorbing structure", "construction of the (almost) covering path(s)", and "conclusion".

above, we need to be able to connect P_A and Q. The Connecting Lemma ensures that we have a sufficiently good connectivity property in our hypergraph to make these connections. In its simplest form, the Connecting Lemma would state that any two $(k-1)$-tuples of vertices can be connected by many paths of relatively short length.

While the proofs of the Connecting Lemma in the early articles on the $(k-1)$-degree were somewhat involved, it is now known that by using a simple adaptation of the proof of the Connecting Lemma in [95], it can be shown quite easily that in a hypergraph H with $\delta_{k-1}(H) \geqslant (1/2 + o(1))n$ any two $(k-1)$-tuples (of vertices) can be connected by many paths of constant length. However, if we just assume a minimum $(k-2)$-degree of $(5/9 + o(1))\binom{n}{2}$, such a general Connecting Lemma, stating that any two $(k-1)$-tuples of vertices can be connected by many paths of constant length, is not true (see [95] for a counterexample). To deal with this, *connectable pairs* (and *connectable* $(k-1)$*-tuples*) were introduced in [95] and the subsequent articles on the $(k-2)$-degree [92, 93]. The Connecting Lemma then states that any two *connectable* $(k-1)$-tuples can be connected by many paths of short length. Note that this forces us to take care in all subsequent steps that we can guarantee connectable tuples in all those positions of our construction at which we later need to make connections.

Let $(a_1, \dots, a_{k-1})$ and $(b_1, \dots, b_{k-1})$ be two connectable $(k-1)$-tuples between which we aim to find many paths of short length. Roughly speaking, we proceed as follows (see also Figures 2.3.1 and 3.3.1). Using the eventual definition of a connectable $(k-1)$-tuple, we can show that there are many tuples $(u, q_1, \dots, q_{2k-4}, w)$ of vertices such that $a_1 \dots a_{k-1}u$ and $b_1 \dots b_{k-1}w$ are edges, $(a_2, \dots, a_{k-1})$ and $(q_1, \dots, q_{k-2})$ have "good connectivity properties" in the link of u, while $(q_{k-1}, \dots, q_{2k-4})$ and $(b_1, \dots, b_{k-2})$ have good connectivity properties in the link of w, and, lastly, $q_1, \dots, q_{2k-4}$ is a walk in the link of uw. These good connectivity properties and the induction hypothesis then yield many paths (of short length) between $(a_2, \dots, a_{k-1})$ and $(q_1, \dots, q_{k-2})$ in the link of u and between $(q_{k-1}, \dots, q_{2k-4})$ and $(b_1, \dots, b_{k-2})$ in the link of w. One can show that this will give rise to many short paths between $(a_1, \dots, a_{k-1})$ and $(b_1, \dots, b_{k-1})$ in H by inserting "different versions" of u respectively w at every k-th position of the aforementioned paths in the links.

This sketch hints at two points. First, the key condition for this approach to work is that there is some kind of connectivity property in the link hypergraphs of some vertices. This can also be seen in the proof of Theorem 1.1.4 in Section 3.3, where in the links of some vertices we essentially have Pósa's degree sequence condition for graphs. The insertion of some vertices $q_1, \dots, q_{2k-4}$ in the middle of the constructed path allows for more flexibility in the proof by distributing the conditions with respect to $(a_1, \dots, a_{k-1})$ and $(b_1, \dots, b_{k-1})$ among u and w. This is not necessary in the basic version of the proof in the setting of the $(k-1)$-degree. Nevertheless, it also turned out to be crucial in the proof of Theorem 1.1.4.

Second, the outline above indicates that to prove the Connecting Lemma given a minimum $(k-2)$-degree - and to define connectable tuples - induction will be helpful. Indeed, this is how we will proceed in the proof of Theorem 1.1.2. However, to handle the induction occurring in the covering part below more efficiently, it is useful to "store" more information than just the hypergraph. Therefore, we will introduce *constellations* in Chapter 2 and prove the Connecting Lemma in this context.

Lastly, let us mention that if there are many paths of fixed length between two $(k-1)$-tuples, then the probability that one particular set of vertices appears in a path chosen uniformly at random is relatively small. Together with similar observations in the absorbing and the covering part, this is basically the reason why all the constructions in an absorbing proof can be performed probabilistically, which is key in the proof of Theorem 1.1.5.

Absorbing

The basic approach to construct an absorbing path is as follows. First, show that for every vertex v, there are many *v-absorbers*. In the simplest case, a v-absorber is a short path A not containing v such that v can be inserted at some "inner" position, meaning that there is a path A' containing v and having the same $(k-1)$-tuples at both ends as A. If there are many such absorbers for every vertex, we can pick a selection of relatively few vertex-tuples which still contains a substantial number of absorbers for every vertex (and only few overlapping pairs of absorbers). Subsequently, we connect the absorbers in this collection to an absorbing path. Into this path, we can now absorb any small set of vertices, by greedily inserting each vertex into a distinct absorber.

The main obstacle in this argument is to find the right structure for a v-absorber. However, Polcyn and Reiher [91] suggested an approach that is somewhat more generic and we use this to prove the minimum $(k-2)$-degree condition for Hamiltonian cycles in Chapter 2. It is based on a result by Erdős [31] which states that the Turán density of a k-partite k-uniform hypergraph is 0. Due to supersaturation, this allows us to make use of any k-partite structure in our absorbers. Then, one can argue that it is indeed possible to extend such a structure in a way that a vertex v can be inserted and that all the end-tuples of paths in this structure are connectable.[9] The k-partite nature of these absorbers entails that we will absorb k vertices at a time.

The difficulty to find the right absorbers in the setting of Theorem 1.1.4 is that there may be pairs of vertices and (even) vertices of very small degree. To overcome this, we introduce a different type of absorbers in that proof, which make use of the fact that with the given degree condition it is possible to "climb up" the degree sequence (in fact, this is essentially also what guarantees the connectivity property in the links which is used in the proof of the Connecting Lemma).

The proof of Theorem 1.1.5 involves several rounds, in each of which we perform the absorption method in a probabilistic way and there are some new obstacles arising here. One is that instead of a Hamiltonian cycle we construct an arbitrary cycle factor, possibly one in which each cycle is of constant length. Therefore, we potentially need to distribute the absorbers in the "good" selection mentioned above among several paths.[10] However, since we do not know beforehand which set of vertices is leftover in the covering step, with the usual approach we could not control which absorbers will be used when absorbing this set. Thus, if the absorbers are distributed among several of these cycles, it would not be possible to control the eventual cycle lengths. To deal with this problem, we consider "meta absorbers" each of which is a path with a large but constant number of "normal" absorbers as subpaths. We can construct these in such a way that for every small set (of fixed size), there exists a way to absorb this set such that exactly one normal absorber is used from each meta absorber. Since each meta absorber is of constant length, we can distribute them among different paths, now knowing how many new vertices will be added to each path later on. A second difficulty that arises in the proof, is that for the subsequent covering step, we want the hypergraph to remain almost vertex-regular after setting aside the absorber structures. To construct the meta absorbers while maintaining an almost regular hypergraph, it turns out to be efficient to use random walks instead of picking each absorber independently at random as in the usual approach.

Covering

In the covering part of proofs via absorption, a path containing almost all vertices is constructed. Due to the Connecting Lemma (and the existence of the reservoir, but we defer the discussion of this to the actual proofs), it is enough to cover almost all vertices

[9] See Figures 2.5.1 and 2.5.2.

[10] To obtain a cycle factor, we first cover almost all vertices by paths of short lengths and subsequently connect these, together with "meta absorbers", to cycles into which we absorb the remaining vertices.

by paths whose lengths are a large constant.

The proofs in the chapters ahead will in fact provide three different approaches to proof the existence of such a covering in some hypergraph H. In the proof of Theorem 1.1.4, we use the weak hypergraph regularity lemma, which is a relatively straightforward generalisation of Szemerédi's regularity lemma for graphs. This provides a partitioning of the vertex set into relatively few partition classes such that between almost any three of those partition classes, the edges of H are distributed quasirandomly. Then the *reduced hypergraph* is defined as the hypergraph which has the partition classes as vertices and whose edges are given by those triples of vertex classes on which H is quasirandom and has some positive density. One can then show (still in the setting of Theorem 1.1.4) that this reduced hypergraph possesses a pair degree condition very similar to the one in H. This enables us to construct a matching in the reduced hypergraph which covers almost all vertices. Next, by a typical quasirandomness argumentation, one sees that each edge in the reduced hypergraph can be "unpacked" to a collection of relatively long (vertex-disjoint) paths in H covering almost all vertices of the three partition classes that form the edge. Thus, the edges of the almost perfect matching in the reduced hypergraph yield an approximate path cover in H as desired.

This approach would be more difficult if in addition we had to take care that all the end-tuples of the paths in the covering are connectable. In the proof of the minimum pair degree condition for Hamiltonian cycles in 4-uniform hypergraphs [92], we introduced a different strategy that was based on the respective 3-uniform result [95] yet involved several new aspects. Since this strategy is inductive in nature, we could lift it to full generality in the proof of Theorem 1.1.2. Although the full proof involves several major technical obstacles, we may give an idea of it here. Take a maximal collection $\mathscr{C}$ of vertex-disjoint paths on M vertices (M-vertex paths) whose end-tuples are connectable and call the set of uncovered vertices U. If $|U|$ is not small enough, we consider so-called blocks, which are the vertex sets of the paths in $\mathscr{C}$ as well as some arbitrary partition of the vertices not covered by $\mathscr{C}$ into sets of size M. We aim to use (the vertices in) M of these blocks together with vertices in U to construct $M+1$ vertex-disjoint M-paths with connectable end-tuples. Subsequently, replacing the paths in $\mathscr{C}$ corresponding to blocks used for this construction by the $M+1$ newly constructed paths, leaves us with a path collection that contains at least one path more than $\mathscr{C}$, a contradiction.

To find M blocks which enable this augmentation, we analyse *societies*, which are sets of M blocks. Given a vertex $u \in U$, we say that a society $\mathscr{S}$ is *useful* for u, if the link (constellation) of u induced on $S = \bigcup_{B \in \mathscr{S}} B$ has certain nice properties. These will allow us to apply induction and cover almost all vertices in S by "good" paths in the link of u (the technicalities involved in this induction called for the definition of constellations to be handled efficiently). We then use a probabilistic argument based on a weighted version of Janson's inequality to show that there is a society which is useful for many vertices in U. By averaging, we obtain a set $U' \subseteq U$ such that in the link of each vertex in U', there exists the same covering of S. Inserting vertices of U' at every k-th position then yields the paths in H we were looking for to augment $\mathscr{C}$. If we prepare a bit more, the properties of the link constellation induced on the vertices of a useful society actually allow us to make sure that the end-tuples of the so constructed paths are indeed connectable in H (and not just in some link constellation).

In the proof of Theorem 1.1.5, more than a path covering of a hypergraph is needed. On the one hand, since here the goal is to construct many edge-disjoint cycle factors, we need to construct many edge-disjoint collections of approximate path coverings. Further, we need to ensure certain quasirandomness conditions of these path covers.

On the other hand, when applying the absorption method to turn each of these approximate path coverings $\mathcal{P}$ into a cycle factor, we need to provide a probabilistic construction of an approximate path covering in a subhypergraph of H.[11] To this end, we can apply the result ensuring many edge-disjoint collections of approximate path coverings (Proposition 4.4.1) and choose one of the path collections at random.

To prove such a covering result, we first consider a fractional cycle decomposition of H (this is a weighted, weaker notion of a cycle decomposition), which exists by an earlier result due to Joos and Kühn [63]. Roughly speaking, we then construct an auxiliary hypergraph whose vertex set consists of the edges of H as well as several disjoint copies of $V(H)$ and whose edges "represent" cycles in H of a fixed length. A matching in this hypergraph will correspond to edge-disjoint collections of cycles in H. At this point, a weighted version of a result by Erhard, Glock, and Joos [29] about quasirandom matchings in hypergraphs can be applied to this auxiliary hypergraph with weights on the edges according to the aforementioned fractional cycle decomposition of H. This will engender a matching in the auxiliary hypergraph with enough quasirandom properties such that the corresponding collection of cycles in H is as desired.

[11] Basically, we need such a covering in the hypergraph $F[V(H) \setminus V(\mathcal{P})]$, mentioned in the general overview above, after setting aside absorbing paths in it.

1.1.2 Covering edge-coloured random graphs with monochromatic trees

In this subsection we combine the search for spanning substructures with Ramsey theory. Ramsey's theorem and Turán's theorem are arguably the two cornerstones of extremal combinatorics.

With Ramsey's theorem in mind and given a positive integer r and some class $\mathcal{C}$ of substructures (like cycles or trees), one may now ask for the minimum number m such that for any r-edge-colouring of K_n, there are m monochromatic copies of elements in $\mathcal{C}$ such that their union covers all vertices of K_n (in this context, we allow the empty graph, single vertices, and edges as trees and cycles). There has been extensive research on problems of this kind and we refer to a review by Gyárfás [54] for an overview.

Here, we are interested in covering with monochromatic trees, in other words, we are looking for the minimum number of monochromatic components that are needed to cover an edge-coloured graph. More precisely, given a graph G and a positive integer r, let $\mathrm{tc}_r(G)$ denote the minimum number m such that in any r-edge-colouring of G, there are m monochromatic trees $T_1, \ldots, T_m$ such that the union of their vertex sets covers $V(G)$, that is,

$$V(G) = V(T_1) \cup \cdots \cup V(T_m).$$

We define $\mathrm{tp}_r(G)$ analogously by requiring the union above to be disjoint (so in particular, $\mathrm{tc}_r(G) \leqslant \mathrm{tp}_r(G)$).

Erdős, Gyárfás, and Pyber [33] conjectured that $\mathrm{tp}_r(K_n) = r - 1$ for all positive integers n and proved the conjecture for $r = 3$ (for $r = 2$ it is easy). Currently, the best known bound is $\mathrm{tp}_r(K_n) \leqslant r$ for sufficiently large n which was proved by Haxell and Kohayakawa [59] and it is not even known whether $\mathrm{tc}_r(K_n) \leqslant r - 1$.

Gyárfás [53] noticed a nice connection between this problem and the famous conjecture by Ryser [61], which asserts that a generalisation of Kőnig's theorem is true for hypergraphs. More precisely, Ryser's conjecture states that if H is an r-partite r-uniform hypergraph, then the smallest size of a vertex cover of H is at most $r - 1$ times the size of the largest matching in H. We will come back to this connection in the sketch of the proof below.

As with many problems in combinatorics which are at first studied for complete graphs, researchers also investigate the random graph $G(n, p)$. For the problem above, this investigation was started by Bal and DeBiasio [6], who proved that if $p \ll \left(\frac{\log n}{n}\right)^{1/r}$, then with high probability (w.h.p.) we have $\mathrm{tc}_r(G(n,p)) \to \infty$ (as $n \to \infty$). They further conjectured that for any $r \geqslant 2$, this is the correct threshold for the event $\mathrm{tp}_r(G(n,p)) \leqslant r$. Kohayakawa, Mota, and Schacht [72] proved that this conjecture holds for $r = 2$, while Ebsen, Mota, and Schnitzer (see also [72]) disproved it for more than two colours.

Subsequently, Bucić, Korándi, and Sudakov [15] proved that for large r, the threshold for the event $\mathrm{tc}_r(G) \leqslant r$ is actually significantly larger than the one conjectured by Bal and DeBiasio. In the quest for the best bounds on p which still guarantee $\mathrm{tc}_r(G(n,p)) \leqslant r$ (before, only bounds directly implied by other results had been known[12]), Bucić, Korándi, and Sudakov showed that if $p \gg \left(\frac{\log n}{n}\right)^{1/2^r}$, then w.h.p. $\mathrm{tc}_r(G(n,p)) \leqslant r$.

In the case of 3-colourings, the results in [15] imply that $\mathrm{tc}_3(G(n,p)) \leqslant 3$ holds w.h.p. if we have $p \gg \left(\frac{\log n}{n}\right)^{1/8}$, and if $p \gg \left(\frac{\log n}{n}\right)^{1/6}$, then w.h.p. $\mathrm{tc}_3(G(n,p)) \leqslant 88$.

Together with Kohayakawa, Mendonça, and Mota, the author [71] proved the following improvement of these results.

Theorem 1.1.7. *If $p = p(n)$ satisfies $p \gg \left(\frac{\log n}{n}\right)^{1/6}$, then with high probability we have*

$$\mathrm{tc}_3\big(G(n,p)\big) \leqslant 3.$$

Let us remark that the bound on $\mathrm{tc}_3(G(n,p))$ is optimal in the sense that if we have $p = 1 - \omega(n^{-1})$, then w.h.p. there is a 3-edge-colouring of $G(n,p)$ for which three monochromatic trees are needed to cover all vertices. To see this, consider three non-adjacent vertices x_1, x_2, and x_3 (which exist w.h.p. for such p), colour the edges incident to x_i with colour i and colour all the remaining edges with any colour.

We now briefly sketch how to proof Theorem 1.1.7. Let $G = G(n,p)$, with $p \gg \left(\frac{\log n}{n}\right)^{1/6}$, and let $\phi : E(G) \to \{\text{red}, \text{ green}, \text{ blue}\}$ be any 3-edge-colouring of G. First note that since we are looking for a covering of the vertices with monochromatic components, it is enough to consider an auxiliary graph F, with $V(F) = V(G)$ and $ij \in E(F)$ if and only if there is, in the colouring ϕ, a monochromatic path in G connecting i and j. Define a 3-edge-colouring ϕ' of F with $\phi'(ij)$ being the colour of any monochromatic path in G connecting i and j. Then any covering of F with monochromatic trees with respect to the colouring ϕ' corresponds to a covering of G with monochromatic trees with respect to the colouring ϕ with the same number of trees.

In our proof, we consider different cases depending on the value of the independence number $\alpha(F)$ of F. If $\alpha(F) = 1$, then F is a complete 3-edge-coloured graph and by the

[12] When we worked on this problem, [15] had not yet appeared on arXiv.org and $\mathrm{tc}_3(G(n,p)) \leqslant 6$ was the best known bound for any sensible p.

aforementioned result by Erdős, Gyárfás, and Pyber [33], there exists a partition of $V(F)$ into 2 monochromatic trees. The remaining proof is divided into the cases $\alpha(F) \geqslant 3$ and $\alpha(F) = 2$.

Case $\alpha(F) \geqslant 3$. In this case there exist three vertices $r, b, g \in V(G)$ such that between any two of them there does not exist any monochromatic path. With high probability, they have a common neighbourhood in G of size at least $np^3/2$. Let X_{rbg} be the largest subset of this common neighbourhood such that for each $i \in \{r, b, g\}$, the edges from i to X_{rbg} in G are all coloured with one colour. Then, since there are no monochromatic paths between any two of r, b, and g, we have $|X_{rbg}| \geqslant np^3/12$ and we may assume that all edges between r and X_{rbg} are red, all between b and X_{rbg} are blue and those between g and X_{rbg} are green. Now notice that all vertices with a neighbour in X_{rbg} are covered by the union of the spanning trees of the red component of r, the blue component of b, and the green component of g. Hence, nothing is left to show if every vertex has a neighbour in X_{rbg}. If this is not the case, we continue by carefully choosing vertices and analysing the possible colourings between these vertices and their common neighbourhood. In such a way, we can first show that five monochromatic trees cover all vertices and subsequently argue that, indeed, three of them suffice.

Case $\alpha(F) = 2$. Let us consider a 3-uniform hypergraph $\mathcal{H}$ defined as follows (this definition is inspired by a construction of Gyárfás [53]). The vertices of $\mathcal{H}$ are the monochromatic components of F and three vertices form a hyperedge if the corresponding three components have a vertex in common. In particular, those three monochromatic components must be of different colours and, hence, $\mathcal{H}$ is a 3-uniform 3-partite hypergraph. Observe that if A is a vertex cover of $\mathcal{H}$, then the monochromatic components associated with the vertices in A cover all the vertices of F. This yields $\mathrm{tc}_3(G) \leqslant \mathrm{tc}_3(F) \leqslant \tau(\mathcal{H})$, where $\tau(\mathcal{H})$ is the covering number of $\mathcal{H}$. On the other hand, observe that each matching M in $\mathcal{H}$ gives rise to an independent set of size $|M|$ in F. Thus, we have $\nu(\mathcal{H}) \leqslant \alpha(F) = 2$, where $\nu(\mathcal{H})$ is the matching number of $\mathcal{H}$. Recall that Ryser's conjecture for $r = 3$ states that for every 3-uniform 3-partite hypergraph $\mathcal{H}$, we have $\tau(\mathcal{H}) \leqslant 2\nu(\mathcal{H})$. As it turns out, $r = 3$ is next to $r = 2$ (Kőnig's theorem) the only known (non-trivial) case in which Ryser's conjecture is true; it was proved by Aharoni [1]. Together with the previous observation, this implies $\mathrm{tc}_3(G) \leqslant 4$.

To show that actually $\mathrm{tc}_3(G) \leqslant 3$, we analyse the hypergraph $\mathcal{H}$ more carefully, reducing the situation to a few possible settings of components covering all vertices. In each of those cases, we can again analyse the possible colourings of edges between certain vertices, inferring that indeed there are 3 monochromatic components which cover all vertices.

1.2 Convex graphon parameters and graph norms

Let us turn our attention to an interesting modern theory within combinatorics, the theory of graphons. This theory is closely related to the well-known conjecture by Sidorenko.[13] Roughly speaking, Sidorenko's conjecture [104, 105] asserts that for any bipartite graph H, the number of copies of H in a graph G is minimised when G is a quasirandom graph. To be more precise, let us introduce some notation and we refer the reader to [81] for further background on graphons and graph limits.

A *graphon* (respectively *signed graphon*) is defined to be a measurable symmetric function $W : [0,1]^2 \to [0,1]$ (respectively $W : [0,1]^2 \to [-1,1]$). Graphons can be seen as a continuous generalisation of (weighted) graphs and, in fact, they appear as limit objects of sequences of weighted graphs. Let $\mathcal{W}$ be the vector space of symmetric (real-valued) bounded measurable functions on $[0,1]^2$.

Given graphs H and G, we are often interested in the number of ways in which we can embed H into G, that is, the number of homomorphisms from H to G (a graph homomorphism from H to G is a map $\phi : V(H) \to V(G)$ such that $\phi(i)\phi(j) \in E(G)$ whenever $ij \in E(H)$). For $W \in \mathcal{W}$, we define the *homomorphism density* of H in W by

$$t_H(W) = \int \prod_{ij \in E(H)} W(x_i, x_j) d\mu^{v(H)},$$

where μ is the Lebesgue measure on $[0,1]$ and $v(H) = |V(H)|$.[14]

Now we can formulate Sidorenko's conjecture.

Conjecture 1.2.1. *Let H be a bipartite graph and let W be a graphon. Then*

$$t_H(W) \geqslant t_{K_2}(W)^{e(H)}. \tag{1.2.1}$$

Given that norms are central in several areas of combinatorics, such as graph limits or additive combinatorics, it is interesting to know if we can define norms by the homomorphism density. Let us write $\|W\|_H = |t_H(W)|^{1/e(H)}$ and $\|W\|_{r(H)} = t_H(|W|)^{1/e(H)}$ (again, we write $e(H) = |E(H)|$) and call a graph H *norming* if $\|\cdot\|_H$ defines a norm on $\mathcal{W}$, and *weakly*

[13] This conjecture was also proposed in a slightly different form by Erdős and Simonovits [35].

[14] The given integral is a short notation for $\int_{[0,1]^{v(H)}} \prod_{ij \in E(H)} W(x_i, x_j) \prod_{i \in V(H)} d\mu(x_i)$

norming if $\|\cdot\|_{r(H)}$ is a norm on $\mathcal{W}$. Lovász [81] and Hatami [58] asked which graphs H are (weakly) norming.

As it turns out, there are strong connections between the theory of norming graphs and Sidorenko's conjecture. In particular, every weakly norming graph satisfies (1.2.1) for every graphon W [58] (we call a graph satisfying (1.2.1) for every graphon W *Sidorenko*).

Much of the work on Sidorenko's conjecture focused on finding more Sidorenko graphs, and although not every Sidorenko graph is weakly norming [20, 74], at times new Sidorenko graphs were found by finding new weakly norming graphs (see, for instance, [22, 58, 81]). Moreover, Conlon and Lee [22] showed that weakly norming graphs can be used as "building blocks" for Sidorenko graphs.

To approach the conjecture, Sidorenko [104, 105] suggested to determine whether $K_{5,5} \setminus C_{10}$ (also called *Möbius ladder*) is Sidorenko or not but despite various partial results on Conjecture 1.2.1 [20–23, 58, 70, 80, 110], this is still not known. In joint work with Lee [79], we showed that this graph is at least not weakly norming. While this does not disprove Sidorenko's conjecture, it underlines the importance of $K_{5,5} \setminus C_{10}$ in the investigation of this conjecture.

In fact, we proved a more general result. For a graph H, we write $H^{\bowtie}$ for the graph with vertex set $\{v_i : v \in V(H), i \in [2]\}$ and edge set

$$\{v_1v_2 : v \in V(H)\} \cup \{u_iv_j : uv \in E(H), \{i,j\} = [2]\}.$$

Observe that $C_5^{\bowtie}$ is isomorphic to $K_{5,5} \setminus C_{10}$. Graphs of this type have been of interest in the theory of weakly norming graphs, for instance, Hatami [58] asked whether $C_{2k}^{\bowtie}$ and the Möbius ladder $C_5^{\bowtie}$ are weakly norming. We answer both these questions.

Theorem 1.2.2. *For every $k > 4$, $C_k^{\bowtie}$ is not weakly norming.*

Our method, which can also be used to prove that certain graphs are not norming, relies on the not very difficult to prove observation that H being weakly norming is equivalent to t_H being convex. More precisely, the following holds.

Theorem 1.2.3. *A graph H is weakly norming if and only if $t_H(\cdot)$ is a convex graphon parameter.*

The reason that this is helpful is that it will provide a computational way to show that a graph is not weakly norming.

Given an $n \times n$ symmetric matrix $A = (a_{ij})$, let U_A be the two-variable symmetric step function on $[0,1]^2$ defined by

$$U_A(x,y) = a_{ij}, \text{ if } (i-1)/n \leqslant x < i/n \text{ and } (j-1)/n \leqslant y < j/n$$

and $U_A = 0$ on the measure-zero set $x = 1$ or $y = 1$ for simplicity. Then $A \mapsto U_A$ is a linear map and

$$t_H(U_A) = n^{-v(H)} \sum_{\phi: V(H) \to [n]} \prod_{uv \in E(H)} a_{\phi(u)\phi(v)}.$$

In other words, $t_H(U_A)$ is $n^{-v(H)}$ times a homogeneous $\binom{n+1}{2}$-variable polynomial of degree $e(H)$. We call the polynomial $P_{H,n}(A)$ for $A \in \mathrm{Sym}_n$, where Sym_n denotes the $\binom{n+1}{2}$-dimensional vector space of $n \times n$ real symmetric matrices. It is not difficult to derive the following formulation of Theorem 1.2.3 in terms of $P_{H,n}$.

Theorem 1.2.4. *A graph H is weakly norming if and only if $P_{H,n}(\cdot)$ is a convex polynomial on the positive orthant for every $n \in \mathbb{N}$.*

Due to rather standard multidimensional analysis, this result on the other hand has the following corollary.

Corollary 1.2.5. *A graph H is weakly norming if and only if the Hessian $\nabla^2 P_{H,n}(A)$ is positive semidefinite for every $A \in \mathrm{Sym}_n$ with positive entries and $n \in \mathbb{N}$.*

To show that a graph H is not weakly norming, it now suffices to find an example of a graph G such that this Hessian is not positive semidefinite for the adjacency matrix of G. While we used a computer program to find such an example initially, we were able to simplify the example and prove that the Hessian is not positive semidefinite by explicitly analysing homomorphisms from H to G. Thus, our eventual proof does not rely on any computer calculations.

As mentioned before, we also use an analogous approach to prove similar results for norming instead of weakly norming.

1.3 On extremal problems concerning traces

The last problem we consider in this thesis is a problem in extremal set theory, which studies the extremal behaviour of families of sets. Let X be some set and consider some family $\mathcal{F} \subseteq \mathscr{P}(X)$ (so $(X, \mathcal{F})$ is a hypergraph[15]). We may now ask: How large can $\mathcal{F}$ be, given that it has a certain property? What gives this subfield a distinctive flavour compared to the problems mentioned above, is that here the considered families are not always uniform and if they are, say, k-uniform, one is often also interested in relatively small vertex sets, that is, $|X|$ being only polynomial in k. Further, the proof techniques are often quite different from those in (extremal) graph theory. We recommend the books by Frankl and Tokushige [42] and by Gerbner and Patkós [48] for an overview of extremal set theory.

Here, we are interested in the notion of traces. For a set X, a family $\mathcal{F} \subseteq \mathscr{P}(X)$, and a set $T \subseteq X$ we define the trace of $\mathcal{F}$ on T to be $\mathcal{F}_{|T} = \{F \cap T : F \in \mathcal{F}\}$. The following problem can be interpreted as Turán problem for traces of arbitrary families. For given positive integers n and s with $n > s$, how large does a family $\mathcal{F}$ on n vertices have to be to guarantee that there is some subset of s vertices on which $\mathcal{F}$ has a "full" trace? To make this precise and to set it into a broader context, let us introduce the following notation. For integers n, m, a, and b, we write $(n, m) \to (a, b)$ if for every family $\mathcal{F} \subseteq \mathscr{P}(X)$ with $|\mathcal{F}| \geqslant m$ and $|X| = n$, there is an a-element set $T \subseteq X$ such that $|\mathcal{F}_{|T}| \geqslant b$. The very general problem is to determine, given n, a, and b, the minimal m such that $(n, m) \to (a, b)$. This problem has also been considered for uniform hypergraphs, namely, what is the maximum number $f_k(n, v, e)$ of edges in a k-uniform hypergraph on n vertices not containing e edges spanned by at most v vertices. The investigation of the uniform version was initiated by Brown, Erdős, and Sós [14, 107], who also conjectured [32, 34] that for every $e \geqslant 3$, $f_3(n, e+3, e) = o(n^2)$, which is among the most important open problems in extremal combinatorics. For a broader overview of the study of traces, we refer the reader to Chapter 8 in [48].

Let us come back to the Turán problem for traces of arbitrary families. Regarding the question asked above, Erdős [101] asked whether the following result holds, which was subsequently proved independently by Sauer [101], Shelah and Perles [103], and Vapnik and Červonenkis [113].

Theorem 1.3.1. *Let $n \geqslant s \geqslant 0$ and m be integers with $m > \sum_{0 \leqslant i < s} \binom{n}{i}$. Then*

$$(n, m) \to (s, 2^s).$$

Note that the bound on m is best possible. Further, although far beyond the scope of

[15]The terms "family" and "hypergraph" are often used interchangeably here when no confusion can arise, that is, we may identify a hypergraph with its edge set.

this thesis, let us mention that this result and the related VC dimension have important applications in machine learning [3].

An even more basic question than the Turán problem is to ask up to which density of edges we can still guarantee that every graph with this density contains a vertex of low degree. While this question is rather easy for graphs, it becomes interesting for abstract simplicial complexes (i.e., hereditary hypergraphs, see below).

This problem was posed in terms of traces by Füredi and Pach [46] and, more recently, by Frankl and Tokushige as Problem 3.8 in [42]:[16]

Problem 1.3.2. *Given non-negative integers n and s, what is the maximum integer $m(n,s)$ such that for every integer $m \leqslant m(n,s)$, we have*

$$(n,m) \to (n-1, m-s).$$

Recall that a family $\mathcal{F} \subseteq \mathscr{P}(X)$ (where X is some set) is said to be *hereditary* if for every $F' \subseteq F \in \mathcal{F}$, we have that $F' \in \mathcal{F}$ (such an $\mathcal{F}$ is also called abstract simplicial complex). In [41], Frankl proved that among families with a fixed number of edges and vertices, the trace is minimised by hereditary families (see Lemma 7.2.1 in Chapter 7). Thus, problems regarding the arrowing notation and in particular Problem 1.3.2 can be reduced to hereditary families (and so given Frankl's result, Theorem 1.3.1 and other previous results on the arrowing notation became easy corollaries).

So Problem 1.3.2 is asking for the maximum integer m such that in every hereditary family on n vertices with at most m edges, there is still a vertex of degree at most s. Conversely, we can say that $m(n,s)+1$ is the minimum number of edges in a hereditary hypergraph on n vertices with minimum (vertex) degree at least $s+1$. The results on Problem 1.3.2 are best formulated as results on $m(s) = \lim_{n\to\infty} \frac{m(n,s)}{n}$ (it is not too difficult to show that this limit exists, see [43]). The investigation of this problem started with Bondy [11] and Bollobás [82] determining $m(0)$ and $m(1)$, respectively. Subsequently, Watanabe [115], [116] and Frankl and Watanabe [43] worked out several more cases for small s. Moreover, Frankl [41] and Frankl and Watanabe [43] proved the first and second part of the following theorem, respectively.

Theorem 1.3.3. *For $d, n \in \mathbb{N}$, we have $m(2^{d-1}-1) = \frac{2^d-1}{d}$ and $m(2^{d-1}-2) = \frac{2^d-2}{d}$.*

In joint work with Piga [89] we made further progress on Problem 1.3.2, solving it for

[16]There have been slightly different versions in use for the arrowing notation and for what we denote by $m(n,s)$. Here we follow the notation in [42].

general $s = 2^{d-1} - c$ as long as c is linearly small in d. More precisely, our main result reads as follows.

Theorem 1.3.4. *Let $d, c, n \in \mathbb{N}$ with $c \leqslant d/4$. Then*

$$m(2^{d-1} - c) = \frac{2^d - c}{d}.$$

We also determined $m(s)$ for some small values of s, one of which had been conjectured by Frankl and Watanabe [43]. Further, we provided a construction showing that the equality in Theorem 1.3.4 does not hold for $c = d$.

It turns out that determining $m(2^{d-1} - c)$ for $c \geqslant 3$ becomes more difficult for a reason. To see why this is the case and how these difficulties can be overcome, let us take a look at the basic idea of the proof. If $d \mid n$, taking n/d disjoint copies of a hereditary family on d vertices with $2^d - c + 1$ edges, gives a family on n vertices with $\frac{n}{d}(2^d - c) + 1$ edges in which each vertex has degree at least $2^{d-1} - c + 1$. Thus, when $d \mid n$, we have $m(n, 2^{d-1} - c) \leqslant \frac{n}{d}(2^d - c)$, which yields $m(2^{d-1} - c) \leqslant \frac{2^d - c}{d}$.

So the main part of the proof is to show that for every hereditary hypergraph $\mathcal{F} = (V, \mathcal{F})$ with $|V| = n$ and minimum degree at least $2^{d-1} - c + 1$, we have that $|\mathcal{F}| \geqslant \frac{n}{d}(2^d - c) + 1$. In the proofs of the identities in Theorem 1.3.3, Frankl and Watanabe used that we have $|\mathcal{F} \setminus \{\varnothing\}| = \sum_{v \in V} \sum_{H \in L_v} \frac{1}{|H|+1}$, where $L_v = \{A \subseteq V : A \cup \{v\} \in \mathcal{F}\}$ is the link of the vertex v. Subsequently, they used a generalised form of the Kruskal-Katona theorem to obtain a lower bound for $\sum_{H \in L_v} \frac{1}{|H|+1}$ that is independent of v. Due to the aforementioned double counting, this in turn yields the lower bound on the number of edges.

For $c \geqslant 3$, there are extremal families which show that a lower bound on $\sum_{H \in L_v} \frac{1}{|H|+1}$ independent of v is not sufficient to provide the desired bound on $|\mathcal{F}|$. To overcome this difficulty, first observe that the double counting argument can be generalised by interpreting $\sum_{H \in L_v} \frac{1}{|H|+1}$ as the weight $w_{\mathcal{F}}(v)$ of a vertex v. We will refer to this weight as *uniform weight* since it can be imagined as uniformly distributing the unit weight of an edge among all of its vertices. In contrast, to prove Theorem 1.3.4, we will at times use non-uniform weights. Moreover, instead of bounding the weight of single vertices, we will sometimes bound the weight of sets of vertices.

The overall structure of the proof can be seen as approximating $\mathcal{F}$ with an extremal family and then showing that deviations of $\mathcal{F}$ from that extremal family engender an increase of the weight of $\mathcal{F}$ (the sum of the weights of all vertices). We proceed by first setting up the "global" structure and afterwards analysing "local" deviations.

To this end, set $V_v = N(v) \cup \{v\}$ for all $v \in V$ (where $N(v) = \{w \in V \setminus \{v\} : \exists A \in \mathcal{F} : \{v, w\} \subseteq A\}$ is the neighbourhood of v) and let $\mathcal{L}$ be a maximal set of vertices w with $|N(w)| \leqslant d - 1$ such that $V_v \cap V_{v'} = \varnothing$ for all $v, v' \in \mathcal{L}$. For all $v \in \mathcal{L}$, we call the set V_v *cluster*. Observe that if the size of the neighbourhood of a vertex is at most $d - 1$, then it has to intersect one of the clusters. For vertices whose neighbourhoods do not intersect any cluster (and which therefore have a neighbourhood of size at least d), we use the uniform weight. To bound these uniform weights, we introduce a "local" lemma which is a close relative to the general form of the Kruskal-Katona theorem mentioned above. Given a vertex of degree at least $2^{d-1} - c + 1$, it provides a lower bound on the uniform weight and, furthermore, the minimum surplus if the link deviates enough from the family which minimises the uniform weight.

The next step is to bound the weight of vertices in the clusters. The difficulty is that the weights of different vertices in a cluster might vary and some weights may be relatively small. To deal with this, instead of bounding the weight of each single vertex, we bound the average weight of the vertices in a cluster. Even if the number of edges inside a cluster is not large enough, $\mathcal{F}$ being hereditary and the minimum degree of $\mathcal{F}$ still provide some lower bound for the number of edges in each cluster. Further, a second local lemma yields that in each cluster with too few edges there are several vertices whose degree with respect to the cluster is smaller than the minimum degree in $\mathcal{F}$. Therefore, there have to be several crossing edges, i.e., edges containing vertices from both the inside and the outside of the cluster. If we use the uniform weight, these crossing edges will contribute enough to the weight of the cluster, even more than needed.

At this point, we still need to bound the weight of vertices with neighbourhoods of size at most $d - 1$ lying outside of any cluster. As mentioned above, the neighbourhood of every such vertex intersects some cluster, meaning every such vertex is contained in a crossing edge. Since distributing the unit weight of crossing edges uniformly among its vertices would contribute more weight than needed to the inside of a cluster, we can assign a larger share to the vertices outside so that both sides will get a share that is big enough.

Organisation

Chapters 2, 3, and 4 are dedicated, respectively, to the proper introductions and proofs of Theorem 1.1.2, Theorem 1.1.4, and Theorem 1.1.5. They essentially consist of the articles [64, 93, 102]. In Chapter 5, essentially consisting of [71], we more carefully introduce and prove Theorem 1.1.7. Chapter 6 deals with "Convex graphon parameter and graph norms". There we prove Theorem 1.2.3 and it basically consists of [79]. Lastly, we discuss the problem on traces of sets in depth in Chapter 7, which includes a proof of Theorem 1.3.4 and essentially consists of [89].

In the appendix, short summaries of this thesis in both English and German are provided and all my publications connected with my PhD studies are listed. In addition, I make a declaration of contributions and a declaration of academic honesty. Further, I will thank various people who have been crucial for this work and for my life.

2. On Hamiltonian cycles in hypergraphs with dense link graphs

2.1 Introduction

Hamiltonian cycles are a central theme in graph theory and extremal combinatorics. Dirac's classic result [27] states that every graph on $n \geqslant 3$ vertices whose minimum degree is at least $\frac{n}{2}$ contains a Hamiltonian cycle. The present work continues the investigation of hypergraph generalisations of Dirac's theorem – an area of research owing many deep insights to Endre Szemerédi.

2.1.1 Hypergraphs and Hamiltonian cycles

For $k \geqslant 2$, a *k-uniform hypergraph* is defined to be a pair $H = (V, E)$ consisting of a (finite) *set of vertices* V and a set

$$E \subseteq V^{(k)} = \{U \subseteq V : |U| = k\}$$

of edges. A k-uniform hypergraph $H = (V, E)$ with n vertices is said to contain a *Hamiltonian cycle* if its vertex set admits a cyclic enumeration $V = \{x_i : i \in \mathbb{Z}/n\mathbb{Z}\}$ such that $\{x_i, x_{i+1}, \ldots, x_{i+k-1}\} \in E$ holds for all $i \in \mathbb{Z}/n\mathbb{Z}$. Observe that this naturally generalises the familiar notion of Hamiltonian cycles in graphs.

In contrast to the graph case, there are several interesting minimum degree notions for hypergraphs. For a k-uniform hypergraph $H = (V, E)$ and a set $S \subseteq V$, the *degree of S in H* is defined by

$$d_H(S) = |\{e \in E : S \subseteq e\}| .$$

Moreover, for an integer i with $1 \leqslant i < k$, the number

$$\delta_i(H) = \min\bigl\{d_H(S)\colon S \in V^{(i)}\bigr\}$$

is called the *minimum i-degree of H*.

The research on minimum i-degree conditions guaranteeing the existence of Hamiltonian cycles in hypergraphs was initiated by Katona and Kierstead [67]. The main problem is to determine, for any two given integers $k \geqslant 2$ and $i \in [k-1]$, the optimal minimum i-degree condition which for k-uniform hypergraphs ensures the existence of a Hamiltonian cycle. Notice that Dirac's aforementioned theorem solves the case $(k, i) = (2, 1)$.

In general, if $i < j$, then a minimum j-degree condition seems to reveal more structural information about a hypergraph than a minimum i-degree condition. For this reason, it is reasonable to suspect that the difficulty of the problem we are interested in increases with $k - i$. The first case, $i = k - 1$, was solved more than a decade ago by Rödl, Ruciński, and Szemerédi [99].

Theorem 2.1.1. *For every integer $k \geqslant 2$ and every $\alpha > 0$, there exists an integer n_0 such that every k-uniform hypergraph H on $n \geqslant n_0$ vertices with $\delta_{k-1}(H) \geqslant \left(\frac{1}{2} + \alpha\right) n$ contains a Hamiltonian cycle.*

Similarly as for Dirac's theorem, slightly unbalanced bipartite hypergraphs show that this result is asymptotically best possible. Our main result addresses the next case, $i = k - 2$.

Theorem 2.1.2. *For every integer $k \geqslant 3$ and every $\alpha > 0$, there exists an integer n_0 such that every k-uniform hypergraph H on $n \geqslant n_0$ vertices with $\delta_{k-2}(H) \geqslant \left(\frac{5}{9} + \alpha\right) \frac{n^2}{2}$ contains a Hamiltonian cycle.*

In previous articles written in collaboration with Ruciński, Schacht, and Szemerédi [92, 95] we solved the cases $k = 3$ and $k = 4$. The general case was also obtained by Lang and Sanhueza-Matamala [77] in independent research. A construction due to Han and Zhao [57] reproduced in the introduction of [92] shows that the number $\frac{5}{9}$ appearing in Theorem 2.1.2 is optimal.

We would like to conclude this subsection by pointing to some problems for future investigations. First and foremost, it remains an intriguing question whether for $k \geqslant 4$, the minimum $(k-3)$-degree condition $\delta_{k-3}(H) \geqslant \left(\frac{5}{8} + o(1)\right) \frac{n^3}{6}$ enforces the existence of a Hamiltonian cycle. Here the number $\frac{5}{8}$ would again match the construction of Han and Zhao [57].

Another possible area of research would be to extend the work of Pósa [94] and Chvátal [18], who in the graph case studied which conditions on the degree sequence (rather than just on the minimum degree) guarantee the existence of Hamiltonian cycles. Such degree sequence versions have recently been obtained for the Hajnal-Szemerédi theorem [55] by Treglown [112] and for Pósa's conjecture (see [38, Problem 9]) by Staden and Treglown [108]. It would be very interesting to find similar theorems for Hamiltonian cycles in hypergraphs. For first results in this direction we refer to [102].

2.1.2 Organisation and overview

We use the *absorption method* developed by Rödl, Ruciński, and Szemerédi and surveyed by Szemerédi himself in [111]. Therefore, the proof decomposes in the usual way into a Connecting Lemma, an Absorbing Path Lemma, and a Covering Lemma.

Very roughly speaking, the Absorbing Path Lemma reduces the task of proving Theorem 2.1.2 to the much easier problem of finding 'almost spanning' cycles in k-uniform hypergraphs H satisfying $\delta_{k-2}(H) \geqslant \left(\frac{5}{9} + \alpha\right) \frac{|V(H)|^2}{2}$. Such an almost spanning cycle is build in two main steps: First, the Covering Lemma asserts that we can cover almost all vertices by means of long paths. Second, the Connecting Lemma allows us to connect these 'pieces' into one long cycle.

In our earlier articles we stored all information about H that became relevant in the course of the proof in various 'setups' and the complexity of these setups got somewhat out of control. To avoid this in the present work, we abandon the setups and replace them by the much more flexible notion of a *constellation* (see Definition 2.2.10 below).

Section 2.2 lays out a systematic treatment of these constellations and contains several auxiliary results that will assist us later. The subsequent Sections 2.3 – 2.6 deal with the main lemmata enumerated above: connecting, absorbing, and covering. Lastly, in Section 2.7 we derive Theorem 2.1.2 from these results.

2.2 Preliminaries

2.2.1 Graphs

In our earlier articles [92, 95] dealing with the 3- and 4-uniform case of Theorem 2.1.2 we inductively reduced connectability properties of the hypergraphs under discussion to connectability properties of their 2-uniform link graphs. Here we pursue the same strategy and the present subsection contains the graph theoretic preliminaries that we require for this purpose. The central notion we work with in this context is taken from [95, Definition 2.2] and reappeared as [92, Definition 2.1].

Definition 2.2.1. *Given $\beta > 0$ and $\ell \in \mathbb{N}$ a graph R is said to be (β, ℓ)-robust if for any two distinct vertices x and y of R the number of x-y-paths of length ℓ is at least $\beta|V(R)|^{\ell-1}$.*

It turns out that every graph whose edge density is larger than 5/9 possesses a robust subgraph containing more than two thirds of its vertices that is quite disconnected from the rest of the graph. The following statement to this effect was proved in [92, Proposition 2.2] (marginally strengthening [95, Proposition 2.3]).

Proposition 2.2.2. *Given α, $\mu > 0$, there exist $\beta > 0$ and an odd integer $\ell \geqslant 3$ such that for sufficiently large n, every n-vertex graph $G = (V, E)$ with $|E| \geqslant \big(\frac{5}{9} + \alpha\big)\frac{n^2}{2}$ contains a (β, ℓ)-robust induced subgraph $R \subseteq G$ satisfying*

(i) $|V(R)| \geqslant \big(\frac{2}{3} + \frac{\alpha}{2}\big)n$,

(ii) *and* $e_G\big(V(R), V \smallsetminus V(R)\big) \leqslant \mu n^2$.

Remark 2.2.3. When using this result we can always assume $\alpha \leqslant 4/9$, for otherwise the hypothesis $|E| \geqslant \big(\frac{5}{9} + \alpha\big)\frac{n^2}{2}$ cannot hold. We shall only apply Proposition 2.2.2 with $\mu \leqslant \frac{\alpha}{4}$. In this case, clause *(ii)* yields

$$e(R) \geqslant \Big(\frac{5}{9} + \frac{\alpha}{2}\Big)\frac{n^2}{2} - \frac{(n - |V(R)|)^2}{2} \overset{(i)}{\geqslant} \Big(\frac{4}{9} + \frac{2}{3}\alpha\Big)\frac{n^2}{2}\,, \tag{2.2.1}$$

where the second inequality relies on $\alpha \leqslant 4/9 < 2/3$. Originally, (2.2.1) was included as a third clause into [92, Proposition 2.2], but it seems preferable to omit this part.

In Section 2.5 below we need to render our absorbers connectable. To this end we shall utilise a consequence of the following graph theoretic lemma.

Lemma 2.2.4. *Let $\alpha > 0$ and let G be a graph with n vertices and at least $\big(\frac{5}{9} + \alpha\big)\frac{n^2}{2}$ edges. If*

$$A = \big\{x \in V(G) \colon d(x) < n/3\big\}$$

and

$$B = \big\{x \in V(G) \colon |N(x) \smallsetminus A| \leqslant \alpha n/3\big\}\,,$$

then

$$e(A \cup B) \leqslant \frac{n^2}{18}\,.$$

Proof. In the special case that $|A| < (\frac{1}{3} - \frac{\alpha}{3})n$, every vertex $x \in B$ satisfies

$$d(x) \leqslant |N(x) \smallsetminus A| + |A| < \frac{n}{3}\,,$$

which yields $B \subseteq A$ and the desired inequality

$$e(A \cup B) = e(A) \leqslant \frac{1}{2}|A|^2 \leqslant \frac{1}{18}n^2\,.$$

So henceforth we may suppose that

$$|A| \geqslant \Big(\frac{1}{3} - \frac{\alpha}{3}\Big)n\,. \tag{2.2.2}$$

Now the definition of A implies

$$\frac{5}{9}n^2 \leqslant 2e(G) = \sum_{x \in V(G)} d(x) \leqslant \frac{1}{3}|A|n + (n - |A|)n = n^2 - \frac{2}{3}|A|n\,,$$

i.e.,

$$|A| \leqslant \frac{2}{3}n\,, \tag{2.2.3}$$

and

$$e(G - A) \geqslant \Big(\frac{5}{9} + \alpha\Big)\frac{n^2}{2} - \frac{1}{3}|A|n\,. \tag{2.2.4}$$

Setting $X = V(G) \smallsetminus (A \cup B)$ we conclude from the definition of B that

$$2e(B \smallsetminus A) + e(B \smallsetminus A, X) = \sum_{x \in B \smallsetminus A} |N(x) \smallsetminus A| \leqslant |B \smallsetminus A| \cdot \frac{\alpha}{3}n \leqslant \frac{\alpha}{3}n^2\,, \tag{2.2.5}$$

which together with (2.2.4) yields

$$\begin{aligned} |X|^2 \geqslant 2e(X) &= 2e(G - A) - 2e(B \smallsetminus A) - 2e(B \smallsetminus A, X) \\ &\geqslant \Big(\frac{5}{9} + \alpha\Big)n^2 - \frac{2}{3}|A|n - \frac{2}{3}\alpha n^2 \geqslant \frac{5}{9}n^2 - \frac{2}{3}|A|n\,. \end{aligned}$$

In view of (2.2.3) this entails

$$|X|^2 \geqslant \frac{4}{9}n^2 - \frac{2}{3}|A|n + \frac{1}{4}|A|^2 = \Big(\frac{2}{3}n - \frac{1}{2}|A|\Big)^2,$$

wherefore

$$|X| \geqslant \frac{2}{3}n - \frac{1}{2}|A|\,. \tag{2.2.6}$$

Next, we claim that

$$\frac{1}{3}|A|n + |B \smallsetminus A||A| + \frac{1}{2}|X|^2 \leqslant \Big(\frac{1}{3} + \frac{\alpha}{6}\Big)n^2\,. \tag{2.2.7}$$

In view of $|A| + |B \smallsetminus A| + |X| = n$ the left side of this estimate rewrites as

$$\frac{1}{3}|A|n + (n - |A| - |X|)|A| + \frac{1}{2}|X|^2 = \frac{4}{3}|A|n - \frac{3}{2}|A|^2 + \frac{1}{2}(|A| - |X|)^2\,.$$

By (2.2.6) and $X \subseteq V(G) \smallsetminus A$ we have

$$\frac{2}{3}n - \frac{3}{2}|A| \leqslant |X| - |A| \leqslant n - 2|A|$$

and, hence,

$$(|A| - |X|)^2 \leqslant \max\Big\{(n - 2|A|)^2, \Big(\frac{2}{3}n - \frac{3}{2}|A|\Big)^2\Big\}\,.$$

So to conclude the proof of (2.2.7) it suffices to observe that

$$\frac{4}{3}|A|n - \frac{3}{2}|A|^2 + \frac{1}{2}(n - 2|A|)^2 = \frac{n^2}{3} + \frac{1}{6}(n - |A|)(n - 3|A|) \overset{(2.2.2)}{\leqslant} \Big(\frac{1}{3} + \frac{\alpha}{6}\Big)n^2$$

and, similarly,

$$\frac{4}{3}|A|n - \frac{3}{2}|A|^2 + \frac{1}{2}\Big(\frac{2}{3}n - \frac{3}{2}|A|\Big)^2 = \frac{n^2}{3} - \Big(\frac{1}{3}n - \frac{1}{2}|A|\Big)^2 - \frac{1}{8}|A|^2 \leqslant \frac{n^2}{3}\,.$$

Having thus established (2.2.7) we appeal to the definition of A again and observe

$$e(A) + e(G) = \sum_{x \in A} d(x) + e(G - A) \leqslant \frac{1}{3}|A|n + e(G - A)\,.$$

Consequently,

$$e(A \cup B) + e(G) \leqslant \frac{1}{3}|A|n + e(B \smallsetminus A, A) + e(B \smallsetminus A) + e(G - A)$$

and (2.2.5) leads to

$$e(A \cup B) + e(G) \leqslant \frac{1}{3}|A|n + |B \smallsetminus A||A| + e(X) + \frac{\alpha}{3}n^2\,.$$

Owing to (2.2.7) we deduce

$$e(A \cup B) + e(G) \leqslant \Big(\frac{1}{3} + \frac{\alpha}{6}\Big)n^2 + \frac{\alpha}{3}n^2 = \Big(\frac{2}{3} + \alpha\Big)\frac{n^2}{2} \leqslant \frac{1}{18}n^2 + e(G)\,,$$

which implies the desired estimate $e(A \cup B) \leqslant \frac{1}{18}n^2$. □

Remark 2.2.5. The set A already had an appearance in [92] and Lemma 2.3 there is roughly equivalent to the weaker estimate $e(A) \leqslant \frac{n^2}{18}$. Concerning the set B one can prove $|B| \leqslant \frac{n}{3}$, but this fact is not going to be exploited in the sequel.

The following consequence of Lemma 2.2.4 will later be generalised to k-uniform hypergraphs (see Lemma 2.2.7) and constitutes the base case of an induction on k.

Corollary 2.2.6. *Let $\alpha > 0$, and let V be a set of n vertices. If G, G' are two graphs with $V(G)$, $V(G') \subseteq V$ and*

$$e(G),\, e(G') \geqslant \Big(\frac{5}{9} + \alpha\Big)\frac{n^2}{2}\,,$$

then there are at least $\frac{\alpha^2}{3}n^3$ triples $(x, y, z) \in V^3$ such that

- *xyz is a walk in G,*
- *$xy \in E(G')$,*
- *and $d_G(y)$, $d_G(z) \geqslant \frac{n}{3}$.*

Proof. By adding some isolated vertices to G and G' if necessary, we may assume that $V(G) = V(G') = V$. The sieve formula yields

$$|E(G) \cap E(G')| \geqslant 2\Big(\frac{5}{9} + \alpha\Big)\frac{n^2}{2} - \frac{n^2}{2} = \Big(\frac{1}{18} + \alpha\Big)n^2\,.$$

Define the sets A and B with respect to G as in Lemma 2.2.4. In view of that lemma itself, there are at least αn^2 unordered pairs $xy \in E(G) \cap E(G')$ for which $x, y \in A \cup B$ fails. Consequently, there are at least αn^2 ordered pairs $(x, y) \in V^2$ such that $xy \in E(G) \cap E(G')$ and $y \notin A \cup B$. For each of them there are, by the definition of B, at least $\frac{\alpha}{3}n$ vertices z with $yz \in E(G)$ and $z \notin A$. Altogether, this yields at least $\frac{\alpha^2}{3}n^3$ triples (x, y, z) with the desired properties. □

2.2.2 Hypergraphs

In this subsection we introduce our terminology and some preliminary results on hypergraphs. When there is no danger of confusion we shall omit several parentheses, braces, and commas. For instance, we write $x_1 \cdots x_k$ for the edge $\{x_1, \ldots, x_k\}$ of a k-uniform hypergraph.

Walks, paths, and cycles

A k-uniform *walk* W of length $\ell \geqslant 0$ is a hypergraph whose vertices can, possibly with repetitions, be enumerated as $(x_1, \ldots, x_{\ell+k-1})$ in such a way that $e \in E(W)$ if and only if $e = x_i \cdots x_{i+k-1}$ for some $i \in [\ell]$. The ordered $(k-1)$-tuples $(x_1, \ldots, x_{k-1})$ and $(x_{\ell+1}, \ldots, x_{\ell+k-1})$ are called the *end-tuples* of W and we say that W is a $(x_1 \cdots x_{k-1})$-$(x_{\ell+1} \cdots x_{\ell+k-1})$-walk. This notion of end-tuples is not symmetric and implicitly fixes a direction of W. Sometimes we refer to $(x_1, \ldots, x_{k-1})$ and $(x_{\ell+1}, \ldots, x_{\ell+k-1})$ as the *starting* $(k-1)$-*tuple* and *ending* $(k-1)$-*tuple* of W, respectively. We call $x_k, \ldots, x_\ell$ the *inner vertices* of W. Counting them with their multiplicities we say for $\ell \geqslant k-1$ that a walk of length ℓ has $\ell - k + 1$ inner vertices. We often identify a walk with the sequence of its vertices $x_1 x_2 \cdots x_{\ell+k-1}$. If the vertices $x_1, \ldots, x_{\ell+k-1}$ are distinct we call the walk W a *path*. For $\ell > k$ a *cycle of length* ℓ is a hypergraph C whose vertices and edges can be represented as $V(C) = \{x_i \colon i \in \mathbb{Z}/\ell\mathbb{Z}\}$ and $E(C) = \{x_i \cdots x_{i+k-1} \colon i \in \mathbb{Z}/\ell\mathbb{Z}\}$.

Link hypergraphs

Given a k-uniform hypergraph $H = (V, E)$ and a set $S \subseteq V$ with $|S| \leqslant k-2$ we define the $(k - |S|)$-uniform *link hypergraph* H_S by $V(H_S) = V(H)$ and

$$E(H_S) = \{e \smallsetminus S \colon S \subseteq e \in E\}\,.$$

Clearly the vertices in S are isolated in H_S and sometimes it is convenient to remove them. In such cases, we write $\overline{H}_S = H_S - S$. For instance, we have $H_\varnothing = \overline{H}_\varnothing = H$ for every hypergraph H. If $S = \{v\}$ consists of a single vertex, we abbreviate $H_{\{v\}}$ to H_v.

A lemma with two hypergraphs

Our next step is to generalise Corollary 2.2.6 to hypergraphs.

Lemma 2.2.7. *Suppose that $k \geqslant 2$, $\alpha > 0$, and that V is a set of n vertices. If H, H' are two k-uniform hypergraphs satisfying*

$$V(H), V(H') \subseteq V$$

and

$$\delta_{k-2}(H),\, \delta_{k-2}(H') \geqslant \left(\frac{5}{9}+\alpha\right)\frac{n^2}{2}\,,$$

then the number of $(2k-1)$*-tuples* $(x_1, \dots, x_{2k-1}) \in V^{2k-1}$ *such that*

- $x_1 \cdots x_{2k-1}$ *is a walk in* H*,*
- $\{x_1, \dots, x_k\} \in E(H')$*,*
- *and* $d_H(x_2, \dots, x_k),\, d_H(x_{k+1}, \dots, x_{2k-1}) \geqslant \frac{n}{3}$

is at least $\left(\frac{\alpha}{2}\right)^{2^{k-1}} n^{2k-1}$.

Proof. For $k = 2$ this follows from Corollary 2.2.6. Proceeding by induction on k, we assume $k \geqslant 3$ and that the assertion holds for $k-1$ in place of k. Construct an auxiliary bipartite graph Γ with vertex classes V and V^{2k-3} by drawing an edge between $x \in V$ and

$$(x_1, \dots, x_{k-2}, x_k, \dots, x_{2k-2}) \in V^{2k-3}$$

if and only if

(*a*) $x_1 \cdots x_{k-2} x_k \cdots x_{2k-2}$ is a walk in $\overline{H}_x$,

(*b*) $\{x_1, \dots, x_{k-2}, x_k\} \in E(\overline{H}'_x)$,

(*c*) $d_{\overline{H}_x}(x_2, \dots, x_{k-2}, x_k) \geqslant \frac{n}{3}$ and $d_{\overline{H}_x}(x_{k+1}, \dots, x_{2k-2}) \geqslant \frac{n}{3}$.

The induction hypothesis, applied to the hypergraphs $\overline{H}_x$ and $\overline{H}'_x$, reveals that every vertex $x \in V$ has at least degree $\left(\frac{\alpha}{2}\right)^{2^{k-2}} n^{2k-3}$ in Γ. Thus

$$e(\Gamma) \geqslant \left(\frac{\alpha}{2}\right)^{2^{k-2}} n^{2k-2}$$

and the Cauchy-Schwarz inequality implies

$$\sum_{\vec{x} \in V^{2k-3}} |N_\Gamma(\vec{x})|^2 \geqslant \frac{e(\Gamma)^2}{n^{2k-3}} \geqslant \left(\frac{\alpha}{2}\right)^{2^{k-1}} n^{2k-1}\,,$$

where $N_\Gamma(\vec{x})$ denotes the neighbourhood of the vertex $\vec{x}$ in Γ. Now if

$$\vec{x} = (x_1, \dots, x_{k-2}, x_k, \dots, x_{2k-2}) \in V^{2k-3} \quad \text{and} \quad x_{k-1}, x_{2k-1} \in N_\Gamma(\vec{x})$$

are arbitrary, then $(x_1, \dots, x_{2k-1})$ has the desired properties. $\square$

Walks in dense hypergraphs

For later use we now state a lower bound on the number of walks of given length in a given dense hypergraph, that is somewhat related to Sidorenko's conjecture [104, 106]. It is well known that this conjecture holds for paths in graphs, i.e., that for $d \in [0,1]$ and $\ell \in \mathbb{N}$ every graph $G = (V, E)$ satisfying $|E| \geqslant d|V|^2/2$ contains at least $d^\ell |V|^{\ell+1}$ walks of length ℓ (see [7] for a proof based on linear algebra and [2, Lemma 3.8] for a different approach using vertex deletions and the tensor power trick). The latter argument generalises in a straightforward manner to partite hypergraphs (see Lemma 2.2.8 below). An alternative proof based on the entropy method was worked out by Fitch [39, Lemma 7] and by Lee [78, Theorems 2.6 and 2.7].

Lemma 2.2.8. *Suppose $k \geqslant 2$, $d \in [0,1]$, and that H is a k-partite k-uniform hypergraph with vertex partition $\{V_i \colon i \in \mathbb{Z}/k\mathbb{Z}\}$. If H has $d \prod_{i \in \mathbb{Z}/k\mathbb{Z}} |V_i|$ edges, then for every $r \geqslant k$ there are at least*

$$d^{r-k+1} \prod_{i \in [r]} |V_i|$$

walks $(x_1, \ldots, x_r)$ in H with $x_1 \in V_1, \ldots, x_k \in V_k$.

Due to a request by the referee, we provide a brief sketch of an argument following the ideas in [2, Lemma 3.8].

Proof of Lemma 2.2.8. The first step is to establish that there are at least $\gamma d^{r-k+1} \prod_{i \in [r]} |V_i|$ such walks, where γ denotes some constant depending only on k and r but not on d and H. To this end we iteratively delete all edges from H containing a $(k-1)$-set of vertices whose degree is small. In each step of the process we ask whether for some $i \in \mathbb{Z}/k\mathbb{Z}$ there is a set $S \in V(H)^{(k-1)}$ with $|S \cap V_j| = 1$ for all $j \neq i$ such that $0 < d(S) < \frac{d}{2k}|V_i|$. If so we delete all edges containing S and continue. At the end of this process we obtain a spanning subhypergraph $H' \subseteq H$ which still has at least $\frac{d}{2} \prod_{i \in \mathbb{Z}/k\mathbb{Z}} |V_i|$ edges. A simple counting argument discloses that H' and, hence, H contains at least $\gamma d^{r-k+1} \prod_{i \in [r]} |V_i|$ walks of the required kind, where $\gamma = \frac{1}{2}\left(\frac{1}{2k}\right)^{r-k}$.

Second, for every $m \in \mathbb{N}$ we consider the k-partite k-uniform hypergraph $H^{\otimes m}$ with vertex classes $V_1^m, \ldots, V_k^m$ and $\{\vec{x}_1, \ldots, \vec{x}_k\} \in E(H^{\otimes m})$ for $\vec{x}_i = (x_{i1}, \ldots, x_{im})$ if and only if $\{x_{1\mu}, \ldots, x_{k\mu}\} \in E(H)$ holds for all $\mu \in [m]$. Clearly $e(H^{\otimes m}) = e(H)^m$ and if Ω denotes the number of walks we are to bound from below, then $H^{\otimes m}$ contains exactly Ω^m walks

$(\vec{x}_1, \dots, \vec{x}_r)$ with $\vec{x}_1 \in V_1^m, \dots, \vec{x}_k \in V_k^m$. Thus the result from the first step yields

$$\Omega^m \geqslant \gamma d^{(r-k+1)m} \prod_{i \in [r]} |V_i|^m$$

or, in other words, $\Omega \geqslant \gamma^{1/m} d^{r-k+1} \prod_{i \in [r]} |V_i|$. In the limit $m \to \infty$ we obtain the desired conclusion. $\square$

By identifying the vertex classes one obtains the following, more standard, non-partite version of this lemma.

Corollary 2.2.9. *For $k \geqslant 2$ and $d \in [0, 1]$ let $H = (V, E)$ be a k-uniform hypergraph. If $|E| \geqslant d|V|^k/k!$, then for every integer $r \geqslant k$ there are at least $d^{r-k+1}|V|^r$ walks $(x_1, \dots, x_r)$ in H.* $\square$

2.2.3 Abstract connectability

Our intended way of using Proposition 2.2.2 is that given a k-uniform hypergraph H satisfying $\delta_{k-2}(H) \geqslant \left(\frac{5}{9} + \alpha\right)|V(H)|^2/2$ we can choose robust subgraphs of all the $\binom{|V(H)|}{k-2}$ link graphs. It will be convenient to collect the data thus arising into a single structure.

Definition 2.2.10. *For $k \geqslant 2$ a k-uniform constellation is a pair*

$$\Psi = \left(H, \left\{R_x \colon x \in V(H)^{(k-2)}\right\}\right)$$

consisting of a k-uniform hypergraph H and a family of induced subgraphs $R_x \subseteq H_x$ of the 2-uniform link hypergraphs that can be formed in H. We write $H(\Psi) = H$ for the underlying hypergraph of a constellation Ψ and use the abbreviations $V(\Psi) = V(H)$, $E(\Psi) = E(H)$ for its vertex set and edge set, respectively. For a constellation Ψ and $x \in V(\Psi)^{(k-2)}$ we denote the subgraph associated with x by $R_x^\Psi = R_x$.

Example 2.2.11. A 2-uniform constellation is determined by its underlying graph H and a distinguished induced subgraph $R_\varnothing \subseteq H_\varnothing = H$.

Notice that so far the induced subgraphs $R_x \subseteq H_x$ are completely arbitrary and at this moment there are no restrictions on their orders, sizes, and connectivity properties. This allows us to study constellations "axiomatically", adding further useful conditions in each of the following subsections. The central connectability notions are definable without any such assumptions and they will be introduced in the present subsection (see Definition 2.2.14 below). Of course one cannot prove a meaningful Connecting Lemma at this level of generality, so our way of organising the material may appear somewhat peculiar on first sight. When establishing the covering lemma in Section 2.6 however, we need to analyse connectability in random subconstellations and for such situations the abstract approach developed here turns out to be advantageous. Subconstellations themselves are defined in the expected way.

Definition 2.2.12. *Let*

$$\Psi = \big(H, \big\{R_x \colon x \in V(H)^{(k-2)}\big\}\big)$$

be a k-uniform constellation, where $k \geqslant 2$. For $X \subseteq V(\Psi)$ we call

$$\Psi[X] = \big(H[X], \big\{R_x[X] \colon x \in X^{(k-2)}\big\}\big)$$

the subconstellation of Ψ induced by X. Moreover, $\Psi - X = \Psi[V(\Psi) \smallsetminus X]$ denotes the constellation obtained from Ψ by removing X.

We can also form link constellations in the obvious way.

Definition 2.2.13. *Let $k \geqslant 2$ and let*

$$\Psi = \big(H, \big\{R_x \colon x \in V(H)^{(k-2)}\big\}\big)$$

be a k-uniform constellation. If $S \subseteq V(\Psi)$ and $|S| \leqslant k-2$, then the $(k-|S|)$-uniform link constellation Ψ_S is defined to be

$$\Psi_S = \big(\overline{H}_S, \big\{R_{x \cup S} - S \colon x \in (V(H) \smallsetminus S)^{(k-2-|S|)}\big\}\big)\,.$$

If $x, y \in V(\Psi)$ are distinct, then in accordance with our convention to omit unnecessary braces and commas, we shall often write Ψ_x and Ψ_{xy} for the link constellations $\Psi_{\{x\}}$ and $\Psi_{\{x,y\}}$, respectively.

Next we tell which $(k-1)$-tuples of vertices of a k-uniform constellation are regarded as being ζ-leftconnectable for a given real number $\zeta > 0$. The definition progresses by recursion on k.

Definition 2.2.14. *Let $k \geqslant 2$, $\zeta > 0$, let*

$$\Psi = \big(H, \big\{R_x \colon x \in V(H)^{(k-2)}\big\}\big)$$

be a k-uniform constellation, and let $\vec{x} = (x_1, \dots, x_{k-1}) \in V(\Psi)^{k-1}$ be a $(k-1)$-tuple of distinct vertices.

(*a*) *If* $k = 2$ *we say that* $\vec{x} = (x_1)$ *is* ζ*-leftconnectable in* Ψ *if* $x_1 \in V(R_\varnothing)$*.*

(*b*) *If* $k \geqslant 3$ *we say that* $\vec{x}$ *is* ζ*-leftconnectable in* Ψ *if*

$$|U^{\Psi}_{\vec{x}}(\zeta)| \geqslant \zeta |V(\Psi)|\,,$$

where

$$U^{\Psi}_{\vec{x}}(\zeta) = \big\{ z \in V(\Psi) \colon x_1 \cdots x_{k-1} z \in E(\Psi) \text{ and } \\ (x_2, \dots, x_{k-1}) \text{ is } \zeta\text{-leftconnectable in } \Psi_z \big\}\,.$$

We remark that this is a "new" concept in the sense that in the earlier articles [92, 95] we managed to work with symmetric notions of connectability. For this reason, we need to be careful when quoting the Connecting Lemma from [95] later.

Example 2.2.15. Let (x_1, x_2) be a pair of distinct vertices from a 3-uniform constellation Ψ and let $\zeta > 0$. According to part (*b*) of Definition 2.2.14 the pair (x_1, x_2) is ζ-leftconnectable in Ψ if and only if $|U^{\Psi}_{(x_1,x_2)}(\zeta)| \geqslant \zeta |V(\Psi)|$. Due to part (*a*) the definition of this set unravels to

$$U^{\Psi}_{(x_1,x_2)}(\zeta) = \big\{ z \in V(\Psi) \colon x_1 x_2 z \in E(\Psi) \text{ and } x_2 \in V(R^{\Psi}_z) \big\}\,.$$

There is a dual notion of rightconnectability obtained by reversing the ordering of the vertices.

Definition 2.2.16. *Let* $k \geqslant 2$*,* $\zeta > 0$*,* Ψ*, and* $\vec{x} \in V(\Psi)^{k-1}$ *be as in Definition 2.2.14.*

(*a*) *If the reverse tuple* $(x_{k-1}, \dots, x_1)$ *is* ζ*-leftconnectable, then* $\vec{x}$ *itself is said to be* ζ*-rightconnectable.*

(*b*) *Further,* $\vec{x}$ *is called* ζ*-connectable if it is* ζ*-leftconnectable and* ζ*-rightconnectable.*

Some readers may react negatively to our choice of the specifiers 'left' and 'right' in these notions, arguing that the definition of leftconnectability of $\vec{x}$ pivots on the right end-segment of $\vec{x}$. The reason for our terminological choice is that the Connecting Lemma (Proposition 2.3.1 below) will assert that under reasonable assumptions every leftconnectable tuple can be connected to every rightconnectable tuple in such a way that the leftconnectable tuple is 'on the left side' in the resulting path, while the rightconnectable tuple is 'on the right side'.

The following observation follows by a straightforward induction from Definition 2.2.14. In later sections we will often use it either tacitly or by referring to 'monotonicity'.

Fact 2.2.17. *For a k-uniform constellation Ψ and $\zeta > \zeta' > 0$ every ζ-leftconnectable $(k-1)$-tuple is also ζ'-leftconnectable. Similarly statements hold for rightconnectability and connectability.*

Proof. It suffices to display the argument for leftconnectability. We argue by induction on k. In the base case $k = 2$ the definition of ζ-leftconnectability does not depend on ζ and there is nothing to prove. Now let $k \geqslant 3$ and suppose that the assertion is true for $k-1$ playing the rôle of k.

Let $\zeta > \zeta' > 0$, let $\Psi = \big(H, \{R_x : x \in V(H)^{(k-2)}\}\big)$ be a k-uniform constellation, and let $\vec{x} = (x_1, \ldots, x_{k-1}) \in V(\Psi)^{k-1}$ be a ζ-leftconnectable $(k-1)$-tuple. We are to prove that $\vec{x}$ is ζ'-leftconnectable as well. To this end we consider the sets

$$U = \{z \in V(\Psi) \colon x_1 \cdots x_{k-1} z \in E(\Psi) \text{ and } (x_2, \ldots, x_{k-1}) \text{ is } \zeta\text{-leftconnectable in } \Psi_z\}$$

and

$$W = \{z \in V(\Psi) \colon \; x_1 \cdots x_{k-1} z \in E(\Psi) \text{ and } (x_2, \ldots, x_{k-1}) \text{ is } \zeta'\text{-leftconnectable in } \Psi_z\}\,.$$

The induction hypothesis discloses $U \subseteq W$ and the assumption that $\vec{x}$ is ζ-leftconnectable means that $|U| \geqslant \zeta |V(\Psi)|$. So altogether we have

$$|W| \geqslant |U| \geqslant \zeta |V(\Psi)| \geqslant \zeta' |V(\Psi)|\,,$$

for which reason $\vec{x}$ is indeed ζ'-leftconnectable. □

Next, we study connectability in subconstellations.

Fact 2.2.18. *Suppose that Ψ is a k-uniform constellation, that $\Psi' = \Psi[X]$ is a subconstellation induced by some $X \subseteq V(\Psi)$ with $|X| \geqslant \frac{1}{2}\big(|V(\Psi)| + k - 2\big)$. If $\vec{x} \in V(\Psi')^{k-1}$ is (2ζ)-leftconnectable in Ψ', then it is ζ-leftconnectable in Ψ as well. Similar statements hold for 'rightconnectability' and 'connectability'.*

Proof. Again we only display the argument for leftconnectability and proceed by induction on k. The base case $k = 2$ is trivial. For the induction step from $k-1$ to k we recall that the assumption entails $|U| \geqslant 2\zeta |V(\Psi')| \geqslant \zeta |V(\Psi)|$, where

$$U = \big\{z \in V(\Psi') \colon x_1 \cdots x_{k-1} z \in E(\Psi') \text{ and } (x_2, \ldots, x_{k-1}) \text{ is } (2\zeta)\text{-leftconnectable in } \Psi'_z\big\}\,.$$

Now consider an arbitrary vertex $z \in U$. Since

$$|V(\Psi'_z)| = |V(\Psi')| - 1 \geqslant \frac{1}{2}\big(|V(\Psi)| + k - 4\big) = \frac{1}{2}\big(|V(\Psi_z)| + k - 3\big)\,,$$

the induction hypothesis is applicable to the constellation Ψ_z, its subconstellation Ψ'_z, and to the (2ζ)-leftconnectable $(k-2)$-tuple $(x_2, \ldots, x_{k-1})$. It follows that

$$U \subseteq \big\{z \in V(\Psi)\colon x_1 \cdots x_{k-1} z \in E(\Psi) \text{ and } (x_2, \ldots, x_{k-1}) \text{ is } \zeta\text{-leftconnectable in } \Psi_z\big\}$$

and together with $|U| \geqslant \zeta |V(\Psi)|$ this shows that $\vec{x}$ is indeed ζ-leftconnectable in Ψ. □

We shall frequently have the situation that for some edge $x_1 \cdots x_k$ of a k-uniform constellation Ψ we know $x_k \in V(R^{\Psi}_{x_1 \cdots x_{k-2}})$ and we would like to conclude from this state of affairs that $(x_2, \ldots, x_k)$ is ζ-leftconnectable in Ψ. While such deductions are invalid in general, it turns out that for small values of ζ there are only few exceptions to this rule of inference. More precisely, we have the following result (cf. [95, Fact 4.1] and [92, Lemma 3.7] for similar statements).

Lemma 2.2.19. *Let $k \geqslant 2$ and $\zeta > 0$ be given. If Ψ is a k-uniform constellation, then there exist at most $(k-2)\zeta |V(\Psi)|^k$ k-tuples $(x_1, \ldots, x_k) \in V(\Psi)^k$ such that*

(a) $\{x_1, \ldots, x_k\} \in E(\Psi)$,

(b) $x_k \in V(R^{\Psi}_{x_1 \cdots x_{k-2}})$,

(c) and $(x_2, \ldots, x_k)$ fails to be ζ-leftconnectable in Ψ.

Proof. We argue by induction on k. In the base case $k = 2$ the demands *(b)* and *(c)* contradict each other and, hence, there are indeed no such pairs. Now let $k \geqslant 3$ and suppose that the lemma is true for $k-1$ in place of k. Define $A \subseteq V(\Psi)^k$ to be the set of all k-tuples satisfying *(a)*–*(c)*, set

$$A' = \big\{(x_1, \ldots, x_k) \in A\colon x_1 \in U^{\Psi}_{(x_2, \ldots, x_k)}(\zeta)\big\}$$

and define

$$A''_x = \big\{(x_2, \ldots, x_k) \in V(\Psi)^{k-1}\colon (x, x_2, \ldots, x_k) \in A \smallsetminus A'\big\}$$

for every $x \in V(\Psi)$. Since

$$|A| = |A'| + \sum_{x \in V(\Psi)} |A''_x|\,,$$

it suffices to show

(*1*) $|A'| \leqslant \zeta |V(\Psi)|^k$

(*2*) and $|A''_x| \leqslant (k-3)\zeta |V(\Psi_x)|^{k-1}$ for every $x \in V(\Psi)$.

Now (*1*) follows from the fact that for $(x_1, \ldots, x_k) \in A' \subseteq A$ we have

$$\left|U^{\Psi}_{(x_2,\ldots,x_k)}(\zeta)\right| < \zeta |V(\Psi)|$$

by (*c*) and the definition of ζ-leftconnectability. For the proof of (*2*) we apply the induction hypothesis to the link constellation Ψ_x. Notice that if $(x_2, \ldots, x_k) \in A''_x$, then

- $\{x_2, \ldots, x_k\} \in E(\Psi_x)$
- and $x_k \in V(R^{\Psi_x}_{x_2 \cdots x_{k-2}})$

follow from (*a*), (*b*), and the definition of Ψ_x. Moreover $(x, x_2, \ldots, x_k) \in A \smallsetminus A'$ yields $x \notin U^{\Psi}_{(x_2,\ldots,x_k)}(\zeta)$, which together with $\{x, x_2, \ldots, x_k\} \in E(\Psi)$ reveals that

$$(x_3, \ldots, x_k) \text{ fails to be } \zeta\text{-leftconnectable in } \Psi_x\,.$$

So altogether the induction hypothesis leads to (*2*) and the induction step is complete. □

We proceed with a similar statement that will ultimately assist us in the construction of the absorbing path.

Lemma 2.2.20. *For $k \geqslant 2$, $\zeta > 0$, and a k-uniform constellation Ψ, there are at most $(k-2)\zeta|V(\Psi)|^{2k-3}$ walks $x_1 \cdots x_{2k-3}$ in $H(\Psi)$ such that*

(*a*) $x_{k-1} \in V(R^{\Psi}_{x_k \cdots x_{2k-3}})$

(*b*) *but $(x_1, \ldots, x_{k-1})$ fails to be ζ-leftconnectable.*

Proof. Again we argue by induction on k. In the base case $k = 2$ condition (*a*) reads $x_1 \in V(R^{\Psi}_{\varnothing})$, whereas (*b*) demands that (x_1) fails to be ζ-leftconnectable in Ψ. As these requirements contradict each other, there are indeed no 1-vertex walks with the required properties.

Now let $k \geqslant 3$ and assume that the lemma is true for $k-1$ instead of k. Let $A \subseteq V(\Psi)^{2k-3}$ be the set of all walks $x_1 \cdots x_{2k-3}$ satisfying (*a*) and (*b*), set

$$A' = \left\{(x_1, \ldots, x_{2k-3}) \in A \colon x_k \in U^{\Psi}_{(x_1,\ldots,x_{k-1})}(\zeta)\right\}$$

and put

$$A''_{x,y} = \big\{(x_2, \dots, x_{k-1}, x_{k+1}, \dots, x_{2k-3}) \in V(\Psi)^{2k-5} \colon \\ (x, x_2, \dots, x_{k-1}, y, x_{k+1}, \dots, x_{2k-3}) \in A \smallsetminus A'\big\}$$

for all $x, y \in V(\Psi)$. In view of

$$|A| = |A'| + \sum_{(x,y) \in V(\Psi)^2} |A''_{x,y}|$$

it suffices to prove

(1) $|A'| \leqslant \zeta |V(\Psi)|^{2k-3}$

(2) and $|A''_{x,y}| \leqslant (k-3)\zeta |V(\Psi_y)|^{2k-5}$ for all $x, y \in V(\Psi)$.

The estimate *(1)* follows from the fact that due to *(b)* every $(x_1, \dots, x_{2k-3}) \in A' \subseteq A$ has the property $\big|U^{\Psi}_{(x_1,\dots,x_{k-1})}(\zeta)\big| < \zeta |V(\Psi)|$. For the proof of *(2)* we intend to apply the induction hypothesis to Ψ_y. Consider any $(2k-5)$-tuple

$$\vec{x} = (x_2, \dots, x_{k-1}, x_{k+1}, \dots, x_{2k-3}) \in A''_{x,y}\,.$$

Since $(x_2, \dots, x_{k-1}, y, x_{k+1}, \dots, x_{2k-3})$ is a walk in $H(\Psi)$, we know that $\vec{x}$ itself is a walk in $H(\Psi_y)$. Moreover, *(a)* rewrites as

$$x_{k-1} \in V(R^{\Psi_y}_{x_{k+1} \cdots x_{2k-3}})\,.$$

Finally, $y \notin U^{\Psi}_{(x, x_2, \dots, x_{k-1})}(\zeta)$ and $\{x, x_2, \dots, x_{k-1}, y\} \in E(\Psi)$ imply that

$$(x_2, \dots, x_{k-1}) \text{ fails to be } \zeta\text{-leftconnectable in } \Psi_y\,.$$

Altogether, the $(2k-5)$-tuples in $A''_{x,y}$ have the required properties for applying the induction hypothesis to Ψ_y. This proves *(2)* and completes the induction step. $\square$

We conclude this subsection by introducing one further notion.

Definition 2.2.21. *Given $k \geqslant 2$, $\zeta > 0$, and a k-uniform constellation*

$$\Psi = \big(H, \big\{R_x \colon x \in V(H)^{(k-2)}\big\}\big)\,,$$

a k-tuple $(x_1, \dots, x_k) \in V(\Psi)^k$ is said to be a ζ-bridge in Ψ if *(a)* $\{x_1, \ldots, x_k\} \in E(\Psi)$,

(b) $(x_1, \ldots, x_{k-1})$ *is ζ-rightconnectable,*

(c) *and* $(x_2, \ldots, x_k)$ *is ζ-leftconnectable.*

Such bridges will help us later to construct connecting paths between given $(k-1)$-tuples of vertices. The fundamental existence result for such bridges (see Corollary 2.2.28 below) asserts, roughly speaking, that under natural assumptions k-uniform constellations contain many ζ-bridges for sufficiently small values of ζ.

2.2.4 On (α, μ)-constellations

In this subsection we study some properties of constellations that can be deduced from the order and size restrictions *(i)* and *(ii)* in Proposition 2.2.2 alone without taking the (β, ℓ)-robustness into account. We are thus led to the following concept.

Definition 2.2.22. *Let $k \geqslant 2$ and $\alpha, \mu > 0$. A k-uniform constellation Ψ is said to be an (α, μ)-constellation if*

$$\delta_{k-2}\big(H(\Psi)\big) \geqslant \Big(\frac{5}{9} + \alpha\Big)\frac{|V(\Psi)|^2}{2}$$

and every $x \in V(\Psi)^{(k-2)}$ satisfies

(a) $|V(R_x^\Psi)| \geqslant \big(\frac{2}{3} + \frac{\alpha}{2}\big)|V(\Psi)|$

(b) *as well as* $e_{H(\Psi_x)}\big(V(R_x^\Psi), V(\Psi) \smallsetminus V(R_x^\Psi)\big) \leqslant \mu|V(\Psi)|^2$.

It turns out that the level of generality provided by this concept is fully appropriate for discussing the key parts of our absorbing mechanism and for constructing an important building block entering the proof of the Connecting Lemma. Before reaching those results we record a couple of easier observations.

Fact 2.2.23. *If Ψ denotes a k-uniform $(\alpha, \frac{\alpha}{9})$-constellation for some $\alpha > 0$, then*

$$e\big(H(\Psi_x)\big) - e(R_x^\Psi) \leqslant \frac{|V(\Psi)|^2}{18}$$

holds for every $x \in V(\Psi)^{(k-2)}$.

Proof. Using both parts of Definition 2.2.22 we obtain

$$e\big(H(\Psi_x)\big) - e(R_x^\Psi) = e_{H(\Psi_x)}\big(V(\Psi) \smallsetminus V(R_x^\Psi)\big) + e_{H(\Psi_x)}\big(V(R_x^\Psi), V(\Psi) \smallsetminus V(R_x^\Psi)\big)$$

$$\leqslant \Big(\frac{1}{3}-\frac{\alpha}{2}\Big)^2\frac{|V(\Psi)|^2}{2}+\frac{\alpha}{9}|V(\Psi)|^2=\Big(\frac{1}{18}+\frac{\alpha^2}{8}-\frac{\alpha}{18}\Big)|V(\Psi)|^2$$

and it remains to observe that the minimum $(k-2)$-degree condition imposed on $H(\Psi)$ is only satisfiable for $\alpha \leqslant \frac{4}{9}$. □

Fact 2.2.24. *Suppose that Ψ is a k-uniform (α,μ)-constellation. If $x \in V(\Psi)^{(k-2)}$ is arbitrary, then there are at most $\frac{2\mu}{\alpha}|V(\Psi)|$ vertices $z \in V(\Psi) \smallsetminus V(R^{\Psi}_x)$ with $d_{H(\Psi_x)}(z) \geqslant \frac{1}{3}|V(\Psi)|-1$.*

Proof. Definition 2.2.22 *(a)* tells us that $|V(\Psi) \smallsetminus V(R^{\Psi}_x)| \leqslant (\frac{1}{3}-\frac{\alpha}{2})|V(\Psi)|$. Consequently, the number of edges that every vertex z from the set

$$Z=\big\{z \in V(\Psi) \smallsetminus V(R^{\Psi}_x)\colon d_{H(\Psi_x)}(z) \geqslant \tfrac{1}{3}|V(\Psi)|-1\big\}$$

sends to $V(R^{\Psi}_x)$ is at least

$$\begin{aligned} d_{H(\Psi_x)}(z)-\big|V(\Psi) \smallsetminus \big(V(R^{\Psi}_x) \cup \{z\}\big)\big| &\geqslant \frac{1}{3}|V(\Psi)|-1-\Big(\frac{1}{3}-\frac{\alpha}{2}\Big)|V(\Psi)|+1 \\ &= \frac{\alpha}{2}|V(\Psi)|\,. \end{aligned}$$

In combination with Definition 2.2.22 *(b)* this yields

$$\frac{\alpha}{2}|V(\Psi)||Z| \leqslant e_{H(\Psi_x)}\big(V(R^{\Psi}_x), V(\Psi) \smallsetminus V(R^{\Psi}_x)\big) \leqslant \mu|V(\Psi)|^2$$

and the upper bound $|Z| \leqslant \frac{2\mu}{\alpha}|V(\Psi)|$ we are aiming for follows. □

Next, there is an obvious monotonicity statement.

Fact 2.2.25. *For $k \geqslant 2$, $\alpha \geqslant \alpha' > 0$, and $\mu' \geqslant \mu > 0$, every k-uniform (α,μ)-constellation is an (α',μ')-constellation as well.* □

Link constellations 'almost' inherit being (α,μ)-constellations, but since we are slightly shrinking the vertex set we need to be careful with clause *(b)* of Definition 2.2.22.

Fact 2.2.26. *Given $k \geqslant 2$, $\alpha > 0$, and $\mu' > \mu > 0$ there exists a natural number n_0 with the following property. If Ψ denotes a k-uniform (α,μ)-constellation having at least n_0 vertices and $S \subseteq V(\Psi)$ with $|S| \leqslant k-2$ is arbitrary, then Ψ_S is a $(k-|S|)$-uniform (α,μ')-constellation.* □

Now we estimate the number of walks of any short length in Ψ, whose starting $(k-1)$-tuple is rightconnectable and whose ending $(k-1)$-tuple is leftconnectable. Later we will use these walks in the proof of the Connecting Lemma thus gaining control over the length of the connections modulo k.

Lemma 2.2.27. *For $k \geqslant 2$ and $\alpha > 0$ let Ψ be a k-uniform $(\alpha, \frac{\alpha}{9})$-constellation. Provided that $|V(\Psi)| \geqslant \frac{k^2}{\alpha}$, there are for every positive integer r at least $\frac{1}{3^{r+1}}|V(\Psi)|^{r+k-1}$ walks $x_1x_2\cdots x_{r+k-1}$ of length r in $H(\Psi)$ starting with a $\frac{1}{k3^{r+1}}$-rightconnectable $(k-1)$-tuple $(x_1, \ldots, x_{k-1})$ and ending with a $\frac{1}{k3^{r+1}}$-leftconnectable $(k-1)$-tuple $(x_{r+1}, \ldots, x_{r+k-1})$.*

Proof. Consider the auxiliary k-partite k-uniform hypergraph $\mathscr{A}$ whose vertex classes $V_1, \ldots, V_k$ are copies of $V(\Psi)$ and whose edges $\{x_1, \ldots, x_k\} \in E(\mathscr{A})$ with

$$x_1 \in V_1, \ldots, x_k \in V_k$$

signify that

(1) $\{x_1, \ldots, x_k\} \in E(\Psi)$,

(2) $x_1x_2 \in E(R^{\Psi}_{x_3\cdots x_k})$,

(3) and $x_{r+k-2}x_{r+k-1} \in E(R^{\Psi}_{x_r\cdots x_{r+k-3}})$,

where the indices in *(3)* are to be read modulo k.

In view of $|V(\Psi)| \geqslant \frac{k^2}{\alpha}$ and $\delta_{k-2}(H(\Psi)) \geqslant \left(\frac{5}{9}+\alpha\right)\frac{|V(\Psi)|^2}{2}$ there are at least

$$\begin{aligned}(|V(\Psi)|-k)^{k-2}\cdot\left(\frac{5}{9}+\alpha\right)|V(\Psi)|^2 &\geqslant \left(\frac{5}{9}+\alpha\right)\left(1-\frac{k}{|V(\Psi)|}\right)^k|V(\Psi)|^k\\ &\geqslant \left(\frac{5}{9}+\alpha\right)\left(1-\frac{k^2}{|V(\Psi)|}\right)|V(\Psi)|^k\\ &\geqslant \left(\frac{5}{9}+\alpha\right)(1-\alpha)|V(\Psi)|^k\\ &\geqslant \frac{5}{9}|V(\Psi)|^k\end{aligned}$$

possibilities $(x_1, \ldots, x_k) \in V_1 \times \cdots \times V_k$ satisfying *(1)*, where in the last estimate we used $\alpha \geqslant 4/9$. Among them, there are by Fact 2.2.23 at most $\frac{1}{9}|V(\Psi)|^k$ violating *(2)* and at most the same number violating *(3)*. Consequently, $e(\mathscr{A}) \geqslant \frac{1}{3}|V(\Psi)|^k$ and Lemma 2.2.8 applied to $\mathscr{A}$ and $d = \frac{1}{3}$ shows that there are at least $\frac{1}{3^r}|V(\Psi)|^{r+k-1}$ walks

$$x_1x_2\cdots x_{r+k-1}$$

of length r in $\mathscr{A}$ with $x_1 \in V_1, \ldots, x_k \in V_k$. Among them, there are by *(2)* and Lemma 2.2.19 applied to $\zeta = \frac{1}{k3^{r+1}}$ at most

$$\frac{k-2}{k3^{r+1}}|V(\Psi)|^{r+k-1} < \frac{1}{3^{r+1}}|V(\Psi)|^{r+k-1}$$

walks for which $(x_1, \ldots, x_{k-1})$ fails to be $\frac{1}{k3^{r+1}}$-rightconnectable. Similarly *(3)* and Lemma 2.2.19 ensure that at most $\frac{1}{3^{r+1}}|V(\Psi)|^{r+k-1}$ of our walks have the defect that $(x_{r+1}, \ldots, x_{r+k-1})$ fails to be $\frac{1}{k3^{r+1}}$-leftconnectable. This leaves us with at least

$$\left(\frac{1}{3^r} - \frac{2}{3^{r+1}}\right)|V(\Psi)|^{r+k-1} = \frac{|V(\Psi)|^{r+k-1}}{3^{r+1}}$$

walks of the desired form. □

Corollary 2.2.28. *Given $k \geqslant 2$ and $\alpha > 0$ let Ψ be a k-uniform $(\alpha, \frac{\alpha}{9})$-constellation. If Ψ has at least $\frac{k^2}{\alpha}$ vertices, then the number of its $\frac{1}{9k}$-bridges is at least $\frac{1}{9}|V(\Psi)|^k$.*

Proof. Plug $r = 1$ into Lemma 2.2.27. □

The following lemma builds a device that will assist us in the inductive proof of the Connecting Lemma in the next section.

Lemma 2.2.29. *Given $k \geqslant 4$, $\alpha > 0$, and $\zeta \in \big(0, \frac{1}{3^{k+2}}\big]$, there exists an integer n_0 such that the following holds for every k-uniform $(\alpha, \frac{\alpha}{10})$-constellation Ψ on $n \geqslant n_0$ vertices.*

If two subsets $U, W \subseteq V(\Psi)$ satisfy $|U|, |W| \geqslant \zeta n$, then there are at least $\zeta^3 n^{2k-2}$ $(2k-2)$-tuples $(u, q_1, \ldots, q_{2k-4}, w) \in V(\Psi)^{2k-2}$ such that

(i) *$u \in U$ and $w \in W$ are distinct,*

(ii) *$q_1 \cdots q_{2k-4}$ is a walk in $H(\Psi_{uw})$,*

(iii) *$(q_1, \ldots, q_{k-2})$ is ζ^3-rightconnectable in Ψ_u,*

(iv) *and $(q_{k-1}, \ldots, q_{2k-4})$ is ζ^3-leftconnectable in Ψ_w.*

Proof. Assuming that n_0 has been chosen sufficiently large for the subsequent arguments, we commence by considering the $(2k-2)$-tuples $(u, q_1, \ldots, q_{2k-4}, w) \in V(\Psi)^{2k-2}$ satisfying *(i)*, *(ii)* as well as the conditions

(v) $(q_1, \ldots, q_{k-3})$ is ζ^3-rightconnectable in Ψ_{uw},

(vi) $(q_k, \ldots, q_{2k-4})$ is ζ^3-leftconnectable in Ψ_{uw}.

First of all, by $|U|, |W| \geqslant \zeta n$ and $n \geqslant n_0 \geqslant 2/\zeta$ there are at least $\frac{1}{2}\zeta^2 n^2$ pairs (u, w) in $U \times W$ with $u \neq w$. For each of them Fact 2.2.26 tells us that Ψ_{uw} is a $(k-2)$-uniform $(\alpha, \frac{\alpha}{9})$-constellation. Applying the case $r = k-1$ of Lemma 2.2.27 to this constellation (with $k-2$ here in place of k there) we learn that the number of $(2k-4)$-tuples $(q_1, \ldots, q_{2k-4}) \in V(\Psi_{uw})^{2k-4}$ obeying *(ii)*, *(v)*, and *(vi)* is at least

$$\frac{1}{3^k}(n-2)^{2k-4} \geqslant \frac{6}{3^{k+2}} n^{2k-4} \geqslant 6\zeta n^{2k-4} .$$

Here we used tacitly that the assumption $\zeta \leqslant \frac{1}{3^{k+2}}$ easily implies $\zeta^3 \leqslant \frac{1}{(k-2)3^k}$.

Summarising, the number of $(2k-2)$-tuples $(u, q_1, \ldots, q_{2k-4}, w)$ satisfying *(i)*, *(ii)*, *(v)*, and *(vi)* is at least $\frac{1}{2}\zeta^2 n^2 \cdot 6\zeta n^{2k-4} = 3\zeta^3 n^{2k-2}$. So it suffices to prove that among all $(2k-2)$-tuples $(u, q_1, \ldots, q_{2k-4}, w) \in V(\Psi)^{2k-2}$ there are

(1) at most $\zeta^3 n^{2k-2}$ with *(ii)*, *(v)*, $\neg$*(iii)*

(2) and at most $\zeta^3 n^{2k-2}$ with *(ii)*, *(vi)*, $\neg$*(iv)*.

For reasons of symmetry we only need to establish *(2)*. To this end it is enough to check that for fixed vertices $w, q_1, \ldots, q_{2k-4} \in V(\Psi)$ the number of vertices u such that

- $\{u, q_{k-1}, \ldots, q_{2k-4}\} \in E(\Psi_w)$,
- *(vi)*, but $\neg$*(iv)*.

is at most $\zeta^3 n$. Now by Definition 2.2.14, the first bullet, and *(vi)* these vertices satisfy $u \in U^{\Psi_w}_{(q_{k-1}, \ldots, q_{2k-4})}(\zeta^3)$ and by $\neg$*(iv)* the latter set has size at most $\zeta^3 |V(\Psi_w)|$. $\square$

The last lemma of this subsection will help us to exchange arbitrary vertices by 'absorbable' ones in Section 2.5. Roughly speaking it asserts that for $\mu \ll \alpha, k^{-1}$, with few exceptions, the links of two vertices in a k-uniform (α, μ)-constellation intersect in a substantial number of connectable $(k-1)$-tuples.

Lemma 2.2.30. *Given $k \geqslant 3$ and $\alpha > 0$ set $\mu = \frac{1}{10k}\left(\frac{\alpha}{2}\right)^{2^{k-3}+1}$. If Ψ denotes a k-uniform (α, μ)-constellation on n vertices and $\zeta > 0$ is arbitrary, then there is a set $X \subseteq V(\Psi)$ of size $|X| \leqslant \frac{\zeta}{\mu} n$ such that for every $a \in V(\Psi)$ and every $x \in V(\Psi) \smallsetminus (X \cup \{a\})$ the number of ζ-connectable $(k-1)$-tuples $(x_1, \ldots, x_{k-1})$ with $\{x_1, \ldots, x_{k-1}\} \in E(\Psi_a) \cap E(\Psi_x)$ is at least $\mu |V(\Psi)|^{k-1}$.*

Proof. Set

$$\eta = \frac{1}{10}\left(\frac{\alpha}{2}\right)^{2^{k-3}} \tag{2.2.8}$$

and $V = V(\Psi)$. Since $\mu = \frac{\alpha\eta}{2k}$, we have

$$\max\left\{\frac{2\mu}{\alpha},\, 2k\mu\right\} \leqslant \eta\,. \tag{2.2.9}$$

The choice of X. With every $x \in V$ we shall associate two exceptional sets, the idea being that on average these sets can be proved to be small. So there will only be few vertices for which one of the exceptional sets is very large and these 'unpleasant vertices' will form the set X. For every vertex not belonging to X, we will then be able to show that its link constellation intersect the link constellations of all other vertices in the desired way.

For an arbitrary $x \in V$ the first of the exceptional sets A_x consists of all $(k-1)$-tuples $(x_1, \ldots, x_{k-1}) \in V^{k-1}$ satisfying

- $\{x_1, \ldots, x_{k-1}, x\} \in E(\Psi)$
- and $x_1 \in V(R^{\Psi}_{x_3 \cdots x_{k-1} x})$
- that fail to be ζ-rightconnectable in Ψ.

We would like to point out that the second bullet does not involve the vertex x_2. Moreover, in the special case $k = 3$ the condition just means that $x_1 \in V(R^{\Psi}_x)$.

The second exceptional set B_x comprises all $(2k-4)$-tuples $(x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_{2k-3})$ in V^{2k-4} such that

- $x_1 \cdots x_{k-1} x x_{k+1} \cdots x_{2k-3}$ is a walk in $H(\Psi)$
- and $x_{k-1} \in V(R^{\Psi}_{x x_{k+1} \cdots x_{2k-3}})$,
- for which $(x_1, \ldots, x_{k-1})$ fails to be ζ-leftconnectable in Ψ.

Now we define

$$X' = \left\{x \in V \colon |A_x| > 2k\mu|V|^{k-1}\right\},$$
$$X'' = \left\{x \in V \colon |B_x| > 2k\mu|V|^{2k-4}\right\},$$

and set $X = X' \cup X''$. By Lemma 2.2.19 and double counting we have

$$2k\mu|X'||V|^{k-1} \leqslant \sum_{x \in X'} |A_x| \leqslant (k-2)\zeta|V|^k\,,$$

whence $|X'| \leqslant \frac{\zeta}{2\mu}|V|$. Similarly, Lemma 2.2.20 yields

$$2k\mu|X''||V|^{2k-4} \leqslant \sum_{x \in X''} |B_x| \leqslant (k-2)\zeta|V|^{2k-3},$$

which shows that $|X''| \leqslant \frac{\zeta}{2\mu}|V|$ holds as well. Altogether we arrive at the desired estimate

$$|X| \leqslant |X'| + |X''| \leqslant \frac{\zeta}{\mu}|V|.$$

For the rest of the proof we fix two distinct vertices $a, x \in V$ with $x \notin X$. We are to show that the number of ζ-connectable $(k-1)$-tuples $(x_1, \ldots, x_{k-1})$ such that

$$\{x_1, \ldots, x_{k-1}\} \in E(\Psi_a) \cap E(\Psi_x)$$

is at least $\mu|V|^{k-1}$. The smallest case $k = 3$ receives a separate treatment.

The special case $k = 3$. We know that both of the graphs $H(\Psi_a)$ and $H(\Psi_x)$ have at least $\left(\frac{5}{9} + \alpha\right)\frac{n^2}{2}$ edges and thus they have at least $\left(\frac{1}{9} + 2\alpha\right)\frac{n^2}{2}$ edges in common. Owing to Fact 2.2.23 this shows that $H(\Psi_a)$ and R_x^Ψ have at least αn^2 common edges or, in other words, that there are at least $2\alpha n^2$ ordered pairs (x_1, x_2) such that $x_1x_2 \in E(\Psi_a) \cap E(R_x^\Psi)$. Due to $x \notin X'$ we have $|A_x| \leqslant 6\mu n^2$ and thus at most $6\mu n^2$ of these pairs fail to be ζ-rightconnectable. By symmetry, at most the same number of pairs under consideration fails to be ζ-leftconnectable. Altogether, this demonstrates that among the ordered pairs (x_1, x_2) with $x_1x_2 \in E(\Psi_a) \cap E(\Psi_x)$ there are at least $(2\alpha - 12\mu)n^2$ which are ζ-connectable. Because of $\mu = \frac{\alpha^2}{120} < \frac{\alpha}{7}$ this is more than what we need.

The general case $k \geqslant 4$. Our first goal is to count ζ-leftconnectable $(k-1)$-tuples in the intersection of $H(\Psi_a)$ and $H(\Psi_x)$ that satisfy a certain minimum degree condition.

Claim 2.2.31. *The number of ζ-leftconnectable $(k-1)$-tuples $(x_1, \ldots, x_{k-1})$ such that*

(1) $\{x_1, \ldots, x_{k-1}\} \in E(\Psi_a) \cap E(\Psi_x)$

(2) and $d(x_2, \ldots, x_{k-1}, x) \geqslant \frac{n}{3}$

is at least $3\eta n^{k-1}$.

Proof. For every vertex $x_{k-1} \in V \smallsetminus \{a, x\}$ we apply Lemma 2.2.7 to the $(k-2)$-uniform hypergraphs $H(\Psi_{xx_{k-1}})$ and $H(\Psi_{ax_{k-1}})$. This yields a lower bound on the number of $(2k-4)$-tuples

$$(x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_{2k-3}) \in V^{2k-4}$$

such that

(*a*) $x_1 \cdots x_{k-1} x x_{k+1} \cdots x_{2k-3}$ is a walk in $H(\Psi)$

(*b*) $\{x_1, \ldots, x_{k-1}\} \in E(\Psi_a)$

(*c*) $d(x_2, \ldots, x_{k-1}, x) \geqslant \frac{n}{3}$

(*d*) and $d(x_{k-1}, x, x_{k+1}, \ldots, x_{2k-3}) \geqslant \frac{n}{3}$.

Notably, there are $n-2$ possibilities for x_{k-1} and for each of them Lemma 2.2.7 yields

$$\left(\frac{\alpha}{2}\right)^{2^{(k-2)-1}} n^{2(k-2)-1} \overset{(2.2.8)}{=} 10\eta n^{2k-5}$$

possibilities for the remaining $2k-5$ vertices. Therefore the number of $(2k-4)$-tuples

$$(x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_{2k-3}) \in V^{2k-4}$$

satisfying (*a*)–(*d*) is at least $10\eta(n-2)n^{2k-5}$.

Because of the minimum $(k-2)$-degree condition Ψ needs to have at least one edge, whence $n \geqslant k \geqslant 4$. As this implies $n - 2 \geqslant \frac{1}{2}n$, the total number of $(2k-4)$-tuples satisfying (*a*)–(*d*) is at least $5\eta n^{2k-4}$.

In view of (*d*) and Fact 2.2.24 applied to $\{x, x_{k+1}, \ldots, x_{2k-3}\}$ here in place of x there we know that all but at most $\frac{2\mu}{\alpha} n^{2k-4}$ of these $(2k-4)$-tuples satisfy

(*e*) $x_{k-1} \in V(R^{\Psi}_{x x_{k+1} \cdots x_{2k-3}})$.

Now $x \notin X''$ yields $|B_x| \leqslant 2k\mu n^{2k-4}$. So at most $2k\mu n^{2k-4}$ of the $(2k-4)$-tuples satisfying (*a*) and (*e*) violate

(*f*) $(x_1, \ldots, x_{k-1})$ is ζ-leftconnecctable.

Summarising, the number of $(2k-4)$-tuples satisfying (*a*)–(*f*) is at least

$$(5\eta - \tfrac{2\mu}{\alpha} - 2k\mu) n^{2k-4} \overset{(2.2.9)}{\geqslant} 3\eta n^{2k-4}\,.$$

Ignoring the vertices $x_{k+1}, \ldots, x_{2k-3}$ as well as the conditions (*d*), (*e*) we arrive at the desired conclusion. □

Now we keep working with the ζ-leftconnectable $(k-1)$-tuples satisfying (*1*) and (*2*) obtained in Claim 2.2.31. According to (*2*) and Fact 2.2.24 applied to $\{x_3, \ldots, x_{k-1}, x\}$ here in place of x there all but at most $\frac{2\mu}{\alpha} n^{k-1}$ of them have the property

(*3*) $x_2 \in V(R^{\Psi}_{x_3\cdots x_{k-1}x})$.

Moreover, by Definition 2.2.22 (*b*) applied to the $(k-2)$-set $\{x_3, \dots, x_{k-1}, x\}$ at most μn^{k-1} tuples of length $k-1$ satisfy (*1*) and (*3*) but not

(*4*) $x_1 \in V(R^{\Psi}_{x_3\cdots x_{k-1}x})$.

Finally, $x \notin X'$ implies $|A_x| \leqslant 2k\mu n^{k-1}$, so among the ζ-leftconnectable $(k-1)$-tuples satisfying (*1*)–(*4*) there are at most $2k\mu n^{k-1}$ for which

(*5*) $(x_1, \dots, x_{k-1})$ is ζ-rightconnectable

fails. In particular, the number of ζ-leftconnectable $(k-1)$-tuples $(x_1, \dots, x_{k-1})$ with (*1*) and (*5*) is at least

$$\left(3\eta - \frac{2\mu}{\alpha} - \mu - 2k\mu\right) n^{k-1} \overset{(2.2.9)}{\geqslant} \mu n^{k-1}\,.$$

Altogether this shows that the number of $(k-1)$-tuples $(x_1, \dots, x_{k-1})$ that are ζ-leftconnectable, ζ-rightconnectable, and satisfy $\{x_1, \dots, x_{k-1}\} \in E(\Psi_a) \cap E(\Psi_x)$ is at least μn^{k-1}. In view of Definition 2.2.16(*b*) this concludes the proof of Lemma 2.2.30. $\square$

The 'connectable' edges in $E(\Psi_a) \cap E(\Psi_x)$ considered in the previous lemma can be used to build paths.

Corollary 2.2.32. *For given $k \geqslant 3$ and $\alpha > 0$ there exists a natural number n_0 such that if $\mu = \frac{1}{10k}\left(\frac{\alpha}{2}\right)^{2^{k-3}+1}$, Ψ is a k-uniform (α, μ)-constellation on $n \geqslant n_0$ vertices, and $\zeta > 0$ then there exists a set $X \subseteq V(\Psi)$ with $|X| \leqslant \frac{\zeta}{\mu} n$ such that the following holds. For every $a \in V(\Psi)$ and every $x \in V(\Psi) \smallsetminus (X \cup \{a\})$ the number of $(k-1)$-uniform paths $b_1 b_2 \cdots b_{2k-2}$ in $H(\Psi_a) \cap H(\Psi_x)$ such that $(b_1, \dots, b_{k-1})$ and $(b_k, \dots, b_{2k-2})$ are ζ-connectable in Ψ is at least $\frac{1}{2}\mu^k n^{2k-2}$.*

Proof. Let X be the set produced by Lemma 2.2.30. Consider two distinct vertices $a, x \in V(\Psi)$ with $x \notin X$. Form an auxiliary $(k-1)$-partite $(k-1)$-uniform hypergraph

$$\mathscr{B} = (V_1 \,\dot\cup\, \dots \,\dot\cup\, V_{k-1}, E_{\mathscr{B}})$$

whose vertex classes are $k-1$ disjoint copies of $V(\Psi)$ and whose edges $\{v_1, \dots, v_{k-1}\} \in E_{\mathscr{B}}$ with $v_i \in V_i$ for $i \in [k-1]$ correspond to ζ-connectable $(k-1)$-tuples $(v_1, \dots, v_{k-1})$ such that $\{v_1, \dots, v_{k-1}\} \in E(\Psi_a) \cap E(\Psi_x)$.

Lemma 2.2.30 tells us that

$$|E_{\mathscr{B}}| \geqslant \mu n^{k-1}\,.$$

Thus Lemma 2.2.8 applied to $\mathscr{B}$ with $(k-1, \mu, 2k-2)$ here in place of (k, d, r) there yields at least $\mu^k n^{2k-2}$ walks $(b_1, \ldots, b_{2k-2})$ in $\mathscr{B}$ with $b_1 \in V_1, \ldots, b_{k-1} \in V_{k-1}$. By the definition of $\mathscr{B}$ each of these walks corresponds to a walk in $H(\Psi_a) \cap H(\Psi_x)$ whose first and last $k-1$ vertices form a ζ-connectable $(k-1)$-tuple in Ψ. At most $O(n^{2k-3})$ of these walks can have repeated vertices and, hence, there are at least

$$\mu^k n^{2k-2} - O(n^{2k-3}) \geqslant \frac{\mu^k}{2} n^{2k-2}$$

paths of the desired from. $\square$

2.2.5 On $(\alpha, \beta, \ell, \mu)$-constellations

This subsection is dedicated to (α, μ)-constellations Ψ whose distinguished graphs R_x^Ψ have the robustness property delivered by Proposition 2.2.2.

Definition 2.2.33. *Let $k \geqslant 2$, $\alpha, \beta, \mu > 0$ and let $\ell \geqslant 3$ be odd. A k-uniform constellation Ψ is said to be an $(\alpha, \beta, \ell, \mu)$-constellation if*

- *(a) it is an (α, μ)-constellation,*
- *(b) and for all $x \in V(\Psi)^{(k-2)}$ and all distinct $y, z \in V(R_x^\Psi)$ the number of y-z-paths in R_x^Ψ of length ℓ is at least $\beta |V(\Psi)|^{\ell-1}$.*

The main result of this subsection shows how to expand sufficiently large k-uniform hypergraphs whose minimum $(k-2)$-degree is at least $\left(\frac{5}{9} + \alpha\right)\frac{n^2}{2}$ for appropriate choices of the parameters to such $(\alpha, \beta, \ell, \mu)$-constellations. Essentially, the proof of this observation proceeds by applying Proposition 2.2.2 to all link graphs.

Fact 2.2.34. *For all $k \geqslant 2$ and $\alpha, \mu > 0$ there exist $\beta = \beta(\alpha, \mu) > 0$ and an odd integer $\ell = \ell(\alpha, \mu) \geqslant 3$ such that for sufficiently large n every k-uniform n-vertex hypergraph H with $\delta_{k-2}(H) \geqslant \left(\frac{5}{9} + \alpha\right)\frac{n^2}{2}$ expands to an $(\alpha, \beta, \ell, \mu)$-constellation.*

Notice that this result is the reason why the study of $(\alpha, \beta, \ell, \mu)$-constellations conducted in the subsequent sections sheds light on Theorem 2.1.2.

Proof of Fact 2.2.34. For α and μ Proposition 2.2.2 delivers some constant $\beta' > 0$ and an odd integer $\ell \geqslant 3$. We contend that $\beta = (2/3)^{\ell-1}\beta'$ and ℓ have the desired property.

To see this, we consider a sufficiently large k-uniform hypergraph H on n vertices satisfying $\delta_{k-2}(H) \geqslant \left(\frac{5}{9} + \alpha\right)\frac{n^2}{2}$. For every $x \in V(H)^{(k-2)}$ Proposition 2.2.2 applies to the link graph H_x and yields a (β', ℓ)-robust induced subgraph $R_x \subseteq H_x$ satisfying

(i) $|V(R_x)| \geqslant \big(\frac{2}{3} + \frac{\alpha}{2}\big)n$

(ii) and $e_{H_x}(V(R_x), V(H) \smallsetminus V(R_x)) \leqslant \mu n^2$.

We shall show that

$$\Psi = \big(H, \big\{R_x \colon x \in V(H)^{(k-2)}\big\}\big)$$

is the desired $(\alpha, \beta, \ell, \mu)$-constellation. By Definition 2.2.22 and (i), (ii) above, Ψ is an (α, μ)-constellation, meaning that part (a) of Definition 2.2.33 holds.

Moving on to the second part we fix an arbitrary $(k-2)$-set $x \subseteq V(H)$ as well as two distinct vertices y, z of R_x. Since R_x is (β', ℓ)-robust, the number of y-z-paths in R_x of length ℓ is indeed at least

$$\beta' |V(R_x)|^{\ell-1} \overset{(i)}{\geqslant} \left(\frac{3}{2}\right)^{\ell-1} \beta \cdot \left(\frac{2}{3} + \frac{\alpha}{2}\right)^{\ell-1} n^{\ell-1} \geqslant \beta n^{\ell-1} . \qquad \square$$

The remainder of this subsection deals with the question to what extent being an $(\alpha, \beta, \ell, \mu)$-constellation is preserved under taking link constellations and removing a small proportion of the vertices. Let us observe that if Ψ denotes a k-uniform $(\alpha, \beta, \ell, \mu)$-constellation, then for each $x \in V(\Psi)^{(k-2)}$ the vertices in x are isolated in H_x, which by Definition 2.2.33 (b) implies that they cannot be vertices of R_x^{Ψ}. Thus we have $V(R_x^{\Psi}) \cap x = \varnothing$ for each $x \in V(\Psi)^{(k-2)}$.

Let us now consider for some $S \subseteq V(\Psi)$ of size $|S| \leqslant k-2$ the $(k-|S|)$-uniform link constellation Ψ_S. For every $x \in V(\Psi_S)^{(k-2-|S|)}$ we have $R_x^{\Psi_S} = R_{S \cup x}^{\Psi} \smallsetminus S = R_{S \cup x}^{\Psi}$. Therefore, Ψ_S inherits the property in Definition 2.2.33 (b) from Ψ and together with Fact 2.2.26 this leads to the following conclusion.

Fact 2.2.35. *Given $k \geqslant 2$, $\alpha, \beta > 0$, $\mu' > \mu > 0$ and an odd integer $\ell \geqslant 3$, there exists a natural number n_0 such that the following holds.*

If Ψ is a k-uniform $(\alpha, \beta, \ell, \mu)$-constellation with at least n_0 vertices and $S \subseteq V(\Psi)$ consists of at most $k-2$ vertices, then the $(k-|S|)$-uniform link constellation Ψ_S is an $(\alpha, \beta, \ell, \mu')$-constellation. $\square$

Next we deal with a similar result allowing vertex deletions as well.

Lemma 2.2.36. *Given $k \geqslant 2$, $\alpha, \beta, \mu > 0$ and an odd integer $\ell \geqslant 3$ set*

$$\vartheta = \min\left\{\frac{\alpha}{4}, \frac{\beta}{2\ell}\right\},$$

and let Ψ be a k-uniform $(\alpha, \beta, \ell, \mu)$-constellation on $n \geqslant 6k$ vertices. If $S, X \subseteq V(\Psi)$ are disjoint, $|S| \leqslant k-2$, and $|X| \leqslant \vartheta n$, then $\Psi_S - X$ is an $\big(\frac{\alpha}{2}, \frac{\beta}{2}, \ell, 2\mu\big)$-constellation.

Proof. Let $\Psi = \big(H, \{R_x \colon x \in V(H)^{(k-2)}\}\big)$ be a k-uniform $(\alpha, \beta, \ell, \mu)$-constellation on $n \geqslant 6k$ vertices. Recall that this means

$$\delta_{k-2}(H) \geqslant \Big(\frac{5}{9} + \alpha\Big)\frac{n^2}{2}\,, \tag{2.2.10}$$

and that for every $x \in V(\Psi)^{(k-2)}$ the graph $R_x \subseteq H_x$ has the following properties:

(i) $|V(R_x)| \geqslant \big(\frac{2}{3} + \frac{\alpha}{2}\big)n$,

(ii) $e_{H_x}\big(V(R_x), V(\Psi) \smallsetminus V(R_x)\big) \leqslant \mu n^2$,

(iii) and for all distinct $y, z \in V(R_x)$ the number of y-z-paths in R_x of length ℓ is at least $\beta n^{\ell-1}$.

Further, let $S, X \subseteq V(\Psi)$ be any disjoint sets such that $|S| \leqslant k-2$ and $|X| \leqslant \vartheta n$. We are to prove that

$$\Psi_\star = \Psi_S - X = \Big(\overline{H}_S - X, \big\{R_{x \cup S} - X \colon x \in \big(V(H) \smallsetminus (S \cup X)\big)^{(k-2-|S|)}\big\}\Big)$$

is a $(k-|S|)$-uniform $\big(\frac{\alpha}{2}, \frac{\beta}{2}, \ell, 2\mu\big)$-constellation, i.e., that its underlying hypergraph satisfies an appropriate minimum degree conditions and that the distinguished subgraphs of its link graphs have properties analogous to *(i)*–*(iii)*.

Because of

$$\begin{aligned}\delta_{k-|S|-2}(\overline{H}_S - X) &\geqslant \delta_{k-2}(H - X) \geqslant \Big(\frac{5}{9} + \alpha\Big)\frac{n^2}{2} - |X|n \\ &\geqslant \Big(\frac{5}{9} + \alpha\Big)\frac{n^2}{2} - \vartheta n^2 \geqslant \Big(\frac{5}{9} + \frac{\alpha}{2}\Big)\frac{n^2}{2} \geqslant \Big(\frac{5}{9} + \frac{\alpha}{2}\Big)\frac{|V(\Psi_\star)|^2}{2}\,,\end{aligned}$$

where we utilised $\vartheta \leqslant \frac{\alpha}{4}$ in the penultimate step, the minimum degree of the hypergraph $H(\Psi_\star) = \overline{H}_S - X$ is indeed as large as we need it to be.

Now let $x \in \big(V(\Psi_\star)\big)^{(k-2-|S|)}$ be arbitrary. Since $x \cup S \in \big(V(\Psi) \smallsetminus X\big)^{(k-2)}$, the above statement *(i)* entails

$$\begin{aligned}|V(R_x^{\Psi_\star})| = |V(R_{x\cup S} - X)| &\geqslant \Big(\frac{2}{3} + \frac{\alpha}{2}\Big)n - |X| \\ &\geqslant \Big(\frac{2}{3} + \frac{\alpha}{2}\Big)n - \vartheta n \geqslant \Big(\frac{2}{3} + \frac{\alpha}{4}\Big)n \geqslant \Big(\frac{2}{3} + \frac{\alpha}{4}\Big)|V(\Psi_\star)|\,,\end{aligned}$$

which shows that the required variant of *(i)* holds for $\Psi_\star$.

Next, the graph $H(\Psi_\star)_x = (\overline{H}_S - X)_x$ is a subgraph of $H_{x\cup S}$, so *(ii)* tells us that

$$e_{H(\Psi_\star)_x}\big(V(R_x^{\Psi_\star}), V(\Psi_\star) \smallsetminus V(R_x^{\Psi_\star})\big) \leqslant e_{H_{x\cup S}}\big(V(R_{x\cup S}), V(\Psi) \smallsetminus V(R_{x\cup S})\big) \leqslant \mu n^2 \,.$$

From $\vartheta \leqslant \frac{\alpha}{4} \leqslant \frac{1}{9}$ and $n \geqslant 6k$ we conclude

$$|V(\Psi_\star)| = n - |X| - |S| \geqslant \Big(1 - \frac{1}{9} - \frac{1}{6}\Big)n = \frac{13}{18}n > \frac{n}{\sqrt{2}}$$

and thus we arrive indeed at

$$e_{H(\Psi_\star)_x}\big(V(R_x^{\Psi_\star}), V(\Psi_\star) \smallsetminus V(R_x^{\Psi_\star})\big) \leqslant 2\mu |V(\Psi_\star)|^2 \,,$$

which concludes the proof that the appropriate modification of *(ii)* holds for $\Psi_\star$. Altogether, we have thereby shown that $\Psi_\star$ is an $\big(\frac{\alpha}{2}, 2\mu\big)$-constellation.

Finally we consider distinct vertices $y, z \in V(R_{x\cup S} - X)$ and recall that by *(iii)* above the number of y-z-paths in $R_{x\cup S}$ is at least $\beta n^{\ell-1}$. At most $(\ell-1)|X|n^{\ell-2} \leqslant \frac{\beta}{2}n^{\ell-1}$ of these paths can have an inner vertex in X and, consequently, $R_{x\cup S} - X$ contains at least $\frac{\beta}{2}n^{\ell-1}$ such paths. Therefore $\Psi_\star$ is indeed an $\big(\frac{\alpha}{2}, \frac{\beta}{2}, \ell, 2\mu\big)$-constellation. □

2.3 The Connecting Lemma

In this section we establish the Connecting Lemma (Proposition 2.3.3). Given an $(\alpha, \beta, \ell, \mu)$-constellation with appropriate parameters this result allows us to connect every leftconnectable $(k-1)$-tuple to every rightconnectable $(k-1)$-tuple by means of a short path. In the course of proving Theorem 2.1.2 the Connecting Lemma gets used $\Omega(n)$ times and, essentially, it allows us to convert an almost spanning path cover into an almost spanning cycle. For some reasons related to our way of employing the absorption method, it will turn out to be enormously helpful later if we can guarantee that the number of left-over vertices outside this almost spanning cycle is a multiple of k. There are several possibilities how one might try to accomplish this and our approach is to prove a version of the Connecting Lemma with absolute control over the length of the connecting path modulo k. When closing the almost spanning cycle by means of a final application of the Connecting Lemma, we will then be able to prescribe in which residue class modulo k the number of left-over vertices is going to be. (For a different way to handle such a situation we refer to recent work of Araújo, Piga, and Schacht [4]).

The following result is implicit in [95, Proposition 2.6] and after stating it we shall briefly explain how it can be derived from the argument presented there.

Proposition 2.3.1. *Depending on $\alpha, \beta, \zeta_\star > 0$ and an odd integer $\ell \geqslant 3$ there exist a constant $\vartheta_\star = \vartheta_\star(\alpha, \beta, \ell, \zeta_\star) > 0$ and a natural number n_0 with the following property.*

If Ψ is a 3-uniform $(\alpha, \beta, \ell, \frac{\alpha}{4})$-constellation on $n \geqslant n_0$ vertices, $\vec{a}, \vec{b} \in V(\Psi)^2$ are two disjoint pairs of vertices such that $\vec{a}$ is $\zeta_\star$-leftconnectable and $\vec{b}$ is $\zeta_\star$-rightconnectable, then the number of $\vec{a}$-$\vec{b}$-paths in $H(\Psi)$ with $3\ell+1$ inner vertices is at least $\vartheta_\star n^{3\ell+1}$. □

Observe that the Setup 2.4 we are assuming in [95, Proposition 2.6] is tantamount to an $(\alpha, \beta, \ell, \frac{\alpha}{4})$-constellation. The connectabilty assumptions in [95] are slightly different. Writing $\vec{a} = (x, y)$ we were using in the proof of [95, Proposition 2.6] that a set called U_{xy} there, and defined to consist of all vertices u with $xy \in E(R_u^\Psi)$, has at least the size $\zeta|V(\Psi)|$. When working with vertices $u \in U_{xy}$, however, we were only using $y \in V(R_u^\Psi)$ and $xyu \in E(\Psi)$. For this reason, the entire proof can also be carried out with the set called $U_{(x,y)}^\Psi(\zeta)$ here, or in other words it suffices to suppose that $\vec{a}$ is ζ-leftconnectable. Similarly, we may assume that $\vec{b}$ is ζ-rightconnectable rather than being ζ-connectable in the sense of [95]. Next we introduce the function giving the number of inner vertices in our connections.

Definition 2.3.2. *Given integers $k \geqslant 3$, $0 \leqslant i < k$, and $\ell \geqslant 3$ we set*

$$f(k, i, \ell) = [4^{k-3}(2\ell + 4) - 2]k + i\,.$$

We are now ready to state the k-uniform Connecting Lemma.

Proposition 2.3.3 (Connecting Lemma)**.** *For all $k \geqslant 3$, $\alpha, \beta, \zeta > 0$, and odd integers $\ell \geqslant 3$ there exist $\vartheta > 0$ and $n_0 \in \mathbb{N}$ with the following property.*

If Ψ is a k-uniform $(\alpha, \beta, \ell, \frac{\alpha}{k+6})$-constellation on $n \geqslant n_0$ vertices, $\vec{a}, \vec{b} \in V(\Psi)^{k-1}$ are two disjoint $(k-1)$-tuples such that $\vec{a}$ is ζ-leftconnectable and $\vec{b}$ is ζ-rightconnectable, and $0 \leqslant i < k$, then the number of $\vec{a}$-$\vec{b}$-paths in $H(\Psi)$ with $f = f(k, i, \ell)$ inner vertices is at least ϑn^f.

The proof of this result occupies the remainder of this section and before we begin we provide a short overview over the main ideas. The plan is to proceed by induction on k. When we reach a certain value of k, most of the work is devoted to showing the weaker assertion (Φ_k) that there exists at least one number $f_\star = f_\star(k, \ell)$ such that the statement of the Connecting Lemma holds for connections with $f_\star$ inner vertices. Once we know (Φ_k) the induction can be completed by putting short 'connectable' walks as obtained by Lemma 2.2.27 in the middle and connecting them with two applications of (Φ_k) to $\vec{a}$ and $\vec{b}$.

The proof of (Φ_k) itself is more complicated and starts by applying Lemma 2.2.29 to $U^{\Psi}_{\vec{a}}(\zeta)$ and $U^{\Psi}_{\vec{b}}(\zeta)$ here in place of U and W there. This yields many $(2k-2)$-tuples $(u, q_1, \ldots, q_{2k-4}, w)$ in $V(\Psi)^{2k-2}$ which, after some reordering, have good chances to end up being middle segments of the desired connections. Applying the induction hypothesis to Ψ_u and Ψ_w we can connect $\vec{a}$ and $\vec{b}$ by many $(k-1)$-uniform paths to these middle segments and it remains to 'augment' these connections to k-uniform paths, which can be done by averaging over many possibilities for u and w, respectively (see Figure 2.3.1).

Proof of Proposition 2.3.3. We proceed by induction on k, keeping α, β, and ℓ fixed.

Choice of constants. Due to monotonicity (see Fact 2.2.17) we may suppose that $\zeta \leqslant \frac{1}{k3^{2k}}$. By recursion on $k \geqslant 3$ we define for every $\zeta \in \big(0, \frac{1}{k3^{2k}}\big]$ a positive real number $\vartheta(k,\zeta)$. Starting with $k=3$ we set

$$\vartheta(3,\zeta) = \zeta\big(\vartheta_\star(\alpha,\beta,\ell,\zeta)\big)^2 \quad \text{for} \quad \zeta \in \big(0, 3^{-7}\big],$$

where $\vartheta_\star(\alpha,\beta,\ell,\zeta)$ is given by Proposition 2.3.1. For $k \geqslant 4$ and $\zeta \in \big(0, \frac{1}{k3^{2k}}\big]$ we stipulate

$$\vartheta(k,\zeta) = \zeta^{6s+1}\big(\vartheta(k-1,\zeta^3)\big)^{4s}, \quad \text{where } s = 4^{k-4}(2\ell+4). \tag{2.3.1}$$

Our goal is to prove the Connecting Lemma with $2\vartheta(k,\zeta)$ playing the rôle of ϑ.

The base case $k=3$. Suppose that Ψ is a sufficiently large 3-uniform $(\alpha,\beta,\ell,\frac{\alpha}{9})$-constellation, $i \in \{0,1,2\}$, the pair $\vec{a} = (a_1,a_2) \in V(\Psi)^2$ is ζ-leftconnectable, $\vec{b} = (b_1,b_2)$ is ζ-rightconnectable, the four vertices a_1, a_2, b_1, and b_2 are distinct, and $\zeta \leqslant \frac{1}{3^7}$. Lemma 2.2.27 applied to $(3, i+2)$ here in place of (k,r) there tells us that there are at least $\frac{1}{3^{i+3}}n^{i+4}$ walks $x_1 \cdots x_{i+4}$ of length $i+2$ in $H(\Psi)$ whose starting pair (x_1,x_2) is ζ-rightconnectable and whose ending pair (x_{i+3},x_{i+4}) is ζ-leftconnectable. Among these walks at least

$$\Big(\frac{1}{3^{i+3}} - \frac{4(i+4)}{n}\Big)n^{i+4} > \frac{n^{i+4}}{3^{i+4}} \geqslant \frac{n^{i+4}}{3^6} \geqslant 3\zeta n^{i+4}$$

avoid $\{a_1,a_2,b_1,b_2\}$.

Now for each of them two applications of Proposition 2.3.1 to the $(\alpha,\beta,\ell,\frac{\alpha}{9})$-constellation Ψ enable us to find in $H(\Psi)$ at least $\vartheta_\star n^{3\ell+1}$ paths $a_1a_2p_1\cdots p_{3\ell+1}x_1x_2$ and at least $\vartheta_\star n^{3\ell+1}$

paths $x_{i+3}x_{i+4}r_1\cdots r_{3\ell+1}b_1b_2$ where $\vartheta_\star = \vartheta_\star(\alpha,\beta,\ell,\zeta)$. Altogether, this reasoning leads to at least $3\zeta\vartheta_\star^2 n^f$ walks

$$a_1a_2p_1\cdots p_{3\ell+1}x_1x_2\cdots x_{i+3}x_{i+4}r_1\cdots r_{3\ell+1}b_1b_2$$

with f inner vertices, where

$$f = 2(3\ell+1)+(i+4) = 6\ell+6+i = f(3,i,\ell)\,.$$

At most $f^2n^{f-1} = o(n^f)$ of these walks fail to be paths and thus the assertion follows.

Induction Step. Suppose $k \geqslant 4$ and that the Connecting Lemma is already known for $k-1$ instead of k. Set

$$t = 2k(s-1)+2 \quad \text{and} \quad \eta = \zeta^{3s}\big(\vartheta(k-1,\zeta^3)\big)^{2s}\,, \tag{2.3.2}$$

where, let us recall, $s = 4^{k-4}(2\ell+4)$ was introduced in (2.3.1) while we chose our constants. Following the plan outlined above, our first step is to prove a Connecting Lemma for connections with t inner vertices.

Claim 2.3.4. *For any two disjoint $(k-1)$-tuples $\vec{a} = (a_1,\dots,a_{k-1})$ and $\vec{b} = (b_1,\dots,b_{k-1})$ such that $\vec{a}$ is ζ-leftconnectable and $\vec{b}$ is ζ-rightconnectable, the number of $\vec{a}$-$\vec{b}$-walks with t inner vertices in $H(\Psi)$ is at least $2\eta n^t$.*

Proof. The connectability assumptions mean that the sets

$$U = \{u \in V(\Psi)\colon a_1\cdots a_{k-1}u \in E(\Psi) \text{ and } (a_2,\dots,a_{k-1}) \text{ is } \zeta\text{-leftconnectable in } \Psi_u\}$$

and

$$W = \{w \in V(\Psi)\colon wb_1\cdots b_{k-1} \in E(\Psi) \text{ and } (b_1,\dots,b_{k-2}) \text{ is } \zeta\text{-rightconnectable in } \Psi_w\}$$

satisfy $|U|, |W| \geqslant \zeta n$. Now by $\frac{\alpha}{k+6} \leqslant \frac{\alpha}{10}$ and Fact 2.2.25 Ψ is an $(\alpha, \frac{\alpha}{10})$-constellation. Combined with $\zeta \leqslant \frac{1}{3^{k+2}}$ and Lemma 2.2.29 this shows that the number of $(2k-2)$-tuples

$$(u,\vec{q},w) = (u,q_1,\dots,q_{2k-4},w) \in U \times V(\Psi)^{2k-4} \times W$$

such that

(*a*) $u \neq w$,

(b) $q_1 \cdots q_{2k-4}$ is a walk in $H(\Psi_{uw})$,

(c) $(q_1, \dots, q_{k-2})$ is ζ^3-rightconnectable in Ψ_u,

(d) and $(q_{k-1}, \dots, q_{2k-4})$ is ζ^3-leftconnectable in Ψ_w.

is at least $\zeta^3 n^{2k-2}$. For later reference we recall that $u \in U$ and $w \in W$ mean

(e) $(a_2, \dots, a_{k-1})$ is ζ-leftconnectable in Ψ_u,

(f) $\{a_1, \dots, a_{k-1}, u\} \in E(\Psi)$,

(g) $(b_1, \dots, b_{k-2})$ is ζ-rightconnectable in Ψ_w,

(h) and $\{w, b_1, \dots, b_{k-1}\} \in E(\Psi)$.

Notice that by Fact 2.2.35 the link constellation of every vertex is a $(k-1)$-uniform $(\alpha, \beta, \ell, \frac{\alpha}{k+5})$-constellation and that $f(k-1, 1, \ell) = (k-1)(s-2)+1$. Now for every $(2k-2)$-tuple $(u, \vec{q}, w)$ satisfying (a)–(h) we apply the induction hypothesis twice with $(\zeta^3, 1)$ here in place of (ζ, i) there. First, by (c) and (e) we can connect $(a_2, \dots, a_{k-1})$ to $(q_1, \dots, q_{k-2})$ in Ψ_u, thus getting at least $2\vartheta(k-1, \zeta^3)(n-1)^{(k-1)(s-2)+1}$

(i) walks $a_2 \cdots a_{k-1} p_1 \cdots p_{(k-1)(s-2)+1} q_1 \cdots q_{k-2}$ in Ψ_u

with $f(k-1, 1, \ell)$ inner vertices. Second, (d) and (g) allow us to connect $(q_{k-1}, \dots, q_{2k-4})$ to $(b_1, \dots, b_{k-2})$ in Ψ_w by at least $2\vartheta(k-1, \zeta^3)(n-1)^{(k-1)(s-2)+1}$

(j) walks $q_{k-1} \cdots q_{2k-4} r_{(k-1)(s-2)+1} \cdots r_1 b_1 \cdots b_{k-2}$ in Ψ_w.

Altogether, the number of $\big((k-1)(2s-2)+2\big)$-tuples

$$(u, \vec{p}, \vec{q}, \vec{r}, w) \in U \times V(\Psi)^{(k-1)(s-2)+1} \times V(\Psi)^{2k-4} \times V(\Psi)^{(k-1)(s-2)+1} \times W$$

with (a)–(j), where

$$\vec{p} = (p_1, \dots, p_{(k-1)(s-2)+1}) \quad \text{and} \quad \vec{r} = (r_{(k-1)(s-2)+1}, \dots, r_1),$$

is at least $4\zeta^3\big(\vartheta(k-1,\zeta^3)\big)^2(n-1)^{(k-1)(2s-4)+2} n^{2k-2} \geqslant 2\zeta^3\big(\vartheta(k-1,\zeta^3)\big)^2 n^{(k-1)(2s-2)+2}$.

Roughly speaking, we plan to derive the $\vec{a}$-$\vec{b}$-paths we are supposed to construct from these $\big((k-1)(2s-2)+2\big)$-tuples by taking many copies of u and w and inserting them in appropriate positions into $(\vec{a}, \vec{p}, \vec{q}, \vec{r}, \vec{b})$. To analyse the number of ways of doing this, we consider the auxiliary 3-partite 3-uniform hypergraph $\mathscr{A}$ with vertex classes $U^\star$, M, and $W^\star$, where $U^\star$ and $W^\star$ are two disjoint copies of $V(\Psi)$, while $M = V(\Psi)^{(k-1)(2s-2)}$.

We represent the vertices in M as sequences

$$\vec{m} = (\vec{p}, \vec{q}, \vec{r}) = (p_1, \dots, p_{(k-1)(s-2)+1}, q_1, \dots, q_{2k-4}, r_{(k-1)(s-2)+1}, \dots, r_1)\,.$$

The edges of $\mathscr{A}$ are defined to be the triples $\{u, \vec{m}, w\}$ with $u \in U \subseteq U^\star$, $\vec{m} \in M$, and $w \in W \subseteq W^\star$, for which the $\big((k-1)(2s-2)+2\big)$-tuple $(u, \vec{m}, w)$ satisfies (*a*)–(*j*). We have just proved that

$$e(\mathscr{A}) \geqslant 2\zeta^3\big(\vartheta(k-1,\zeta^3)\big)^2 n^{(k-1)(2s-2)+2} = 2\zeta^3\big(\vartheta(k-1,\zeta^3)\big)^2 |U^\star||M||W^\star|\,. \tag{2.3.3}$$

By the (ordered) bipartite link graph of a vertex $\vec{m} \in M$ we mean the set of pairs

$$\mathscr{A}_{\vec{m}} = \big\{(u,w) \in U \times W \colon u\vec{m}w \in E(\mathscr{A})\big\}\,.$$

The convexity of the function $x \longmapsto x^s$ on $\mathbb{R}_{\geqslant 0}$ yields

$$\begin{aligned}\sum_{\vec{m}\in M} |\mathscr{A}_{\vec{m}}|^s &\geqslant |M|\left(\frac{e(\mathscr{A})}{|M|}\right)^s \overset{(2.3.3)}{\geqslant} n^{(k-1)(2s-2)}\big(2\zeta^3\big(\vartheta(k-1,\zeta^3)\big)^2 n^2\big)^s \\ &\geqslant 2\zeta^{3s}\big(\vartheta(k-1,\zeta^3)\big)^{2s} n^{k(2s-2)+2} \overset{(2.3.2)}{=} 2\eta n^t\,.\end{aligned} \tag{2.3.4}$$

In other words, the number of t-tuples

$$(\vec{u}, \vec{p}, \vec{q}, \vec{r}, \vec{w}) \in U^s \times V(\Psi)^{(k-1)(s-2)+1} \times V(\Psi)^{2k-4} \times V(\Psi)^{(k-1)(s-2)+1} \times W^s$$

with

$$(u_1, w_1), \dots, (u_s, w_s) \in \mathscr{A}_{\vec{m}}\,,$$

where

$$\vec{u} = (u_1, \dots, u_s)\,, \quad \vec{w} = (w_1, \dots, w_s)\,, \quad \text{and} \quad \vec{m} = (\vec{p}, \vec{q}, \vec{r}) \in M\,,$$

is at least $2\eta n^t$. So to conclude the proof of Claim 2.3.4 it suffices to show that for every such t-tuple the sequence

$$a_1 \cdots a_{k-1} u_1 p_1 \cdots p_{k-1} u_2 p_k \cdots p_{2k-2} u_3 \cdots u_{s-2} p_{(k-1)(s-3)+1} \cdots p_{(k-1)(s-2)} u_{s-1} p_{(k-1)(s-2)+1}$$
$$q_1 \cdots q_{k-2} u_s w_s q_{k-1} \cdots q_{2k-4}$$

$$r_{(k-1)(s-2)+1}w_{s-1}r_{(k-1)(s-2)}\cdots r_{(k-1)(s-3)+1}w_{s-2}\cdots w_3r_{2k-2}\cdots r_kw_2r_{k-1}\cdots r_1w_1b_1\cdots b_{k-1}$$

indicated in Figure 2.3.1 is an $\vec{a}$-$\vec{b}$-walk in $H(\Psi)$.

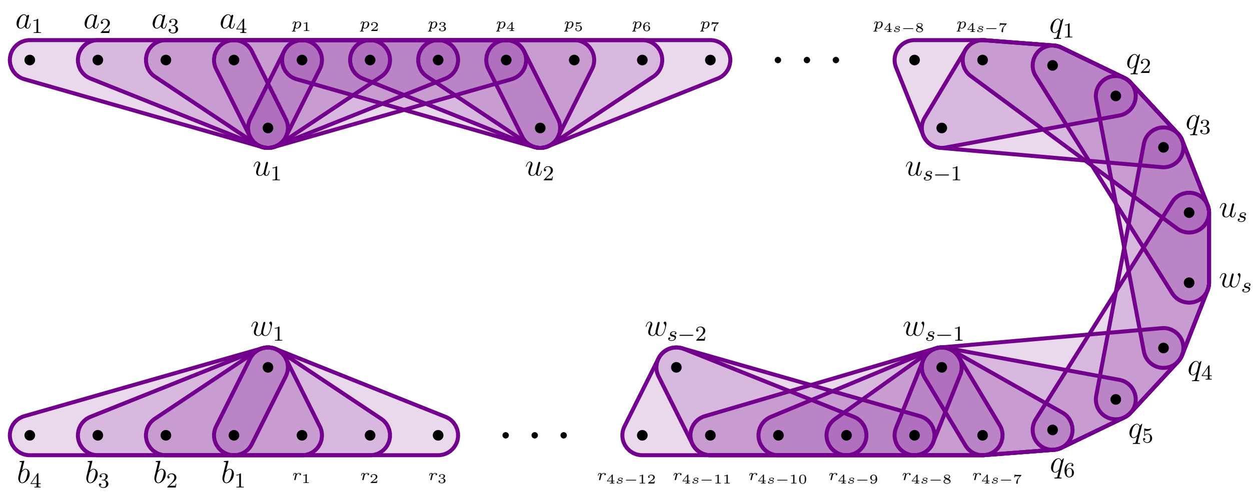


Figure 2.3.1: Connecting (a_1, a_2, a_3, a_4) and (b_1, b_2, b_3, b_4) in a 5-uniform constellation.

We shall now argue that this follows from the fact that for each $j \in [s]$ the conditions (*a*)–(*j*) hold for u_j and w_j here in place of u and w there.

The first of the required edges is provided by the case $u = u_1$ of (*f*). Together with (*i*) this shows that the initial segment

$$a_1a_2\cdots a_{k-1}u_1p_1\cdots p_{k-1}u_2p_k\cdots p_{2k-2}u_3\cdots u_{s-2}p_{(k-1)(s-3)+1}\cdots p_{(k-1)(s-2)}$$
$$u_{s-1}p_{(k-1)(s-2)+1}q_1\cdots q_{k-2}u_s$$

is a walk in $H(\Psi)$. Similarly, by (*h*) and (*j*) the terminal segment

$$w_sq_{k-1}\cdots q_{2k-4}r_{(k-1)(s-2)+1}w_{s-1}r_{(k-1)(s-2)}\cdots r_{(k-1)(s-3)+1}w_{s-2}\cdots w_3$$
$$r_{2k-2}\cdots r_kw_2r_{k-1}\cdots r_1w_1b_1\cdots b_{k-2}b_{k-1}$$

is a walk in $H(\Psi)$. Finally, the middle part

$$q_1\cdots q_{k-2}u_sw_sq_{k-1}\cdots q_{2k-4}$$

is a walk in $H(\Psi)$, because by (*b*) we know that $q_1\cdots q_{2k-4}$ is a walk in $H(\Psi_{u_sw_s})$. $\square$

Returning to the induction step, we consider $i \in \{0, 1, \ldots, k-1\}$, a ζ-leftconnectable

$(k-1)$-tuple $\vec{a} \in V(\Psi)^{k-1}$, and a ζ-rightconnectable $(k-1)$-tuple $\vec{b}$ such that $\vec{a}$ and $\vec{b}$ have no vertices in common. Plugging $r = i + k - 3$ into Lemma 2.2.27 we obtain at least $\frac{1}{3^{i+k-2}} n^{i+2k-4}$ walks $x_1 \cdots x_{i+2k-4}$ of length $i + k - 3$ in $H(\Psi)$ that start with a ζ-rightconnectable $(k-1)$-tuple and end with a ζ-leftconnectable $(k-1)$-tuple. Of these walks, at least

$$\Big(\frac{1}{3^{i+k-2}} - \frac{2(k-1)(i+2k-4)}{n}\Big) n^{i+2k-4} > \frac{n^{i+2k-4}}{3^{i+k-1}} > \frac{n^{i+2k-4}}{3^{2k}} > \zeta n^{i+2k-4}$$

have no common vertices with $\vec{a}$ and $\vec{b}$. For each such walk, Claim 2.3.4 tells us that we can connect $\vec{a}$ to $(x_1, \ldots, x_{k-1})$ in at least $2\eta n^t$ ways by a walk with t inner vertices, and the same applies to connections from $(x_{i+k-2}, \ldots, x_{i+2k-4})$ to $\vec{b}$.

Altogether this reasoning leads to $4\zeta\eta^2 n^f = 4\vartheta(k, \zeta) n^f$ walks in $H(\Psi)$ from $\vec{a}$ to $\vec{b}$ with f inner vertices, where

$$\begin{aligned} f &= 2t + (i + 2k - 4) = 2(2ks - 2k + 2) + (i + 2k - 4) \\ &= (4s - 2)k + i = [4^{k-3}(2\ell + 4) - 2]k + i = f(k, i, \ell)\,. \end{aligned}$$

As usual, at most $O(n^{f-1})$ of these walks can fail to be paths. So, in particular, there exist at least $2\vartheta(k, \zeta) n^f$ paths from $\vec{a}$ to $\vec{b}$ possessing f inner vertices. This completes the induction step and, hence, the proof of the Connecting Lemma. □

2.4 Reservoir Lemma

In this section we discuss a standard device occurring in many applications of the absorption method: the reservoir. The problem addressed by the Reservoir Lemma is that while the Connecting Lemma delivers many connections for any two disjoint connectable $(k-1)$-tuples, it gives us no control where the inner vertices of these connections are. Thus it might happen that each of these connections has an inner vertex which is ‘unavailable’ to us, because we already assigned a different rôle to it in the Hamiltonian cycle we are about to construct. To avoid this problem, one fixes a small random subset of the vertex set, called the reservoir, and decides that the vertices in the reservoir will only be used for the purpose of connecting pairs of $(k-1)$-tuples by means of short paths.

Proposition 2.4.1 (Reservoir Lemma). *Suppose that $k \geqslant 3$, $\alpha, \beta, \xi, \zeta_{\star\star} > 0$, and that $\ell \geqslant 3$ is an odd integer. If $\vartheta_{\star\star} = \vartheta(k, \alpha, \beta, \ell, \zeta_{\star\star})$ is provided by Proposition 2.3.3, then* *there exists some $n_0 \in \mathbb{N}$ such that for every k-uniform $(\alpha, \beta, \ell, \frac{\alpha}{k+6})$-constellation Ψ on $n \geqslant n_0$ vertices there exists a subset $\mathcal{R} \subseteq V(\Psi)$ with the following properties.*

(i) We have $\frac{1}{2}\xi n \leqslant |\mathcal{R}| \leqslant \xi n$.

(ii) For all pairs of disjoint $(k-1)$-tuples $\vec{a}, \vec{b} \in V(\Psi)^{k-1}$ such that $\vec{a}$ is $\zeta_{\star\star}$-leftconnectable and $\vec{b}$ is $\zeta_{\star\star}$-rightconnectable in Ψ, and for every $i \in [0, k)$, the number of $\vec{a}$-$\vec{b}$-paths in $H(\Psi)$ with $f = f(k, i, \ell)$ inner vertices all of which belong to $\mathcal{R}$ is at least $\frac{1}{2}\vartheta_{\star\star}|\mathcal{R}|^f$.

Since the proof of this result is quite standard, we will only provide a brief sketch here. It suffices to prove that the binomial random subset $\mathcal{R} \subseteq V(\Psi)$ including every vertex independently with probability $\frac{3}{4}\xi$ a.a.s. has the properties *(i)* and *(ii)*. Now *(i)* is a straightforward consequence of Chernoff's inequality. As there are only polynomially many possibilities for $(\vec{a}, \vec{b}, i)$ in *(ii)*, it suffices to show that for each of them the probability that there are fewer than $\frac{1}{2}\vartheta_{\star\star}|\mathcal{R}|^f$ paths of the desired form is at most $\exp\big(-\Omega(n)\big)$. This can in turn be established by applying the Azuma-Hoeffding inequality to the at least $\vartheta_{\star\star}n^f$ such paths in $V(\Psi)^f$ delivered by Proposition 2.3.3. For further details we refer to [92, Proposition 4.1], where we gave a full account of the argument for $k = 4$.

Let us emphasise again that the set $\mathcal{R}$ provided by Proposition 2.4.1 is called the *reservoir*. The connections in *(ii)* whose inner vertices belong to $\mathcal{R}$ are called paths *through* $\mathcal{R}$.

In the proof of Theorem 2.1.2 we shall repeatedly connect suitable tuples through the reservoir. Whenever such a connection is made, some of the vertices of the reservoir are used and the part of $\mathcal{R}$ still available for further connections shrinks. Although the reservoir gets used $\Omega(|V(\Psi)|)$ times, we shall be able to keep an appropriate version of property *(ii)* of the reservoir intact throughout this process.

Corollary 2.4.2. *Let a sufficiently large k-uniform $(\alpha, \beta, \ell, \frac{\alpha}{k+6})$-constellation Ψ as well as a reservoir $\mathcal{R} \subseteq V(\Psi)$ as in Proposition 2.4.1 be given. Moreover, let $\mathcal{R}' \subseteq \mathcal{R}$ be an arbitrary set with $|\mathcal{R}'| \leqslant \frac{\xi\vartheta_{\star\star}}{4^k k\ell}n$. If $\vec{a}, \vec{b} \in V(\Psi)^{k-1}$ are two disjoint $(k-1)$-tuples such that $\vec{a}$ is $\zeta_{\star\star}$-leftconnectable and $\vec{b}$ is $\zeta_{\star\star}$-rightconnectable, then for every $i \in [0, k)$ there is an $\vec{a}$-$\vec{b}$-path through $\mathcal{R} \smallsetminus \mathcal{R}'$ with $f(k, i, \ell)$ inner vertices.*

Proof. Set $f = f(k, i, \ell)$ and recall that $f(k, i, \ell) = (4^{k-3}(2\ell + 4) - 2)k + i < 4^{k-2}k\ell$. So the lower bound in Proposition 2.4.1 *(i)* together with the bound on $|\mathcal{R}'|$ yields

$$|\mathcal{R}'| \leqslant \frac{\vartheta_{\star\star}|\mathcal{R}|}{4^{k-1}k\ell} \leqslant \frac{\vartheta_{\star\star}|\mathcal{R}|}{4f}\,.$$

Consider all $\vec{a}$-$\vec{b}$-paths through $\mathcal{R}$ with f inner vertices. On the one hand, there are at least $\frac{\vartheta_{\star\star}}{2}|\mathcal{R}|^f$ of them due to Proposition 2.4.1 *(ii)*. On the other hand, there are at most

$$f|\mathcal{R}'||\mathcal{R}|^{f-1} \leqslant \frac{\vartheta_{\star\star}}{4}|\mathcal{R}|^f$$

such paths having an inner vertex in $\mathcal{R}'$. Consequently, at least $\frac{\vartheta_{\star\star}}{2}|\mathcal{R}|^f - \frac{\vartheta_{\star\star}}{4}|\mathcal{R}|^f > 0$ of our paths have all their inner vertices in $\mathcal{R} \smallsetminus \mathcal{R}'$. □

2.5 The absorbing path

2.5.1 Overview

In this section we establish that for $\mu \ll \alpha$ every sufficiently large $(\alpha, \beta, \ell, \mu)$-constellation contains an *absorbing path* P_A, whose main property is that it can 'absorb' an arbitrary but not too large set of vertices whose cardinality is a multiple of k. Thus the problem of proving Theorem 2.1.2 gets reduced to the simpler task of finding an almost spanning cycle containing the absorbing path and missing a number of vertices that is divisible by k. In order to have a realistic chance to include the absorbing path into such a cycle we make sure that its first and last $(k-1)$-tuple is connectable. Moreover, we will need to be able to work outside a forbidden 'reservoir set' that later will have been selected in advance.

Proposition 2.5.1 (Absorbing Path Lemma)**.** *Given $k \geqslant 3$, $\alpha > 0$, $\beta > 0$, and an odd integer $\ell \geqslant 3$ there exist constants $\zeta = \zeta(\alpha, k)$, $\vartheta_\star = \vartheta_\star(k, \alpha, \beta, \ell, \zeta) > 0$ and an integer n_0 with the following property.*

Suppose that Ψ is a k-uniform $(\alpha, \beta, \ell, \mu)$-constellation with $\mu = \frac{1}{10k}\left(\frac{\alpha}{2}\right)^{2^{k-3}+1}$ on $n \geqslant n_0$ vertices. If $\mathcal{R} \subseteq V(\Psi)$ with $|\mathcal{R}| \leqslant \vartheta_\star^2 n$ is arbitrary, then there exists a path $P_A \subseteq H(\Psi) - \mathcal{R}$ such that

(i) $|V(P_A)| \leqslant \vartheta_\star n$,

(ii) *the starting and ending $(k-1)$-tuple of P_A are ζ-connectable,*

(iii) *and for every subset $Z \subseteq V(\Psi) \smallsetminus V(P_A)$ with $|Z| \leqslant 2\vartheta_\star^2 n$ and $|Z| \equiv 0 \pmod{k}$, there exists a path $Q \subseteq H(\Psi)$ with $V(Q) = V(P_A) \cup Z$ having the same end-$(k-1)$-tuples as P_A.*

Our absorbers will be analogous to those in [92] and we refer to [92, Section 5.1] for further motivation. Here we will only recall that the absorbers have two kinds of main components reflecting the following observations.

- A complete k-partite subhypergraph S of $H(\Psi)$ whose vertex classes $\{x_i, x_{i+k}, x_{i+2k}\}$ are of size 3 (where $i \in [k]$) contains a spanning path $P = x_1 \dots x_{3k}$. Moreover, S also contains the path $P' = x_1 \dots x_k x_{2k+1} \dots x_{3k}$, which has the same first and last $(k-1)$-tuple as P. Thus if the absorbing path contains P' as a subpath but avoids the vertices $x_{k+1}, \dots, x_{2k}$, then it can absorb these vertices simultaneously (see Figure 2.5.1a). However, not every k-element subset of $V(\Psi)$ is *absorbable* in this manner.

- If the links of two vertices a and x intersect in a $(k-1)$-uniform path $b_1 \dots b_{2k-2}$, then we can form two k-uniform paths in $H(\Psi)$, namely $P_a = b_1 \dots b_{k-1} a b_k \dots b_{2k-1}$ and $P_x = b_1 \dots b_{k-1} x b_k \dots b_{2k-1}$ (see Figure 2.5.1b). Now if the absorbing path contains P_x, then we can remove x and insert a instead. We call such a structure an (a, x)-*exchanger*.

Now the plan for absorbing an arbitrary set $\{a_1, \dots, a_k\}$ of k vertices is that we will find an 'absorbable' set $\{x_1, \dots, x_k\}$ such that for every $i \in [k]$ there is an (a_i, x_i)-exchanger. The main difficulty in executing this strategy is that we need to pay a lot of attention to connectability issues, because ultimately we need to connect all parts of the absorbers we are about to construct to the rest of the Hamiltonian cycle we intend to exhibit. For this reason, the definition of absorbers reads as follows.

Definition 2.5.2. *Suppose that $k \geqslant 3$, $\alpha, \mu, \zeta > 0$, that Ψ is a k-uniform (α, μ)-constellation, and that $\vec{a} = (a_1, \dots, a_k) \in V(\Psi)^k$ is a k-tuple consisting of distinct vertices. We say that*

$$\vec{A} = (\vec{u}, \vec{x}, \vec{w}, \vec{b}_1, \dots, \vec{b}_k) \in V(\Psi)^{2k^2+k}$$

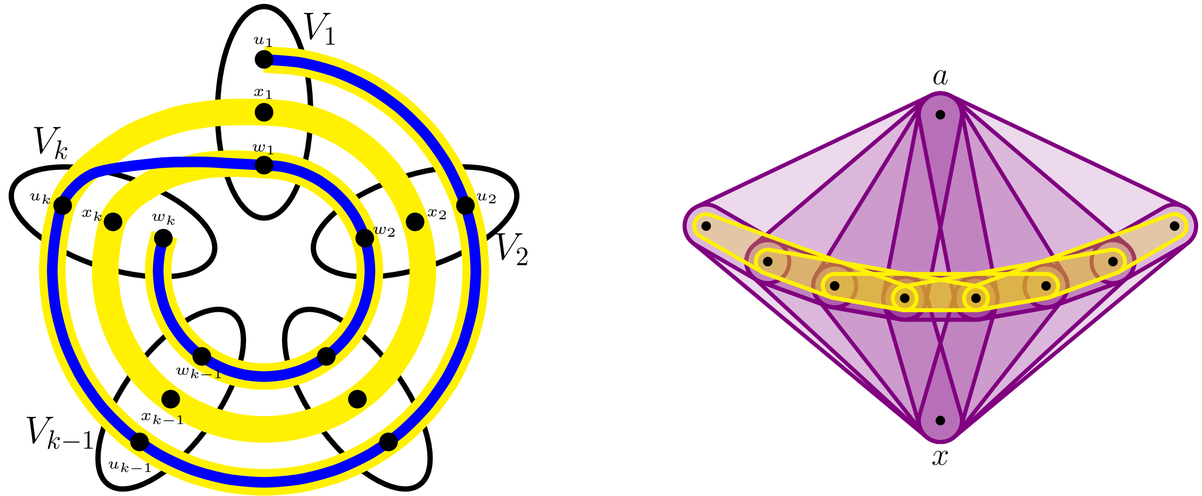
with $\vec{u} = (u_1, \dots, u_k)$, $\vec{x} = (x_1, \dots, x_k)$, $\vec{w} = (w_1, \dots, w_k)$, and $\vec{b}_i = (b_{i1}, \dots, b_{i(2k-2)})$ for $i \in [k]$ is an $(\vec{a}, \zeta)$-absorber in Ψ, if

(a) all $2k^2 + k$ vertices of $\vec{A}$ are distinct and different from those in $\vec{a}$,

(b) $\vec{u}\vec{x}\vec{w}$ and $\vec{u}\vec{w}$ are paths in $H(\Psi)$,

(c) $(u_1, \dots, u_{k-1})$ is ζ-rightconnectable and $(w_2, \dots, w_k)$ is ζ-leftconnectable in Ψ,

(d) and for every $i \in [k]$ the $(2k-2)$-tuple $\vec{b}_i$ is a path in $H(\Psi_{a_i}) \cap H(\Psi_{x_i})$ whose first and last $(k-1)$-tuple is ζ-connectable in Ψ.



(a) The $K_k^{(k)}(3)$ with two paths (b) A 5-uniform (a,x)-exchanger

Figure 2.5.1: The building blocks of an absorber.

We conclude this subsection with an explicit description how these absorbers are going to be utilised (see Figure 2.5.2). Suppose to this end that for some k-tuple $\vec{a} = (a_1, \dots, a_k)$ consisting of k distinct vertices and some $(\vec{a}, \zeta)$-absorber $(\vec{u}, \vec{x}, \vec{w}, \vec{b}_1, \dots, \vec{b}_k)$ it turns out that the paths

$$\vec{u}\vec{w} \quad \text{and} \quad b_{i1} \cdots b_{i(k-1)} x_i b_{ik} \cdots b_{i(2k-2)} \quad \text{for } i \in [k] \tag{2.5.1}$$

end up being subpaths of the absorbing path P_A we are about to construct, while $a_1, \dots, a_k$ are not in $V(P_A)$. We may then replace for each $i \in [k]$ the path

$$b_{i1} \cdots b_{i(k-1)} x_i b_{ik} \cdots b_{i(2k-2)} \quad \text{by the path} \quad b_{i1} \cdots b_{i(k-1)} a_i b_{ik} \cdots b_{i(2k-2)} \,,$$

and then

$$\vec{u}\vec{w} \quad \text{by} \quad \vec{u}\vec{x}\vec{w} \,.$$

In this manner we transform P_A into a new path Q with $V(Q) = V(P_A) \cup \{a_1, \dots, a_k\}$ having the same first and last $(k-1)$-tuple as P_A. We say in this situation that Q arises from P_A by *absorbing* $\{a_1, \dots, a_k\}$. The $k+1$ paths enumerated in (2.5.1) are called the *pre-absorption paths* of the absorber $(\vec{u}, \vec{x}, \vec{w}, \vec{b}_1, \dots, \vec{b}_k)$. So there is one pre-absorption path with $2k$ vertices, namely $\vec{u}\vec{w}$, and there are k pre-absorption paths with $2k-1$ vertices having a vertex x_i in the middle.

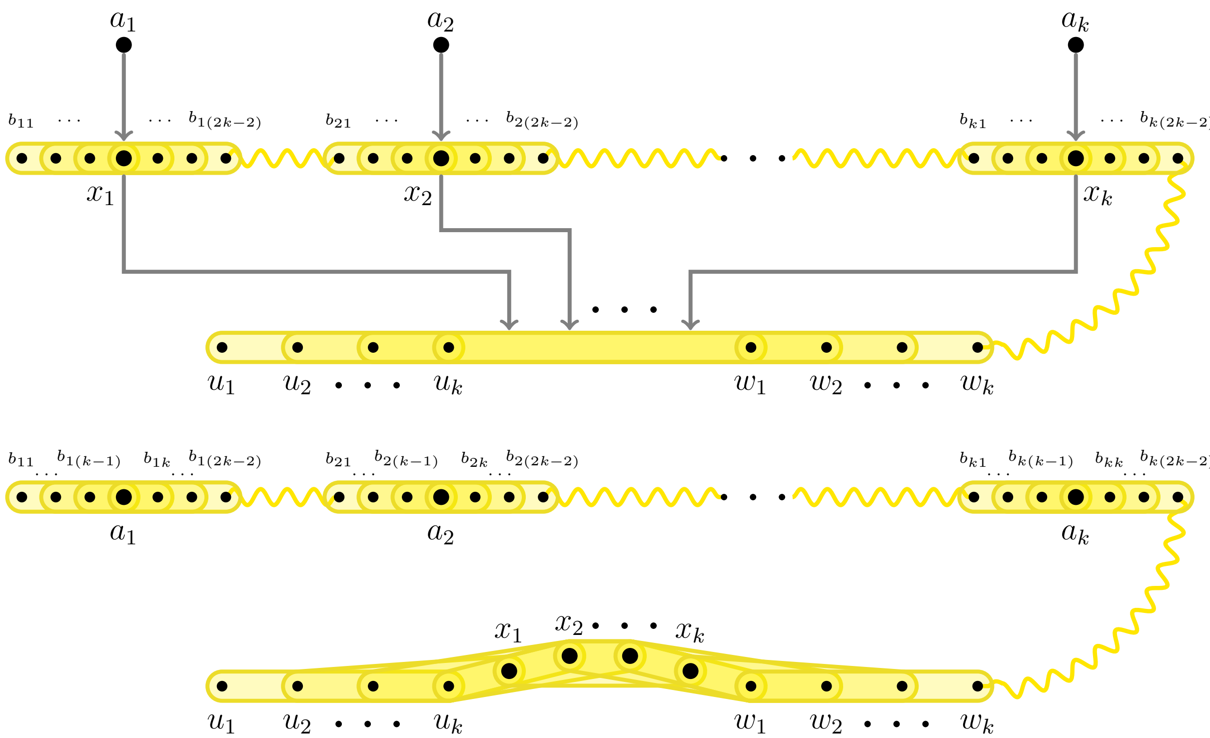


Figure 2.5.2: Absorber for $(a_1, \ldots, a_k)$ before and after absorption.

2.5.2 Construction of the building blocks

We commence with the first part $(\vec{u}, \vec{x}, \vec{w})$ of our absorbers consisting of $3k$ vertices. As we have already indicated, we shall find $(3k)$-tuples satisfying clause *(b)* of Definition 2.5.2 by looking for complete k-partite subhypergraphs of $H(\Psi)$ whose vertex classes are of size three.

Let us recall for this purpose that by a classic result of Erdős [31] the Turán density of every k-partite k-uniform hypergraph vanishes. This means that, given a k-partite k-uniform hypergraph F and a constant $\varepsilon > 0$, every sufficiently large k-uniform hypergraph H satisfying $|E(H)| \geqslant \varepsilon |V(H)|^k$ contains a copy of F. Due to the so-called 'supersaturation' phenomenon later discovered by Erdős and Simonovits [35], the same assumption actually implies that H contains $\Omega\big(|V(H)|^{|V(F)|}\big)$ copies of F. For later reference, we record this fact as follows.

Lemma 2.5.3. *Given a k-partite k-uniform hypergraph F and $\varepsilon > 0$, there are a constant $\xi > 0$ and a natural number n_0 such that every k-uniform hypergraph H on $n \geqslant n_0$ vertices with at least εn^k edges contains at least $\xi n^{|V(F)|}$ copies of F.* $\square$

We shall now apply this result to $F = K_k^{(k)}(3)$, the complete k-partite hypergraph with vertex classes of size 3, and to an auxiliary hypergraph whose edges are derived from bridges (recall Definition 2.2.21). This will establish the following statement, whose conditions *(i)* and *(ii)* coincide with *(b)* and *(c)* from Definition 2.5.2.

Lemma 2.5.4. *For every $k \geqslant 2$ there exists $\xi = \xi(k) > 0$ such that for every $\alpha > 0$ there is an integer n_0 with the following property.*

For every k-uniform $(\alpha, \frac{\alpha}{9})$-constellation Ψ on $n \geqslant n_0$ vertices the number of $(3k)$-tuples $(\vec{u}, \vec{x}, \vec{w}) \in V(\Psi)^k \times V(\Psi)^k \times V(\Psi)^k$ such that writing $\vec{u} = (u_1, \ldots, u_k)$, $\vec{x} = (x_1, \ldots, x_k)$, and $\vec{w} = (w_1, \ldots, w_k)$

(i) both $\vec{u}\vec{x}\vec{w}$ and $\vec{u}\vec{w}$ are k-uniform paths in Ψ,

(ii) $(u_1, \ldots, u_{k-1})$ is $\frac{1}{9k}$-rightconnectable and $(w_2, \ldots, w_k)$ is $\frac{1}{9k}$-leftconnectable in Ψ

is at least ξn^{3k}.

Proof. Throughout the argument we assume that $\xi \ll k^{-1}$ is sufficiently small and that $n_0 \gg \alpha^{-1}, \xi^{-1}$ is sufficiently large. Let Ψ be a k-uniform $(\alpha, \frac{\alpha}{9})$-constellation on $n \geqslant n_0$ vertices. Construct an auxiliary k-partite k-uniform hypergraph $\mathscr{B} = (V_1 \cup \ldots \cup V_k, E_{\mathscr{B}})$ whose vertex classes are k disjoint copies of $V(\Psi)$ and whose edges $\{v_1, \ldots, v_k\} \in E_{\mathscr{B}}$ with $v_i \in V_i$ for $i \in [k]$ correspond to the $\frac{1}{9k}$-bridges $(v_1, \ldots, v_k)$ of Ψ. Corollary 2.2.28 tells us that

$$|E_{\mathscr{B}}| \geqslant \frac{1}{9} n^k = \frac{1}{9k^k} |V(\mathscr{B})|^k .$$

So Lemma 2.5.3 applied to $\mathscr{B}$ and $F = K_k^{(k)}(3)$ leads to $\Omega(n^{3k})$ copies of $K_k^{(k)}(3)$ in $\mathscr{B}$, where the implied constant only depends on k. In other words, for some constant $\xi = \xi(k)$ depending only on k there are at least ξn^{3k} tuples $(\vec{u}, \vec{x}, \vec{w}) \in V(\Psi)^k \times V(\Psi)^k \times V(\Psi)^k$ such that, writing $\vec{u} = (u_1, \ldots, u_k)$, $\vec{x} = (x_1, \ldots, x_k)$, and $\vec{w} = (w_1, \ldots, w_k)$, we have a copy of $K_k^{(k)}(3)$ in $\mathscr{B}$ with $u_i, x_i, w_i \in V_i$ for all $i \in [k]$. Clearly, these $(3k)$-tuples satisfy the demand *(i)* of the lemma and, since $\vec{u}$ and $\vec{w}$ are $\frac{1}{9k}$-bridges, they have property *(ii)* as well (cf. Definition 2.2.21). □

Armed with this result and with Corollary 2.2.32 we can now prove that if $\zeta, \mu \ll \alpha, k^{-1}$, then for every k-tuple $\vec{a}$ of distinct vertices from a sufficiently large (α, μ)-constellation the number of $(\vec{a}, \zeta)$-absorbers is at least $\Omega(n^{2k^2+k})$.

Lemma 2.5.5. *For every $k \geqslant 3$ and $\alpha > 0$ there exist constants $\zeta = \zeta(\alpha, k)$ and $\xi = \xi(\alpha, k)$ as well as an integer n_0 with the following property.*

If Ψ is a k-uniform (α, μ)-constellation on $n \geqslant n_0$ vertices, where $\mu = \frac{1}{10k}\left(\frac{\alpha}{2}\right)^{2^{k-3}+1}$, and $\vec{a} \in V(\Psi)^k$ is an arbitrary k-tuple of distinct vertices, then the number of $(\vec{a}, \zeta)$-absorbers in Ψ is at least ξn^{2k^2+k}.

Proof. Starting with the constant $\xi'' = \xi''(k) > 0$ provided by Lemma 2.5.4 we set

$$\xi' = \frac{\mu^k}{2}, \quad \zeta = \frac{\xi''\mu}{7k}, \quad \text{and} \quad \xi = \frac{1}{4}(\xi')^k\xi'' \tag{2.5.2}$$

and we suppose that n_0 is sufficiently large.

In order to show that ζ and ξ have the desired property, we consider a k-uniform (α, μ)-constellation Ψ on $n \geqslant n_0$ vertices as well as a k-tuple $\vec{a} = (a_1, \ldots, a_k) \in V(\Psi)^k$ consisting of distinct vertices. The set $X \subseteq V(\Psi)$ delivered by Corollary 2.2.32 (with the same meaning of Ψ, α, μ, and ζ as here) satisfies

$$|X| \leqslant \frac{\zeta}{\mu}n \overset{(2.5.2)}{=} \frac{\xi''}{7k}n\,. \tag{2.5.3}$$

By $\mu \leqslant \frac{\alpha}{9}$, $\zeta \leqslant \frac{1}{9k}$, and monotonicity, Lemma 2.5.4 yields at least $\xi'' n^{3k}$ paths $(\vec{u}, \vec{x}, \vec{w})$ in $V(\Psi)^{3k}$ with the properties *(i)* and *(ii)* of that lemma. Since the number of these paths sharing a vertex with $X \cup \{a_1, \ldots, a_k\}$ can be bounded from above by

$$3k(|X| + k)n^{3k-1} \overset{(2.5.3)}{\leqslant} 3k\frac{\xi''}{7k}n^{3k} + 3k^2n^{3k-1} < \frac{\xi''}{2}n^{3k}\,,$$

there are at least $\frac{\xi''}{2}n^{3k}$ such paths avoiding both X and $\vec{a}$. Now it suffices to establish that each of them participates in at least $\frac{1}{2}(\xi')^k n^{2k^2-2k}$ absorbers.

For the rest of the proof we fix some such path $(\vec{u}, \vec{x}, \vec{w}) \in V(\Psi)^{3k}$ and, as usual, we write $\vec{x} = (x_1, \ldots, x_k)$. Now we apply Corollary 2.2.32 for every $i \in [k]$ to the vertices a_i and x_i, thus obtaining $\xi' n^{2k-2}$ paths $\vec{b}_i = (b_{i1}, \ldots, b_{i(2k-2)}) \in V(\Psi)^{2k-2}$ in $H(\Psi_{a_i}) \cap H(\Psi_{x_i})$ whose first and last $(k-1)$-tuples are ζ-connectable in Ψ. Altogether, this yields $(\xi')^k n^{2k^2-2k}$ possibilities for $(\vec{b}_1, \ldots, \vec{b}_k)$ and for most of them $(\vec{u}, \vec{x}, \vec{w}, \vec{b}_1, \ldots, \vec{b}_k)$ is an $(\vec{a}, \zeta)$-absorber. The only exceptions occur when some of these $2k^2 + k$ vertices coincide, but this can happen in at most $(2k^2 + k)(2k^2 - 2k)n^{(2k-2)k-1} < \frac{1}{2}(\xi')^k n^{2k^2-2k}$ ways. Thus $(\vec{u}, \vec{x}, \vec{w})$ is indeed extendable in at least $\frac{1}{2}(\xi')^k n^{2k^2-2k}$ distinct ways to an $(\vec{a}, \zeta)$-absorber. $\square$

2.5.3 Construction of the absorbing path

After these preparations the Absorbing Path Lemma can be shown in a rather standard fashion. The argument starts by observing that a random selection of $(2k^2+k)$-tuples contains, with high probability, for every k-tuple $\vec{a}$ a positive proportion of $(\vec{a},\zeta)$-absorbers. Moreover, if we generate $\Theta(n)$ such random tuples with a small implied constant, then most of them will be disjoint to all others and it remains to connect the paths they consist of by means of the Connecting Lemma.

Proof of Proposition 2.5.1. Given to us are $k \geqslant 3$, $\alpha, \beta > 0$, an odd integer $\ell \geqslant 3$, and $\mu = \frac{1}{10k}\left(\frac{\alpha}{2}\right)^{2^{k-3}+1}$. Let $\zeta = \zeta(\alpha, k) > 0$ and $\xi = \xi(\alpha, k) > 0$ be the constants supplied by Lemma 2.5.5, let $\vartheta = \vartheta(k, \alpha, \beta, \ell, \zeta)$ be provided by Proposition 2.3.3, define an auxiliary constant by

$$\gamma = \min\left\{\frac{\xi}{48k^2M^2}, \frac{\vartheta}{8kM^2}\right\}, \quad \text{where} \quad M = 4^{k-2}k\ell \geqslant 12k\,, \tag{2.5.4}$$

and finally set

$$\vartheta_\star = 4kM\gamma\,.$$

We contend that ζ and $\vartheta_\star$ have the desired properties.

To verify this we consider a k-uniform $(\alpha, \beta, \ell, \mu)$-constellation Ψ on n vertices, where n is sufficiently large, as well as an arbitrary subset $\mathcal{R} \subseteq V(\Psi)$ whose size is at most $\vartheta_\star^2 n$. Let

$$t = 2k^2 + k < 3k^2$$

be the length of our absorbers. Since the desired absorbing path needs to be disjoint to $\mathcal{R}$, only the absorbers avoiding $\mathcal{R}$ are relevant in the sequel. For every k-tuple $\vec{a} \in V(\Psi)^k$ consisting of distinct vertices we denote the collection of appropriate absorbers by

$$\mathscr{A}(\vec{a}) = \left\{\vec{A} \in (V(\Psi) \smallsetminus \mathcal{R})^t \colon \vec{A} \text{ is an } (\vec{a}, \zeta)\text{-absorber}\right\}.$$

Lemma 2.5.5 tells us that the total number of $(\vec{a},\zeta)$-absorbers is at least ξn^t and by subtracting those which meet $\mathcal{R}$ we obtain

$$|\mathscr{A}(\vec{a})| \geqslant \xi n^t - t|\mathcal{R}|n^{t-1} \geqslant (\xi - t\vartheta_\star^2)n^t \geqslant \frac{\xi}{2}n^t\,. \tag{2.5.5}$$

Let

$$\mathscr{A} = \bigcup\big\{\mathscr{A}(\vec{a})\colon \vec{a} \in V(\Psi)^k \text{ consists of } k \text{ distinct vertices}\big\} \subseteq \big(V(\Psi) \smallsetminus \mathcal{R}\big)^t$$

be the set of all relevant absorbers. The probabilistic argument we have been alluding to earlier leads to the following result.

Claim 2.5.6. *There is a set $\mathscr{B} \subseteq \mathscr{A}$ of mutually disjoint absorbers of size $|\mathscr{B}| \leqslant 2\gamma n$ satisfying $|\mathscr{A}(\vec{a}) \cap \mathscr{B}| \geqslant \vartheta_\star^2 n$ for every k-tuple $\vec{a} \in V(\Psi)^k$ consisting of distinct vertices.*

Proof. Let $\mathscr{A}_p \subseteq \mathscr{A}$ be a random subset including every absorber in $\mathscr{A}$ independently with probability $p = \gamma n^{1-t}$. As $|\mathscr{A}_p|$ is binomially distributed with expectation $p|\mathscr{A}| \leqslant pn^t = \gamma n$, Markov's inequality yields

$$\mathbb{P}\big(|\mathscr{A}_p| \geqslant 2\gamma n\big) \leqslant \mathbb{P}\big(|\mathscr{A}_p| \geqslant 2p|\mathscr{A}|\big) \leqslant \frac{1}{2}\,. \tag{2.5.6}$$

Next we observe that the set

$$\big\{\{\vec{A}, \vec{A'}\} \in \mathscr{A}^{(2)}\colon \vec{A} \text{ and } \vec{A'} \text{ share a vertex}\big\}$$

of *overlapping pairs of absorbers* has at most the cardinality $t^2 n^{2t-1}$. So the expected size of its intersection with $\mathscr{A}_p^{(2)}$ is at most $p^2 t^2 n^{2t-1} = \gamma^2 t^2 n$. Since

$$\gamma t \leqslant 3k^2\gamma \leqslant \tfrac{1}{4}\vartheta_\star\,,$$

a further application of Markov's inequality reveals

$$\mathbb{P}\Big(\big|\big\{\{\vec{A}, \vec{A'}\} \in \mathscr{A}_p^{(2)}\colon\ \vec{A} \text{ and } \vec{A'} \text{ share a vertex}\big\}\big| \geqslant \tfrac{1}{4}\vartheta_\star^2 n\Big) \leqslant \frac{1}{4}\,. \tag{2.5.7}$$

Finally, for every k-tuple $\vec{a} \in V(\Psi)^k$ of distinct vertices the random variable $|\mathscr{A}_p \cap \mathscr{A}(\vec{a})|$ is binomially distributed with expectation $p|\mathscr{A}(\vec{a})|$. By (2.5.5) we know that

$$p|\mathscr{A}(\vec{a})| \geqslant \frac{1}{2}\gamma\xi n \geqslant 24k^2M^2\gamma^2 n = \tfrac{3}{2}\vartheta_\star^2 n$$

and, therefore, Chernoff's inequality yields

$$\mathbb{P}\big(|\mathscr{A}_p \cap \mathscr{A}(\vec{a})| \leqslant \tfrac{5}{4}\vartheta_\star^2 n\big) \leqslant e^{-\Omega(n)} < \frac{1}{4n^k}\,.$$

As there are at most n^k possibilities for $\vec{a}$, the union bound leads to

$$\mathbb{P}\big(|\mathscr{A}_p \cap \mathscr{A}(\vec{a})| \leqslant \tfrac{5}{4}\vartheta_\star^2 n \text{ holds for some } \vec{a}\big) < \frac{1}{4}\,. \tag{2.5.8}$$

Taken together, the probabilities estimated in (2.5.6) – (2.5.8) amount to less than 1. Thus there exists a deterministic set $\mathscr{B}_\star \subseteq \mathscr{A}$ of size $|\mathscr{B}_\star| \leqslant 2\gamma n$ containing at most $\frac{1}{4}\vartheta_\star^2 n$ pairs of overlapping absorbers and satisfying $|\mathscr{B}_\star \cap \mathscr{A}(\vec{a})| \geqslant \frac{5}{4}\vartheta_\star^2 n$ for all k-tuples $\vec{a} \in V(\Psi)^k$ of distinct vertices.

Now it suffices to check that a maximal subcollection $\mathscr{B} \subseteq \mathscr{B}_\star$ of mutually disjoint absorbers has the desired properties. The upper bound $|\mathscr{B}| \leqslant |\mathscr{B}_\star| \leqslant 2\gamma n$ is clear and due to $|\mathscr{B}_\star \smallsetminus \mathscr{B}| \leqslant \frac{1}{4}\vartheta_\star^2 n$ we have

$$|\mathscr{B} \cap \mathscr{A}(\vec{a})| \geqslant \tfrac{5}{4}\vartheta_\star^2 n - \tfrac{1}{4}\vartheta_\star^2 n = \vartheta_\star^2 n$$

for every $\vec{a}$. □

It remains to connect the absorbers we have just selected into a path. Recall that every member of $\mathscr{B}$ possesses $k+1$ pre-absorptions paths introduced in the last paragraph of Subsection 2.5.1. Each of these paths has at most $2k$ vertices, starts with a ζ-rightconnectable $(k-1)$-tuple, and ends with a ζ-leftconnectable $(k-1)$-tuple. In fact, most of the pre-absorptions paths even have ζ-connectable end-tuples (see Definition 2.5.2 *(d)*).

Setting $r = (k+1)|\mathscr{B}| \leqslant 4k\gamma n$, let $P_1, \ldots, P_r$ be the pre-absorption paths of the absorbers in $\mathscr{B}$ enumerated in such a way that the end-tuples of P_1 and P_r are ζ-connectable. We shall construct our absorbing path P_A to be of the form

$$P_A = P_1 C_1 P_2 C_2 \cdots P_{r-1} C_{r-1} P_r\,,$$

where $C_1, \ldots, C_{r-1}$ are connections that will be provided by Proposition 2.3.3. Since we intend to use the Connecting Lemma with $i = 0$, each of these connections is going to have

$$f = f(k, 0, \ell) = [4^{k-3}(2\ell + 4) - 2]k \leqslant M - 2k$$

vertices, which will yield

$$|V(P_A)| \leqslant r\big(2k + (M - 2k)\big) = rM \leqslant 4kM\gamma n\,. \tag{2.5.9}$$

We will determine the connections $C_1, \ldots, C_{r-1}$ one by one. When choosing C_j for some $j \in [r-1]$, the Connecting Lemma (Proposition 2.3.3) offers us at least ϑn^f possible ways to connect P_j with P_{j+1} by means of a path with f inner vertices. As we need to avoid both the already constructed parts of P_A and the set $\mathcal{R}$, there are at most

$$f(|\mathcal{R}| + 4kM\gamma n)n^{f-1} < (M\vartheta_\star^2 + 4kM^2\gamma)n^f \overset{(2.5.4)}{<} 8kM^2\gamma n^f \overset{(2.5.4)}{\leqslant} \vartheta n^f$$

potential connections we cannot use, and thus the choice of C_j is indeed possible. This concludes the description of the construction of P_A and it remains to check that the path we just defined has all required properties.

Condition *(i)* follows from (2.5.9) and *(ii)* is guaranteed by our choice of the enumeration $P_1, \ldots, P_r$. For the proof of *(iii)* we consider any set $Z \subseteq V(\Psi) \smallsetminus V(P_A)$ satisfying $|Z| \leqslant 2\vartheta_\star^2 n$ and $|Z| \equiv 0 \pmod{k}$. Let $\vec{a}_1, \ldots, \vec{a}_z \in V(\Psi)^k$ with $z = \frac{|Z|}{k} \leqslant \vartheta_\star^2 n$ be disjoint k-tuples with the property that every vertex from Z occurs in exactly one of them. By Claim 2.5.6 we can find distinct absorbers $\vec{A}_1, \ldots, \vec{A}_z \in \mathscr{B}$ such that $\vec{A}_j$ is a $(\vec{a}_j, \zeta)$-absorber for every $j \in [z]$. It remains to utilise these absorbers one by one. $\square$

2.6 Covering

The aim of this section is to prove that under natural assumptions on the parameters almost all vertices of every large k-uniform $(\alpha, \beta, \ell, \mu)$-constellation can be covered by long paths whose first and last $(k-1)$-tuples are connectable. Before formulating the precise statements let us give an overview of the argument, which will proceed by induction on k.

In the induction step from $k-1$ to k we study a largest possible collection $\mathscr{C}$ of mutually vertex-disjoint M-vertex paths with connectable end-tuples and we denote the set of currently uncovered vertices by U. If U is not small enough already, i.e., if $|U| = \Omega(|V(\Psi)|)$, then we partition $V(\Psi)$ into sets of size M, the so-called blocks, such that the vertex set of each path in $\mathscr{C}$ is one such block. Next, we show by probabilistic arguments that there is a special selection of M blocks, called a useful society below, such that their union S has the property that for 'many' vertices $u \in U$ the induction hypothesis applies to $\Psi_u[S]$. For such vertices u we can then find $M+1$ (actually even more) long disjoint $(k-1)$-uniform paths in $\Psi_u[S]$ starting and ending with connectable $(k-2)$-tuples.

In fact, for some still not too small set $U'' \subseteq U'$ these paths will coincide for all $u \in U''$, meaning that inserting vertices from U'' at every k^{th} position will yield $M+1$ paths in Ψ with connectable end-tuples (see Figure 2.6.2). This allows us to take the original paths contained in S out of $\mathscr{C}$ and to add the newly constructed paths instead, thus increasing the size of $\mathscr{C}$. The following covering principle lies at the heart of this inductive argument.

Definition 2.6.1. *For $k \geqslant 3$ the statement $\heartsuit_k$ asserts that given $\alpha, \beta, \vartheta_\star > 0$ and an odd integer $\ell \geqslant 3$ there exists a constant $\zeta_{\star\star} > 0$ such that for every $M_0 \in \mathbb{N}$ there exist a natural number $M \geqslant M_0$ with $M \equiv -1 \pmod{k}$ and the following property:*

For every sufficiently large k-uniform $(\alpha, \beta, \ell, \frac{4\alpha}{17^k})$-constellation Ψ we can cover all but at most $\vartheta_\star^2|V(\Psi)|$ vertices by mutually vertex-disjoint M-vertex paths whose first and last $(k-1)$-tuples are $\zeta_{\star\star}$-connectable.

For the base case $k = 3$ we quote [92, Lemma 2.14]. One needs to be a little bit careful here, because [92] uses a slightly different notion of $\zeta_{\star\star}$-connectable pairs in 3-uniform hypergraphs. However, every pair that is $\zeta_{\star\star}$-connectable in the sense of [92] is $\zeta_{\star\star}$-connectable in the sense of Definition 2.2.16 as well and, therefore, [92, Lemma 2.14] is strictly stronger than $\heartsuit_3$.

Fact 2.6.2. *The assertion $\heartsuit_3$ holds.* $\square$

There is one issue with the inductive proof of $\heartsuit_k$ sketched above: when applying the induction hypothesis to a $(k-1)$-uniform constellation of the form $\Psi_u[S]$, where S is the vertex set of a useful society, we would prefer to get a covering of almost all vertices in S by paths of length $\Omega(\sqrt{|S|})$ rather than $\Omega(1)$, but prima facie $\heartsuit_{k-1}$ does not seem to deliver this. For this reason we also have to deal with the following statement capable of providing coverings by very long paths.

Definition 2.6.3. *For $k \geqslant 3$ the covering principle $\spadesuit_k$ asserts that given $\alpha, \beta, \xi > 0$ and an odd integer $\ell \geqslant 3$, there exists an infinite arithmetic progression $P \subseteq k\mathbb{N}$ with the following property.*

If Ψ is a k-uniform $(\alpha, \beta, \ell, \frac{\alpha}{17^k})$-constellation, $M \in P$, and $\mathfrak{B} \subseteq V(\Psi)^k$ is a collection of ξ-bridges in Ψ with $|\mathfrak{B}| \geqslant \xi|V(\Psi)|^k$, then all but at most $\lfloor \xi|V(\Psi)| \rfloor + M$ vertices of Ψ can be covered with mutually disjoint M-vertex paths starting and ending with bridges from $\mathfrak{B}$.

Observe that for a fixed k-uniform $(\alpha, \beta, \ell, \frac{\alpha}{17^k})$-constellation Ψ we can apply $\spadesuit_k$ with every $M \in P$. For a larger value of M we have to cover fewer vertices, but, on the other hand, we need to cover them with longer paths. Thus there is no obvious monotonicity in M.

Now we plan to establish the implication $\heartsuit_{k-1} \Rightarrow \spadesuit_{k-1} \Rightarrow \heartsuit_k$, thus decomposing the induction step of the proof of $\heartsuit_k$ into two simpler tasks. They will be treated in Lemma 2.6.4 and Lemma 2.6.9, respectively.

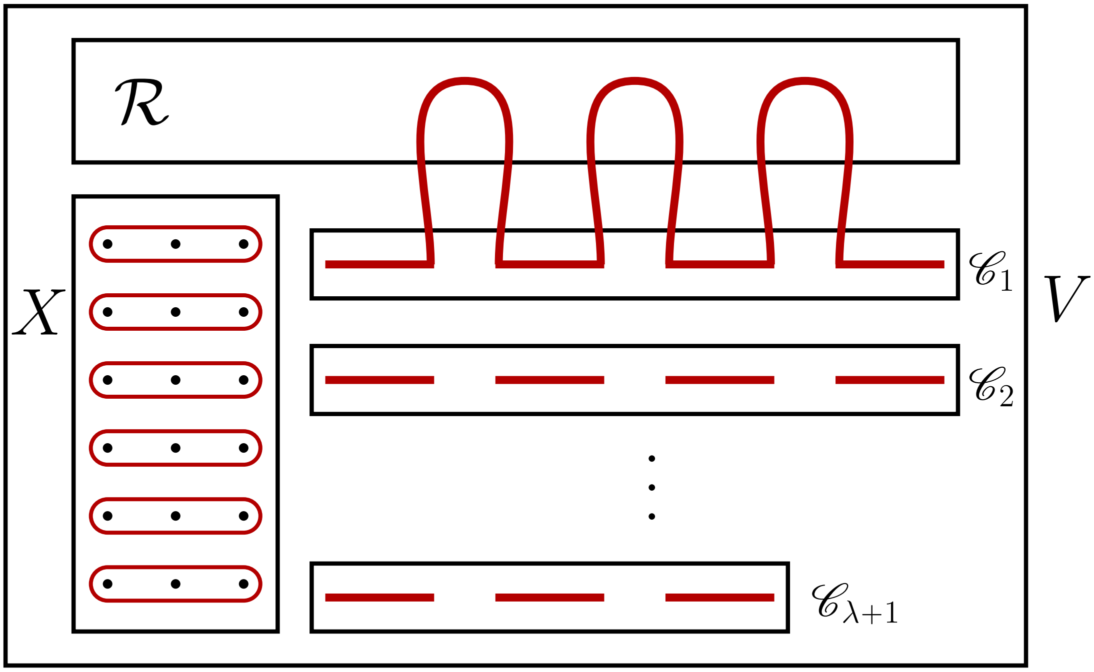


Figure 2.6.1: The case $k = 3$ of Lemma 2.6.4. The set X of vertices is reserved for bridges.

Lemma 2.6.4. *If $k \geqslant 3$ and $\heartsuit_k$ holds, then so does $\spadesuit_k$.*

The idea behind the proof of this implication is the following (see Figure 2.6.1). Given an appropriate constellation Ψ, our first step is to take out a reservoir set $\mathcal{R}$. Next we decide which bridges from $\mathfrak{B}$ are going to appear at the ends of the paths we are supposed to construct. After these choices are made, we apply $\heartsuit_k$ to the constellation obtained from Ψ by removing $\mathcal{R}$ and the vertices reserved for the bridges, thus getting a covering of almost all remaining vertices with 'short' paths. Now we partition the set of these paths into groups of size p, where p denotes an arbitrary natural number. For each group we connect all its paths through the reservoir. Moreover, we connect the ends of the resulting paths to some of the bridges that have been put aside. In this manner we obtain a covering of almost all vertices of Ψ with longer paths, whose precise length depends linearly on p. Thus by varying p we can reach an arithmetic progression of possible lengths for the paths in the new covering.

Proof of Lemma 2.6.4. Let α, β, $\xi > 0$ and an odd integer $\ell \geqslant 3$ be given. Choose some auxiliary constants obeying the hierarchy

$$\alpha, \beta, \xi, k^{-1}, \ell^{-1} \gg \vartheta_\star \gg \zeta_{\star\star} \gg \vartheta_{\star\star} \gg M^{-1} \gg n_0^{-1},$$

where M is an integer with $M \equiv -1 \pmod{k}$.

We contend that

$$P = \{M' \in k\mathbb{N} \colon M' > n_0 \text{ and } M' \equiv f(k,0,\ell) + 2k \pmod{M + f(k,0,\ell)}\}$$

has the property demanded by $\spadesuit_k$.

By Definition 2.3.2 the number $f(k,0,\ell)$ is divisible by k and, consequently, P is indeed an infinite arithmetic progression. Now let Ψ be a k-uniform $(\alpha, \beta, \ell, \frac{\alpha}{17^k})$-constellation with n vertices, let $M' \in P$ be arbitrary, and let $\mathfrak{B} \subseteq V(\Psi)^k$ be a set of ξ-bridges in Ψ with $|\mathfrak{B}| \geqslant \xi |V(\Psi)|^k$. We are to cover all but at most $\xi|V(\Psi)| + M'$ vertices of Ψ by mutually disjoint M'-vertex paths starting and ending with bridges from $\mathfrak{B}$. If $|V(\Psi)| \leqslant M'$, then the empty set is such a collection of paths. Thus, we may assume that $|V(\Psi)| > M' > n_0$.

Let $\mathcal{R} \subseteq V(\Psi)$ with $|\mathcal{R}| \leqslant \vartheta_\star n$ be the reservoir set provided by Proposition 2.4.1 with $\vartheta_\star$, $\frac{\zeta_{\star\star}}{2}$ here in place of ξ, $\zeta_{\star\star}$ there. For later use we record that due to $\vartheta_{\star\star} \ll \vartheta_\star, k^{-1}, \ell^{-1}$ the case $i = 0$ of Corollary 2.4.2 yields:

($\star$) If $\mathcal{R}' \subseteq \mathcal{R}$ is an arbitrary set with $|\mathcal{R}'| \leqslant \vartheta_{\star\star}^2 |V(\Psi)|$, the $(k-1)$-tuple $\vec{a} \in V(\Psi)^{k-1}$ is $\frac{\zeta_{\star\star}}{2}$-leftconnectable, and $\vec{b} \in V(\Psi)^{k-1}$ is $\frac{\zeta_{\star\star}}{2}$-rightconnectable and disjoint to $\vec{a}$, then there is an $\vec{a}$-$\vec{b}$-path through $\mathcal{R} \smallsetminus \mathcal{R}'$ with $f(k,0,\ell)$ inner vertices.

Let $b_1, \dots, b_r$ be a maximal sequence of bridges from $\mathfrak{B}$ that are mutually disjoint and disjoint to $\mathcal{R}$. Since the selected bridges and $\mathcal{R}$ together involve $kr + |\mathcal{R}|$ vertices, the maximality implies

$$k(kr + |\mathcal{R}|)|V(\Psi)|^{k-1} \geqslant |\mathfrak{B}| \geqslant \xi |V(\Psi)|^k\,,$$

whence

$$r \geqslant \frac{(\xi - k\vartheta_\star)|V(\Psi)|}{k^2} \geqslant \vartheta_\star |V(\Psi)|\,. \tag{2.6.1}$$

Set $x = \lfloor \vartheta_\star |V(\Psi)| \rfloor$ and let X be the set of vertices constituting $b_1, \dots, b_x$. Lemma 2.2.36 reveals that $\Psi' = \Psi - (X \cup \mathcal{R})$ is an $\left(\frac{\alpha}{2}, \frac{\beta}{2}, \ell, \frac{2\alpha}{17^k}\right)$-constellation. Therefore, the principle $\heartsuit_k$ yields a family $\mathscr{C}$ of disjoint M-vertex paths in Ψ' which together cover all but at most $\vartheta_\star^2 |V(\Psi')|$ vertices of Ψ' and whose end-tuples are $\zeta_{\star\star}$-connectable in Ψ'. For later use we remark that owing to Fact 2.2.18 the end-tuples of the paths in $\mathscr{C}$ are $\frac{\zeta_{\star\star}}{2}$-connectable in Ψ.

By the definition of P there is a natural number p such that

$$M' = \big(M + f(k,0,\ell)\big)p + f(k,0,\ell) + 2k\,.$$

Fix an arbitrary partition $\mathscr{C} = \mathscr{C}_1 \cup \ldots \cup \mathscr{C}_{\lambda+1}$ with $|\mathscr{C}_1| = \cdots = |\mathscr{C}_\lambda| = p > |\mathscr{C}_{\lambda+1}|$.

Now we declare our strategy for constructing vertex-disjoint paths $P_1, \dots, P_\lambda \subseteq H(\Psi)$ witnessing the conclusion of $\spadesuit_k$. For every $j \in [\lambda]$ we first intend to form a path P'_j by connecting the p paths in $\mathscr{C}_j$ through the reservoir $\mathcal{R}$. Subsequently, we plan to derive P_j from P'_j by connecting its ends with two bridges from the list $b_1, \dots, b_x$, say with b_{2j-1} and b_{2j}. For all $p+1$ connections required for this construction of P_j, we want to appeal to $(\star)$. Clearly, if the paths $P_1, \dots, P_\lambda$ can be constructed, then each of them will consist of M' vertices.

Altogether, we are aiming for $(p+1)\lambda$ connections that require a total number of

$$(p+1)f(k,0,\ell)\lambda$$

vertices from the reservoir. If this number is less than $\vartheta_{\star\star}^2 n$, then repeated applications of $(\star)$ allow us to choose our connections disjointly. Since $M \gg \vartheta_{\star\star}^{-1} \gg k, \ell$, we have indeed

$$(p+1)f(k,0,\ell)\lambda \leqslant 2p \cdot 4^k k\ell \cdot \frac{|V(\Psi)|}{Mp} = \frac{2 \cdot 4^k k\ell |V(\Psi)|}{M} < \vartheta_{\star\star}^2 |V(\Psi)|\,.$$

Similarly,

$$2\lambda \leqslant \frac{2|V(\Psi)|}{Mp} \leqslant \frac{2|V(\Psi)|}{M} \leqslant \vartheta_\star |V(\Psi)|$$

proves that we have sufficiently many bridges at our disposal.

Altogether, the vertex-disjoint paths $P_1, \dots, P_\lambda \subseteq H(\Psi)$ can indeed be constructed. The number of vertices of Ψ they fail to cover can be bounded from above by

$$\begin{aligned}|X| + |\mathcal{R}| + \Big|V(\Psi') \smallsetminus \bigcup_{P \in \mathscr{C}} V(P)\Big| + \Big|\bigcup_{P \in \mathscr{C}_{\lambda+1}} V(P)\Big| &\leqslant kx + \vartheta_\star |V(\Psi)| + \vartheta_\star^2 |V(\Psi)| + Mp \\ &\leqslant Mp + \big((k+1)\vartheta_\star + \vartheta_\star^2\big)|V(\Psi)| \\ &\leqslant M' + \xi |V(\Psi)|\,,\end{aligned}$$

which concludes the proof of $\spadesuit_k$. □

The proof of our next result involves some probabilistic arguments based on the following consequence of Janson's inequality (see [92, Corollary A.3]).

Lemma 2.6.5. *Let $m \geqslant k$ and M be positive integers, and let $\eta \in (0, \frac{1}{2k})$. Suppose that V is a finite set and that*

$$V = B_1 \cup \dots \cup B_\nu \cup Z$$

is a partition with $|B_1| = \dots = |B_\nu| = M < \eta|V|$, $|Z| < \eta|V|$, and $\nu \geqslant m$. Let

$\mathscr{S} \subseteq \{B_1, \ldots, B_\nu\}$ *be an* m*-element subset chosen uniformly at random and set* $S = \bigcup \mathscr{S}$. *Further, let* ξ *be a real number with* $\max(8k^2\eta, 16k^2/m) < \xi < 1$.

(*a*) *If* $Q \subseteq V^k$ *has size* $|Q| = d|V|^k$, *then*

$$\mathbb{P}\big(\big||Q \cap S^k| - d(Mm)^k\big| \geqslant \xi(Mm)^k\big) \leqslant 12\sqrt{m}\exp\left(-\frac{\xi^2 m}{48k^{2k+2}}\right).$$

(*b*) *Similarly, if* G *denotes a* k*-uniform hypergraph with vertex set* V *and* $d|V|^k/k!$ *edges, then*

$$\mathbb{P}\big(\big|e_G(S) - d(Mm)^k/k!\big| \geqslant \xi(Mm)^k/k!\big) \leqslant 12\sqrt{m}\exp\left(-\frac{\xi^2 m}{48k^{2k+2}}\right). \qquad \square$$

This has the following consequence on random subconstellations.

Lemma 2.6.6. *Given* $k \geqslant 2$, $\alpha, \beta, \mu, \xi > 0$, *and an odd integer* $\ell \geqslant 3$ *there exists a natural number* M_0 *such that the following holds for every* $M \geqslant M_0$. *If* Ψ *is a sufficiently large* k*-uniform* $(\alpha, \beta, \ell, \mu)$*-constellation,*

$$V(\Psi) = B_1 \cup \ldots \cup B_\nu \cup B'$$

is a partition with $|B_1| = \ldots = |B_\nu| = M$ *and* $|B'| < 2M$, *and* $\mathfrak{B} \subseteq V(\Psi)^k$ *is a set of* ξ*-bridges in* Ψ *of size* $|\mathfrak{B}| \geqslant \xi|V(\Psi)|^k$, *then there are at least* $\frac{3}{4}\binom{\nu}{M}$ *sets* $\mathscr{S} \subseteq \{B_1, \ldots, B_\nu\}$ *of size* M *such that their union* $S = \bigcup \mathscr{S}$ *has the properties that* $\Psi[S]$ *is a* $(\frac{\alpha}{2}, \frac{\beta}{2}, \ell, 2\mu)$*-constellation and*

$$\mathfrak{B}_\star = \big\{\vec{x} \in \mathfrak{B} \cap S^k \colon \vec{x} \text{ is a } \tfrac{\xi}{2}\text{-bridge in } \Psi[S]\big\}$$

has at least the size $|\mathfrak{B}_\star| \geqslant \frac{\xi}{2}|S|^k$.

Proof. Let $M_0 \gg \alpha^{-1}, \beta^{-1}, \mu^{-1}, \xi^{-1}, k, \ell$ be sufficiently large. We call the sets $B_1, \ldots, B_\nu$ *blocks*. Choose a set $\mathscr{S} \subseteq \{B_1, \ldots, B_\nu\}$ of M blocks uniformly at random among all $\binom{\nu}{M}$ possibilities. We shall prove that the probability that $S = \bigcup \mathscr{S}$ fails to have the desired properties is at most $\exp\big(-\Omega(M)\big)$, where the implied constant only depends on α, β, μ, ξ, k, and ℓ. Hence, by choosing M_0 sufficiently large, this probability can be pushed below $\frac{1}{4}$, as desired. It will be convenient to set $V' = V \smallsetminus B'$. For $y \in V'$ we denote the unique block containing y by B_y.

Claim 2.6.7. *The event that* $\Psi[S]$ *fails to be a* $(\frac{\alpha}{2}, \frac{\beta}{2}, \ell, 2\mu)$*-constellation has at most the probability* $\exp\big(-\Omega(M)\big)$.

Proof. We begin by estimating the probability of the unfortunate event $\mathfrak{U}$ that $\Psi[S]$ fails to be a $(\frac{\alpha}{2}, 2\mu)$-constellation. For an arbitrary set $x \in (V')^{(k-2)}$ we define

$$\mathscr{Z}_x = \{B_y \colon y \in x\}\,, \quad t_x = |\mathscr{Z}_x| \in [k-2]\,, \text{ and } \quad Z_x = \bigcup \mathscr{Z}_x\,.$$

Further, we consider the conditional probabilities

$$P_1(x) = \mathbb{P}\left(e_{\Psi_x}(S \smallsetminus Z_x) < \left(\frac{5}{9} + \frac{2\alpha}{3}\right)\frac{(M-t_x)^2M^2}{2} \;\middle|\; x \in S^{(k-2)}\right),$$
$$P_2(x) = \mathbb{P}\left(\left|V(R_x^\Psi[S])\right| < \left(\frac{2}{3} + \frac{\alpha}{3}\right)(M-t_x)M \;\middle|\; x \in S^{(k-2)}\right)\,,$$

and

$$P_3(x) = \mathbb{P}\left(e_{\Psi_x[S]}\left(V(R_x^\Psi[S]), S \smallsetminus V(R_x^\Psi[S])\right) > 2\mu(M-t_x)^2M^2 \;\middle|\; x \in S^{(k-2)}\right)$$

and observe that

$$\mathbb{P}(\mathfrak{U}) \leqslant \sum_{x \in (V')^{(k-2)}} \mathbb{P}(x \in S^{(k-2)})\big(P_1(x) + P_2(x) + P_3(x)\big)\,. \tag{2.6.2}$$

So if we manage to prove

$$P_1(x), P_2(x), P_3(x) \leqslant \exp\big(-\Omega(M)\big)\,, \tag{2.6.3}$$

then

$$\mathbb{P}(\mathfrak{U}) \leqslant (M^2)^{k-2}\exp\big(-\Omega(M)\big) \leqslant \exp\big(-\Omega(M)\big) \tag{2.6.4}$$

will follow. Thus our next goal is to establish (2.6.3).

To this end, we will repeatedly apply Lemma 2.6.5 with

$$M - t_x\,, \frac{kM}{n}\,, B' \cup Z_x\,, \nu - t_x\,, \text{ and } \min\{\tfrac{\alpha}{6}, \mu\}$$

here in place of

$$m\,, \eta\,, Z\,, \nu\,, \text{ and } \xi$$

there and relocating the elements of $\mathscr{Z}_x$ to the exceptional set of the partition.

First, the minimum degree condition imposed on $H(\Psi)$ implies that the graph $H(\Psi_x)$ has at least $\left(\frac{5}{9} + \alpha\right)\frac{|V(\Psi)|^2}{2}$ edges. So Lemma 2.6.5 *(b)* applied with 2 and $H(\Psi_x)$ here in place of k and G there yields $P_1(x) \leqslant \exp\bigl(-\Omega(M)\bigr)$.

Second, we know that $|V(R_x^\Psi)| \geqslant \bigl(\frac{2}{3}+\frac{\alpha}{2}\bigr)|V(\Psi)|$, since Ψ is an (α,μ)-constellation. Hence, applying Lemma 2.6.5 *(a)* with 1 and $V(R_x^\Psi)$ here instead of k and Q there entails $P_2(x) \leqslant \exp\bigl(-\Omega(M)\bigr)$.

Lastly, from Ψ being a (α,μ)-constellation it also follows that

$$e_{\Psi_x}\bigl(V(R_x^\Psi), V \smallsetminus V(R_x^\Psi)\bigr) \leqslant \mu|V(\Psi)|^2\,.$$

Hence, Lemma 2.6.5 *(b)* applied to the bipartite subgraph of $H(\Psi_x)$ between $V(R_x^\Psi)$ and its complement tells us that $P_3(x) \leqslant \exp\bigl(-\Omega(M)\bigr)$. This concludes the proof of (2.6.3) and, hence, of (2.6.4). An analogous proof allows us to transfer part *(b)* of Definition 2.2.33 from Ψ to $\Psi[S]$ and we omit the details. □

It remains to prove that the event $|\mathfrak{B}_\star| \geqslant \frac{\xi}{2}|S|^k$ has high probability as well. Here we start with the estimate

$$\mathbb{P}\bigl(|\mathfrak{B}_\star| \leqslant \tfrac{\xi}{2}|S|^k\bigr) \leqslant \mathbb{P}\bigl(|\mathfrak{B} \cap S^k| \leqslant \tfrac{\xi}{2}|S|^k\bigr) + \mathbb{P}(\neg\mathfrak{E})\,,$$

where $\mathfrak{E}$ denotes the event that every ξ-bridge $\vec{x} \in \mathfrak{B} \cap S^k$ is a $\frac{\xi}{2}$-bridge in $\Psi[S]$. Another application of Lemma 2.6.5 *(a)* tells us that the first summand is at most $\exp\bigl(-\Omega(M)\bigr)$ and thus it remains to prove that

$$\mathbb{P}(\neg\mathfrak{E}) \leqslant \exp\bigl(-\Omega(M)\bigr)\,. \tag{2.6.5}$$

Towards this goal we analyse how connectability transfers to $\Psi[S]$.

Claim 2.6.8. *If $k' \in [k-1]$, $z, z' \in (V')^{(k-1-k')}$, and $\vec{x} \in \bigl(V' \smallsetminus (z \cup z')\bigr)^{k'}$ is a ξ-leftconnectable tuple in Ψ_z, then*

$$\mathbb{P}\bigl(\vec{x} \text{ fails to be } \tfrac{\xi}{2}\text{-leftconnectable in } \Psi_z[S] \mid \vec{x} \in S^{k'} \text{ and } z' \subseteq S\bigr) \leqslant \exp\bigl(-\Omega(M)\bigr)\,.$$

Proof. We argue by induction on k'. In the base case $k' = 1$ the probability under consideration vanishes. This is because a 1-tuple $\vec{x} = (x)$ is ξ-leftconnectable in Ψ_z if and only if $x \in V(R_z^\Psi)$. Moreover, if $x \in S \smallsetminus z$, then (x) is $\frac{\xi}{2}$-leftconnectable in $\Psi_z[S]$ if and only if $x \in R_z^{\Psi[S]}$. Due to $R_z^{\Psi[S]} = R_z^\Psi[S]$ these two statements are equivalent to each other.

For the induction step from $k'-1$ to k' we write $\vec{x} = (x_1, \dots, x_{k'})$ and recall that the

ξ-leftconnectability of $\vec{x}$ in Ψ_z means that $|U| \geqslant \xi|V(\Psi_z)|$, where

$$U = \big\{u \in V(\Psi_z)\colon x_1 \cdots x_{k'} u \in E(\Psi_z) \text{ and } (x_2, \ldots, x_{k'}) \text{ is } \xi\text{-leftconnectable in } \Psi_{zu}\big\}\,.$$

Assuming $\vec{x} \in S^{k'}$ the analogous set whose size decides whether $\vec{x}$ is $\frac{\xi}{2}$-leftconnectable in $\Psi_z[S]$ either contains $U \cap S$ as a subset, or it does not. Accordingly, if $\vec{x}$ fails to be $\frac{\xi}{2}$-leftconnectable in $\Psi_z[S]$, then either $|U \cap S| \leqslant \frac{\xi}{2}|V(\Psi_z[S])|$ or the event $\mathfrak{A}$ that for some $u \in S \cap U$ the $(k'-1)$-tuple $(x_2, \ldots, x_{k'})$ fails to be $\frac{\xi}{2}$-leftconnectable in $\Psi_{zu}[S]$ occurs. For this reason, it suffices to prove

$$\mathbb{P}\big(|U \cap S| \leqslant \tfrac{\xi}{2}|S| \mid \vec{x} \in S^{k'} \text{ and } z' \subseteq S\big) \leqslant \exp\big(-\Omega(M)\big) \tag{2.6.6}$$

$$\text{and} \quad \mathbb{P}\big(\mathfrak{A} \mid \vec{x} \in S^{k'} \text{ and } z' \subseteq S\big) \leqslant \exp\big(-\Omega(M)\big)\,. \tag{2.6.7}$$

Now (2.6.6) follows in the usual way from Lemma 2.6.5 *(a)*. To prove (2.6.7) we observe that the induction hypothesis yields

$$\mathbb{P}\big((x_2, \ldots, x_{k'}) \text{ fails to be } \tfrac{\xi}{2}\text{-leftconnectable in } \Psi_{zu}[S] \mid (x_2, \ldots, x_{k'}) \in S^{k'-1}\,, \\ \text{and } (z' \cup \{x_1\}) \subseteq S\big) \leqslant \exp\big(-\Omega(M)\big)$$

for every $u \in U$, whence

$$\begin{aligned} \mathbb{P}(\mathfrak{A} \mid \vec{x} \in S^{k'} \text{ and } z' \subseteq S) &\leqslant \sum_{u \in U} \mathbb{P}(u \in S) \exp\big(-\Omega(M)\big) \\ &\leqslant M^2 \exp\big(-\Omega(M)\big) \leqslant \exp\big(-\Omega(M)\big)\,. \end{aligned}$$ □

By applying the case $k' = k-1$ of Claim 2.6.8 to all ξ-leftconnectable $(k-1)$-tuples in Ψ we obtain

$$\mathbb{P}\big(\text{Some } \vec{x} \in S^{k-1} \text{ that is } \xi\text{-leftconnectable in } \Psi \\ \text{fails to be } \tfrac{\xi}{2}\text{-leftconnectable in } \Psi[S]\big) \leqslant \exp\big(-\Omega(M)\big)\,.$$

By symmetry the same holds for rightconnectability as well and, therefore,

$$\mathbb{P}\big(\text{Some } \xi\text{-bridge } \vec{x} \in S^k \text{ fails to be a } \tfrac{\xi}{2}\text{-bridge in } \Psi[S]\big) \leqslant \exp\big(-\Omega(M)\big)\,.$$

In other words, we have thereby proved (2.6.5) and, hence, Lemma 2.6.6. □

The next lemma shows how to ascend from $(k-1)$-uniform coverings to k-uniform coverings.

Lemma 2.6.9. *For every $k \geqslant 4$ the covering principle $\spadesuit_{k-1}$ implies $\heartsuit_k$.*

Proof. Let $\alpha, \beta, \vartheta_\star > 0$, and an odd integer $\ell \geqslant 3$ be given. Without loss of generality we may assume that $\vartheta_\star \ll \alpha, \beta, k^{-1}, \ell^{-1}$. Pick a sufficiently small constant

$$\zeta_{\star\star} \ll \vartheta_\star \,. \tag{2.6.8}$$

The statement $\spadesuit_{k-1}$ applied to $\frac{\alpha}{2}$, $\frac{\beta}{2}$, ℓ, $\frac{\zeta_{\star\star}}{2}$ here in place of α, β, ℓ, ξ there delivers an infinite arithmetic progression $P \subseteq (k-1)\mathbb{N}$. Choose $M \gg \zeta_{\star\star}^{-1}$ such that $\frac{k-1}{k}(M+1) \in P$ and notice that $M \equiv -1 \pmod{k}$ is clear.

Now let Ψ be a $(\alpha, \beta, \ell, \frac{4\alpha}{17^k})$-constellation on n vertices, where n is sufficiently large. We are to prove that all but at most $\vartheta_\star^2|V(\Psi)|$ vertices of Ψ can be covered by vertex-disjoint M-vertex paths starting end ending with $\zeta_{\star\star}$-connectable $(k-1)$-tuples. Let

$$\mathscr{P} = \big\{P \subseteq H(\Psi) \colon P \text{ is a } k\text{-uniform } M\text{-vertex path} \\ \text{whose first and last } (k-1)\text{-tuple is } \zeta_{\star\star}\text{-connectable}\big\}$$

be the collection of all paths that might occur in such a covering, and let $\mathscr{C} \subseteq \mathscr{P}$ be a maximal subcollection of vertex-disjoint paths from $\mathscr{P}$. Further, let

$$U = V(\Psi) \smallsetminus \bigcup_{P \in \mathscr{C}} V(P)$$

be the set of uncovered vertices. We may assume that

$$|U| > \vartheta_\star^2 |V(\Psi)| \,, \tag{2.6.9}$$

since otherwise nothing is left to show. Now roughly speaking the strategy is to find a set $S \subseteq V(\Psi)$ of size M^2 meeting at most M paths from $\mathscr{C}$ such that for 'many' vertices $u \in U$ we can apply $\spadesuit_{k-1}$ to the $(k-1)$-uniform constellation $\Psi_u[S]$, thus getting at least $M+1$ vertex-disjoint paths with $\frac{k-1}{k}(M+1)$ vertices. These paths will agree for many vertices $u \in U$ and can then be augmented to k-uniform paths engendering a contradiction to the maximality of $\mathscr{C}$. In the intended application of $\spadesuit_{k-1}$ we are allowed to specify a set of bridges $\mathfrak{B}$ that we potentially would like to see at the ends of the paths we obtain. Since we ultimately aim at generating paths in $\mathscr{P}$ and, hence, paths starting and ending with $\zeta_{\star\star}$-connectable $(k-1)$-tuples, it seems advisable to let $\mathfrak{B}$ be the set of $\frac{\zeta_{\star\star}}{2}$-bridges in $\Psi_u[S]$ that are $\zeta_{\star\star}$-connectable in Ψ. This choice of $\mathfrak{B}$ is only permissible if $|\mathfrak{B}|$ is sufficiently large (i.e., at least $\frac{\zeta_{\star\star}}{2}|S|^{k-1}$). Our way of ensuring this in sufficiently many cases exploits that for fixed $u \in U$ and a random choice of $S \subseteq V(\Psi)$ Lemma 2.6.6 tells us that the $\zeta_{\star\star}$-bridges in Ψ_u are likely to be $\frac{\zeta_{\star\star}}{2}$-bridges in $\Psi_u[S]$. Thus it suffices to focus on vertices $u \in U$ which are not in the set

$$U_{\text{bad}} = \left\{u \in U\colon \text{at most } \tfrac{1}{20}n^{k-1} \text{ of the } \zeta_{\star\star}\text{-bridges in } \Psi_u \text{ are } \zeta_{\star\star}\text{-connectable in } \Psi\right\}.$$

The next claim states that this set is indeed small.

Claim 2.6.10. *We have $|U_{\text{bad}}| \leqslant 40\zeta_{\star\star}n$.*

Proof. Set

$$\Pi = \Big\{(x_1, \dots, x_{k-1}, u) \in V(\Psi)^{k-1} \times U_{\text{bad}}\colon (x_1, \dots, x_{k-1}) \text{ is a } \zeta_{\star\star}\text{-bridge in } \Psi_u \\ \text{but not } \zeta_{\star\star}\text{-connectable in } \Psi\Big\}.$$

For every $u \in U_{\text{bad}}$ Corollary 2.2.28 tells us that the number of $\zeta_{\star\star}$-bridges $(x_1, \dots, x_{k-1})$ in Ψ_u is at least $\frac{1}{9}(n-1)^{k-1} > \frac{1}{10}n^{k-1}$ and by the definition of U_{bad} at least $\frac{1}{20}n^{k-1}$ among them fail to be $\zeta_{\star\star}$-connectable in Ψ. This proves that

$$|\Pi| \geqslant \frac{1}{20}n^{k-1}|U_{\text{bad}}|.$$

On the other hand, an upper bound on $|\Pi|$ can be obtained as follows. Let Π_{left} be the set of k-tuples in Π for which $(x_1, \dots, x_{k-1})$ fails to be $\zeta_{\star\star}$-leftconnectable and define Π_{right} similarly with respect to rightconnectability. As a $(k-1)$-tuple that is not $\zeta_{\star\star}$-leftconnectable in Ψ can only be a $\zeta_{\star\star}$-bridge in Ψ_u for less than $\zeta_{\star\star}n$ vertices u, we have $|\Pi_{\text{left}}| \leqslant \zeta_{\star\star}n^k$. The same upper bound can be proved for $|\Pi_{\text{right}}|$ and because of $\Pi = \Pi_{\text{left}} \cup \Pi_{\text{right}}$ this yields $|\Pi| \leqslant 2\zeta_{\star\star}n^k$. Combining the two bounds on $|\Pi|$ we obtain indeed $|U_{\text{bad}}| \leqslant 40\zeta_{\star\star}n$. ◻

Because of our choice of $\zeta_{\star\star}$ in (2.6.8) this yields $|U_{\text{bad}}| \leqslant \frac{1}{2}\vartheta_\star^2 n$, which combined with (2.6.9) implies

$$|U \smallsetminus U_{\text{bad}}| \geqslant \frac{1}{2}\vartheta_\star^2 n. \tag{2.6.10}$$

Next we will partition the vertex set into blocks some of which will later be selected randomly for hosting the augmentation of $\mathscr{C}$. Form a partition

$$V(\Psi) = B_1 \mathbin{\dot\cup} \ldots \mathbin{\dot\cup} B_\nu \mathbin{\dot\cup} B', \tag{2.6.11}$$

with $|B_1| = \cdots = |B_\nu| = M > |B'|$, where the first $|\mathscr{C}|$ classes $B_1, \ldots, B_{|\mathscr{C}|}$ are the vertex sets of the paths in the collection $\mathscr{C}$, and $B_{|\mathscr{C}|+1}, \ldots, B_\nu$ are arbitrary disjoint M-sets making (2.6.11) true. The sets $B_1, \ldots, B_\nu$ are called *blocks*. A *society* is a set of M blocks. We point out that

$$\text{if } \mathscr{S} \text{ is a society and } S = \bigcup \mathscr{S}, \text{ then } |S| = M^2. \tag{2.6.12}$$

Definition 2.6.11. *A society $\mathscr{S}$ with $S = \bigcup \mathscr{S}$ is called useful for a vertex $u \in U$ if*

(*1*) $u \notin S$,

(*2*) $\Psi_u[S]$ *is a* $(k-1)$*-uniform* $(\frac{\alpha}{2}, \frac{\beta}{2}, \ell, \frac{\alpha/2}{17^{k-1}})$*-constellation.*

(*3*) *The number of $(k-1)$-tuples in S^{k-1} that are $\frac{\zeta_{\star\star}}{2}$-bridges in $\Psi_u[S]$ and $\zeta_{\star\star}$-connectable in Ψ is at least $\frac{\zeta_{\star\star}}{2}|S|^{k-1}$.*

The next claim explains the naming of useful societies: $\Psi_u[S]$ contains $M+1$ "suitable" paths.

Claim 2.6.12. *If a society $\mathscr{S}$ is useful for $u \in U$ and $S = \bigcup \mathscr{S}$, then there is a collection $\mathscr{W}$ of mutually disjoint $(k-1)$-uniform paths in $\Psi_u[S]$ with the following properties.*

(*i*) *Every path in $\mathscr{W}$ has $\frac{k-1}{k}(M+1)$ vertices.*

(*ii*) *Every path in $\mathscr{W}$ starts and ends with a $(k-1)$-tuple that is $\zeta_{\star\star}$-connectable in Ψ.*

(*iii*) $|\mathscr{W}| \geqslant M+1$.

Proof. By Definition 2.6.11 (*3*) and (2.6.12) the set

$$\Xi = \left\{\vec{e} \in S^{k-1} \colon \vec{e} \text{ is } \zeta_{\star\star}\text{-connectable in } \Psi \text{ and a } \tfrac{\zeta_{\star\star}}{2}\text{-bridge in } \Psi_u[S]\right\}$$

satisfies $|\Xi| \geqslant \frac{\zeta_{\star\star}}{2}(M^2)^{k-1}$. Now we apply $\spadesuit_{k-1}$ to $\Psi_u[S]$, Ξ, $\frac{\zeta_{\star\star}}{2}$, and $\frac{k-1}{k}(M+1)$ here in place of Ψ, $\mathfrak{B}$, ξ, and M there – which is permissible due to the selection of parameters in the beginning of the proof of Lemma 2.6.9.

This application of $\spadesuit_{k-1}$ yields a collection $\mathscr{W}$ of mutually disjoint $(k-1)$-uniform paths in $\Psi_u[S]$ that covers all but at most $\frac{\zeta_{\star\star}}{2}|S| + \frac{k-1}{k}(M+1)$ vertices of S such that each path starts and ends with a bridge from Ξ. Since each bridge in Ξ is a $\zeta_{\star\star}$-connectable tuple in Ψ, it remains to check that $|\mathscr{W}| \geqslant M+1$. Because of $M \gg \zeta_{\star\star}^{-1} \gg k$ we have indeed

$$|\mathscr{W}| \geqslant \frac{(1-\zeta_{\star\star}/2)M^2 - \frac{k-1}{k}(M+1)}{\frac{k-1}{k}(M+1)} \geqslant \frac{(1-\zeta_{\star\star})M(M+1)}{(1-\zeta_{\star\star})M} = M+1\,. \qquad \square$$

Lemma 2.6.6 implies that some society is useful for many vertices.

Claim 2.6.13. *There exists a society $\mathscr{S}$ that is useful for $\frac{2}{3}|U \smallsetminus U_{bad}|$ vertices in $U \smallsetminus U_{bad}$.*

Proof. By double counting it suffices to establish that for every vertex $u \in U \smallsetminus U_{\rm bad}$ at least $\frac{2}{3}$ of all societies are useful. Fix an arbitrary such vertex u and suppose first that $u \notin B'$. Without loss of generality we may assume that $u \in B_\nu$. We plan to apply Lemma 2.6.6 with $(k-1, \frac{\alpha}{4\cdot 17^{k-1}}, \zeta_{\star\star})$ here in place of (k, μ, ξ) there to the $(k-1)$-uniform constellation Ψ_u, the partition

$$V(\Psi_u) = B_1 \charfusion[\mathbin]{\cup}{\cdot} \ldots \charfusion[\mathbin]{\cup}{\cdot} B_{\nu-1} \charfusion[\mathbin]{\cup}{\cdot} (B_\nu \cup B' \smallsetminus \{u\})\,,$$

and the set

$$\mathfrak{B}_u = \left\{\vec{x} \in V(\Psi_u)^{k-1} \colon \vec{x} \text{ is } \zeta_{\star\star}\text{-connectable in } \Psi \text{ and a } \zeta_{\star\star}\text{-bridge in } \Psi_u\right\}.$$

Notice that Fact 2.2.35 tell us that Ψ_u is indeed an $(\alpha, \beta, \ell, \frac{\alpha}{4\cdot 17^{k-1}})$-constellation. Moreover, $u \notin U_{\rm bad}$ implies $|\mathfrak{B}_u| \geqslant \frac{1}{20}n^{k-1} > \zeta_{\star\star}|V(\Psi_u)|^{k-1}$. So all assumptions of Lemma 2.6.6 hold and we conclude that at least $\frac{3}{4}\binom{\nu-1}{M} > \frac{2}{3}\binom{\nu}{M}$ societies are useful for u. The case $u \in B'$ is similar. $\square$

For the remainder of this proof we fix a society $\mathscr{S}$ that is useful for at least $\frac{2}{3}|U \smallsetminus U_{\rm bad}|$ vertices in $U \smallsetminus U_{\rm bad}$ and set $S = \bigcup \mathscr{S}$. Claim 2.6.12 informs us that for every $u \in U$, for which $\mathscr{S}$ is useful, there is a collection $\mathscr{W}_u$ of $M+1$ mutually vertex disjoint $(k-1)$-uniform paths in $\Psi_u[S]$ consisting of $\frac{k-1}{k}(M+1)$ vertices each, which start and end with $\zeta_{\star\star}$-connectable $(k-1)$-tuples.

Since there are at most $(M^2)!$ possibilities to order the vertices in S, there has to exist a subset $U' \subseteq U \smallsetminus U_{\rm bad}$ such that $\mathscr{W}_u = \mathscr{W}$ is the same for every $u \in U'$ and

$$|U'| \geqslant \frac{\frac{2}{3}|U \smallsetminus U_{\rm bad}|}{(M^2)!} \overset{(2.6.10)}{\geqslant} \frac{\vartheta_\star^2 n}{3(M^2)!} \geqslant \frac{(M-(k-1))(M+1)}{k}\,.$$

Now, for every path in $\mathscr{W}$ put $\frac{M-(k-1)}{k}$ distinct vertices from U' aside and insert them at every k-th position into the path from $\mathscr{W}$ (see Figure 2.6.2).

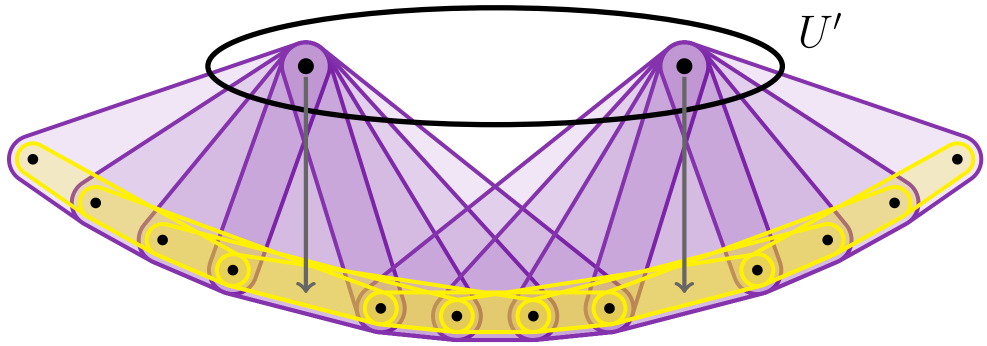


Figure 2.6.2: Augmenting a yellow $\frac{4}{5}(M+1)$-vertex path to a lila M-vertex path.

Since the starting and ending $(k-1)$-tuples of every path in $\mathscr{W}$ are $\zeta_{\star\star}$-connectable in Ψ and the insertion of the additional vertices increases their length to $\frac{k-1}{k}(M+1)+\frac{M-(k-1)}{k} = M$, the resulting $M+1$ paths are elements of $\mathscr{P}$. Hence, the collection $\mathscr{C}$ can be augmented by removing the at most M paths whose blocks lie in $\mathscr{S}$ and adding the $M+1$ newly constructed paths instead. As this contradicts the maximality of $\mathscr{C}$, the assumption (2.6.9) must have been false. This concludes the proof of Lemma 2.6.9. $\square$

Finally, we arrive at the main result of this section.

Proposition 2.6.14. *For every $k \geqslant 3$ the statement $\heartsuit_k$ holds.*

Proof. We argue by induction on k, the base case being provided by Fact 2.6.2. The Lemmata 2.6.4 and 2.6.9 show that $\heartsuit_{k-1} \Rightarrow \spadesuit_{k-1} \Rightarrow \heartsuit_k$, which is the induction step. $\square$

2.7 The proof of Theorem 2.1.2

The results in the foregoing sections routinely imply Theorem 2.1.2, but for the sake of completeness we provide the details.

Proof of Theorem 2.1.2. Given $k \geqslant 3$ and $\alpha > 0$ we choose some auxiliary constants fitting into the hierarchy

$$\alpha, k^{-1} \gg \mu \gg \beta, \ell^{-1} \gg \zeta_\star \gg \vartheta_\star \gg \zeta_{\star\star} \gg \vartheta_{\star\star} \gg M^{-1} \gg n_0^{-1}, \tag{2.7.1}$$

where $\ell \geqslant 3$ is an odd integer and $M \equiv -1 \pmod{k}$.

Now let $H = (V, E)$ be a k-uniform hypergraph on $n \geqslant n_0$ vertices satisfying the minimum $(k-2)$-degree condition $\delta_{k-2}(H) \geqslant (\frac{5}{9} + \alpha)\frac{n^2}{2}$. By Fact 2.2.34 and $\alpha \gg \mu \gg \beta, \ell^{-1}$ there exists an $(\alpha, \beta, \ell, \mu)$-constellation Ψ with underlying hypergraph H.

Stage A. We set aside a reservoir set $\mathcal{R}$ of size $|\mathcal{R}| \leqslant \vartheta_\star^2 n$ provided by Proposition 2.4.1. Let us recall that by Corollary 2.4.2 and $\vartheta_{\star\star} \ll \vartheta_\star, k^{-1}, \ell^{-1}$

(*1*) for every set $\mathcal{R}' \subseteq \mathcal{R}$ of at most $\vartheta_{\star\star}^2 n$ "forbidden" vertices, every $\zeta_{\star\star}$-leftconnectable $(k-1)$-tuple $\vec{a}$, every $\zeta_{\star\star}$-rightconnectable $(k-1)$-tuple $\vec{b}$ that is disjoint to $\vec{a}$, and every $i \in [0, k)$, there is an $\vec{a}$-$\vec{b}$-path through $\mathcal{R} \smallsetminus \mathcal{R}'$ with $f(k, i, \ell)$ inner vertices.

Stage B. Next, we choose an absorbing path avoiding $\mathcal{R}$. More precisely, Proposition 2.5.1 yields a path $P_A \subseteq H - \mathcal{R}$ with the properties that

(*2*) $|V(P_A)| \leqslant \vartheta_\star n$,

(*3*) the starting and ending $(k-1)$-tuple of P_A are $\zeta_{\star\star}$-connectable,

(*4*) and for every subset $Z \subseteq V \smallsetminus V(P_A)$ with $|Z| \leqslant 2\vartheta_\star^2 n$ and $|Z| \equiv 0 \pmod{k}$, there is a path $Q \subseteq H$ with $V(Q) = V(P_A) \cup Z$ having the same end-$(k-1)$-tuples as P_A.

Stage C. We proceed by covering almost all vertices belonging neither to $\mathcal{R}$ nor to P_A by long paths. To this end we set $X = \mathcal{R} \cup V(P_A)$ and consider the constellation $\Psi' = \Psi - X$. Since $|X| \leqslant \vartheta_\star^2 n + \vartheta_\star n \leqslant 2\vartheta_\star n$, Lemma 2.2.36 tells us that Ψ' is an $\big(\frac{\alpha}{2}, \frac{\beta}{2}, \ell, 2\mu\big)$-constellation. So the covering principle $\heartsuit_k$ defined in Definition 2.6.1 and proved in Proposition 2.6.14 applies to Ψ', $2\zeta_{\star\star}$ here in place of Ψ, $\zeta_{\star\star}$ there. In other words, in Ψ' there exists a collection $\mathscr{C}$ of mutually disjoint M-vertex paths whose end-tuples are $(2\zeta_{\star\star})$-connectable in Ψ' such that

$$\Big| V(\Psi') \smallsetminus \bigcup_{P \in \mathscr{C}} V(P) \Big| \leqslant \vartheta_\star^2 n\,.$$

Due to Fact 2.2.18, the end-tuples of the paths in $\mathscr{C}$ are $\zeta_{\star\star}$-connectable in Ψ.

Stage D. Now we want to connect the paths in $\mathscr{C}$ and P_A, thus obtaining one long path T with $\zeta_{\star\star}$-connectable end-tuples. This is to be done by means of $|\mathscr{C}|$ connections through the reservoir, iteratively using (*1*) with $i = 0$. Altogether these connections require

$$|\mathscr{C}| f(k, 0, \ell) \leqslant \frac{4^k \ell k n}{M} \leqslant \vartheta_{\star\star}^2 n$$

vertices from the reservoir. So $|\mathscr{C}|$ successive applications of (*1*) indeed allow us to construct this long path T (see Figure 2.7.1).

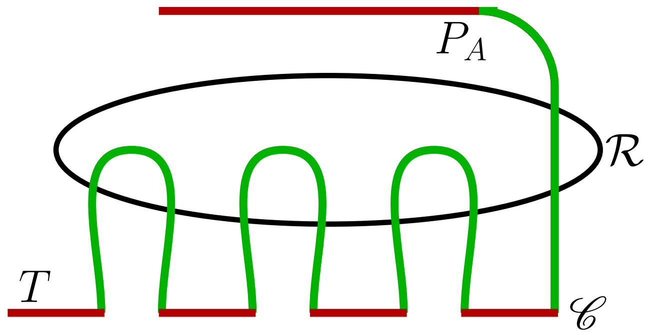


Figure 2.7.1: The situation after Stage D.

Stage E. Moreover, we can still use *(1)* one more time in order to connect the end-tuples of T, thus creating one long cycle C. For this last connection we use $f(k, i, \ell)$ inner vertices, where $i \in [0, k)$ is determined by the congruence $i \equiv n - |V(T)| \pmod{k}$. The current situation is depicted in Figure 2.7.2.

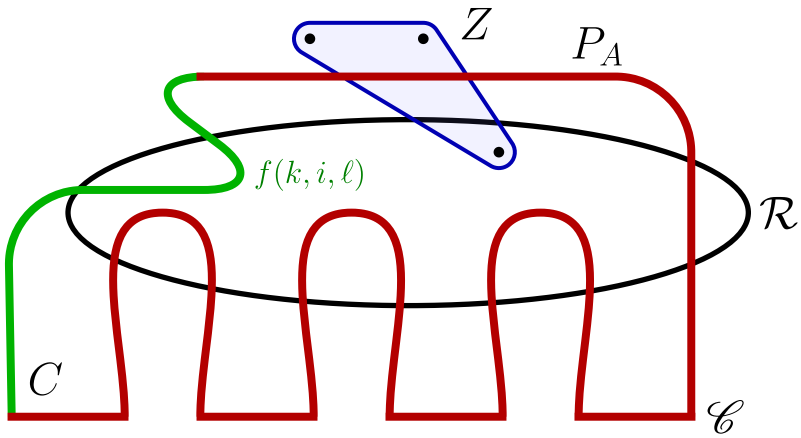


Figure 2.7.2: The situation after Stage E. The dots in Z represent sets of k vertices each.

Our choice of i guarantees that the set $Z = V(\Psi) \smallsetminus V(C)$ satisfies

$$|Z| \equiv n - |V(T)| - f(k, i, \ell) \equiv 0 \pmod{k}.$$

Furthermore, Z has at most the size

$$|Z| \leqslant |\mathcal{R}| + \Big|V(\Psi') \smallsetminus \bigcup_{P \in \mathscr{C}} V(P)\Big| \leqslant 2\vartheta_\star^2 n.$$

Stage F. Taken together, the last two displayed formulae and *(4)* show that Z can be absorbed by P_A, i.e., that there exists a path Q with $V(Q) = V(P_A) \cup Z$ having the same end-tuples as P_A. Upon replacing the subpath P_A of C by Q we obtain the desired

Hamiltonian cycle in H (see Figure 2.7.3). $\square$

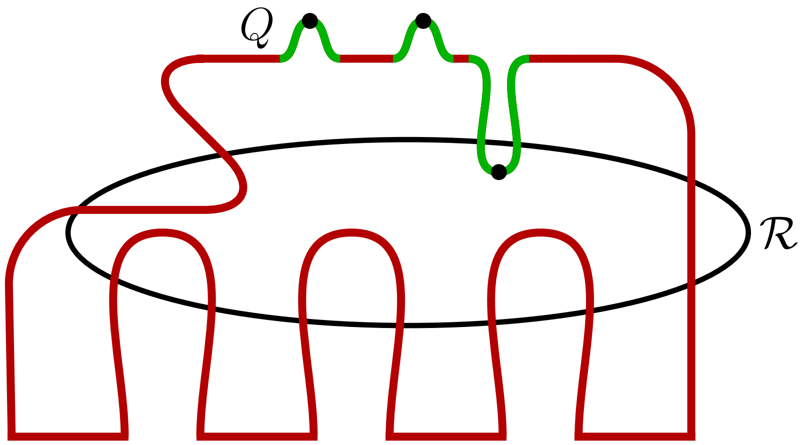


Figure 2.7.3: The situation after Stage F.

3. A pair degree condition for Hamiltonian cycles in 3-uniform hypergraphs

3.1 Introduction

The search for conditions ensuring the existence of Hamiltonian cycles in graphs has been one of the main themes in graph theory. For graphs, several classic results exist, starting with the tight condition by Dirac [27] stating that every graph $G = (V, E)$ on at least 3 vertices and with minimum degree $\delta(G) \geqslant |V|/2$ contains a Hamiltonian cycle. Pósa [94] improved this result to a condition on the degree sequence:

Theorem 3.1.1. *Let $G = ([n], E)$ be a graph on $n \geqslant 3$ vertices with degree sequence $d(1) \leqslant \cdots \leqslant d(n)$. If $d(i) \geqslant i+1$ for all $i < (n-1)/2$ and if furthermore $d(\lceil n/2 \rceil) \geqslant \lceil n/2 \rceil$ when n is odd, then G contains a Hamiltonian cycle.*

Finally, Chvátal [18] achieved an even stronger result: A graph $G = ([n], E)$ on $n \geqslant 3$ vertices with degree sequence $d(1) \leqslant \cdots \leqslant d(n)$ contains a Hamiltonian cycle if for all $i < \frac{n}{2}$ we have: $d(i) \leqslant i \Rightarrow d(n-i) \geqslant n-i$. On the other hand, for any sequence $a_1 \leqslant \cdots \leqslant a_n < n$ not satisfying this condition, there exists a graph on vertex set $[n]$ with $a_i \leqslant d(i)$, for all $i \in [n]$, that does not contain a Hamiltonian cycle. The aim of this article is to take a first step towards a generalisation of Chvátal's result to more general structures, namely hypergraphs, by proving an analogue of Pósa's result above for 3-uniform hypergraphs.

A *k-uniform hypergraph* (or *k-graph*) is a pair (V, E) consisting of a (vertex) set V and an (edge) set $E \subseteq V^{(k)}$. We sometimes write $v(H) = |V(H)|$ and $e(H) = |E(H)|$. In the following let $H = (V, E)$ be a 3-graph. For $U \subseteq V$, we define $H[U] := (U, E(U))$ with $E(U) := \{e \in E : e \subseteq U\}$. For vertices $v, w \in V$, we denote by $d(v, w) := |\{x \in V : vwx \in E\}|$ the *pair degree*, where for convenience we write an edge as vwx instead of $\{v, w, x\}$. In addition, it is also common to study the *vertex degree* $d(v) := |\{e \in E : v \in e\}|$. The minimum pair degree is $\delta_2(H) := \min_{vw \in V^{(2)}} d(v, w)$ and the minimum vertex degree is $\delta_1(H) := \min_{v \in V} d(v)$.

Often it is useful to consider something like a 2-uniform projection of H with respect to a vertex $v \in V$; we define the *link graph* L_v of v as the graph $(V, \{xy : xyv \in E\})$.

We will follow the definition of paths and cycles in [95], suggested by Katona and Kierstead in [67]. A 3-graph P is a *tight path* of length ℓ, if $|V(P)| = \ell + 2$ and there is an ordering of the vertices $V(P) = \{x_1, \dots, x_{\ell+2}\}$ such that $E(P) = \{x_i x_{i+1} x_{i+2} : i \in [\ell]\}$. The tuple (x_1, x_2) is the *starting pair* of P, the tuple $(x_{\ell+1}, x_{\ell+2})$ is the *ending pair* of P, both are the *end-pairs* of P and we say that P is a tight (x_1, x_2)-$(x_{\ell+1}, x_{\ell+2})$-path. All other vertices of P are called *internal*. We sometimes identify a path with the sequence of its vertices $x_1, \dots, x_{\ell+2}$. Accordingly, a *tight cycle* C of length $\ell \geqslant 4$ consists of a path $x_1, \dots, x_\ell$ of length $\ell - 2$ together with the two hyperedges $x_{\ell-1}x_\ell x_1$ and $x_\ell x_1 x_2$. A *tight walk* of length ℓ is a hypergraph W with $V(W) = \{x_1, \dots, x_{\ell+2}\}$, where the x_i are not necessarily distinct, and $E(W) = \{x_i x_{i+1} x_{i+2} : i \in [\ell]\}$. Note that the length of a path, a cycle or a walk is the number of its edges and we will use this convention for cycles, paths, and walks in graphs as well.

One might also consider degree conditions for loose Hamiltonian cycles in k-uniform hypergraphs, in which consecutive edges intersect in less than $k - 1$ vertices. Loose Hamiltonian cycles were for instance studied in [16, 25, 56, 76]. From now on, we only consider tight paths and cycles and consequently we may omit the prefix "tight".

In recent years, there has been some progress to achieve Dirac like results for hypergraphs. Rödl, Ruciński, and Szemerédi [98] started by showing that for $\alpha > 0$, there is some n_0 such that every 3-graph on $n \geqslant n_0$ vertices with minimum pair degree at least $(\frac{1}{2} + \alpha)n$ contains a Hamiltonian cycle. Actually, in [100] they improved the result to the following.

Theorem 3.1.2. *Let H be a 3-graph on n vertices, where n is sufficiently large. If H satisfies $\delta_2(H) \geqslant \lfloor n/2 \rfloor$, then H has a Hamiltonian cycle. Moreover, for every n, there exists an n-vertex 3-graph H_n such that $\delta_2(H_n) = \lfloor n/2 \rfloor - 1$ and H_n does not have a Hamiltonian cycle.*

More recently, Reiher, Rödl, Ruciński, Schacht, and Szemerédi [95] proved the following asymptotically optimal result.

Theorem 3.1.3. *For every $\alpha > 0$, there is an $n_0 \in \mathbb{N}$ such that every 3-graph H on $n \geqslant n_0$ vertices with $\delta_1(H) \geqslant \left(\frac{5}{9} + \alpha\right)\frac{n^2}{2}$ contains a Hamiltonian cycle.*

Since the first version of this article, this has been generalised to all k independently by Lang and Sanhueza-Matamala [77] and by Polcyn, Reiher, Rödl, and myself [93].

In this work, we study a new pair degree condition that forces large 3-graphs to contain a Hamiltonian cycle. Call a matrix $(d_{ij})_{ij}$ *Hamiltonian* if every 3-graph $H = ([n], E)$ with $d(i,j) \geqslant d_{ij}$, for all $ij \in [n]^{(2)}$, contains a Hamiltonian cycle. It would be very desirable to get a result for 3-graphs similar to the one by Chvátal for degree sequences in graphs, that is, a characterisation of all Hamiltonian matrices. For the graph case, Pósa's result (Theorem 3.1.1) was a step towards the characterisation by Chvátal. In a sense, our main result can be seen as a 3-uniform (asymptotic) analogue of the theorem by Pósa.

Theorem 3.1.4 (Main result)**.** *For $\alpha > 0$, there exists an $n_0 \in \mathbb{N}$ such that for all $n \in \mathbb{N}$ with $n \geqslant n_0$, the following holds. If $H = ([n], E)$ is a 3-graph with $d(i,j) \geqslant \min\left(i, j, \frac{n}{2}\right) + \alpha n$ for all $ij \in [n]^{(2)}$, then H contains a (tight) Hamiltonian cycle.*

This result strengthens the asymptotic version of Theorem 3.1.2 achieved in [98].

Let us remark that recently there have also been related results on degree sequences in graphs. For example, Treglown [112] gave a degree sequence condition that forces the graph to contain a clique factor and Staden and Treglown [108] proved a degree sequence condition that forces the graph to contain the square of a Hamiltonian cycle. Since the first version of this article, Bowtell and Hyde [12] obtained a degree sequence condition for perfect matchings in 3-graphs.

Note that in the proof (and the proofs of the lemmas) we can always assume $\alpha \ll 1$. Before we start with the outline of the proof of Theorem 3.1.4 in the next section, we give the following examples showing that our result is asymptotically optimal in some regard.

Example 3.1.5. *(i) Consider the partition $X \dot\cup Y = [n]$ with $X = \left[\left\lceil \frac{n+1}{3} \right\rceil\right]$ and let H be the hypergraph on $[n]$ containing all triples $e \in V^{(3)}$ with $|e \cap X| \neq 2$.*

Then we have $d(i,j) \geqslant \min\left(i, j, \frac{n}{2}\right) - 1$ for all $ij \in [n]^{(2)}$. However, if there was a Hamiltonian cycle in H, it would contain at least one edge with two vertices from X. But such an edge can only lie in a cycle in which all vertices are from $X \subsetneq [n]$. Hence, H does not contain a Hamiltonian cycle.

(ii) Next, look at the partition $X \dot\cup Y = [n]$ with $X = \left[\left\lfloor \frac{n}{2} \right\rfloor\right]$ and let H be the hypergraph on $[n]$ containing all triples $e \in V^{(3)}$ such that $|e \cap Y| \neq 2$.

Then for all $ij \in [n]^{(2)}$, we have $d(i,j) \geqslant \frac{n}{2} - 2$. But a similar argument as above shows that H does not contain a Hamiltonian cycle.

The two examples show that Theorem 3.1.4 does not hold when replacing the degree condition with $d(i,j) \geqslant \min\left(i, j, \frac{n}{2}\right) - 1$ (not even when replacing it with $d(i,j) \geqslant \min(i,j) - 1$)

and neither when replacing it with $d(i,j) \geqslant \min\left(i, j, \frac{n}{2}-2\right)$. Note that this means that Theorem 3.1.4 cannot (asymptotically) be improved on by decreasing the requirement on the degree of every pair and neither by "capping" at a lower value than at $\frac{n}{2}-1$. However, it is not yet a Chvátal like characterisation of all Hamiltonian matrices. For instance, it is easy to see that there are Hamiltonian matrices with $d_{ij} = 0$ for $\Omega(n^2)$ choices of $i, j \in [n]$.

In the following, we will omit rounding issues if they are not important, e.g., we will assume that αn etc. are natural numbers. Further, for $A, B \subseteq \mathbb{R}_+$, we write that a statement $\mathfrak{S}$ holds for all $a \in A$ and $b \in B$ with $a \ll b$, to say that for every $b \in B$, there exists an $a_0 \in \mathbb{R}_+$ such that for all $a \in A$ with $a \leqslant a_0$, the statement $\mathfrak{S}$ holds.

Organisation

In the next section we give an overview of the proof, state the auxiliary results for each step and finally deduce the main result Theorem 3.1.4 from these. Sections 3.3-3.6 are devoted to the proofs of the auxiliary results. In the end, we collect some interesting related problems in Section 3.7.

3.2 Overview and final proof

The proof of Theorem 3.1.4 uses the absorption method introduced by Rödl, Ruciński, and Szemerédi in [98], which helps to reduce the problem of finding a Hamiltonian cycle to the problem of constructing a cycle containing almost all vertices.

This strategy proceeds by constructing a cycle containing almost all vertices of the hypergraph H and a special subpath into which we can "absorb" any small set of vertices, meaning we can integrate the left-over vertices into this subpath to obtain a Hamiltonian cycle. For that, we use that for every vertex $v \in V(H)$, there exist many absorbers in H, a structure consisting of several paths which can be restructured into paths containing v while keeping the same end-pairs. Then, utilising the probabilistic method, we can construct an absorbing path, a path containing many absorbers for every vertex. Lastly, we build a long path in the remainder of H, consisting of almost all vertices, and connect it with the absorbing path to a cycle into which the left-over vertices can be absorbed.

For these constructions we often need to connect two paths, that is, find a path between their end-pairs. Hence, we will begin by showing that we can connect every pair of pairs of vertices by a large number of paths with a fixed length.

Lemma 3.2.1 (Connecting Lemma)**.** *Let $\alpha, \vartheta > 0$, $n, L \in \mathbb{N}$ with $1/n \ll \vartheta \ll 1/L \ll \alpha$. If $H = ([n], E)$ is a 3-graph with $d(i,j) \geqslant \min\left(i, j, \frac{n}{2}\right) + \alpha n$, for all $ij \in [n]^{(2)}$, then for all disjoint ordered pairs of distinct vertices $(x,y), (w,z) \in [n]^2$, the number of paths of length L in H connecting (x,y) and (w,z) is at least ϑn^{L-2}.*

See Section 3.3 for the proof of Lemma 3.2.1.

Later, we will use this result whenever we need to connect different paths that have been constructed before. However, when we want to connect paths after almost all the vertices are covered by paths, we need to ensure that there still exist paths, disjoint to all previously built paths. To this end, we will take a special selection of vertices - the reservoir - aside, with the property that for every pair of pairs of vertices, we still have many paths of fixed length connecting them, where all internal vertices of those paths are vertices of the reservoir. The existence of such a set will be shown by the probabilistic method.

Lemma 3.2.2 (Reservoir Lemma)**.** *Let $\alpha, \vartheta > 0$ and $n, L \in \mathbb{N}$ such that $1/n \ll \vartheta \ll 1/L \ll \alpha$. If $H = ([n], E)$ is a 3-graph satisfying $d(i,j) \geqslant \min\left(i, j, \frac{n}{2}\right) + \alpha n$, for all $ij \in [n]^{(2)}$, then there exists a reservoir set $\mathcal{R} \subseteq [n]$ with $\frac{\vartheta^2}{2} n \leqslant |\mathcal{R}| \leqslant \vartheta^2 n$ such that for all disjoint ordered pairs of distinct vertices $(x,y), (w,z) \in [n]^2$, there are at least $\vartheta \left|\mathcal{R}\right|^{L-2}/2$ paths of length L in H which connect (x,y) and (w,z) and whose internal vertices all belong to $\mathcal{R}$.*

It follows that removing a few vertices from the reservoir will not destroy its connectability property.

Lemma 3.2.3 (Preservation of the Reservoir)**.** *Let $\alpha, \vartheta > 0$ and $n, L \in \mathbb{N}$ such that $1/n \ll \vartheta \ll 1/L \ll \alpha$. If $H = ([n], E)$ is a 3-graph satisfying $d(i,j) \geqslant \min\left(i, j, \frac{n}{2}\right) + \alpha n$, for all $ij \in [n]^{(2)}$, $\mathcal{R}$ is given by Lemma 3.2.2, and $\mathcal{R}' \subseteq \mathcal{R}$ with $|\mathcal{R}'| \leqslant 2\vartheta^4 n$, then for all disjoint ordered pairs of distinct vertices $(x,y), (w,z) \in [n]^2$, there is an (x,y)-(w,z)-path of length L in H with all internal vertices belonging to $\mathcal{R} \setminus \mathcal{R}'$.*

See Section 3.4 for the proof of Lemma 3.2.2 and Lemma 3.2.3.

The proof will continue with the definition of the absorbers and we will show that for each vertex, there are many absorbers. We make use of this fact when we show that a small random selection of tuples still contains many absorbers for every $v \in V(H)$. With the Connecting Lemma we can afterwards connect all the small paths in that selection to a path that can absorb any small set of vertices.

Lemma 3.2.4 (Absorbing Path)**.** *Let $\alpha, \vartheta > 0$ and $n, L \in \mathbb{N}$ such that $1/n \ll \vartheta \ll 1/L \ll \alpha$. If $H = ([n], E)$ is a 3-graph satisfying $d(i,j) \geqslant \min\left(i, j, \frac{n}{2}\right) + \alpha n$, for all $ij \in [n]^{(2)}$, and $\mathcal{R}$ is given by Lemma 3.2.2, then there exists a path $P_A \subseteq H \setminus \mathcal{R}$ with $v(P_A) \leqslant \vartheta n$ and with the (absorbing) property that for each $X \subseteq [n]$ with $|X| \leqslant 2\vartheta^2 n$, there is a path with vertex set $X \cup V(P_A)$ and the same end-pairs as P_A.*

See Section 3.5 for the proof of Lemma 3.2.4.

By using weak hypergraph regularity and then an explicit result to obtain an almost perfect matching in the reduced hypergraph, we show in Section 3.6 that in every hypergraph H satisfying the degree condition in Theorem 3.1.4, there exists a path which contains almost all vertices of H (see Proposition 3.2.5).

Proposition 3.2.5 (Long Path)**.** *Let $\alpha, \vartheta > 0$ and $n, L \in \mathbb{N}$ such that $1/n \ll \vartheta \ll 1/L \ll \alpha$. Let $H = ([n], V)$ be a 3-graph with $d(i,j) \geqslant \min\left(i, j, \frac{n}{2}\right) + \alpha n$, for all $ij \in [n]^{(2)}$, let $\mathcal{R}$ be as in Lemma 3.2.2, and P_A as in Lemma 3.2.4.*

Then there exists a path $Q \subseteq H \setminus P_A$ such that

$$v(Q) \geqslant \left(1 - 2\vartheta^2\right) n - v\left(P_A\right)$$

and $|V(Q) \cap \mathcal{R}| \leqslant \vartheta^4 n$.

See Section 3.6 for the proof of Proposition 3.2.5.

Now we are ready to prove our main result, Theorem 3.1.4 (see also Figure 3.2.1).

Proof of Theorem 3.1.4. Let $\alpha, \vartheta > 0$ and $n, L \in \mathbb{N}$ such that $1/n \ll \vartheta \ll 1/L \ll \alpha$. Now let $H = ([n], E)$ be a 3-graph satisfying the degree condition $d(i,j) \geqslant \min\left(i, j, \frac{n}{2}\right) + \alpha n$ for all $ij \in [n]^{(2)}$. Lemmas 3.2.2, 3.2.4, and Proposition 3.2.5 provide a reservoir $\mathcal{R}$, an absorbing path $P_A \subseteq H \setminus \mathcal{R}$ and a long path $Q \subseteq H \setminus P_A$ with $|\mathcal{R} \cap V(Q)| \leqslant \vartheta^4 n$. Let (a,b) and (c,d) be the end-pairs of P_A and let (r,s) and (t,u) be the end-pairs of Q (note that they are disjoint since we have $Q \subseteq H \setminus P_A$). Since $|\mathcal{R} \cap V(Q)| \leqslant \vartheta^4 n$ and $P_A \subseteq H \setminus \mathcal{R}$ and by Lemma 3.2.3, we can choose a path P_1 of length L connecting (t,u) and (a,b) with all internal vertices in $\mathcal{R} \setminus (V(Q) \cup V\left(P_A\right))$ and, by the hierarchy of constants, we also find a path P_2 of length L connecting (c,d) and (r,s) with all internal vertices in $\mathcal{R} \setminus (V(Q) \cup V\left(P_A\right) \cup V\left(P_1\right))$. That leaves us with a cycle C in H which satisfies $v(C) \geqslant (1 - 2\vartheta^2)\, n$ and $P_A \subseteq C$. The absorbing property of P_A guarantees that for $X := [n] \setminus V(C)$, there exists a path P'_A with $V\left(P'_A\right) = V\left(P_A\right) \cup X$ which has the same end-pairs as P_A (which are connected to Q) and hence there is a Hamiltonian cycle in H. ∎

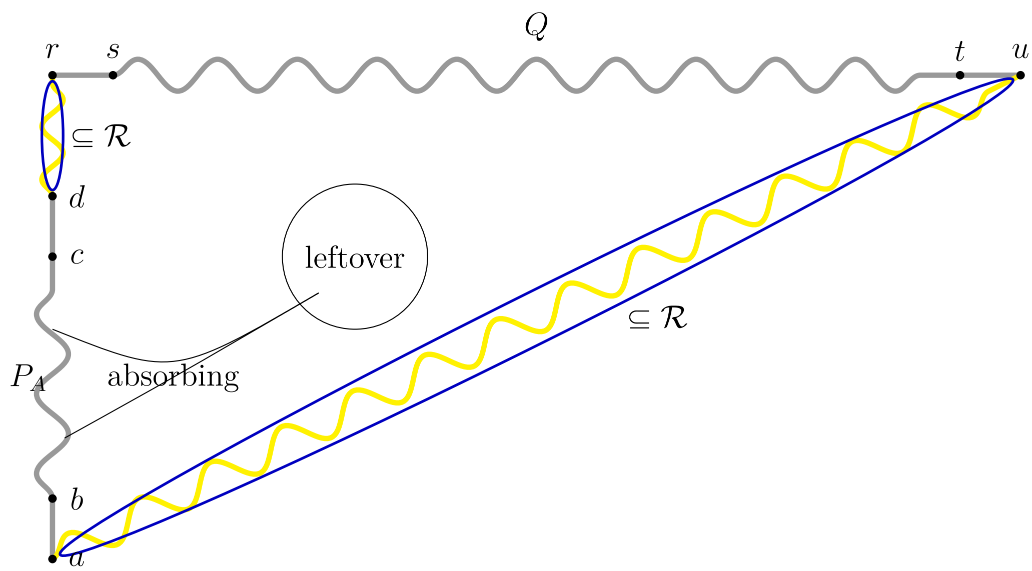


Figure 3.2.1: Overview of the proof

3.3 Connecting Lemma

Before we start with the actual proof of Lemma 3.2.1, let us take a look at the strategy. Say, we want to connect two (ordered) pairs (x, y) and (w, z) in a hypergraph H satisfying the condition in Theorem 3.1.4. One can easily reduce the case of both pairs being arbitrary to that of both having pair degree at least $\frac{n}{2} + \alpha n$ by "climbing up" in the degree sequence (see the beginning of the proof). Then $N((x, y), (w, z))$, the set of common neighbours of (x, y) and (w, z), is non-empty because of the high pair degrees of (x, y) and (w, z). If we were able to find many (2-uniform) y-w-paths in the link graphs of elements in $N((x, y), (w, z))$, we could subsequently insert the elements of $N((x, y), (w, z))$ at every third position of such a path, thereby obtaining a 3-uniform walk.

So we could indeed connect two pairs if the link graphs of vertices in $N((x, y), (w, z))$ would inherit the right degree condition, i.e., if the vertices would be large (regarded as elements of $\mathbb{N}$). However, since we cannot control how large the elements in $N((x, y), (w, z))$ are, the degree condition that the link graphs of vertices in $N((x, y), (w, z))$ inherit may not be strong enough to let us connect two vertices by "climbing up" the degree sequence. The idea to insert a middle pair (a, b), as done in [95], overcomes this problem. If (a, b) has some large common neighbours with (x, y) and some with (w, z), we can find enough (x, y)-(w, z) walks passing through (a, b) by applying the strategy explained above (now we can connect vertices in the link graphs by "climbing up" the degree sequence). The number of those walks will depend on the number of large common neighbours that (a, b) has with each (x, y) and (w, z). So roughly speaking, if the sum over all (a, b) of large common neighbours of (a, b) and (x, y) and of (a, b) and (w, z) is large, we can indeed prove the Connecting Lemma. This last point (in its accurate form) will follow from the observation that two link graphs of large vertices have many common edges.

Note that this strategy can be used in the seemingly different settings of our pair degree condition and the minimum vertex degree condition in [95], since in both cases we have "well connected" subgraphs in every link graph and each two of these subgraphs intersect in many edges: In [95] those subgraphs are the *robust subgraphs* and in our case we can just consider the link graphs of large vertices. After the first version of this article, this idea has also been used extensively in [93].

Proof of Lemma 3.2.1. Observe that when we show that there exists an $L \in \mathbb{N}$ and a $\vartheta > 0$ such that the statement of Lemma 3.2.1 holds for these, it easily follows that it holds for all $L \in \mathbb{N}$ and $\vartheta > 0$ with $1/n \ll \vartheta \ll 1/L \ll \alpha \ll 1$. Hence, let the hierarchy and H be given as described in the lemma and let $(x, y), (w, z) \in [n]^2$ be two disjoint ordered pairs of distinct vertices.

First, we will show that it is possible to "climb up" along the degree sequence in (compared to n) few steps, starting from the pairs (x, y) and (w, z) and ending with pairs of vertices $\geqslant \frac{n}{2}$.

In the second step, we will connect these two by utilising an analogous "climb up" argument in the link graphs of neighbours of a pair and slipping in an additional connective pair. We first look for walks rather than paths and conclude by remarking that many of them will actually be paths.

First Step

By induction on $\ell \geqslant 3$, we will prove the following statement: There exist at least $\left(\frac{\alpha}{5}\right)^{\ell-2} n^{\ell-2}$ walks $x_1 = x, x_2 = y, x_3, \ldots, x_\ell$ such that for $i \geqslant 3$ we have:

$$x_i \geqslant \min\left(\frac{\alpha}{4}n(i-2), \frac{n}{2}\right) + \frac{\alpha}{4}n \tag{3.3.1}$$

We will first show the statement for $\ell = 3$ and $\ell = 4$ and then deduce it for any $\ell \geqslant 5$ given that it holds for $\ell - 1$.

$\ell = 3$: By the degree condition on H we have $d(x, y) \geqslant \min\left(1, 2, \frac{n}{2}\right) + \alpha n$. Hence, there exist at least $\frac{\alpha}{5}n$ possible vertices x_3 such that x_1, x_2, x_3 is a walk and $x_3 \geqslant \frac{\alpha}{4}n + \frac{\alpha}{4}n$.

$\ell = 4$: Let x_1, x_2, x_3 be one of those $\frac{\alpha}{5}n$ walks satisfying the condition (3.3.1) that we get by the previous case. We then have $d(x_2, x_3) \geqslant \min\left(1, \frac{\alpha}{2}n, \frac{n}{2}\right) + \alpha n$, so there exist at least $\frac{\alpha}{5}n$ possible vertices x_4 such that x_1, x_2, x_3, x_4 is a walk and $x_i \geqslant \frac{\alpha}{4}n(i-2) + \frac{\alpha}{4}n$ for $i = 3, 4$.

$\ell \geqslant 5$: Let $x_1, x_2, x_3, \ldots, x_{\ell-1}$ be one of the $\left(\frac{\alpha}{5}\right)^{\ell-3} n^{\ell-3}$ walks satisfying, for $i \geqslant 3$,

$$x_i \geqslant \min\left(\frac{\alpha}{4}n(i-2), \frac{n}{2}\right) + \frac{\alpha}{4}n$$

that we get by induction. Then our pair degree condition entails

$$d(x_{\ell-2}, x_{\ell-1}) \geqslant \min\left(\frac{\alpha}{4}n(\ell-4) + \frac{\alpha}{4}n, \frac{n}{2}\right) + \alpha n$$

which in turn gives rise to at least $\frac{\alpha}{5}n$ possible vertices x_ℓ such that $x_1, x_2, \ldots, x_\ell$ build a walk and we have $x_i \geqslant \min\left(\frac{\alpha}{4}n(i-2), \frac{n}{2}\right) + \frac{\alpha}{4}n$ for all $i \in [\ell], i \geqslant 3$.

This leaves us with $\left(\frac{\alpha}{5}\right)^{\frac{2}{\alpha}} n^{\frac{2}{\alpha}}$ possibilities for walks

$$x_1 = x, x_2 = y, x_3, \ldots, x_{\frac{2}{\alpha}+2}$$

with $x_{\frac{2}{\alpha}+1}, x_{\frac{2}{\alpha}+2} \geqslant \frac{n}{2}$ and an analogous argument for (w, z) with just as many possibilities for walks

$$z_1 = z, z_2 = w, z_3, \ldots, z_{\frac{2}{\alpha}+2}$$

with $z_{\frac{2}{\alpha}+1}, z_{\frac{2}{\alpha}+2} \geqslant \frac{n}{2}$.

Second Step

Let m be the smallest even number $\geqslant \frac{1}{\alpha} + 1$. It now suffices to show that for some $\vartheta' > 0$ with $1/n \ll \vartheta' \ll \alpha$ we have the following. For all ordered pairs $(x', y'), (w', z') \in [n]^2$ for which the vertices within each pair are distinct and $x', y', w', z' \geqslant \frac{n}{2}$, the number of (x', y')-(w', z') walks with $3m+4$ internal vertices is at least $\vartheta' n^{3m+4}$.

Since $d(x', y') \geqslant \frac{n}{2} + \alpha n$, there exists a set $U_{x'y'} = \{u_1, \ldots, u_{\alpha n}\} \subseteq [n] \smallsetminus [n/2]$ such that $x'y' \in E(L_{u_i})$, for all $i \in [\alpha n]$ (recall that L_{u_i} denotes the link graph of u_i). Similarly, there exists $U_{w'z'} = \{v_1, \ldots, v_{\alpha n}\} \subseteq [n] \smallsetminus [n/2]$ such that $w'z' \in E(L_{v_i})$, for all $i \in [\alpha n]$.

For $(a, b) \in [n]^2$, let $I_{ab} = \{i \in [\alpha n] : ab \in E(L_{u_i}) \cap E(L_{v_i})\}$. Since all vertices $\geqslant \frac{n}{2}$ (apart from u_i, v_i) have in both L_{u_i} and L_{v_i} at least $\frac{n}{2} + \alpha n$ neighbours, and therefore $2\alpha n$ vertices that they are adjacent to in both L_{u_i} and L_{v_i}, there are at least $\frac{\alpha n^2}{4}$ edges in $L_{v_i} \cap L_{u_i}$.

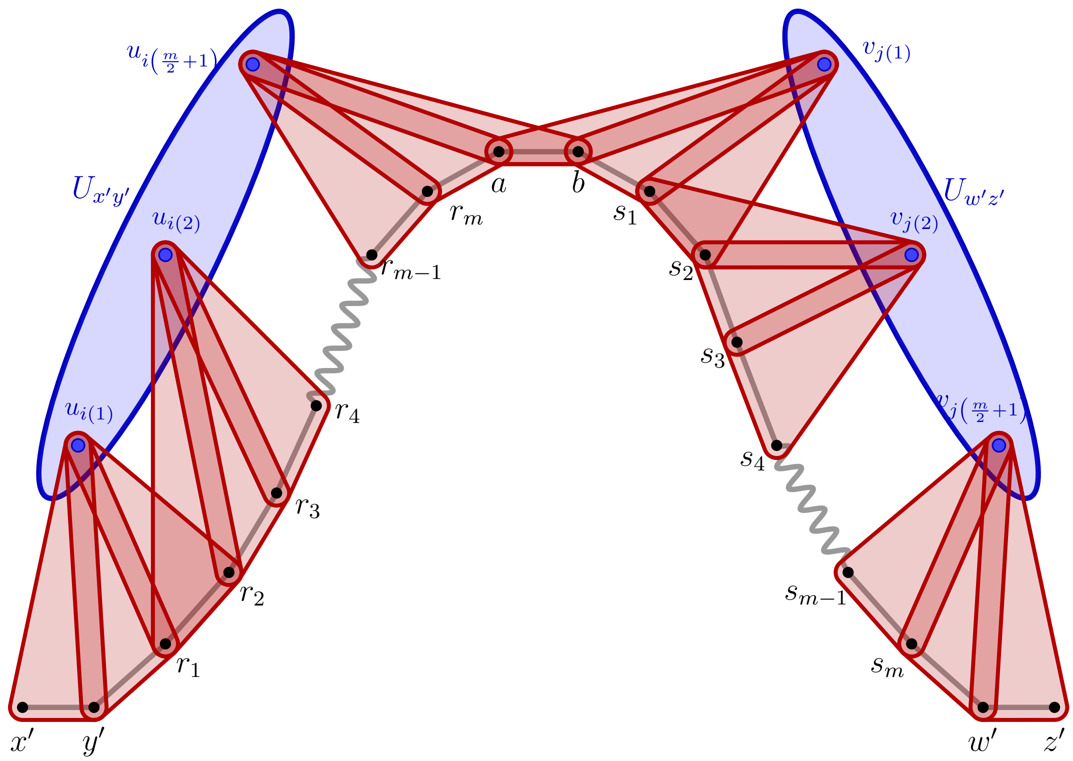


Figure 3.3.1: Idea of the second step, the picture is similar to [95, Fig. 4.1]

Thus, by double counting we have

$$\sum_{(a,b)\in[n]^2} |I_{ab}| \geqslant \sum_{i\in[\alpha n]} |E(L_{v_i}) \cap E(L_{u_i})| \geqslant \frac{\alpha n^2}{4} \alpha n\,.$$

Next, for fixed $(a,b) \in [n]^2$, we find a lower bound on the number L_{ab} of 3-uniform walks of the form

$$x'y'u_{i(1)}r_1r_2u_{i(2)}\ldots u_{i\left(\frac{m}{2}\right)}r_{m-1}r_mu_{i\left(\frac{m}{2}+1\right)}ab$$

where $y'r_1r_2\ldots r_{m-1}r_ma$ is a 2-uniform walk in $L_{u_{i(k)}}$ and $i(k) \in I_{ab}$, for all $k \in \left[\frac{m}{2}+1\right]$.

To this goal, first observe that for all $i \in [\alpha n]$, the number of $y'a$-walks of length $m+1$ in L_{u_i} is at least $\left(\frac{\alpha}{3}\right)^m n^m$. Indeed, since $u_i \geqslant n/2$, we know that for $j \in [n]$, we have $d_{L_{u_i}}(j) \geqslant \min\left(j, \frac{n}{2}\right) + \alpha n$. Therefore, there are at least $\left(\frac{\alpha n}{2}\right)^{m-1}$ walks of length $m-1$ starting in a in which each vertex is either at least $\frac{n}{2}+\frac{\alpha n}{2}$ or at least $\frac{\alpha n}{2}$ larger than the preceding vertex. Since we set $m \geqslant 1/\alpha + 1$, each of these walks ends in a vertex $\geqslant \frac{n}{2}$ and for at least $\left(\frac{\alpha n}{3}\right)^{m-1}$ of them the last vertex is distinct from y'. For each such walk T with its last vertex $a'_T \neq y'$, there are $2\alpha n$ possibilities for common neighbours of y' and a'_T (note that the degrees in L_{u_i} of both y' and a'_T are at least $\frac{n}{2}+\alpha n$). In total, that gives us at least $\left(\frac{\alpha n}{3}\right)^m$ $y'a$-walks of length $m+1$ in L_{u_i}.

Now for $\vec{r} \in [n]^m$, we set $D_{ab}(\vec{r}) := \{i \in I_{ab} : y'\vec{r}a \text{ is a walk in } L_{u_i}\}$. Again by double counting and by the previous observation we infer

$$\sum_{\vec{r}\in[n]^m} |D_{ab}(\vec{r})| = \sum_{i\in I_{ab}} \left|\left\{\vec{r} \in [n]^m : y'\vec{r}a \text{ is a walk in } L_{u(i)}\right\}\right| \geqslant |I_{ab}| \left(\frac{\alpha}{3}\right)^m n^m.$$

Note that for each $\vec{r} \in [n]^m$ that is a $y'a$-walk in $L_{u_{i(k)}}$ for every $k \in \left[\frac{m}{2}+1\right]$, we have that

$$x'y'u_{i(1)}r_1r_2u_{i(2)}\dots u_{i\left(\frac{m}{2}\right)}r_{m-1}r_mu_{i\left(\frac{m}{2}+1\right)}ab$$

is a 3-uniform $(x'y')$-(ab)-walk of length $m+\frac{m}{2}+3$ in H. Hence, with Jensen's inequality we derive:

$$L_{ab} \geqslant \sum_{\vec{r}\in[n]^m} |D_{ab}(\vec{r})|^{\frac{m}{2}+1} \geqslant n^m \left(\sum \frac{1}{n^m}|D_{ab}(\vec{r})|\right)^{\frac{m}{2}+1} \geqslant n^m \left(|I_{ab}|\left(\frac{\alpha}{3}\right)^m\right)^{\frac{m}{2}+1}.$$

We define R_{ab} analogously as the number of 3-uniform walks of the form

$$abv_{j(1)}s_1s_2v_{j(2)}\dots v_{j\left(\frac{m}{2}\right)}s_{m-1}s_mv_{j\left(\frac{m}{2}+1\right)}w'z',$$

where $bs_1s_2\dots s_{m-1}s_mw'$ is a 2-uniform walk in $L_{v_{j(k)}}$ and $j(k) \in I_{ab}$, for all $k \in \left[\frac{m}{2}+1\right]$, and get the same lower bound by an analogous argument.

At last, let W be the number of $(x'y')$-$(w'z')$-walks of length $3m+6$ in H. We apply Jensen's inequality a second time to obtain:

$$\begin{aligned}
W \geqslant& \sum_{(a,b)\in[n]^2} L_{ab}R_{ab}\\
\geqslant& n^{2m}\left(\frac{\alpha}{3}\right)^{m^2+2m} \sum_{(a,b)\in[n]^2} |I_{ab}|^{m+2}\\
\geqslant& n^{2m}\left(\frac{\alpha}{3}\right)^{m^2+2m} n^2\left(\frac{1}{n^2}\frac{\alpha^2n^3}{4}\right)^{m+2}\\
\geqslant& \left(\frac{\alpha}{3}\right)^{m^2+2m}\left(\frac{\alpha^2}{4}\right)^{m+2} n^{3m+4}\\
\geqslant& \left(\frac{\alpha^2}{4}\right)^{m^2+3m+2} n^{3m+4}.
\end{aligned}$$

In total, putting together the walks connecting (x,y) and (x',y'), (x',y') and (w',z')

and (w', z') and (w, z) we get that the number of (x, y)-(w, z)-walks of length $2 \cdot \frac{2}{\alpha} + 3m + 6$ in H is at least

$$\left(\left(\frac{\alpha}{5}\right)^{\frac{2}{\alpha}} n^{\frac{2}{\alpha}}\right)^2 \times \left(\frac{\alpha^2}{4}\right)^{m^2+3m+2} n^{3m+4} \geqslant \alpha^{m^3} n^{\frac{4}{\alpha}+3m+4}.$$

Since only $\mathcal{O}\left(n^{\frac{4}{\alpha}+3m+3}\right)$ of these fail to be a path, we are done. □

3.4 Reservoir

In this section, we will prove the existence of a small set, the reservoir, such that any two pairs of vertices can be connected by paths with all internal vertices lying in the reservoir. The probabilistic proof of this lemma as done in [95] works in almost the same way with different conditions as soon as the Connecting Lemma is provided. We will state two inequalities first that we will need for the probabilistic method.

Lemma 3.4.1 (Chernoff, see for instance Cor. 2.3 in [62])**.** *Let $X_1, X_2, \ldots, X_m$ be a sequence of m independent random variables $X_i :\to \{0, 1\}$ with $\mathbb{P}(X_i = 1) = p$ and $\mathbb{P}(X_i = 0) = 1-p$. Then we have for $\delta \in (0, 1)$:*

- $\mathbb{P}\left(\sum_{i\in[m]} X_i \geqslant (1+\delta)\, pm\right) \leqslant \exp\left(-\frac{\delta^2}{3} pm\right)$
- $\mathbb{P}\left(\sum_{i\in[m]} X_i \leqslant (1-\delta)\, pm\right) \leqslant \exp\left(-\frac{\delta^2}{2} pm\right)$

Lemma 3.4.2 (Azuma-Hoeffding, McDiarmid, Cor. 2.27 in [62] and Thm. 1 in [85])**.** *Suppose that $X_1, \ldots, X_m$ are independent random variables taking values in $\Lambda_1, \ldots, \Lambda_m$ and let $f : \Lambda_1 \times \cdots \times \Lambda_m \to \mathbb{R}$ be a measurable function. Moreover, suppose that for certain real numbers $c_1, \ldots, c_m \geqslant 0$, we have that if $J, J' \in \prod \Lambda_i$ differ only in the k-th coordinate, then $|f(J) - f(J')| \leqslant c_k$. Then the random variable $X := f(X_1, \ldots, X_m)$ satisfies*

$$\mathbb{P}\left(|X - \mathbb{E}(X)| \geqslant t\right) \leqslant 2\exp\left(-\frac{2t^2}{\sum c_i^2}\right)$$

We are now ready to prove Lemma 3.2.2.

Proof of Lemma 3.2.2. Let α, L, ϑ, n, and H be given as in the statement. We choose a random subset $\mathcal{R} \subseteq [n]$, where we select each vertex independently with probability $p = \left(1 - \frac{1}{10L}\right)\vartheta^2$. Since $|\mathcal{R}|$ is now binomially distributed, we can apply Chernoff's inequality (Lemma 3.4.1) and utilise the hierarchy to obtain

$$\mathbb{P}\left(|\mathcal{R}| < \vartheta^2 n/2\right) \leqslant \mathbb{P}\left(|\mathcal{R}| < \frac{2}{3}\mathbb{E}\left(\mathcal{R}\right)\right) \leqslant \exp\left(-\frac{(1/3)^2}{2}pn\right) < 1/3. \tag{3.4.1}$$

We also have $\vartheta^2 n \geqslant (1 + c(L))\,\mathbb{E}\left(|\mathcal{R}|\right)$ for some small $c(L) \in (0,1)$ not depending on n and therefore, again by Chernoff we get for large n:

$$\mathbb{P}\left(|\mathcal{R}| > \vartheta^2 n\right) \leqslant \mathbb{P}\left(|\mathcal{R}| \geqslant (1 + c(L))\,\mathbb{E}\left(\mathcal{R}\right)\right) \leqslant \exp\left(-\frac{c(L)^2}{3}pn\right) < 1/3 \tag{3.4.2}$$

By Lemma 3.2.1, we have that for all disjoint ordered pairs of distinct vertices (x, y) and (w, z), the number of (x, y)-(w, z)-paths of length L in H is at least ϑn^{L-2}. Let $X = X\left((x, y), (w, z)\right)$ denote the random variable counting the number of those (x, y)-(w, z)-paths in H that are of length L and have all internal vertices in $\mathcal{R}$. We then have $\mathbb{E}(X) \geqslant p^{L-2}\vartheta n^{L-2}$.

Now we apply the Azuma-Hoeffding inequality (Lemma 3.4.2) (with $X_1, \ldots, X_n$ being the indicator variables for the events "$1 \in \mathcal{R}$",…,"$n \in \mathcal{R}$") which gives us, since the presence or absence of one particular vertex in $\mathcal{R}$ affects X by at most $(L-2)\,n^{L-3}$, that

$$\begin{aligned}
\mathbb{P}\left(X \leqslant \frac{2}{3}\vartheta (pn)^{L-2}\right) \leqslant & \mathbb{P}\left(X \leqslant \frac{2}{3}\mathbb{E}(X)\right) \\
\leqslant & 2\exp\left(-\frac{2\left(p^{L-2}\vartheta n^{L-2}\right)^2}{9n\left((L-2)n^{L-3}\right)^2}\right) \\
= & \exp\left(-\Omega(n)\right).
\end{aligned}$$

By the union bound, also the probability that there is a pairs of pairs for which the respective number of connecting paths with all internal vertices in $\mathcal{R}$ is less than $\frac{2}{3}\vartheta(pn)^{L-2}$ can be bounded from above by

$$\exp\left(-\Omega(n)\right) \times n^4 < 1/3 \tag{3.4.3}$$

for n large. Moreover, recalling our hierarchy we have

$$\frac{2}{3}\vartheta p^{L-2} n^{L-2} = \left(1 - \frac{1}{10L}\right)^{L-2} \frac{2}{3}\vartheta\left(\vartheta^2 n\right)^{L-2} \geqslant \frac{\vartheta}{2}\left(\vartheta^2 n\right)^{L-2}$$

which together with (3.4.2) and (3.4.3) implies the following: With probability $> 1/3$ the chosen set $\mathcal{R}$ satisfies $|\mathcal{R}| \leqslant \vartheta^2 n$ and has the property that for all disjoint ordered pairs of distinct vertices (x, y) and (w, z) there exist at least $\frac{\vartheta}{2}|\mathcal{R}|^{L-2}$ paths of length L in H that connect those pairs and have all their internal vertices in $\mathcal{R}$. Therefore, combining this with (3.4.1) ensures that there indeed exists a version of $\mathcal{R}$ that has all the required properties of our reservoir set. □

It is not hard now to show the preservation of the reservoir, Lemma 3.2.3.

Proof of Lemma 3.2.3. Let $H, \mathcal{R}, \mathcal{R}'$ be as in the statement of the Lemma. Consider any two disjoint ordered pairs of distinct vertices (x, y) and (w, z). We have

$$|\mathcal{R}'| \leqslant 2\vartheta^4 n \leqslant \vartheta^{3/2}\frac{\vartheta^2}{2}n \leqslant \vartheta^{3/2}|\mathcal{R}|$$

by the lower bound we get from Lemma 3.2.2. Since every particular vertex in $\mathcal{R}'$ is an internal vertex of at most $(L-2)|\mathcal{R}|^{L-3}$ of the (x, y)-(w, z)-paths of length L in H with all internal vertices from $\mathcal{R}$, the Reservoir Lemma tells us that there are at least

$$\frac{\vartheta}{2}|\mathcal{R}|^{L-2} - |\mathcal{R}'|(L-2)|\mathcal{R}|^{L-3} \geqslant \frac{\vartheta}{2}|\mathcal{R}|^{L-2} - \vartheta^{3/2}(L-2)|\mathcal{R}|^{L-2} > 0$$

such (x, y)-(w, z)-paths with all internal vertices in $\mathcal{R} \setminus \mathcal{R}'$. □

3.5 Absorbing path

In this section, we will construct a short (absorbing) path P_A that can "absorb" every small set of arbitrary vertices: For each small set $X \subseteq V$, we can build a path P'_A with $V(P'_A) = V(P_A) \cup X$ which has the the same end-pairs as P_A. Later, it will then suffice to find a cycle containing P_A and almost all vertices, and subsequently absorb the remaining vertices into P_A. Since we already have a Connecting Lemma, actually the only step left will be to find a long path.

In order to construct such an absorbing path, one first has to find many *absorbers* for each vertex v: In our case, an absorber is a "cascade" of small paths that allows us to build a new such cascade of paths with the same end-pairs, containing all vertices of the first two paths and in addition the "absorbed" vertex v (see Definition 3.5.1). This makes sure that we can maintain the path structure of P_A when absorbing a vertex since the linking pairs remain unchanged. Once we know that for every vertex v, there exist many such v-absorbers in H, the probabilistic method provides a small set of disjoint paths with the property that for every vertex v, this set contains many v-absorbers. Lastly, we will simply connect all these paths via the Connecting Lemma and note that then we can absorb a small set of vertices by greedily inserting each vertex into a different absorber.

To construct the absorbers, we again utilize that we can "climb up" the degree sequence. More precisely, we define the following "absorbers".

Definition 3.5.1. *Let $\alpha > 0$, $n \in \mathbb{N}$, set $s = s(\alpha) = 2 \cdot \frac{1}{\alpha}$, and let $H = ([n], E)$ be a 3-graph.*[1] *For $x \in [n]$, a $4s$-tuple*

$$(v_1, w_1, y_1, z_1, \ldots, v_s, w_s, y_s, z_s) \in [n]^{4s}$$

of distinct vertices is called (x, α)-absorber (in H) if

1. *$v_1 w_1 x y_1 z_1$ is a path in H,*
2. *for $i \in [s-1]$, we know that $v_i w_i y_{i+1} z_{i+1}$ and $v_{i+1} w_{i+1} y_i z_i$ are paths in H, and*
3. *$v_s w_s y_s z_s$ is a path in H.*

When α is not important, we omit it in the notation, then simply speaking of x-absorbers. Note that we can absorb x into an x-absorber $(v_1, w_1, y_1, z_1, \ldots, v_s, w_s, y_s, z_s)$ as follows, see also Figure 3.5.1. Before absorption, we consider the paths $v_i w_i y_{i+1} z_{i+1}$ and $v_{i+1} w_{i+1} y_i z_i$, for all *odd* $i \in [s]$. After absorption, we consider the path $v_1 w_1 x y_1 z_1$, the paths $v_i w_i y_{i+1} z_{i+1}$ and $v_{i+1} w_{i+1} y_i z_i$ for all *even* $i \in [s-2]$, and the path $v_s w_s y_s z_s$. Note that the (ordered) end-pairs of the considered paths are the same before and after absorption.

Lemma 3.5.2 (Many Absorbers). *Let $1/n \ll \vartheta \ll \alpha \ll 1$. If $H = ([n], E)$ is a 3-graph with $d(i,j) \geqslant \min\left(i, j, \frac{n}{2}\right) + \alpha n$ for all $ij \in [n]^{(2)}$ and $\mathcal{R}$ is a reservoir set given by Lemma 3.2.2, then for every $x \in [n]$, the number of (x, α)-absorbers in $([n] \smallsetminus \mathcal{R})^{4s(\alpha)}$ is at least $(\frac{\alpha n}{3})^{4s(\alpha)}$.*

Proof of Lemma 3.5.2. Let $1/n \ll \vartheta \ll \alpha \ll 1$, let H be as in the statement, and let $x \in [n]$. There are at least $\frac{n}{3}$ possibilities to choose a vertex $w_1 \in [n] \smallsetminus (\mathcal{R} \cup \{x\})$ with $w_1 \geqslant \min(x + \frac{\alpha n}{2}, \frac{n}{2})$. Then, there are at least $\frac{\alpha n}{3}$ choices for a vertex $v_1 \in N(w_1, x) \smallsetminus \mathcal{R}$ with $v_1 \geqslant \min(x + \frac{\alpha n}{2}, \frac{n}{2})$ since $|N(w_1, x)| \geqslant \min(w_1, x, \frac{n}{2}) + \alpha n$ and $w_1 \geqslant \min(x + \frac{\alpha n}{2}, \frac{n}{2})$. Similarly, there are at least $\frac{\alpha n}{3}$ choices for a vertex $y_1 \in N(w_1, x) \smallsetminus (\mathcal{R} \cup \{v_1\})$ with $y_1 \geqslant \min(x + \frac{\alpha n}{2}, \frac{n}{2})$ and at least $\frac{\alpha n}{3}$ choices for a vertex $z_1 \in N(x, y_1) \smallsetminus (\mathcal{R} \cup \{v_1, w_1\})$ with $z_1 \geqslant \min(x + \frac{\alpha n}{2}, \frac{n}{2})$.

[1] Recall that in our convention $\frac{1}{\alpha}$ is an integer and, hence, s is an even integer.

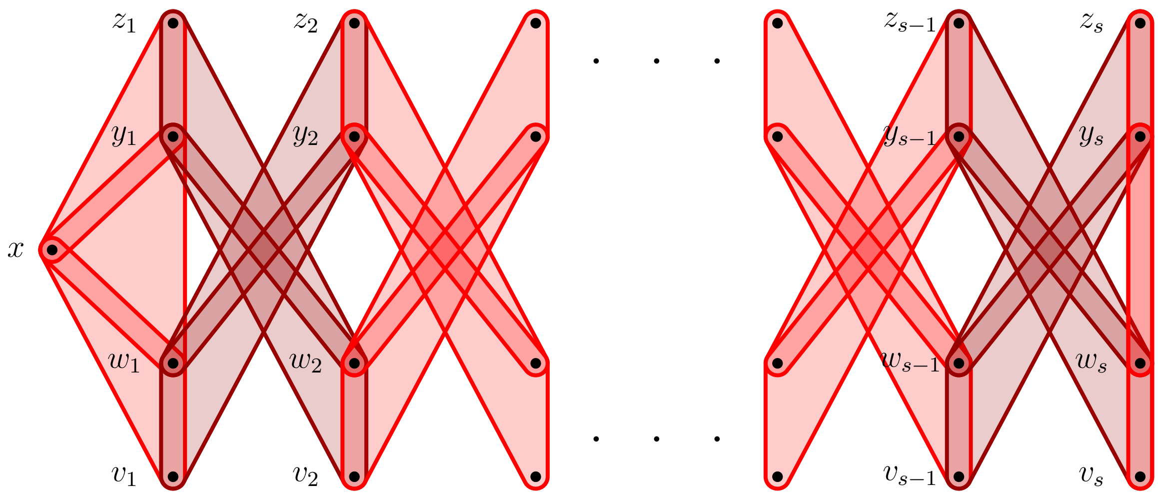


Figure 3.5.1: Structure of the absorbers with hyperedges used before absorption of x in dark red and hyperedges used after absorption of x in light red.

Now assume that for some $i \in [s-2]$, vertices v_j, w_j, y_j, and z_j have already been selected, for all $j \in [i]$, in such a way that all edges required by Definition 3.5.1 are present and $v_j, w_j, y_j, z_j \geqslant \min(x + j\frac{\alpha n}{2}, \frac{n}{2})$ for all $j \in [i]$, and denote the set containing all these vertices, all vertices from $\mathcal{R}$, and x by A_i. Note that for all $i \in [s-2]$, we have $|A_i| \leqslant \frac{\alpha n}{7}$. Therefore, there are at least $\frac{\alpha n}{3}$ choices for a vertex $w_{i+1} \in N(y_i, z_i) \smallsetminus A_i$ with $w_{i+1} \geqslant \min(x + (i+1)\frac{\alpha n}{2}, \frac{n}{2})$. Further, there are at least $\frac{\alpha n}{3}$ choices for a vertex $v_{i+1} \in N(w_{i+1}, y_i) \smallsetminus A_i$ with $v_{i+1} \geqslant \min(x+(i+1)\frac{\alpha n}{2}, \frac{n}{2})$. Similarly, there are at least $\frac{\alpha n}{3}$ choices for a vertex $y_{i+1} \in N(v_i, w_i) \smallsetminus (A_i \cup \{v_{i+1}, w_{i+1}\})$ with $y_{i+1} \geqslant \min(x+(i+1)\frac{\alpha n}{2}, \frac{n}{2})$ and at least $\frac{\alpha n}{3}$ choices for a vertex $z_{i+1} \in N(w_i, y_{i+1}) \smallsetminus (A_i \cup \{v_{i+1}, w_{i+1}\})$ with $z_{i+1} \geqslant \min(x+(i+1)\frac{\alpha n}{2}, \frac{n}{2})$.

Assume that v_j, w_j, y_j, and z_j have been selected for all $j \in [s-1]$ such that all edges required by Definition 3.5.1 are present and $v_j, w_j, y_j, z_j \geqslant \min(x+j\frac{\alpha n}{2}, \frac{n}{2})$, for all $j \in [s-1]$, and denote the set containing all these vertices, all vertices from $\mathcal{R}$, and x by A_{s-1}. Then there are at least $\frac{\alpha n}{3}$ choices for a vertex $w_s \in N(y_{s-1}, z_{s-1}) \smallsetminus A_{s-1}$ with $w_s \geqslant \min(x+s\frac{\alpha n}{2}, \frac{n}{2})$ and at least $\frac{\alpha n}{3}$ choices for a vertex $y_s \in N(v_{s-1}, w_{s-1}) \smallsetminus (A_{s-1} \cup \{w_s\})$ with $y_s \geqslant \min(x + s\frac{\alpha n}{2}, \frac{n}{2})$. Note that by the choice of s we have $v_{s-1}, w_{s-1}, y_{s-1}, z_{s-1}, w_s, y_s \geqslant \min((s-1)\frac{\alpha n}{2}, \frac{n}{2}) = \frac{n}{2}$. Thus, we know that

$$|N(w_s, y_{s-1}) \cap N(w_s, y_s)| \geqslant \frac{n}{2} + \alpha n + \frac{n}{2} + \alpha n - n \geqslant 2\alpha n$$

and so there are at least αn choices for $v_s \in N(w_s, y_{s-1}) \cap N(w_s, y_s) \smallsetminus A_{s-1}$ and, similarly, we know that there are at least αn choices for $z_s \in N(w_{s-1}, y_s) \cap N(w_s, y_s) \smallsetminus (A_{s-1} \cup \{v_s\})$.

Observe that if the vertices $v_1, w_1, y_1, z_1, \ldots, v_s, w_s, y_s, z_s$ are chosen in the respective neighbourhoods as described above, they form an (x, α)-absorber. Hence, the number of (x, α)-absorbers is indeed at least $(\frac{\alpha}{3}n)^{4s(\alpha)}$. □

We are now ready to prove Lemma 3.2.4.

Proof of Lemma 3.2.4. The proof proceeds in two steps. First, we will use the probabilistic method, showing that with positive probability a randomly chosen set of $4s$-tuples contains many absorbers for every vertex while being not too large. In the second, part we connect all those paths using the Connecting Lemma.

Let $1/n \ll \vartheta \ll \alpha$, let $L \in \mathbb{N}$ be given by the Connecting Lemma, let $s = s(\alpha)$, and let $H, \mathcal{R}$ be given as in the statement.

Let $\mathcal{X} \subseteq ([n] \setminus \mathcal{R})^{4s}$ be a random selection in which each $4s$-tuple in $([n] \setminus \mathcal{R})^{4s}$ is included independently with probability $p := \frac{\vartheta^2 3^{4s+2}}{\alpha^{4s} n^{4s-1}}$. Then $\mathbb{E}[|\mathcal{X}|] \leqslant pn^{4s} = \frac{\vartheta^2 3^{4s+2}}{\alpha^{4s}} n$ and by Markov's inequality we get

$$\mathbb{P}\left(|\mathcal{X}| > 2\frac{\vartheta^2 3^{4s+2}}{\alpha^{4s}} n\right) \leqslant \frac{1}{2}. \tag{3.5.1}$$

Calling two distinct $4s$-tuples *overlapping* if they contain a common vertex, we observe that there are at most $(4s)^2 n^{8s-1}$ ordered pairs of overlapping $4s$-tuples. Let us denote the number of overlapping pairs with both of their tuples occurring in $\mathcal{X}$ by D. We then get $\mathbb{E}[D] \leqslant (4s)^2 n^{8s-1} p^2 = (4s)^2 \big(\frac{\vartheta^2 3^{4s+2}}{\alpha^{4s}}\big)^2 n$ and Markov yields

$$\mathbb{P}\left[D > \vartheta^2 n\right] \leqslant \mathbb{P}\left[D > 64s^2\left(\frac{\vartheta^2 3^{4s+2}}{\alpha^{4s}}\right)^2 n\right] \leqslant \frac{1}{4} \tag{3.5.2}$$

since $1/n \ll \vartheta \ll \alpha$.

Next, we focus on the number of absorbers contained in $\mathcal{X}$. For $x \in [n]$, let A_x denote the set of all (x, α)-absorbers. Lemma 3.5.2 gives that for every $x \in [n]$,

$$\mathbb{E}[|A_x \cap \mathcal{X}|] \geqslant \left(\frac{\alpha n}{3}\right)^{4s} p = 9\vartheta^2 n.$$

Since $|A_x \cap \mathcal{X}|$ is binomially distributed, we may apply Chernoff's inequality to get for every $x \in [n]$,

$$\mathbb{P}\left(|A_x \cap \mathcal{X}| \leqslant 3\vartheta^2 n\right) \leqslant \exp\left(-\frac{\left(\frac{2}{3}\right)^2}{2} 9\vartheta^2 n\right) < \frac{1}{5n}. \tag{3.5.3}$$

Hence, by the union bound and (3.5.1), (3.5.2) and (3.5.3), there exists a selection $\mathcal{F}_* \subseteq ([n] \smallsetminus \mathcal{R})^{4s}$ with:

- $|\mathcal{F}_*| \leqslant \frac{2\vartheta^2 3^{4s+2}}{\alpha^{4s}} n$
- $\mathcal{F}_*$ contains at most $\vartheta^2 n$ overlapping pairs
- $\mathcal{F}_*$ contains at least $3\vartheta^2 n$ x-absorbers, for every $x \in [n]$

For each overlapping pair, we delete one of its $4s$-tuples and thus, for every $x \in [n]$, we lose at most $\vartheta^2 n$ x-absorbers. Furthermore, we delete every $4s$-tuple $A \in \mathcal{F}_*$ for which there does not exist an $x \in [n]$ such that A is an x-absorber. Note that now every remaining tuple has all edges present as in Definition 3.5.1 and all its vertices are distinct. This deletion process gives rise to an $\mathcal{F} \subseteq ([n] \smallsetminus \mathcal{R})^{4s}$ satisfying:

- $|\mathcal{F}| \leqslant \frac{2\vartheta^2 3^{4s+2}}{\alpha^{4s}} n$,
- for every $4s$-tuple $A \in \mathcal{F}$ there is an $x \in [n]$ such that A is an x-absorber, in particular, all the vertices in A are distinct and there are edges present as in Definition 3.5.1, and
- for every $x \in [n]$, there are at least $2\vartheta^2 n$ x-absorbers in $\mathcal{F}$.

Next, we want to connect the elements in $\mathcal{F}$ to a path utilising the Connecting Lemma. Let $\mathcal{G}$ be the set consisting of all the paths $v_i w_i y_{i+1} z_{i+1}$ and $v_{i+1} w_{i+1} y_i z_i$ for odd i and for each $(v_1, w_1, y_1, z_1, \ldots, v_s, w_s, y_s, z_s) \in \mathcal{F}$:

$$\mathcal{G} = \bigcup_{(v_1,w_1,y_1,z_1,\ldots,v_s,w_s,y_s,z_s)\in\mathcal{F}} \left\{ v_{i+j} w_{i+j} y_{i+1-j} z_{i+1-j} : i \in [s] \text{ odd}, j \in \{0,1\} \right\}$$

We then have $|\mathcal{G}| = 2|\mathcal{F}| \leqslant \frac{4\vartheta^2 3^{4s+2}}{\alpha^{4s}} n$. Let $\mathcal{G}^* \subseteq \mathcal{G}$ be a maximal subset such that there exists a path $P^* \subseteq H - \mathcal{R}$ with:

- P^* contains all paths in $\mathcal{G}^*$ as subpaths
- $V(P^*) \cap \bigcup_{P \in \mathcal{G} \smallsetminus \mathcal{G}^*} V(P) = \varnothing$
- P^* satisfies $v(P^*) = (L+2)(|\mathcal{G}^*| - 1) + 4$.

First assume $\mathcal{G}^* \subsetneq \mathcal{G}$, and let $Q^* \in \mathcal{G} \smallsetminus \mathcal{G}^*$. Notice that recalling $1/n \ll \vartheta \ll \alpha, 1/L \ll 1$, we have

$$v(P^*) + \Big| \bigcup_{P \in \mathcal{G} \smallsetminus \mathcal{G}^*} V(P) \Big| + |\mathcal{R}| \leqslant (L+2) \frac{4\vartheta^2 3^{4s+2}}{\alpha^{4s}} n + \vartheta^2 n \leqslant \frac{\vartheta n}{2(L-2)}. \tag{3.5.4}$$

Now Lemma 3.2.1 tells us that there are at least ϑn^{L-2} paths of length L connecting the ending-pair (a,b) of P^* with the starting-pair (b,c) of Q^* (which are disjoint by the choice of P^*). By (3.5.4), at least half of those are disjoint to $\mathcal{R} \cup \bigcup_{P \in \mathcal{G} \setminus (\mathcal{G}^* \cup \{Q^*\})} V(P)$ and (apart from the end-pairs) disjoint to $V(P^*)$ and $V(Q^*)$. Hence, there exists a path P^{**} starting with P^* and ending with Q^* whose vertex set is disjoint to $\mathcal{R} \cup \bigcup_{P \in \mathcal{G} \setminus (\mathcal{G}^* \cup \{Q^*\})} V(P)$ and for which we further have

$$v(P^{**}) = v(P^*) + L - 2 + v(Q^*) = 4 + (L+2)\left(|\mathcal{G}^* \cup \{Q^*\}| - 1\right).$$

Therefore, $\mathcal{G}^* \cup \{Q^*\}$ contradicts the maximality of $\mathcal{G}^*$ and thus, $\mathcal{G}^* = \mathcal{G}$. Further, for $P_A := P^*$, the hierarchy $1/n \ll \vartheta \ll \alpha, 1/L \ll 1$ gives us the required bound on $v(P_A)$:

$$v(P_A) \leqslant 4 + (L+2)\frac{4\vartheta^2 3^{4s+2}}{\alpha^{4s}} n \leqslant \vartheta n.$$

Lastly, the structure and the number of the absorbers in P_A ensure the absorbing property: Let $X \subseteq [n]$ with $|X| \leqslant 2\vartheta^2 n$. For each $x \in X$, we can choose one x-absorber $(v_1, w_1, y_1, z_1, \ldots, v_s, w_s, y_s, z_s)$ from $\mathcal{F}$ such that all chosen absorbers are distinct, since for every $x \in V$, the number of x-absorbers in $\mathcal{F}$ is at least $2\vartheta^2 n$. For every $x \in X$, we then "open" P_A at the paths $v_{i+j}w_{i+j}y_{i+1-j}z_{i+1-j}$ for $i \in [s]$ odd and $j \in \{0,1\}$ and reconnect it to a path containing x by instead considering the paths $v_{i+j}w_{i+j}y_{i+1-j}z_{i+1-j}$, for all even $i \in [s]$ and $j \in \{0,1\}$, and the paths $v_1w_1xy_1z_1$ and $v_sw_sy_sz_s$. That leaves us with a path P' which satisfies $V(P') = V(P_A) \cup X$ and has the same end-pairs as P_A. □

3.6 Long path

In this section, we will prove the existence of a path that contains almost all vertices. To do so, we will need a weak form of the hypergraph regularity method which we will therefore introduce briefly.

Let $H = (V, E)$ be a 3-graph and $V_1, V_2, V_3 \subseteq V$; we write

$$E(V_1, V_2, V_3) = \{(v_1, v_2, v_3) \in V_1 \times V_2 \times V_3 : v_1v_2v_3 \in E\}$$

and $e(V_1, V_2, V_3) = |E(V_1, V_2, V_3)|$. Further, we write

$$H(V_1, V_2, V_3) = (V_1 \dot\cup V_2 \dot\cup V_3, E(V_1, V_2, V_3)).$$

For $\delta > 0, d \geqslant 0$ and $V_1, V_2, V_3 \subseteq V$, we say that $H(V_1, V_2, V_3)$ is *weakly* (δ, d)-*quasirandom* if for all $U_1 \subseteq V_1, U_2 \subseteq V_2, U_3 \subseteq V_3$, we have that

$$|e(U_1, U_2, U_3) - d|U_1||U_2||U_3|| \leqslant \delta |V_1||V_2||V_3|.$$

We say that $H(V_1, V_2, V_3)$ is weakly δ-quasirandom if it is weakly (δ, d)-quasirandom for some $d \geqslant 0$. For brevity, we might also say that V_1, V_2, V_3 are weakly (δ, d)-quasirandom (or δ-quasirandom) (in H). Lastly, since we only look at weak quasirandomness in this section, we may omit the prefix "weakly".

The regularity lemma is a strong tool in extremal combinatorics. While the full generalisation to hypergraphs is more involved than the version for graphs, there is also a light version for hypergraphs that can already be useful and indeed it is for us:

Lemma 3.6.1 (Weak Hypergraph Regularity Lemma)**.** *For $\delta > 0, t_0 \in \mathbb{N}$, there exists a $T_0 \in \mathbb{N}$ such that for every 3-graph $H = ([n], E)$ with $n \geqslant t_0$, there exist an integer t with $t_0 \leqslant t \leqslant T_0$ and a partition $[n] = V_0 \dot\cup V_1 \dot\cup \ldots \dot\cup V_t$ such that:*

- $|V_0| \leqslant \delta n$ *and* $|V_1| = \cdots = |V_t|$
- *for $i \geqslant 1$, we have* $\max(V_i) \leqslant \max(V_{i+1})$ *and* $\max(V_i) - \min(V_i) \leqslant \frac{n}{t_0}$
- *there are at most δt^3 sets $ijk \in [t]^{(3)}$ such that the "triplet" V_i, V_j, V_k, also written as V^{ijk}, is not δ-quasirandom in H.*

For a proof of Lemma 3.6.1 see for instance [17, 40, 109]. One can get the slight extra requirement on the ordering of the vertices by dividing the vertex set in intervals of length $\frac{n}{t_0}$ and afterwards going on with the proof refining those sets. This has been remarked before, e.g., by Reiher, Rödl, and Schacht in [96].

We will regularise H and then observe that a quasirandom triplet V^{ijk} with positive density can almost be covered with not too short disjoint paths. Thus, we can think of the situation as a reduced hypergraph with the partition classes as vertices and edges encoding those "good triplets" that in H we can almost cover with paths. At that point we will notice that the degree condition can almost be transferred to the reduced hypergraph. In Lemma 3.6.3, we prove that this degree condition will ensure an almost perfect matching in the reduced hypergraph. But that means that in H almost all vertices can be covered with paths, which we can then connect through the reservoir to a long path in H.

Lemma 3.6.2 (Good Triplets)**.** *For $\xi > 0, d > 0, \delta > 0, n \in \mathbb{N}$ with $\frac{d\xi^3 - \delta}{2} n \geqslant 1$, the following holds. Let $H = (U \dot\cup V \dot\cup W, E)$ with $|U|, |V|, |W| = n$ be a 3-graph and suppose* *that U, V, W are (δ, d)-quasirandom in H. Then at least $(1-\xi)3n$ vertices of H can be covered by vertex-disjoint paths of length at least $\frac{d\xi^3-\delta}{2}n-2$.*

Proof of Lemma 3.6.2. For convenience set $c=\frac{d\xi^3-\delta}{6}n$. Let $\mathcal{P}$ be a maximal set of vertex-disjoint paths of length $3c-2$ in H, where each path takes alternatingly vertices from each partition class, i.e., each path is of the form

$$u_1v_1w_1u_2v_2w_2\ldots u_cv_cw_c$$

with $u_i\in U, v_i\in V, w_i\in W$.

Assume that $|V|-|\bigcup_{P\in\mathcal{P}}V(P)|>3\xi n$. Then the sets

$$U':=U\smallsetminus\bigcup_{P\in\mathcal{P}}V(P), V':=V\smallsetminus\bigcup_{P\in\mathcal{P}}V(P), W':=W\smallsetminus\bigcup_{P\in\mathcal{P}}V(P)$$

satisfy $|U'|, |V'|, |W'|>\xi n$.

Next, we will delete all the edges that contain vertex pairs of small pair degree. With the edges that still remain after this process we can build a path of the required length.

We start with $F_1=H[U',V',W']$ and set F_{i+1}, for $i\geqslant 1$, as the hypergraph obtained from F_i by deleting all edges containing a vertex pair xy with $d^{\times}_{F_i}(x,y)\leqslant c$, where $d^{\times}_{F_i}(x,y)=|\{e\in E(F_i): x,y\in e, |e\cap U'|=|e\cap V'|=|e\cap W'|=1\}|$. This process stops with a hypergraph F_j in which for all $x,y\in V(F_j)$, we either have $d^{\times}_{F_j}(x,y)=0$ or $d^{\times}_{F_j}(x,y)\geqslant c$. The deletion condition guarantees

$$e^{\times}(F_1)-e^{\times}(F_j)\leqslant 3cn^2\,,$$

with $e^{\times}(F_i)=|\{e\in E(F_i): |e\cap U'|=|e\cap V'|=|e\cap W'|=1\}|$, and the quasirandomness of U, V, W gives that $e^{\times}(F_1)=e(U',V',W')\geqslant(d\xi^3-\delta)n^3$. Thus, there still exists an edge $u_1v_1w_1$ in F_j with $u_1\in U'$, $v_1\in V'$ and $w_1\in W'$. But this means that there is a path of length $3c-2$ in F_j: Let $P^*=u_1v_1w_1\ldots u_kv_kw_k$ be a maximal path in F_j with $u_i\in U', v_i\in V'$ and $w_i\in W'$, for all $i\in[k]$ (note that $k\geqslant 1$). Assuming $k<c$ for a contradiction, less than c vertices of U' appear in P^*. But since v_kw_k is contained in the edge $u_kv_kw_k\in E^{\times}(F_j)$, we actually have that $d^{\times}_{F_j}(v_k,w_k)\geqslant c$, whence there is a $u_{k+1}\in U'\smallsetminus V(P^*)$ such that P^*u_{k+1} is a path in F_j.

The same argument applied to w_ku_{k+1} gives a $v_{k+1}\in V'$ such that $P^*u_{k+1}v_{k+1}$ is a path in F_j and finally applying the argument to $u_{k+1}v_{k+1}$ gives rise to a $w_{k+1}\in W'$ such that the path $P^*u_{k+1}v_{k+1}w_{k+1}$ exists in F_j and thus contradicts the maximality of P^*, telling us that P^* actually contains an alternating path of length $3c - 2$. That, on the other hand, gives us another alternating path of length at least $3c - 2$ that is vertex-disjoint to all paths in $\mathcal{P}$ and, therefore, contradicts the maximality of $\mathcal{P}$. So we indeed have $|V| - |\bigcup_{P \in \mathcal{P}} V(P)| \leqslant 3\xi n$. $\square$

As mentioned before, we later want to find an almost perfect matching in a reduced hypergraph whose edges represent "good" triplets as in Lemma 3.6.2. Then "translating back" those edges in the matching will give us a set of (not too many) paths in H which almost covers all vertices. To find an almost perfect matching in a hypergraph satisfying the pair degree condition in Theorem 3.1.4 for almost all pairs, we look at a maximal matching in which the sum of the vertices not contained in it is also maximal. This should give us the best chance to enlarge the matching if too many vertices would be left over, deriving a contradiction. A similar maximisation idea has also been used in [112] when a degree sequence condition was given for a graph. The following Lemma will later guarantee the existence of an almost perfect matching in the reduced hypergraph.

Lemma 3.6.3 (Matching)**.** *Let $1/n \ll \alpha, \beta$. If $H = ([n], E)$ is a 3-graph, G_H is a graph on vertex set $[n]$ with maximum degree $\Delta(G_H) \leqslant \beta n$ and H satisfies $d(i,j) \geqslant \min\left(i, j, \frac{n}{2}\right) + \alpha n$, for all $ij \in [n]^{(2)}$ with $ij \notin E(G_H)$, then H has a matching M with $v(M) \geqslant (1 - 3\beta) n$.*

Proof of Lemma 3.6.3. Without restriction let $\alpha \ll 1$ and $\beta < 1/3$ and let H, G_H be given as in the statement. For matchings $M_1, M_2 \subseteq H$ of maximal size, we write $M_1 \prec M_2$ if $[n] \smallsetminus V(M_1) \leqslant_{\text{lex}} [n] \smallsetminus V(M_2)$, where $\leqslant_{\text{lex}}$ is the usual lexicographic order on $\mathscr{P}([n])$, i.e., $A \leqslant_{\text{lex}} B$ if $\min A \triangle B \in A$. Now, let $M \subseteq H$ be a matching of maximal size which is (subject to being of maximal size) maximal with respect to $\prec$. Assuming the statement is false, gives an $A \subseteq [n] \smallsetminus V(M)$ with $|A| \geqslant 3\beta n$. Let us call a pair *true* if it is not an edge in G_H. Since $\Delta(G_H) \leqslant \beta n$, we can find $2\beta n$ distinct vertices $v_1, \ldots, v_{\beta n}, w_1, \ldots, w_{\beta n} \in A$ such that all the pairs $v_i w_i$ are true. Without restriction assume that $v_i < w_i$. Notice that all the neighbours of each such pair lie inside $V(M)$, otherwise adding the respective edge to M would lead to a larger matching. In the following, we will show two properties and afterwards deduce the statement from them.

Firstly, we have that for each $v_i w_i$, there are at least $\frac{\alpha n}{3}$ edges in M in which $v_i w_i$ has at least two neighbours: Let us first consider a pair $v_i w_i$ with $v_i \leqslant \frac{n}{2}$. For any edge abc of the matching with $a \in N(v_i, w_i)$, we have that $\min\{b, c\} \leqslant v_i$ as otherwise $E(M) \smallsetminus \{abc\} \cup \{a v_i w_i\}$ would be the edge set of a matching M' with the same size as M but with $M \prec M'$, contradicting our choice of M. This means that in each edge of M which contains only one neighbour of v_iw_i there is one vertex $\leqslant v_i$. Thus, (and since all those edges are disjoint), at most v_i neighbours of v_iw_i can lie in edges that contain no further neighbour of v_iw_i. Hence, recalling $d\,(v_i, w_i) \geqslant v_i + \alpha n$, at least $\frac{\alpha n}{3}$ edges in M contain at least two neighbours of v_iw_i.

For a pair v_iw_i with $v_i \geqslant n/2$, there exist at least $\frac{\alpha n}{3}$ edges in M containing more than one neighbour of v_iw_i as well since $d\,(v_i, w_i) \geqslant \frac{n}{2} + \alpha n$ but $e(M) \leqslant n/3$.

Secondly, note that any edge of M that contains at least two neighbours of one true pair v_iw_i cannot contain a neighbour of any other true pair v_jw_j: Assume for contradiction there were true pairs v_iw_i and v_jw_j together with an edge $abc \in E(M)$ such that $a \in N\,(v_i, w_i)$ and $|\{abc\} \cap N\,(v_j, w_j)| \geqslant 2$. Then b or c, without restriction b, is a neighbour of v_jw_j and $E(M) \smallsetminus \{abc\} \cup \{av_iw_i, bv_jw_j\}$ is the edge set of a matching in H contradicting the maximal size of M.

Summarised, for each of the βn true pairs v_iw_i in $[n] \smallsetminus V(M)$, we get a set of at least $\frac{\alpha n}{3}$ edges in M that contain more than one neighbour of the respective pair and thus all those sets of edges are pairwise disjoint. Therefore, we have $\frac{\alpha n}{3} \times \beta n$ distinct edges in M which is a contradiction to $1/n \ll \alpha, \beta$. So M was indeed a matching satisfying $v(M) \geqslant (1 - 3\beta)\,n$. $\square$

We are now ready to prove Proposition 3.2.5. For that we will apply the Weak Regularity Lemma to H (actually to a slightly smaller subgraph), obtain a pair degree condition for the reduced hypergraph and hence find a matching in it by the previous Lemma. Lastly, we will "unfold" the edges of that matching to paths in H by Lemma 3.6.2 and connect these to a long path by the Connecting Lemma.

Proof of Proposition 3.2.5. Let α, ϑ be given as in the Proposition and set $\alpha' = \alpha - \vartheta - \vartheta^2$. Next choose ξ, δ, t_0 such that we have $1/t_0 \ll \delta \ll \xi \ll \vartheta \ll \alpha'$. Applying the Weak Regularity Lemma 3.6.1 to δ and t_0 gives us a T_0 and by the hierarchy in the Proposition, we may assume $1/n \ll 1/T_0$. Now let H, $\mathcal{R}$, and P_A be given as in the statement. Notice that $H' = H\big[[n] \smallsetminus (\mathcal{R} \cup V\,(P_A))\big]$ after a renaming of the vertices can be seen as a 3-graph $H' = ([n'], E')$ with $n' \geqslant (1 - \vartheta^2 - \vartheta)\,n$ and satisfying the usual degree condition: $d(i,j) \geqslant \min\left(i, j, \frac{n'}{2}\right) + \alpha' n'$ for all $ij \in [n']^{(2)}$.

For H', the statement of the Weak Regularity Lemma provides an integer $t \in [t_0, T_0]$ and a partition $V = V_0 \dot\cup V_1 \dot\cup V_2 \dot\cup \ldots \dot\cup V_t$ satisfying all three points of Lemma 3.6.1. Setting $m = |V_1| = \cdots = |V_t|$, we have that $\frac{n'}{t} \geqslant m \geqslant \frac{1-\delta}{t} n'$ and recall that $|V_0| \leqslant \delta n'$. Note that for $v_i \in V_i$, we have $v_i \geqslant i \cdot m - \frac{n'}{t_0}$. Summarised, we have the following hierarchy:

$$\frac{1}{n'} \ll \frac{1}{T_0}, \frac{1}{t}, \frac{1}{t_0} \ll \delta \ll \xi \ll \vartheta \ll \alpha' \ll 1 \tag{3.6.1}$$

Let us write $e^{\times}\left(V^{ijk}\right) = |\{e \in E' : |e \cap V_i| = |e \cap V_j| = |e \cap V_k| = 1\}|$ for the number of crossing edges in V^{ijk} and we call a triplet V^{ijk} dense, if $e^{\times}\left(V^{ijk}\right) \geqslant \frac{\alpha' m^3}{2}$.

Now we will show that we can almost "transfer" the pair degree condition to a reduced hypergraph. We will do this in two steps: First, we show that every pair V_iV_j belongs to many dense triplets V^{ijk}, and second, we show that we can almost keep that up when restricting ourselves to quasirandom triplets.

Claim 3.6.4. *For every $ij \in [t]^{(2)}$, there are at least $\min\left(i, j, \frac{t}{2}\right) + \frac{\alpha' t}{3}$ many $k \in [t] - \{i, j\}$ such that V^{ijk} is a dense triplet.*

Proof. Suppose there is a pair V_iV_j, $ij \in [t]^{(2)}$, belonging to less than $\min\left(i, j, \frac{t}{2}\right) + \frac{\alpha' t}{3}$ dense triplets V^{ijk}. Let S be the set of hyperedges in H' that contain one vertex in V_i, one in V_j and a third vertex outside of $V_i \dot\cup V_j$. By invoking the pair degree condition of H' and with the hierarchy (3.6.1), we get that

$$\begin{aligned}|S| \geqslant & m^2\left[\min\left(i \cdot m - \frac{n'}{t_0}, j \cdot m - \frac{n'}{t_0}, \frac{n'}{2}\right) + \alpha' n' - 2m\right]\\ > & \frac{n'^3}{t^2}\left(\min\left(\frac{i}{t}, \frac{j}{t}, \frac{1}{2}\right) + \frac{6}{7}\alpha'\right)\end{aligned}$$

We will derive a contradiction by finding a smaller upper bound on $|S|$. To this aim, we split S into two parts. By S_1 let us denote the set of those edges in S that lie in a dense triplet V^{ijk}, for some $k \in [t] \setminus \{i, j\}$, (we say an edge e lies or is in V^{ijk} if we have $|e \cap V_i| = |e \cap V_j| = |e \cap V_k| = 1$). Since in one triplet there are at most m^3 edges and by assumption V_iV_j does not belong to many dense triplets, we get

$$|S_1| < \left(\min\left(i, j, \frac{t}{2}\right) + \frac{\alpha' t}{3}\right) m^3 \leqslant \frac{n'^3}{t^2}\left(\min\left(\frac{i}{t}, \frac{j}{t}, \frac{1}{2}\right) + \frac{\alpha'}{3}\right)$$

Let $S_2 = S \setminus S_1$ be the set of edges in S lying in triplets that are not dense. There are less than $\frac{\alpha'}{2} m^3$ crossing edges in each triplet that is not dense and V_iV_j belongs to at most t triplets. Hence

$$|S_2| < \frac{\alpha'}{2} m^3 \times t \leqslant \frac{n'^3}{t^2}\frac{\alpha'}{2}.$$

Summarised, we have

$$\frac{n'^3}{t^2}\left(\min\left(\frac{i}{t}, \frac{j}{t}, \frac{1}{2}\right) + \frac{6}{7}\alpha'\right) < |S| = |S_1| + |S_2| < \frac{n'^3}{t^2}\left(\min\left(\frac{i}{t}, \frac{j}{t}, \frac{1}{2}\right) + \frac{5\alpha'}{6}\right),$$

which is a contradiction. $\square$

From the Weak Regularity Lemma we also get that in total at most δt^3 triplets V^{ijk} are not δ-quasirandom.

Let us now complete the "reduction" of the hypergraph and notice that we can find an almost perfect matching in the reduced hypergraph. Denote by D the hypergraph on the vertex set $[t]$ with ijk being an edge if and only if the triplet V^{ijk} is dense. Let, on the other hand, IR be the hypergraph on the vertex set $[t]$ with ijk being an edge if and only if V^{ijk} is not weakly δ-quasirandom in H'. In the following, we will remove a few vertices in such a way that $D - IR$ induced on the remaining vertices satisfies our pair degree condition for almost all pairs.

We call a pair $ij \in [t]^2$ *malicious pair* if it belongs to more than $\sqrt{\delta}t$ edges of IR. Since $e(IR) \leqslant \delta t^3$, there are at most $3\sqrt{\delta}t^2$ malicious pairs. Let B be the graph on vertex set $[t]$ in which the edges are given by the malicious pairs. We call a vertex i *malicious vertex* if $d_B(i) > \delta^{1/4}t$, i.e., if it belongs to many malicious pairs. The upper bound on the number of malicious pairs implies that there are at most $6\delta^{1/4}t$ malicious vertices. Now we remove these malicious vertices and set $D' := D\big[[t] \smallsetminus \{v \in [t] : v \text{ malicious}\}\big]$ and $B' = B\big[[t] \smallsetminus \{v \in [t] : v \text{ malicious}\}\big]$.

The reduced hypergraph we looked for is now $K = D' - IR$, in which edges encode dense, δ-quasirandom triplets. In K, every pair $ij \in V(K)^{(2)}$ with $ij \notin E(B')$ satisfies

$$d_K(i,j) \geqslant \min\left(i, j, \frac{t}{2}\right) + \left(\frac{\alpha'}{3} - 6\delta^{1/4} - \sqrt{\delta}\right)t \geqslant \min\left(i, j, \frac{t}{2}\right) + \frac{\alpha'}{4}t.$$

Thus, we have that the graph G_K on vertex set $V(K)$ with ij being an edge if and only if ij does not satisfy the degree condition $d_K(i,j) \geqslant \min\left(i, j, \frac{v(K)}{2}\right) + \frac{\alpha'}{4}v(K)$ is a subgraph of B'. Therefore, and since $v(K) \geqslant (1 - 6\delta^{1/4})t$, we have

$$\Delta\left(G_K\right) \leqslant \Delta\left(B'\right) \leqslant \delta^{1/4}t \leqslant 2\delta^{1/4}v(K)$$

and we can apply Lemma 3.6.3 to K with $\frac{\alpha'}{4}$ in place of α and $2\delta^{1/4}$ instead of β and obtain a matching M in K covering all but at most $6\delta^{1/4}t$ vertices of K.

Finally, notice that each triplet V^{ijk} with ijk being an edge in K is (δ, d_{ijk})-quasirandom with $d_{ijk} \geqslant \frac{\alpha'}{2} - \delta \geqslant \frac{\alpha'}{3}$. Hence, we may apply Lemma 3.6.2 (with ξ as in (3.6.1), $d_{ijk} \geqslant \frac{\alpha'}{3}$ in place of d and δ as δ) to each of the triplets V^{ijk} that corresponds to an edge in M. Doing so and recalling the definition of H', we notice that in H we can cover at least

$$n - \left(\left(\delta + 6\delta^{1/4} + 6\delta^{1/4} + \xi\right)n' + |\mathcal{R}| + v\left(P_A\right)\right) \geqslant n - \left(2\vartheta^2 n + v\left(P_A\right)\right)$$

vertices with paths of length at least $\frac{\frac{\alpha'}{3}\xi^3-\delta}{2}m-2$ that are all disjoint to $\mathcal{R}$ and $V(P_A)$. We can connect all those at most $\frac{3t}{\frac{\alpha'}{3}\xi^3-\delta}$ paths in H through $\mathcal{R}$ to a path Q by Lemma 3.2.3 since until we connect the last one we have still only used at most

$$(L-2)\cdot\frac{3t}{\frac{\alpha'}{3}\xi^3-\delta}<\vartheta^4 n$$

vertices from $\mathcal{R}$ (recall the hierarchy (3.6.1)). In fact, we have that Q has at most a small intersection with $\mathcal{R}$, that is, $|V(Q)\cap\mathcal{R}|\leqslant\vartheta^4 n$ and it covers many vertices, i.e., $v(Q)\geqslant(1-2\vartheta^2)\,n-v(P_A)$. Hence, Q is a path satisfying the claims in the statement. □

3.7 Concluding remarks

We would like to finish by pointing to some related problems. Firstly, as mentioned in the introduction, our result can be seen as a stepping stone towards a complete characterisation of those pair degree matrices that force a 3-graph to contain a Hamiltonian cycle.

Further, it seems possible to generalise our proof without too much effort for k-uniform hypergraphs $H=([n],E)$ with n large satisfying the $(k-1)$-degree condition

$$d_{k-1}(i_1,\dots,i_{k-1})\geqslant\min\left(i_1,\dots,i_{k-1},\frac{n}{2}\right)+\alpha n\,,$$

where $d_{k-1}(i_1,\dots,i_{k-1})=|\{\,e\in E:\{i_1,\dots,i_{k-1}\}\subseteq e\,\}|$.

Another very interesting problem is to get a similar result for the vertex degree, strengthening the result by Reiher, Rödl, Ruciński, Schacht, and Szemerédi in [95]: Does every 3-graph $H=([n],E)$ with $d(i)\geqslant\min\left(\max\left(i,\gamma n\right),\frac{5}{9}n\right)+\alpha n$ for some $\gamma<5/9$ contain a Hamiltonian cycle if n is large? The proof of Theorem 3.1.3 in [95] depends on the existence of *robust subgraphs* for every vertex, for which one needs the factor 5/9.

Lastly, one could try to improve Theorem 3.1.4 by weakening the pair degree condition to $d(i,j)\geqslant\min\left(i,j,\frac{n}{2}\right)$, i.e., without the additional αn term, as Rödl, Ruciński, and Szemerédi did for the minimum pair degree condition in [100].

4. Decomposing hypergraphs into cycle factors

4.1 Introduction

Decompositions are a very active branch of extremal combinatorics. One of the earliest results regarding decompositions of graphs is Walecki's theorem, which states that a complete graph on an odd number of vertices has a decomposition into (edge-disjoint) Hamiltonian cycles. In recent years, there have been many breakthroughs in the area of decompositions, such as the verification of the existence of designs [50, 68], the resolution of the Oberwolfach problem [49], and the proof of Ringel's conjecture [69, 86].

A classic result by Dirac states that a graph on $n \geqslant 3$ vertices with minimum degree at least $n/2$ contains a Hamiltonian cycle. It is natural to ask how many edge-disjoint Hamiltonian cycles exist in this setting. Nash-Williams [87] showed that there are $\lfloor 5n/224 \rfloor$ edge-disjoint Hamiltonian cycles. As the union of edge-disjoint Hamiltonian cycles is an even-regular spanning subgraph, there are at most $r/2$ edge-disjoint Hamiltonian cycles where r is the largest even integer for which there exists an r-regular spanning subgraph. Thus, for a graph G, we define $\mathrm{reg}_2(G)$ to be the largest even integer r such that G contains a spanning r-regular subgraph and set $\mathrm{reg}_2(n, \delta) = \min\{\mathrm{reg}_2(G) \colon |V(G)| = n, \delta(G) = \delta\}$. Csaba, Kühn, Lo, Osthus, and Treglown [24] improved the result by Nash-Williams by showing that all large graphs G on n vertices contain at least $\mathrm{reg}_2(n, \delta(G))/2$ edge-disjoint Hamiltonian cycles. Kühn, Lapinskas, and Osthus [75] conjectured that this can be strengthened as each single G may have $\mathrm{reg}_2(G)/2$ edge-disjoint Hamiltonian cycles provided $\delta(G) \geqslant n/2$. They also asked whether an approximate version is true; namely, that any graph G on n vertices with $\delta(G) \geqslant (1/2 + o(1))n$ contains $(1/2 - o(1))\,\mathrm{reg}_2(G)$ edge-disjoint Hamiltonian cycles. Subsequently, this was proved by Ferber, Krivelevich, and Sudakov [36].

The main result of this paper implies an analogous statement for k-uniform hypergraphs for $k \geqslant 2$. To state our results, we recall some terminology. For an integer $k \geqslant 2$, a hypergraph H is called *k-uniform hypergraph* or *k-graph* if all its edges have size k. We call a k-graph whose vertex set has a cyclic ordering such that its edge set consists of all sets of k consecutive vertices in this ordering a *(tight) cycle* (we only consider tight cycles in this article). The *length* of a cycle C is defined as the number of edges in C. As usual, a *Hamiltonian cycle* in H is a cycle containing all vertices of H. Let $\delta_{k-1}(H) = \min |\{e \in E(H) : \mathbf{x} \subseteq e\}|$ where the minimum is taken over all $(k-1)$-sets $\mathbf{x} \subseteq V(H)$. In analogy to the above, we define $\mathrm{reg}_k(H)$ as the largest integer r divisible by k such that H contains a spanning subgraph F in which each vertex of F belongs to exactly r edges of F.

Dirac's result was first generalised to hypergraphs by Rödl, Ruciński, and Szemerédi in [98–100]. They showed that any k-graph H on n vertices with $\delta_{k-1}(H) \geqslant (\frac{1}{2} + o(1))n$ contains a Hamiltonian cycle. Observe that, trivially, H contains at most $\mathrm{reg}_k(H)/k$ edge-disjoint Hamiltonian cycles. Our main result implies that H indeed asymptotically contains that many edge-disjoint Hamiltonian cycles. More precisely, it yields the following strengthening of the result by Rödl, Ruciński, and Szemerédi.

Theorem 4.1.1. *For all integers $k \geqslant 2$ and all $\varepsilon > 0$, there exists an integer n_0 such that every k-graph H on $n \geqslant n_0$ vertices with $\delta_{k-1}(H) \geqslant (1/2 + \varepsilon)n$ contains $(1-\varepsilon)\,\mathrm{reg}_k(H)/k$ edge-disjoint Hamiltonian cycles.*

This asymptotically solves (a much stronger version of) a conjecture due to Glock, Kühn, and Osthus [51, Conjecture 6.6] which states that if in addition to the assumptions in Theorem 4.1.1, we assume that each vertex is contained in the same number of edges and $k \mid n$, then H has a decomposition into perfect matchings. Observe that in this case a Hamiltonian cycle contains k edge-disjoint perfect matchings. A similar observation also applies to all other notions of cycles in hypergraphs, for instance, loose cycles.

In fact, there is no need to restrict our attention only to Hamiltonian cycles. We call a k-graph $\mathcal{C}$ a *cycle factor* (with respect to H) if $\mathcal{C}$ is a union of vertex-disjoint cycles and has the same number of vertices as H. The *girth* of a cycle factor is the length of its shortest cycle.

Theorem 4.1.2. *For all integers $k \geqslant 2$ and all $\varepsilon > 0$, there exist integers n_0 and L such that every k-graph H on $n \geqslant n_0$ vertices with $\delta_{k-1}(H) \geqslant (1/2 + \varepsilon)n$ contains edge-disjoint copies of any given cycle factors $\mathcal{C}_1, \dots, \mathcal{C}_{r'}$, where $r' \leqslant (1-\varepsilon)\,\mathrm{reg}_k(H)/k$, whose girths are at least L.*

To the best of our knowledge, under the above condition not even the existence of a single given cycle factor was known previously. Note that L has to grow as k grows and ε shrinks, but we do not require any dependence on n.

As it turns out, we can restrict our attention to essentially vertex-regular k-graphs. To this end, we call a k-graph H on n vertices *ϱ-almost r-regular* for some $\varrho, r \geqslant 0$ if $d_H(v) = (1 \pm \varrho)r$ for all $v \in V(H)$ and ϱ-almost regular if it is ϱ-almost r-regular for some $r \geqslant 0$. Note that ϱ-almost r-regular k-graphs may simply be the disjoint union of two cliques, say, and thus they may not even contain a single Hamiltonian cycle. To avoid such scenarios, we work with the following fairly weak quasirandomness property. Given a $(k-1)$-set $\mathbf{x} = \{x_1, \ldots, x_{k-1}\} \in \binom{V(H)}{k-1}$, we write $N_H(\mathbf{x})$ for the neighbourhood of $\mathbf{x}$ in H, that is, the set $\{v \in V(H) : \{v, x_1, \ldots, x_{k-1}\} \in E(H)\}$. Define H to be *η-intersecting* if for all $\mathbf{x}, \mathbf{y} \in \binom{V(H)}{k-1}$, we have $|N(\mathbf{x}) \cap N(\mathbf{y})| \geqslant \eta n$. Considering a complete graph on n vertices where we delete the edges of a clique on $(1 - 1/k + o(1))n$ vertices, implies that being $(1/k - o(1))$-intersecting alone is not sufficient to ensure the existence of a Hamiltonian cycle either.

The following theorem is our main result and Theorems 4.1.1 and 4.1.2 follow from it.

Theorem 4.1.3. *For all integers $k \geqslant 2$ and all $\eta, \varepsilon > 0$, there exist integers L and n_0, and $\varrho > 0$ such that every η-intersecting ϱ-almost r-regular k-graph H on $n \geqslant n_0$ vertices contains edge-disjoint copies of any given cycle factors $\mathcal{C}_1, \ldots, \mathcal{C}_{r'}$, where $r' \leqslant (1-\varepsilon)r/k$, whose girths are at least L.*

A result of Ferber, Krivelevich, and Sudakov [37] implies that any k-graph H on n vertices with $\delta_{k-1}(H) \geqslant (1/2 + o(1))n$, contains an $n^{-1/2}$-almost r-regular spanning subgraph F for some $r > \frac{1}{8}\binom{n}{k-1}$. This F may not be η-intersecting for some $\eta > 0$, but the next result shows that this is not an obstacle for the application of Theorem 4.1.3.

Lemma 4.1.4. *For all integers k and all $\varepsilon > 0$, there exist an integer n_0 and $\eta > 0$ such that every k-graph H on $n \geqslant n_0$ vertices with $\delta_{k-1}(H) \geqslant (\frac{1}{2} + \varepsilon)n$ which contains a ϱ-almost r-regular spanning subgraph for some $\varrho \in [0, 1/2]$ and $r \geqslant 0$ also contains an η-intersecting $(\varrho + n^{-1/3})$-almost r'-regular spanning subgraph for some $r' \geqslant \max\{(1-\varepsilon)r, \frac{1}{8}\binom{n}{k-1}\}$.*

In particular, Theorem 4.1.3 together with Lemma 4.1.4 implies Theorem 4.1.2 (and thereby Theorem 4.1.1). In addition, for k-graphs H on n vertices with $\delta_{k-1}(H) \geqslant (1/2 + o(1))n$, we have $\mathrm{reg}_k(H) = (1 - o(1))r'$, where r' is the largest integer such that H contains an $o(1)$-almost r'-regular spanning subgraph.

Theorem 4.1.3 for $k = 2$, is an implication of a bandwidth theorem for approximate decompositions proved only recently by Condon, Kim, Kühn, and Osthus [19], explicitly mentioned as a statement in [51].

We prove Theorem 4.1.3 in Section 4.7 and Lemma 4.1.4 in Section 4.3.5.

Let us proceed with a few comments relating our results to results in the literature. In comparison to graphs, k-graphs H with $k \geqslant 3$ exhibit a significantly more divers complexity landscape. To see this, we define the degree $d_H(\mathbf{x})$ of a set $\mathbf{x} \subseteq V(H)$ as the number of edges containing it, where $|\mathbf{x}| \in [k-1]$. We say H is ϱ-almost ℓ-degree regular if for some r, we have $d_H(\mathbf{x}) = (1 \pm \varrho)r$ for all $\mathbf{x} \in \binom{V(H)}{\ell}$. It is easy to see that if H is ϱ-almost ℓ-degree regular, then it is also ϱ-almost ℓ'-degree regular for all $\ell' \leqslant \ell$. Similar relations also hold for the minimum ℓ-degree $\delta_\ell(H)$ of H. Consequently, lower bounds on $\delta_\ell(H)$ become stronger the larger ℓ gets; similarly, the assumption of ℓ-degree regularity also becomes stronger the larger ℓ gets. Not surprisingly, Dirac's result was first extended to hypergraphs with a lower bound on $\delta_{k-1}(H)$. An asymptotically sharp lower bound for the existence of a Hamiltonian cycle in terms of $\delta_{k-2}(H)$ was only proved recently [77, 93, 95] and analogous results for $\delta_\ell(H)$ where $\ell \leqslant k-3$ seem to be very difficult and at the moment out of reach.

With this in mind, the first natural step when extending the results of Ferber, Krivelevich, and Sudakov in [36] to hypergraphs, is to consider k-graphs which are $o(1)$-almost $(k-1)$-degree regular. In fact, progress towards this first step was made by Frieze and Krivelevich [45] as well as Ferber, Krivelevich, and Sudakov [37]. However, their nice approaches to reduce the problem to decomposition problems in graphs only works for weaker notions of cycles (often called loose cycles) and in particular fails for tight cycles as we consider them in this paper. Bal and Frieze [5] constructed decompositions of k-graphs into tight Hamiltonian cycles by reducing the problem to a decomposition problem in digraphs to the expense of an even more restrictive notion than $o(1)$-almost $(k-1)$-degree regularity. Unfortunately, this approach does not work when we only assume that the k-graphs are $o(1)$-almost ℓ-degree regular for some $\ell < k-1$.

In contrast to this, Theorem 4.1.3 only assumes that the k-graph is almost 1-degree regular and therefore is considerably less restrictive than the above mentioned results; in fact, our regularity assumption is necessary since every k-graph with a decomposition into Hamiltonian cycles is 1-degree regular. Further, note that our second assumption that H is $o(1)$-intersecting is much weaker than the assumed form of quasirandomness in [5] and is still implied by the minimum degree condition assumed in [37]. Although the assumptions of our main theorem are substantially weaker than those in the aforementioned results, it yields a stronger output by providing approximate decompositions into copies of any given cycle factors of not too small girth. These differences in both the assumptions and the results gave rise to a conceptually different proof compared to [5, 36, 37, 45].

4.2 Proof sketch of Theorem 4.1.3

Suppose we are in the setting of Theorem 4.1.3; that is, suppose H is an η-intersecting ϱ-almost r-regular k-graph in which we aim to find edge-disjoint copies of the cycle factors. Our argumentation is built on three stages which are described in Sections 4.4–4.6.

With some foresight, we set aside a thin randomly selected spanning subgraph F of H; in particular, F is η'-intersecting and ϱ'-almost regular for some $\eta' < \eta$ and $\varrho' > \varrho$.

In the first stage, we only consider the k-graph $H' = H - F$. For large L (but which does not grow with n) the k-graph H' has a fractional decomposition into cycles of length L, by a recent result in [63] (see Theorem 4.3.4). Next, we exploit a result about hypergraph matchings with pseudorandom properties [29] (see Theorem 4.3.5 and Corollary 4.3.6) to turn this fractional decomposition of H' into edge-disjoint collections $\mathcal{P}_1, \ldots, \mathcal{P}_r$ of vertex-disjoint paths of length L such that $V(H) \smallsetminus \bigcup_{P \in \mathcal{P}_i} V(P)$ is very small for each $i \in [r]$ and $E(H') \smallsetminus \bigcup_{i \in [r]} E(\mathcal{P}_i)$ is very small as well (see Proposition 4.4.1). This completes the first stage.

The second stage in our approach deals with the question of how one can turn a single $\mathcal{P}_i$ into a particular cycle factor $\mathcal{C}_i$ (see Lemma 4.6.3). For this we use the edges in F. One might hope that one can proceed similarly as Rödl, Ruciński, and Szemerédi in [98–100] to join up paths and absorb the remaining vertices to obtain the desired cycle factor. However, as the cycles in $\mathcal{C}_i$ may be very short, we cannot utilize an absorbing path of length $o(n)$ into which we could incorporate any small set of remaining vertices, simply because there may not exist a cycle in the desired cycle factor which is long enough to contain such a path. If we split the absorbing path into subpaths and distributed these among numerous cycles in the cycle factor, we would have too little control over how many vertices each cycle actually incorporates and, thereby, over the resulting cycle lengths.

To overcome this, we prepare by grouping small absorbing elements (paths on $2k$ vertices) into more powerful absorbers, which we call *blocks* (see Section 4.5). The crucial property is that in the end, regardless of the leftover vertices, each block absorbs exactly one vertex. Thus, a cycle of length $\ell - b$ obtained by connecting paths in $\mathcal{P}_i$ and b blocks turns into a cycle of length ℓ during the absorption of the leftover vertices. Hence, keeping this predictable change of lengths in mind, we can construct an almost spanning collection of vertex-disjoint cycles in such a way that the absorption of the leftover vertices engenders a cycle factor as desired.

The third stage in our argumentation deals with the task of repeating the second stage for every $i \in [r]$. Proceeding in a greedy fashion, iteratively considering each $i \in [r]$, may quickly ruin the quasirandom properties of F. Therefore, we actually provide all tools from the second stage as probabilistic constructions. With this, we can ensure a fairly uniform use of the edges in F when applying the arguments of the second stage iteratively for each $i \in [r]$. By using Freedman's inequality, one observes that with positive probability this process terminates successfully before significantly spoiling the quasirandomness of F (see Proposition 4.6.4).

4.3 Preliminaries

4.3.1 Notation

For $n \in \mathbb{N}_0$, we set $[n] = \{1, \dots, n\}$ and $[n]_0 = \{0, \dots, n\}$. For a set A, we say that A is a *k-set* if $|A| = k$; we write $\binom{A}{k}$ for the set of k-sets that are subsets of A and $A^{\underline{k}}$ for the set of tuples $(x_1, \dots, x_k) \in A^k$ with $x_i \neq x_j$ for all $i \neq j$. We often use $\mathbf{x}, \mathbf{y}$ to refer to sets and $\vec{\mathbf{x}}, \vec{\mathbf{y}}$ when considering tuples; however, if the tuple arises from ordering the vertices of an edge, then we often use $\vec{e}, \vec{f}$. We may subsequently drop the arrow to denote the set of elements of a tuple, so that for a tuple $\vec{\mathbf{x}} = (x_1, \dots, x_k)$, we have $\mathbf{x} = \{x_1, \dots, x_k\}$. An *ordering* of a k-set $\mathbf{x} = \{x_1, \dots, x_k\}$ is a sequence $x_1 \dots x_k$ without repetitions.

For non-negative reals $\alpha, \beta, \delta, \delta'$, we write $\alpha = (1 \pm \delta)\beta$ to mean $(1-\delta)\beta \leqslant \alpha \leqslant (1+\delta)\beta$ and we write $(1 \pm \delta)\alpha = (1 \pm \delta')\beta$ to mean $(1-\delta')\beta \leqslant (1-\delta)\alpha \leqslant (1+\delta)\alpha \leqslant (1+\delta')\beta$. We write $\alpha \ll \beta$ to mean that there is a non-decreasing function $\alpha_0\colon (0,1] \to (0,1]$ such that for any $\beta \in (0,1]$, the subsequent statement holds for all $\alpha \in (0, \alpha_0(\beta)]$. Hierarchies with more constants are defined similarly and should be read from the right to the left. Constants in hierarchies will always be reals in $(0,1]$. Moreover, if $1/n$ appears in a hierarchy, this implicitly means that n is a positive integer. We ignore rounding issues when they do not affect the argument.

Whenever we use k to refer to the uniformity of a hypergraph, we tacitly assume that $k \geqslant 2$. Let H be a k-graph on n vertices. We write $V(H)$ for the vertex set and $E(H)$ for the edge set of H and we define $\vec{E}(H) = \{\vec{e} \in V(H)^{\underline{k}}\colon e \in E(H)\}$. For $j \in [k-1]$ and $\mathbf{x} = \{x_1, \dots, x_j\} \in \binom{V(H)}{j}$, we write $d_H(\mathbf{x})$ or $d_H(x_1 \dots x_j)$ for the j-degree $|\{e \in$

$E(H)\colon \mathbf{x} \subseteq e\}|$ of $\mathbf{x}$, $\delta_j(H)$ for the minimum j-degree $\min\{d_H(\mathbf{x})\colon \mathbf{x} \in \binom{V(H)}{j}\}$ of H and $\Delta_j(H)$ for the maximum j-degree $\max\{d_H(\mathbf{x})\colon \mathbf{x} \in \binom{V(H)}{j}\}$ of H. We define $\delta(H) = \delta_{k-1}(H)$ and $\Delta(H) = \Delta_{k-1}(H)$. The k-graph H is *vertex-regular* if there is an $r \geqslant 0$ such that $d_H(v) = r$ holds for all $v \in V(H)$.

For $U \subseteq V(H)$, we write $H[U]$ for the induced subgraph $(U, \{e \in E(H)\colon e \subseteq U\})$ and if X is a set, we define $H - X = H[V(H) \setminus X]$. For two k-graphs H_i with vertex set V_i and edge set E_i for $i \in [2]$, we define $H_1 - H_2 = (V_1, E_1 \setminus E_2)$ and $H_1 \cap H_2 = (V_1 \cap V_2, E_1 \cap E_2)$ and we write $H_1 \subseteq H_2$ to indicate that H_1 is a subgraph of H_2.

A *walk* W in H is a sequence $w_1 \dots w_\ell$ of vertices of H such that $\{w_i, \dots, w_{i+k-1}\}$ is an edge of H for all $i \in [\ell - k + 1]$; we say that W is an ℓ-walk. The *length* of W is $\ell - k + 1$ and if $\ell \geqslant k$, the walk W is a walk from $(w_1, \dots, w_k)$ to $(w_{\ell-k+1}, \dots, w_\ell)$. The walk W is *self-avoiding* if no vertex of H appears twice in W.

A k-graph P on is called a *path* if there is an ordering $v_1 \dots v_\ell$ of its vertex set such that a k-set forms an edge of P if and only if its elements appear consecutively in $v_1 \dots v_\ell$. We say that P is an *ℓ-path* and the *length* of P is $|E(P)|$. A cycle of length ℓ is also called an *ℓ-cycle*. Sometimes P is identified with the sequence $v_1 \dots v_\ell$. Further, we call the tuples $(v_1, \dots, v_i)$ and $(v_{\ell-k+1}, \dots, v_\ell)$ with $i \in [\ell]$ *end-tuples* of P and whenever $\vec{\mathbf{x}}$ is an end-tuple of P, the set $\mathbf{x}$ is an *end-set* of P. End-tuples of P that are k-tuples are also called *ordered end-edges* of P and end-sets of P that are k-sets are also called *end-edges* of P. For end-tuples $\vec{\mathbf{s}}$ and $\vec{\mathbf{t}}$ of P with $\mathbf{s} \cap \mathbf{t} = \varnothing$, the graph P is also called an $\vec{\mathbf{s}}$–$\vec{\mathbf{t}}$-path. For ordered end-edges $\vec{s}$ and $\vec{t}$ of P with $s \cap t = \varnothing$, we sometimes arbitrarily fix a direction of P by saying that $\vec{\mathbf{s}}$ is the *ordered starting edge* and $\vec{\mathbf{t}}$ the *ordered ending edge* of P.

A *matching* $\mathcal{M}$ in H is a set of disjoint edges of H and $\mathcal{M}$ is *perfect* if all vertices of H belong to an edge in $\mathcal{M}$. We also treat a perfect matching $\mathcal{M}$ in a bipartite graph G with bipartition $\{U, V\}$ as a bijection $\mu\colon A \to B$, this means that μ is the bijection with $\{u, \mu(u)\} \in \mathcal{M}$ for all $u \in U$. A *perfect fractional matching* in H is a function $\omega\colon E(H) \to [0,1]$ with $\sum_{e \in E(H)\colon v \in e} \omega(e) = 1$ for all $v \in V(H)$.

Sometimes we identify a set $\mathcal{H}$ of k-graphs on disjoint vertex sets with the k-graph with vertex set $\bigcup_{H \in \mathcal{H}} V(H)$ and edge set $\bigcup_{H \in \mathcal{H}} E(H)$; in this case we refer to $\mathcal{H}$ as a *collection* of k-graphs. For a collection $\mathcal{H}$ of k-graphs, we use $|\mathcal{H}|$ to refer to the size of $\mathcal{H}$ as a set.

4.3.2 Concentration inequalities

We use the following versions of Chernoff's, McDiarmid's and Freedman's inequality.

Lemma 4.3.1 (Chernoff's inequality)**.** *Suppose $X_1, \dots, X_n$ are independent Bernoulli random variables and let $X = \sum_{i \in [n]} X_i$. Then, for all $\delta > 0$,*

$$\mathbb{P}[X \geqslant (1+\delta)\mathrm{ex}[X]] \leqslant \exp\Big(-\frac{\delta^2}{2+\delta}\mathrm{ex}[X]\Big),$$

and, if $0 < \delta < 1$, then

$$\mathbb{P}[X \leqslant (1-\delta)\mathrm{ex}[X]] \leqslant \exp\Big(-\frac{\delta^2}{2}\mathrm{ex}[X]\Big).$$

Lemma 4.3.2 (McDiarmid's inequality)**.** *Suppose $X_1, \dots, X_n$ are independent random variables and let $f \colon \mathrm{Im}(X_1) \times \dots \times \mathrm{Im}(X_n) \to \mathbb{R}$. Assume for all $i \in [n]$ that changing the i-th coordinate of $\vec{\mathbf{x}} \in \mathrm{dom}(f)$ changes $f(\vec{\mathbf{x}})$ by at most $c_i > 0$. Then, for all $\mu > 0$,*

$$\mathbb{P}[|f(X_1, \dots, X_n) - \mathrm{ex}[f(X_1, \dots, X_n)]| \geqslant \mu] \leqslant 2\exp\left(-\frac{2\mu^2}{\sum_{i \in [n]} c_i^2}\right).$$

Lemma 4.3.3 (Freedman's inequality [44])**.** *Suppose $X_1, \dots, X_n$ are Bernoulli random variables and let $\mu > 0$ with $\sum_{i \in [n]} \mathrm{ex}[X_i \mid X_1, \dots, X_{i-1}] \leqslant \mu$. Then,*

$$\mathbb{P}\Big[\sum_{i \in [n]} X_i \geqslant 2\mu\Big] \leqslant \exp\Big(-\frac{\mu}{6}\Big).$$

4.3.3 Fractional cycle decompositions

Suppose H and F are a k-graphs and $\mathcal{F}$ is the set of copies of F in H. We say a function $\omega \colon \mathcal{F} \to [0,1]$ with $\sum_{F' \in \mathcal{F} \colon e \in E(F')} \omega(F') = 1$ for all $e \in E(H)$ is a *fractional F-decomposition* of H. We are only interested in the case where F is a cycle and need the following result from [63].

Theorem 4.3.4 ([63, Theorem 1.4])**.** *Suppose $1/n \ll 1/L \ll \eta, 1/k$. Suppose H is an η-intersecting k-graph on n vertices with edge set E and C_L is a k-uniform cycle of length L. Then, there is a fractional C_L-decomposition ω of H with*

$$\frac{|E|}{\Delta(H)^L} \leqslant \omega(C) \leqslant \frac{3|E|}{\delta(H)^L}$$

for all L-cycles C in H.

4.3.4 Matchings in hypergraphs

It is nowadays well-known that essentially vertex-regular hypergraphs admit almost perfect matchings provided each pair of vertices is only contained in few edges [90, 97]. In fact, these matchings can be chosen in a way that they exhibit pseudorandom properties, which is very useful for applications. The following result provides such matchings.

Theorem 4.3.5 ([29, Theorem 1.2])**.** *Suppose $1/\Delta \ll \delta, 1/k$. Let $\varepsilon = \frac{\delta}{50k^2}$. Suppose H is a k-graph with edge set E, $\Delta_1(H) \leqslant \Delta$, and $\Delta_2(H) \leqslant \Delta^{1-\delta}$ as well as $|E| \leqslant \exp(\Delta^{\varepsilon^2})$. Suppose that $\mathcal{W}$ is a set of at most $\exp(\Delta^{\varepsilon^2})$ functions from E to $\mathbb{R}_{\geqslant 0}$. Then, there exists a matching $\mathcal{M}$ in H such that $\sum_{e\in\mathcal{M}} \omega(e) = (1 \pm \Delta^{-\varepsilon})(\sum_{e\in E} \omega(e))/\Delta$ for all $\omega \in \mathcal{W}$ with $\sum_{e\in E} \omega(e) \geqslant \max_{e\in E} \omega(e)\Delta^{1+\delta}$.*

It is straightforward to turn Theorem 4.3.5 into a result about edge sets in weighted hypergraphs in the case where the functions in $\mathcal{W}$ are $\{0, 1\}$-valued. It can be obtained from Theorem 4.3.5 by modelling the edge weights as corresponding numbers of edges in an auxiliary $(k + 1)$-graph; see, for example [69], where a similar statement is deduced from Theorem 4.3.5.

Corollary 4.3.6. *Suppose $1/c \ll \delta, 1/k$. Let $\varepsilon = \frac{\delta}{200(k+1)^2}$. Suppose H is a k-graph with vertex set V and edge set E and $\omega\colon E \to [1/c, 1]$ is a function with $\sum_{e\in E\colon v\in e} \omega(e) \leqslant 1$ and $\sum_{e\in E\colon u,v\in e} \omega(e) \leqslant 1/c^\delta$ for all distinct $u, v \in V$ as well as $\sum_{e\in E} \omega(e) \leqslant \exp(c^{\varepsilon^2})$. Suppose that $\mathcal{E}$ is a family of at most $\exp(c^{\varepsilon^2})$ subsets of E with $\sum_{e\in E'} \omega(e) \geqslant c^\delta$ for all $E' \in \mathcal{E}$. Then, there exists a matching $\mathcal{M}$ in H with $|\mathcal{M} \cap E'| = (1 \pm c^{-\varepsilon}) \sum_{e\in E'} \omega(e)$ for all $E' \in \mathcal{E}$.*

Proof. We will apply Theorem 4.3.5 to an auxiliary hypergraph obtained by replacing every edge in H with essentially $\omega(e)M$ copies of itself, where $M = (c - 1)c$ is a sufficiently large convenient multiplicity. As we want to avoid multihypergraphs, we simply increase the uniformity by 1 and add some dummy vertices. More precisely, consider the auxiliary $(k+1)$-graph H_M with vertex set $V \cup (E \times [M])$ whose edges are the sets $\{v_1, \dots, v_k, (e, i)\}$ with $\{v_1, \dots, v_k\} = e$ and $i \leqslant \lceil M\omega(e)\rceil$. Let $E_M = E(H_M)$. Observe that there is a correspondence of matchings in H_M and matchings in H, namely, for every matching $\mathcal{M}_M$ in H_M, those edges of H that are subsets of edges in $\mathcal{M}_M$ form a matching in H.

We will verify that the given properties of H and ω translate to properties of H_M that allow an application of Theorem 4.3.5 that in turn yields the desired matching in H via the aforementioned correspondence. Let $\Delta_M = c^2$. Since $1/c \leqslant \omega(e)$ holds for all $e \in E$, we obtain $d_H(v) \leqslant c$ and $d_H(uv) \leqslant c^{1-\delta}$ for all distinct $u, v \in V$ as well as $|E| \leqslant c\exp(c^{\varepsilon^2})$. This implies

$$d_{H_M}(v) = \sum_{e \in E\colon v \in e} \lceil M\omega(e) \rceil \leqslant M + d_H(v) \leqslant \Delta_M$$

and

$$d_{H_M}(uv) = \sum_{e \in E\colon u,v \in e} \lceil M\omega(e) \rceil \leqslant \frac{M}{c^{\delta}} + d_H(uv) \leqslant \Delta_M^{1-\delta/2} \leqslant \Delta_M^{1-\delta/4}$$

as well as

$$|E_M| = \sum_{e \in E} \lceil M\omega(e) \rceil \leqslant M\exp(c^{\varepsilon^2}) + |E| \leqslant c^2\exp(c^{\varepsilon^2}) \leqslant \exp(\Delta_M^{\varepsilon^2}).$$

Furthermore, observe that $|\mathcal{E}| \leqslant \exp(c^{\varepsilon^2}) \leqslant \exp(\Delta_M^{\varepsilon^2})$. For $E' \in \mathcal{E}$, let E'_M be the set of edges $\{v_1, \ldots, v_k, (e,i)\}$ of H_M with $e \in E'$. We have

$$|E'_M| = \sum_{e \in E'} \lceil M\omega(e) \rceil \geqslant Mc^{\delta} \geqslant c^2 c^{\delta/2} = \Delta_M^{1+\delta/4}.$$

An application of Theorem 4.3.5, with Δ_M, $\delta/4$, H_M, the set of indicator functions of the sets E'_M playing the roles of Δ, δ, H, $\mathcal{W}$, yields a matching $\mathcal{M}_M$ in H_M with

$$|\mathcal{M}_M \cap E'_M| = (1 \pm \Delta_M^{-\varepsilon})\frac{\sum_{e \in E'} \lceil M\omega(e) \rceil}{\Delta_M}.$$

Since we have $1/c \leqslant \omega(e)$ for all $e \in E$ and thus $|E'| \leqslant c\sum_{e \in E'} \omega(e)$ for all $E' \in \mathcal{E}$, this implies

$$|\mathcal{M}_M \cap E'_M| \leqslant (1+\Delta_M^{-\varepsilon})\frac{|E'| + M\sum_{e \in E'} \omega(e)}{\Delta_M} \leqslant (1+\Delta_M^{-\varepsilon})\frac{c+M}{\Delta_M}\sum_{e \in E'} \omega(e) \leqslant (1+c^{-\varepsilon})\sum_{e \in E'} \omega(e)$$

and

$$|\mathcal{M}_M \cap E'_M| \geqslant (1-\Delta_M^{-\varepsilon})\frac{M\sum_{e \in E'} \omega(e)}{\Delta_M} \geqslant (1-c^{-\varepsilon})\sum_{e \in E'} \omega(e)$$

for all $E' \in \mathcal{E}$. Thus, the edges in H which are subsets of edges in $\mathcal{M}_M$ form a matching in H with the desired properties. □

We also need the following result from [52] and another lemma concerning perfect fractional matchings in hypergraphs. For a k-graph with edge set E and $C > 0$, we say $\omega\colon E \to \mathbb{R}_{>0}$ is *C-balanced* if $\frac{\max_{e \in E}\omega(e)}{\min_{e \in E}\omega(e)} \leqslant C$ and we say ω is *balanced* if it is 2-balanced.

Lemma 4.3.7 ([52, Lemma 4.2])**.** *Let $1/n \ll 1/C \ll \gamma, 1/k$. Suppose H is a k-graph on n vertices with $\delta(H) \geqslant \left(\frac{1}{2}+\gamma\right)n$. Then there exists a C-balanced perfect fractional matching in H.*

Lemma 4.3.8. *Suppose $1/n \ll \varrho \ll \eta, 1/k$. Suppose H is an η-intersecting ϱ-almost regular k-graph on n vertices. Then, there exists a balanced perfect fractional matching ω in H.*

Proof. The strategy of the proof is as follows. We start with a uniform wight distribution $\omega_0\colon E(H) \to [0,1]$ and show that ω_0 can be turned into a perfect fractional matching as desired through a series of minor modifications.

Let $V = V(H)$ and $E = E(H)$. For $\omega\colon E \to [0,1]$ and $v \in V$, we define $\omega(v) = \sum_{e\in E\colon v\in e}\omega(e)$. Let $\omega_0\colon E \to [0,1]$ with $\omega_0(e) = \frac{n}{k|E|}$ for all $e \in E$ and $\xi\colon V \to \mathbb{R}$ with $\xi(v) = \omega_0(v) - 1$ for all $v \in V$. We have $\sum_{v\in V}\omega_0(v) = n$ and thus

$$\sum_{v\in V}\xi(v) = 0. \tag{4.3.1}$$

We wish to choose the series of modifications of ω_0 such that it mimics redistributing the deviations $\xi(v)$ from the target weight 1 uniformly across all vertices. We achieve this by defining every modification as a manipulation of the weights on edges of suitable walks of length 2 as follows. For a self-avoiding walk $W = v_1 \dots v_{k+1}$ in H, $\omega\colon E \to \mathbb{R}$, and $a \in \mathbb{R}$, we say that $\omega'\colon E \to \mathbb{R}$ is the function *obtained from ω by using W with weight a* if $\omega'(e) = \omega(e) - a$ for $e = \{v_1, \dots, v_k\}$, $\omega'(e) = \omega(e) + a$ for $e = \{v_2, \dots, v_{k+1}\}$, and $\omega'(e) = \omega(e)$ otherwise. Hence $\omega'(v_1) = \omega(v_1) - a$, $\omega'(v_{k+1}) = \omega(v_{k+1}) + a$, and $\omega'(v) = \omega(v)$ for all $v \in V \setminus \{v_1, v_{k+1}\}$.

Since H is η-intersecting, the number of self-avoiding walks $v_1 \dots v_{k+1}$ with $v_1 = s$ and $v_{k+1} = t$ is at least $\eta n^{k-1}/2$ for all distinct $s, t \in V$. For distinct $s, t \in V$, let $W_{s,t}$ be a set of $\eta n^{k-1}/2$ self-avoiding walks $v_1 \dots v_{k+1}$ in H with $v_1 = s$ and $v_{k+1} = t$. Let $v_1^1 \dots v_{k+1}^1, \dots, v_1^m \dots v_{k+1}^m$ be an ordering of $\bigcup_{s,t\in V\colon s\neq t} W_{s,t}$ and for $i \in [m]$, let ω_i be the function obtained from ω_{i-1} by using $v_1^i \dots v_{k+1}^i$ with weight $\frac{2\xi(v_1^i)}{\eta n^k}$. Let $\omega = \omega_m$. From (4.3.1) we conclude that

$$\omega(v) = \omega_0(v) - (n-1)\cdot\frac{\eta n^{k-1}}{2}\cdot\frac{2\xi(v)}{\eta n^k} + \sum_{u\in V\setminus\{v\}}\frac{\eta n^{k-1}}{2}\cdot\frac{2\xi(u)}{\eta n^k} = 1$$

holds for all $v \in V$. Thus, it suffices to show $\omega(e) = (1 \pm 1/3)\omega_0(e)$ for all $e \in E$.

Since for all $e \in E$, there are at most $2k!n$ self-avoiding walks $v_1 \dots v_{k+1}$ in H with $e \in$

$\{\{v_1,\ldots,v_k\},\{v_2,\ldots,v_{k+1}\}\}$, transitioning from ω_0 to ω changed the weight on e by at most $2k!n\cdot\frac{2\max_{v\in V}|\xi(v)|}{\eta n^k}$. Observe that

$$\omega_0(v)=d_H(v)\frac{n}{k|E|}=d_H(v)\frac{n}{\sum_{u\in V}d_H(u)}=1\pm 4\varrho$$

and hence $|\xi(v)|\leqslant 4\varrho$ holds for all $v\in V$. Thus we obtain

$$\omega(e)=\left(1\pm 2k!n\cdot\frac{8\varrho}{\eta n^k}\cdot\frac{k|E|}{n}\right)\omega_0(e)=\left(1\pm\frac{1}{3}\right)\omega_0(e)$$

for all $e\in E$. $\square$

4.3.5 Almost regular spanning subgraphs with intersecting neighbourhoods

In this subsection we prove Lemma 4.1.4. We use the following statement which follows from [37, Theorem 1.2] by considering the union of perfect matchings in an induced subgraph obtained after removing at most $k-1$ vertices to make the number of vertices divisible by k and subsequently adding edges for the previously removed vertices.

Lemma 4.3.9 ([37, Theorem 1.2])**.** *Suppose $1/n\ll\varepsilon,1/k$. Suppose H is a k-graph on n vertices with $\delta(H)\geqslant(\frac{1}{2}+\varepsilon)$. Then H contains an $n^{-1/2}$-almost r-regular spanning subgraph for some $r\geqslant(1+3\varepsilon/2)\frac{1}{8}\binom{n}{k-1}$.*

Proof of Lemma 4.1.4. Suppose $1/n\ll\eta\ll\varepsilon,1/k$. Suppose H is a k-graph on n vertices with $\delta(H)\geqslant(\frac{1}{2}+\varepsilon)n$ that contains a ϱ-almost r-regular spanning subgraph for some $\varrho\in[0,1/2]$ and $r\geqslant 0$. If $r\geqslant(1+3\varepsilon/2)\frac{1}{8}\binom{n}{k-1}$, let $\varrho_F=\varrho$, $r_F=r$, otherwise let $\varrho_F=n^{-1/2}$ and choose $r_F\geqslant(1+3\varepsilon/2)\frac{1}{8}\binom{n}{k-1}$ such that there exists a ϱ_F-almost r_F-regular spanning subgraph of H, which is possible by Lemma 4.3.9. Let F denote a ϱ_F-almost r_F-regular spanning subgraph of H. In order to obtain a random spanning subgraph F' of H that has the desired properties with positive probability, we construct its edge set $E(F')$ by essentially choosing the edges of F while ensuring that each edge of H is included with positive probability.

By Lemma 4.3.7 there is a $1/\eta^{1/3}$-balanced perfect fractional matching ω in H. Let $\omega_{\max}=\max_{e\in E(H)}\omega(e)$. Construct the edge set of the random spanning subgraph F' of H as follows.

For all $e \in E(H)$, include e in $E(F')$ independently at random with probability p_e, where

$$p_e = \begin{cases} (1-\varepsilon) + \frac{\varepsilon\omega(e)}{\omega_{\max}} & \text{if } e \in E(F) \\ \frac{\varepsilon\omega(e)}{\omega_{\max}} & \text{if } e \notin E(F). \end{cases}$$

Fix $v \in V(H)$ and $\mathbf{x}, \mathbf{y} \in \binom{V(H)}{k-1}$. Let $r' = (1-\varepsilon)r_F + \frac{\varepsilon}{\omega_{\max}} \geqslant n^{k-1}/(9k!)$. Clearly, $\mathrm{ex}[d_{F'}(v)] = (1 \pm \varrho_F)r'$. Since $p_e \geqslant \varepsilon\eta^{1/3}$ and since H is 2ε-intersecting, we obtain $\mathrm{ex}[|N_{F'}(\mathbf{x}) \cap N_{F'}(\mathbf{y})|] \geqslant \varepsilon^2\eta^{2/3} \cdot 2\varepsilon n \geqslant 2\eta n$. Using Chernoff's inequality (Lemma 4.3.1) and the union bound shows that F' is as desired with positive probability. $\square$

4.3.6 Different types of degrees

For a k-graph with vertex set V, the following lemma shows that whenever a set $U \subseteq V$ meets the neighbourhood $N_H(\mathbf{x})$ in roughly $\vartheta d_H(\mathbf{x})$ vertices for all $\mathbf{x} \in \binom{V}{k-1}$, then all vertex-degrees decrease by about a factor of ϑ^{k-1} when transitioning to the subgraph induced by U.

Lemma 4.3.10. *Suppose H is a k-graph with vertex set V. Let $\vartheta \in (0,1)$ and $\varepsilon \in \left[0, \frac{1-\vartheta}{8k^2}\right]$. Suppose $U \subseteq V$ is a set with $d_{H[U \cup \mathbf{x}]}(\mathbf{x}) = (1 \pm \varepsilon)\vartheta d_H(\mathbf{x})$ for all $\mathbf{x} \in \binom{V}{k-1}$. Then $d_{H[U \cup \{v\}]}(v) = (1 \pm 8k^3\varepsilon)\vartheta^{k-1} d_H(v)$ for all $v \in V$.*

Proof. Fix $v \in V$. Let $V' = V \setminus \{v\}$ and $U' = U \setminus \{v\}$. For $i \in [k-1]_0$, let m_i denote the number of edges $e \in E(H)$ with $v \in e$ and $|e \cap U'| = i$. Then our task is to estimate $m_{k-1} = d_{H[U \cup \{v\}]}(v)$. To this end, we inductively relate m_i and m_{i-1} for all $i \in [k-1]$.

For $i \in [k-1]$, we have

$$\begin{aligned} im_i &= \sum_{\mathbf{x} \in \binom{U'}{i-1}, \mathbf{y} \in \binom{V' \setminus U'}{k-i-1}} d_{H[U \cup \{v\} \cup \mathbf{x} \cup \mathbf{y}]}(\{v\} \cup \mathbf{x} \cup \mathbf{y}) \\ &= (1 \pm \varepsilon)\vartheta \sum_{\mathbf{x} \in \binom{U'}{i-1}, \mathbf{y} \in \binom{V' \setminus U'}{k-i-1}} d_H(\{v\} \cup \mathbf{x} \cup \mathbf{y}) \\ &= (1 \pm \varepsilon)\vartheta((k-i)m_{i-1} + im_i) \end{aligned}$$

and hence,

$$m_{i-1} = \left(1 \pm \frac{4\varepsilon}{1-\vartheta}\right)\frac{1-\vartheta}{\vartheta}\frac{i}{k-i}m_i.$$

From this, we inductively conclude that

$$m_i = \left(1 \pm \frac{4k^2\varepsilon}{1-\vartheta}\right)\frac{(1-\vartheta)^{k-1-i}}{\vartheta^{k-1-i}}\binom{k-1}{i}m_{k-1}.$$

Recall that $\sum_{i=0}^{k-1}\binom{k-1}{i}(1-\vartheta)^{k-1-i}\vartheta^i = ((1-\vartheta)+\vartheta)^{k-1} = 1$. Therefore,

$$\begin{aligned} d_H(v) - m_{k-1} = \sum_{i=0}^{k-2} m_i &= \left(1 \pm \frac{4k^2\varepsilon}{1-\vartheta}\right)\frac{m_{k-1}}{\vartheta^{k-1}}\sum_{i=0}^{k-2}\binom{k-1}{i}(1-\vartheta)^{k-1-i}\vartheta^i \\ &= \left(1 \pm \frac{4k^2\varepsilon}{1-\vartheta}\right)\frac{1-\vartheta^{k-1}}{\vartheta^{k-1}}m_{k-1} \end{aligned}$$

and thus,

$$m_{k-1} = \left(1 \pm \frac{8k^2(1-\vartheta^{k-1})\varepsilon}{1-\vartheta}\right)\vartheta^{k-1}d_H(v) = (1 \pm 8k^3\varepsilon)\vartheta^{k-1}d_H(v)$$

which completes the proof. □

4.3.7 Many paths in intersecting k-graphs

We say that a walk $W = w_1 \dots w_\ell$ in a k-graph is *internally self-avoiding* if the walks $w_1 \dots w_{\ell-k}$ and $w_{k+1} \dots w_\ell$ are self-avoiding. We use the following result from [63].

Lemma 4.3.11 ([63, Lemma 2.3])**.** *Suppose $1/n \ll \alpha \ll 1/\ell \ll \eta, 1/k$. Suppose H is an η-intersecting k-graph on n vertices with vertex set V. Then, for all $\vec{s}, \vec{t} \in \overrightarrow{E}(H)$, the number of internally self-avoiding ℓ-walks from $\vec{s}$ to $\vec{t}$ in H is at least $\alpha n^{\ell-2k}$.*

4.4 Approximate decomposition into path coverings

In this section we use Corollary 4.3.6 to turn fractional decompositions provided by Theorem 4.3.4 into approximate decompositions of almost vertex-regular k-graphs into almost spanning collections of paths. These collections of paths form the basis for the construction of cycle factors in Section 4.6.

We need to keep track of how k-sets that may be edges in a larger graph are distributed with respect to the paths constructed below. Suppose H is a k-graph on n vertices with vertex set V and $\mathcal{P}$ is a collection of paths in H. Let $\mathbf{e} \in \binom{V}{k}$. We use the following terminology to classify $\mathbf{e}$ with respect to $\mathcal{P}$, where we often tacitly assume that the collection of paths with respect to which we classify a k-set is obvious from the context. We say that $\mathbf{e}$ is

(i) *j-ending* (or of type j-end) if j is the maximal integer for which there is a j-set $\mathbf{x} \subseteq \mathbf{e}$ and a path $P \in \mathcal{P}$ such that $\mathbf{x}$ is an end-set of P;

(ii) *ending* if it is j-ending for some $j \in [k]$;

(iii) *leftover* (or of type lo) if it is neither ending nor a subset of $V(\mathcal{P})$;

(iv) *j-concentrated* (or of type j-con) if it is neither ending nor leftover and j is the maximal integer for which there is a j-set $\mathbf{x} \subseteq \mathbf{e}$ and a path $P \in \mathcal{P}$ such that $\mathbf{x} \subseteq V(P)$;

(v) *concentrated* if it is j-concentrated for some $j \in [k]$.

We denote the set of types by $\mathcal{T} = \mathcal{T}_{\text{end}} \cup \{\text{lo}\} \cup \mathcal{T}_{\text{con}}$ where $\mathcal{T}_{\text{end}} = \{1\text{-end}, \dots, k\text{-end}\}$ and $\mathcal{T}_{\text{con}} = \{1\text{-con}, \dots, k\text{-con}\}$. Note that given $\mathcal{P}$ and $\mathbf{e}$, the k-set $\mathbf{e}$ has a unique type with respect to $\mathcal{P}$. Given multiple collections $\mathcal{P}_1, \dots, \mathcal{P}_r$ of paths, for $i \in [r]$, we use $\tau(\mathbf{e}, i)$ to denote the type of $\mathbf{e}$ with respect to $\mathcal{P}_i$, and for $\tau \in \mathcal{T}$, we set $\mathcal{I}_\tau(\mathbf{e}) = \{i \in [r] \colon \tau(\mathbf{e}, i) = \tau\}$, where we tacitly assume that the index set and the collection of paths that belongs to a given index are obvious from the context.

Proposition 4.4.1. *Suppose $1/n \ll \varrho, 1/L \ll \eta, \mu, 1/k$. Let H be an η-intersecting k-graph on n vertices with vertex set V such that there is an integer r with $kr \leqslant d_H(v) \leqslant (1+\varrho)kr$ for all $v \in V$. Then, there exist edge-disjoint collections $\mathcal{P}_1, \dots, \mathcal{P}_r$ of L-paths in H with $|V(\mathcal{P}_i)| \geqslant (1-\mu)n$ for all $i \in [r]$ such that the following holds for all $\mathbf{e} \in \binom{V}{k}$.*

(i) *$|\mathcal{I}_{\text{lo}}(\mathbf{e})| \leqslant \mu r$;*

(ii) *$|\mathcal{I}_{j\text{-con}}(\mathbf{e})| \leqslant n^{k-j}/\eta^{2L}$ for all $j \in [k-1]$ and $|\mathcal{I}_{k\text{-con}}(\mathbf{e})| \leqslant n/\eta^{2L}$;*

(iii) *$|\mathcal{I}_{j\text{-end}}(\mathbf{e})| \leqslant n^{k-j}/L^{1/2}$ for all $j \in [k-1]$.*

Note that the proof also yields edge-disjoint collections $\mathcal{C}_1, \dots, \mathcal{C}_r$ of L-cycles instead of the paths with $|V(\mathcal{C}_i)| \geqslant (1-\mu)n$ for all $i \in [r]$ (without the properties (i)–(iii)); we use this in the proof of Lemma 4.6.3.

Proof of Proposition 4.4.1. First we argue that it suffices to find collections $\mathcal{C}_1, \dots, \mathcal{C}_r$ of L-cycles in H with properties similar to those of collections of paths in the statement; then we obtain such collections of cycles by applying Corollary 4.3.6 in an auxiliary hypergraph that represents a fractional cycle decomposition given by Theorem 4.3.4.

Suppose $\mathcal{C}_1, \ldots, \mathcal{C}_r$ are edge-disjoint collections of L-cycles in H. Note that for all $j \in [k-1]$ and $\mathbf{x} \subseteq \binom{V}{j}$, there are at most n^{k-j} collections $\mathcal{C}_i$ with $i \in [r]$ where the elements of $\mathbf{x}$ appear consecutively in a cycle in $\mathcal{C}_i$ and for all $i \in [r]$, there is at most one cycle $C \in \mathcal{C}_i$ such that the elements of $\mathbf{x}$ appear consecutively in C. Now, for all $i \in [r]$ and every cycle in $C \in \mathcal{C}_i$, delete $k-1$ consecutive edges uniformly at random and independently of the edges deleted in the other cycles to obtain collections of paths $\mathcal{P}_1, \ldots, \mathcal{P}_r$. Then the expected value of the random variable counting the number of collections $\mathcal{P}_i$ with $i \in [r]$ where $\mathbf{x}$ is an end-set of a path in $\mathcal{P}_i$ is at most $2n^{k-j}/L$. Hence Chernoff's inequality (Lemma 4.3.1) entails that it is possible to delete $k-1$ consecutive edges of every cycle in $\mathcal{C}_i$ for all $i \in [r]$ to obtain collections of paths $\mathcal{P}_1, \ldots, \mathcal{P}_r$ such that for all $j \in [k-1]$ and $\mathbf{x} \in \binom{V}{j}$, there are at most $3n^{k-j}/L$ collections $\mathcal{P}_i$ with $i \in [r]$ where $\mathbf{x}$ is an end-set of a path in $\mathcal{P}_i$. For such collections $\mathcal{P}_1, \ldots, \mathcal{P}_r$, $j \in [k-1]$ and $\mathbf{e} \in \binom{V}{k}$, we have $\mathcal{I}_{j\text{-end}}(\mathbf{e}) \leqslant \binom{k}{j} \cdot 3n^{k-j}/L \leqslant n^{k-j}/L^{1/2}$ and thus it suffices to obtain edge-disjoint collections $\mathcal{C}_1, \ldots, \mathcal{C}_r$ of L-cycles in H with $|V(\mathcal{C}_i)| \geqslant (1-\mu)n$ for all $i \in [r]$ such that the following holds.

- $|\{i \in [r] \colon v \notin V(\mathcal{C}_i)\}| \leqslant \mu r/k$ for all $v \in V$;
- $|\{i \in [r] \colon \exists C \in \mathcal{C}_i \colon \mathbf{x} \subseteq V(C)\}| \leqslant n^{k-j}/(\eta^{2L}\binom{k}{j})$ for all $j \in [k-1]$ and $\mathbf{x} \in \binom{V}{j}$.

Let $E = E(H)$ and let $\mathcal{C}_L(H)$ denote the set of L-cycles in H. From Theorem 4.3.4 we obtain a fractional L-cycle decomposition ω of H with

$$\frac{|E|}{n^L} \leqslant \omega(C) \leqslant \frac{3|E|}{\eta^L n^L} \tag{4.4.1}$$

for all $C \in \mathcal{C}_L(H)$. Consider the $2L$-graph $H^\star$ with vertex set $(V \times [r]) \cup E$ where we add for each $C \in \mathcal{C}_L(H)$ and all $i \in [r]$, the edge $e^\star_{C,i} = \{(v,i) \colon v \in V(C)\} \cup E(C)$ (note that the edge $e^\star_{C,i}$ uniquely identifies C and i). Let $E^\star = E(H^\star)$ and $\Gamma = (1+\varrho)r \geqslant \Delta(H)/k$ and let $\omega^\star \colon E^\star \to [0,1]$ be the edge weight function with $\omega^\star(e^\star_{C,i}) = \omega(C)/\Gamma$ for all $e^\star \in E^\star$. Here, $\omega^\star$ is a representation of ω that is normalized such that for all $v \in V$ and $i \in [r]$, we have

$$\sum_{\substack{e^\star \in E^\star \colon\\ (v,i) \in e^\star}} \omega^\star(e^\star) = \frac{1}{k} \cdot \sum_{\substack{e \in E(H) \colon\\ v \in e}} \sum_{\substack{C \in \mathcal{C}_L(H) \colon\\ e \in E(C)}} \frac{\omega(C)}{\Gamma} = \frac{1}{k} \cdot \frac{d_H(v)}{\Gamma} \in \left[\frac{1}{1+\varrho}, 1\right]. \tag{4.4.2}$$

Note that there is a correspondence of matchings in $H^\star$ to edge-disjoint collections of L-cycles in H, namely a matching $\mathcal{M}^\star$ in $H^\star$ corresponds to the collections $\mathcal{C}_1, \ldots, \mathcal{C}_r$, where $\mathcal{C}_i = \{C \in \mathcal{C}_L(H) \colon e^\star_{C,i} \in \mathcal{M}^\star\}$ for all $i \in [r]$. This will allow us to obtain collections of cycles with the desired properties from Corollary 4.3.6. We now introduce appropriate parameters including suitable subsets of $E^\star$ and check that the conditions necessary for a suitable application of Corollary 4.3.6 hold in this setting.

Let $\delta = \frac{1}{3L}$, $\varepsilon = \frac{\delta}{900L^2}$, and $c = n^L$. From (4.4.1) we obtain (with some room to spare)

$$\frac{1}{c} \leqslant \frac{\eta}{2k!\Gamma n^{L-k}} \leqslant \frac{|E|}{\Gamma n^L} \leqslant \frac{\omega(C)}{\Gamma} \leqslant \frac{3|E|}{\eta^L \Gamma n^L} \leqslant \frac{3}{\eta^L \Gamma n^{L-k}} \leqslant \frac{1}{n^{L-3/2}} c^{-\delta} \tag{4.4.3}$$

for all $C \in \mathcal{C}_L(H)$. To complete our analysis of the 1-degrees in $H^\star$, we observe that in addition to (4.4.2), for all $e \in E$, we have

$$\sum_{\substack{e^\star \in E^\star: \\ e \in e^\star}} \omega^\star(e^\star) = r \cdot \sum_{\substack{C \in \mathcal{C}_L(H): \\ e \in E(C)}} \frac{\omega(C)}{\Gamma} = r \cdot \frac{1}{\Gamma} \leqslant 1. \tag{4.4.4}$$

Let us now consider the 2-degrees in $H^\star$. For distinct $u, v \in V$ and distinct $e, f \in E$, we have

$$|\{C \in \mathcal{C}_L(H) \colon u, v \in V(C)\}| \leqslant Ln^{L-2}, \quad |\{C \in \mathcal{C}_L(H) \colon v \in V(C) \wedge e \in E(C)\}| \leqslant k!n^{L-k}$$
$$\text{and} \quad |\{C \in \mathcal{C}_L(H) \colon e, f \in E(C)\}| \leqslant 2k!n^{L-k-1}.$$

Using (4.4.3), this yields for all $i \in [r]$, that

$$\sum_{\substack{e^\star \in E^\star: \\ (u,i),(v,i) \in e^\star}} \omega^\star(e^\star) = \sum_{\substack{C \in \mathcal{C}_L(H): \\ u,v \in V(C)}} \frac{\omega(C)}{\Gamma} \leqslant \frac{L}{n^{1/2}} c^{-\delta} \leqslant c^{-\delta}, \tag{4.4.5}$$

$$\sum_{\substack{e^\star \in E^\star: \\ (v,i),e \in e^\star}} \omega^\star(e^\star) = \sum_{\substack{C \in \mathcal{C}_L(H): \\ v \in V(C) \wedge e \in E(C)}} \frac{\omega(C)}{\Gamma} \leqslant \frac{k!}{n^{k-3/2}} c^{-\delta} \leqslant c^{-\delta}, \quad \text{and} \tag{4.4.6}$$

$$\sum_{e^\star \in E^\star : e,f \in e^\star} \omega^\star(e^\star) = r \cdot \sum_{\substack{C \in \mathcal{C}_L(H): \\ e,f \in E(C)}} \frac{\omega(C)}{\Gamma} \leqslant \frac{2k!r}{n^{k-1/2}} c^{-\delta} \leqslant c^{-\delta}. \tag{4.4.7}$$

Furthermore, observe that

$$\sum_{e^\star \in E(H^\star)} \omega^\star(e^\star) = r \cdot \sum_{C \in \mathcal{C}_L(H)} \frac{\omega(C)}{\Gamma} = r \cdot \frac{|E|}{\Gamma L} \leqslant \exp(c^{\varepsilon^2}) \tag{4.4.8}$$

gives an upper bound for the total weight on the edges of $H^\star$. For $i \in [r]$, $v \in V$, $j \in [k-1]$, and $\mathbf{x} \in \binom{V}{j}$, we define edge-sets as follows.

$$E_i^\star = \{e_{C,i'}^\star \in E^\star : i' = i\}, \quad E_v^\star = \{e_{C,i'}^\star \in E^\star : v \in V(C)\}$$
$$\text{and} \quad E_{\mathbf{x}}^\star = \{e_{C,i'}^\star \in E^\star : \mathbf{x} \subseteq V(C)\}.$$

We have

$$\sum_{e^\star \in E_i^\star} \omega^\star(e^\star) = \sum_{C \in \mathcal{C}_L(H)} \frac{\omega(C)}{\Gamma} = \frac{|E|}{\Gamma L} \geqslant \frac{rn}{\Gamma L} \geqslant c^\delta \tag{4.4.9}$$

and

$$\sum_{e^\star \in E_v^\star} \omega^\star(e^\star) = r \cdot \frac{1}{k} \cdot \sum_{\substack{e \in E(H)\colon \\ v \in e}} \sum_{\substack{C \in \mathcal{C}_L(H)\colon \\ e \in E(C)}} \frac{\omega(C)}{\Gamma} = r \cdot \frac{1}{k} \cdot \frac{d_H(v)}{\Gamma} \geqslant \frac{r}{1+\varrho} \geqslant c^\delta. \tag{4.4.10}$$

Since Lemma 4.3.11 implies $|\{C \in \mathcal{C}_L(H)\colon d_C(\mathbf{x}) \geqslant 1\}| \geqslant n^{L-j-1/3}$, using (4.4.3) we obtain

$$\sum_{e^\star \in E_{\mathbf{x}}^\star} \omega^\star(e^\star) \geqslant r \cdot \sum_{\substack{C \in \mathcal{C}_L(H)\colon \\ d_C(\mathbf{x}) \geqslant 1}} \frac{\omega(C)}{\Gamma} \geqslant r \cdot n^{L-j-1/3} \cdot \frac{\eta}{2k!\Gamma n^{L-k}} \geqslant n^{k-j-1/2} \geqslant c^\delta. \tag{4.4.11}$$

Thus, since $r + n + \sum_{j \in [k-1]} \binom{n}{j} \leqslant \exp(c^{\varepsilon^2})$ holds and by (4.4.2)–(4.4.11) we may apply Corollary 4.3.6 to obtain a matching $\mathcal{M}^\star$ in $H^\star$ with

$$|\mathcal{M}^\star \cap E_i^\star| \geqslant (1 - c^{-\varepsilon})\frac{|E|}{\Gamma L} \geqslant (1 - c^{-\varepsilon})\frac{rn}{\Gamma L} \geqslant (1-\mu)\frac{n}{L},$$

$$|\mathcal{M}^\star \cap E_v^\star| \geqslant (1 - c^{-\varepsilon})\frac{r}{k} \cdot \frac{d_H(v)}{\Gamma} \geqslant (1 - c^{-\varepsilon})\frac{r^2}{\Gamma} \geqslant \left(1 - \frac{\mu}{k}\right) r,$$

and

$$|\mathcal{M}^\star \cap E_{\mathbf{x}}^\star| \leqslant (1 + c^{-\varepsilon}) \sum_{e^\star \in E_{\mathbf{x}}^\star} \omega^\star(e^\star)$$

for all $i \in [r]$, $v \in V$, $j \in [k-1]$, and $\mathbf{x} \in \binom{V}{j}$. Since we have $|\{C \in \mathcal{C}_L(H)\colon \mathbf{x} \subseteq V(C)\}| \leqslant L^j n^{L-j}$, (4.4.3) implies

$$|\mathcal{M}^\star \cap E_{\mathbf{x}}^\star| \leqslant (1 + c^{-\varepsilon})r \cdot \sum_{\substack{C \in \mathcal{C}_L(H)\colon \\ \mathbf{x} \subseteq V(C)}} \frac{\omega(C)}{\Gamma} \leqslant (1 + c^{-\varepsilon})r \cdot L^j n^{L-j} \cdot \frac{3}{\eta^L \Gamma n^{L-k}} \leqslant \frac{1}{\eta^{2L}\binom{k}{j}} n^{k-j}$$

and thus, choosing $\mathcal{C}_i$ as the collection of L-cycles C in H with $e_{C,i}^\star \in \mathcal{M}^\star$ yields collections of cycles in H with the desired properties. $\square$

4.5 Ingredients for absorption

Suppose we are given a k-graph H on n vertices. In this section we construct a collection of paths $\mathcal{P}$ in H whose lengths do not grow with n such that any small set of vertices X can be absorbed into these paths; that is, for every path $P \in \mathcal{P}$, there is a new path P' with the same end-tuples as P and $V(P') \subseteq V(P) \cup X$ such that the new paths form a collection of paths $\mathcal{P}'$ with $V(\mathcal{P}') = V(P) \cup X$.

There are two main novelties in our setting. Firstly, we choose $\mathcal{P}$ randomly in an extremely uniform way such that $V(\mathcal{P})$ behaves like a uniformly chosen vertex set of size $|V(\mathcal{P})|$. Secondly, we can control how many vertices each path in $\mathcal{P}$ will absorb if an adversary determines a set of vertices to absorb. Here the main difficulty is that the lengths of the paths in $\mathcal{P}$ do not grow with n. To the best of our knowledge, this problem has not been dealt with in the literature so far.

4.5.1 Random walks and vertex absorbers

Let H be a k-graph on n vertices with vertex set V and edge set E. For $x \in V$, a path $A = a_1 \ldots a_{2k}$ in H is called an *x-absorber* if $a_1 \ldots a_k x a_{k+1} \ldots a_{2k}$ is also a path in H. In what follows, we describe how certain random walks contain many vertex-disjoint x-absorbers for all $x \in V$ simultaneously and at the same time ensure that the set of vertices visited by these random walks behaves like a vertex set chosen uniformly at random among all vertex sets of the same size.

In what follows, a, i, ℓ, L, $t_\star$ are always positive integers. Let $W = w_1 \ldots w_L$ be a walk. For $i \in [L/(2k+\ell)]$, we define

$$A_i^\ell(W) = w_{(2k+\ell)(i-1)+1} \ldots w_{(2k+\ell)(i-1)+2k} \quad \text{and} \quad \mathcal{A}^\ell(W) = \left\{ A_i^\ell(W) \colon i \in \left[\frac{L}{2k+\ell} \right] \right\}.$$

We may think of A_i^ℓ as the i-th potential absorber in W when requiring ℓ vertices between absorbers whereas $\mathcal{A}^\ell(W)$ is the set of all these potential absorbers in W. To gain control over where absorbed vertices will be placed, we consider absorbers in groups which we call *blocks*. Towards the end of the section, our construction yields a set of blocks $\mathcal{B}$ and during the absorption, every block $B \in \mathcal{B}$ will absorb exactly one vertex. More precisely, we say that a walk $B = b_1 \ldots b_{a(2k+\ell)}$ in H is an *(a,ℓ)-block* in H; that is, B can be split into a consecutive walks that consist of a $2k$-walk followed by an ℓ-walk. For $i \in [L/(a(2k+\ell))]$, we define

$$B_i^{a,\ell}(W) = w_{a(2k+\ell)(i-1)+1} \dots w_{a(2k+\ell)i} \quad \text{and} \quad \mathcal{B}^{a,\ell}(W) = \left\{ B_i^{a,\ell}(W) \colon i \in \left[\left\lfloor \frac{L}{a(2k+\ell)} \right\rfloor \right] \right\}$$

where $B_i^{a,\ell}(W)$ can be considered as the i-th (a,ℓ)-block in W whereas $\mathcal{B}^{a,\ell}(W)$ is the set of all these (a,ℓ)-blocks in W. Later, we will choose absorbers and blocks randomly. In contrast to other approaches existing in the literature, we build our absorbing structure via random walks whose distributions are given by perfect fractional matchings to ensure a very uniform distribution of the vertices visited by these random walks.

We introduce the convention that sequences $s_m \dots s_n$ with $m > n$ are considered as the empty sequence which we identify with $\varnothing$. For $\omega \colon E \to \mathbb{R}_{\geqslant 0}$, $j \in [k]$, and $v_1, \dots, v_j \in V$, we define $\omega(v_1 \dots v_j) = \sum_{e \in E \colon v_1, \dots, v_j \in e} \omega(e)$ and we set $\omega(\varnothing) = \sum_{e \in E} \omega(e)$.

Let $\omega \colon E \to \mathbb{R}_{>0}$. We say a sequence of V-valued random variables $X_1 \dots X_{t_\star}$ is a random walk in H *with parameters* (L, ω), or simply an (L,ω)-*random walk* in H, if its distribution is given by

$$\begin{aligned} \mathbb{P}[X_t = v_t \mid X_1 \dots X_{t-1} = v_1 \dots v_{t-1}] &= \mathbb{P}[X_t = v_t \mid X_{t-m} \dots X_{t-1} = v_{t-m} \dots v_{t-1}] \\ &= \begin{cases} \frac{\omega(v_{t-m} \dots v_t)}{(k-m)\omega(v_{t-m} \dots v_{t-1})} & \text{if } v_t \notin \{v_{t-m}, \dots, v_{t-1}\} \\ 0 & \text{if } v_t \in \{v_{t-m}, \dots, v_{t-1}\} \end{cases} \end{aligned} \tag{4.5.1}$$

for all $t \in [t_\star]$, $m = \min\{k-1, t-1 \mod L\}$, and $v_1, \dots, v_t \in V$ with $\mathbb{P}[X_1 \dots X_{t-1} = v_1 \dots v_{t-1}] > 0$. This is indeed a probability distribution because for all $j \in [k-1]$ and $v_1, \dots, v_j \in V$, we have

$$\sum_{v \in V \setminus \{v_1, \dots, v_j\}} \omega(v_1 \dots v_j v) = (k-j)\omega(v_1 \dots v_j).$$

Whenever we consider an (L,ω)-random walk in a k-graph H, we assume that L is a positive integer and $\omega \colon E(H) \to \mathbb{R}_{>0}$. Observe that if $X_1 \dots X_{t_\star}$ is an (L,ω)-random walk in a k-graph H, then also $X_{L(s-1)+1} \dots X_{Ls}$ is an (L,ω)-random walk in H for all $s \in [t_\star/L]$.

Observation 4.5.1. *Suppose H is a k-graph. Suppose $X_1 \dots X_{t_\star}$ is an (L,ω)-random walk in H. Then, the random walks $X_1 \dots X_{Ls}$ and $X_{Ls+1} \dots X_{t_\star}$ are independent for all positive integers s.*

Recall that $\omega \colon E \to \mathbb{R}_{>0}$ is *balanced* if $\frac{\max_{e \in E} \omega(e)}{\min_{e \in E} \omega(e)} \leqslant 2$. For an (L,ω)-random walk $X_1 \dots X_{t_\star}$ in H for some balanced ω, the balancedness of ω allows us to bound the probability of the event that $X_t = X_{t'}$ for some distinct $t, t' \in [t_\star]$; the union bound yields the following observation.

Observation 4.5.2. *Let $1/n \ll \eta, 1/k$. Suppose H is a k-graph on n vertices with $\delta(H) \geqslant \eta n$. Suppose $X_1 \dots X_{t_\star}$ with $t_\star \leqslant 2n^{1/3}$ is an (L, ω)-random walk in H for some balanced ω. Then, we have*

$$\mathbb{P}[X_1 \dots X_{t_\star} \textit{ is self-avoiding}] \geqslant 1 - n^{-1/4}.$$

The following lemma shows that for an (L, ω)-random walk $X_1 \dots X_{t_\star}$ in H not only the transition probabilities are determined by ω, but also the probability of $X_1 \dots X_{t_\star}$ consecutively visiting a sequence of vertices can be easily computed in terms of ω.

Lemma 4.5.3. *Suppose H is a k-graph on n vertices with vertex set V. Suppose $X_1 \dots X_L$ is an (L, ω)-random walk in H. Let $t \in [L]$ and $j \in [\min\{k, t\}]$. Then, we have*

$$\mathbb{P}[X_{t-j+1} \dots X_t = v_{-j+1} \dots v_0] = \frac{(k-j)!\omega(v_{-j+1} \dots v_0)}{k!\omega(\varnothing)} \tag{4.5.2}$$

for all $v_{-j+1}, \dots, v_0 \in V$.

Proof. We prove the statement by induction on t. If $t = 1$, then the statement is true by choice of X_1. Next assume that (4.5.2) is true for a $t \in [L-1]$ and all $j \in [\min\{k, t\}]$.

Given such a $t \in [L-1]$, let $j \in [\min\{k, t+1\}]$ and $m = \min\{k-1, t\}$ (hence $j \leqslant m+1$) as well as $v_{-j+1}, \dots, v_0 \in V$ be given. Furthermore, let $U = V \setminus \{v_{-j+1}, \dots, v_0\}$ and $h = m - j + 1 \geqslant 0$. Now we establish (4.5.2) with $t+1$ instead of t. We compute

$$\begin{aligned}
&\mathbb{P}[X_{t-j+2} \dots X_{t+1} = v_{-j+1} \dots v_0] \\
=& \sum_{(v_{-m}, \dots, v_{-j}) \in U^{\underline{h}}} \mathbb{P}[X_{t-m+1} \dots X_{t+1} = v_{-m} \dots v_0] \\
=& \sum_{(v_{-m}, \dots, v_{-j}) \in U^{\underline{h}}} \frac{\omega(v_{-m} \dots v_0)\mathbb{P}[X_{t-m+1} \dots X_t = v_{-m} \dots v_{-1}]}{(k-m)\omega(v_{-m} \dots v_{-1})} \qquad (4.5.3) \\
=& \sum_{(v_{-m}, \dots, v_{-j}) \in U^{\underline{h}}} \frac{\omega(v_{-m} \dots v_0) \cdot (k-m)!\omega(v_{-m} \dots v_{-1})}{(k-m)\omega(v_{-m} \dots v_{-1}) \cdot k!\omega(\varnothing)} \qquad (4.5.4) \\
=& \frac{(k-m-1)!}{k!\omega(\varnothing)} \sum_{(v_{-m}, \dots, v_{-j}) \in U^{\underline{h}}} \omega(v_{-m} \dots v_0) \\
=& \frac{(k-m-1)!}{k!\omega(\varnothing)} \sum_{\substack{e \in E: \\ v_{-j+1}, \dots, v_0 \in e}} \binom{k-j}{h} \cdot h! \cdot \omega(e)
\end{aligned}$$

$$= \frac{(k-j)!\omega(v_{-j+1}\dots v_0)}{k!\omega(\varnothing)},$$

where we used (4.5.1) for (4.5.3) and the induction hypothesis for (4.5.4). □

The next lemma shows that whenever $X_1 \dots X_{t_\star}$ is an (L,ω)-random walk in H for some balanced ω and x a vertex in a slightly larger k-graph, then after a few steps $X_1 \dots X_{t_\star}$ has a decent chance of producing an x-absorber in $2k$ consecutive steps. This follows easily, because as an η-intersecting k-graph, H contains sufficiently many suitable x-absorbers, Lemma 4.3.11 guarantees that there are sufficiently many ways for the random walk to arrive at such an x-absorber independent of the starting conditions and the balancedness of ω entails that every walk extending an already chosen initial segment of the random walk occurs with sufficiently large probability.

Observation 4.5.4. *Let $1/n \ll \alpha \ll \nu_+, \ell \ll \eta, 1/k$ and $L \geqslant 2k$. Suppose H_+ is an η-intersecting k-graph on at most $(1+\nu_+)n$ vertices with vertex set V_+ and H is an induced subgraph of H_+ on n vertices. Suppose $X_1 \dots X_L$ is an (L,ω)-random walk in H for some balanced ω. Then, for all $t \in \{-\ell, \dots, L-\ell-2k\}$ and $x \in V_+$, we have*

$$\mathbb{P}[X_{t+\ell+1}\dots X_{t+\ell+2k} \text{ is an } x\text{-absorber in } H_+ \mid X_1 \dots X_t] \geqslant \alpha.$$

Measuring the impact of removing the vertices of an (L,ω)-random walk in H from H is one of the core objectives in this subsection. The following lemma shows that for each $(k-1)$-set of vertices of H, its neighbourhood is essentially visited as often as expected. Via Lemma 4.3.10, this transfers to the vertex degrees appropriately.

Lemma 4.5.5. *Let $1/n \ll \eta, 1/k, 1/L$. Suppose H is a k-graph on n vertices with vertex set V and $\delta(H) \geqslant \eta n$. Suppose $X_1 \dots X_{t_\star}$ with $n^{1/3}/2 \leqslant t_\star \leqslant 2n^{1/3}$ is an (L,ω)-random walk in H for some balanced perfect fractional matching ω in H. Let $U = V \setminus \{X_t : t \in [t_\star]\}$. Then, for all $\mathbf{x} \in \binom{V}{k-1}$, we have*

$$\mathbb{P}\left[d_{H[U\cup\mathbf{x}]}(\mathbf{x}) = (1 \pm n^{-31/40})\left(1-\frac{t_\star}{n}\right)d_H(\mathbf{x})\right] \geqslant 1-\exp(-n^{1/14}). \qquad (4.5.5)$$

Proof. Since Lemma 4.5.3 yields $\mathbb{P}[X_t = v] = 1/n$ for all $t \in [t_\star]$ and $v \in V$ and since for all $\ell \in [L]$, the random variables $X_{L(s-1)+\ell}$ with $s \in [t_\star/L]$ are mutually independent by Observation 4.5.1, the statement is a consequence of Chernoff's inequality (Lemma 4.3.1).

Let us turn to the details. Fix $\mathbf{x} \in \binom{V}{k-1}$. To see that (4.5.5) holds, we will show that

$$\mathbb{P}\left[|N_H(\mathbf{x}) \cap \{X_t : t \in [t_\star]\}| = (1 \pm n^{-1/8})\frac{t_\star}{n}|N_H(\mathbf{x})|\right] \geqslant 1 - \exp(-n^{1/14}). \qquad (4.5.6)$$

To this end, for $t \in [L]$, let Y_t denote the indicator random variable of the event that $X_t \in N_H(\mathbf{x})$. Note that Lemma 4.5.3 implies $\mathbb{P}[Y_t = 1] = |N_H(\mathbf{x})|/n$. For $\ell \in [L]$, let $N_\ell = \sum_{s \in [t_\star/L]} Y_{L(s-1)+\ell}$. Crucially, note that N_ℓ is the sum of the independent random variables $Y_{L(s-1)+\ell}$ with $s \in [t_\star/L]$. Furthermore, for $t \in [t_\star]$, let Z_t denote the indicator random variable of the event that $X_t \in \{X_{t'} : t' \in [t-1]\} \cap N_H(\mathbf{x})$ and let $Z = \sum_{t \in [t_\star]} Z_t$. Observe that

$$|N_H(\mathbf{x}) \cap \{X_t : t \in [t_\star]\}| = \sum_{\ell \in [L]} N_\ell - Z.$$

Let us estimate N_ℓ and Z. For all $\ell \in [L]$, Chernoff's inequality (Lemma 4.3.1) entails

$$\mathbb{P}\left[N_\ell = \left(1 \pm \frac{n^{-1/8}}{2}\right)\frac{|N_H(\mathbf{x})|}{n}\frac{t_\star}{L}\right] \geqslant 1 - \exp(-n^{1/13}).$$

Furthermore, from $t_\star \leqslant 2n^{1/3}$, $\delta(H) \geqslant \eta n$, (4.5.1), and the balancedness of ω, we obtain

$$\mathbb{P}[Z_t = 1 \mid Z_1, \ldots, Z_{t-1}] \leqslant n^{-1/2}$$

for all $t \in [t_\star]$. This shows that Z is stochastically dominated by a binomial random variable with parameters $t_\star$ and $n^{-1/2}$ and thus Chernoff's inequality (Lemma 4.3.1) implies

$$\begin{aligned}\mathbb{P}\left[Z \geqslant \frac{n^{-1/8}}{2}\frac{t_\star}{n}|N_H(\mathbf{x})|\right] &\leqslant \mathbb{P}\left[Z \geqslant \left(1 + \frac{\eta}{3}n^{3/8}\right)n^{-1/2}t_\star\right] \\ &\leqslant \exp(-n^{1/6}).\end{aligned}$$

The union bound yields (4.5.6). □

4.5.2 Building the absorbing structure

The following result verifies the existence of few vertex-disjoint paths whose union contains many vertex absorbers. Moreover, there is in fact a "uniform" probabilistic construction of these paths. To state the next result, we introduce the following terminology concerning paths. For a path $P = v_1 \ldots v_\ell$, we define the boundary ∂P as $\partial P = P[\{v_1, \ldots, v_k, v_{\ell-k+1}, \ldots, v_\ell\}]$ if $\ell \geqslant 2k+1$ and $\partial P = P$ otherwise. Furthermore, for a collection of paths $\mathcal{P}$, we define $\partial\mathcal{P} = \bigcup_{P\in\mathcal{P}} \partial P$.

Lemma 4.5.6. *Let $1/n \ll \varrho \ll 1/L \ll 1/a \ll \nu_+, 1/\ell \ll \eta, 1/k$. Suppose H_+ is an η-intersecting k-graph on at most $(1+\nu_+)n$ vertices with vertex set V_+ and H is an induced ϱ-almost regular subgraph of H_+ on n vertices with edge set E. Let $\vartheta = 1/a$ and for all $x \in V_+$, denote the set of x-absorbers in H_+ by $\mathcal{A}_x$. Then, there is a probabilistic construction of a collection $\mathcal{P}$ of L-paths in H together with a set $\mathcal{B} \subseteq \bigcup_{P\in\mathcal{P}} \mathcal{B}^{a,\ell}(P)$ such that the following holds.*

(i) $|\mathcal{P}| \leqslant \vartheta^2 n/L$;

(ii) $H - V(\mathcal{P})$ *is* 2ϱ*-almost regular;*

(iii) $|\{B \in \mathcal{B} \colon \mathcal{A}^\ell(B) \cap \mathcal{A}_x \neq \varnothing\}| \geqslant 3\vartheta^4 n$ *for all* $x \in V_+$ *(in particular,* $|\mathcal{B}| \geqslant 3\vartheta^4 n$*);*

(iv) $|\{x \in V_+ \colon \mathcal{A}^\ell(B) \cap \mathcal{A}_x = \varnothing\}| \leqslant \vartheta^4 n$ *for all* $B \in \mathcal{B}$*;*

(v) $\mathbb{P}[e \in E(\mathcal{P})] \leqslant \vartheta \frac{1}{n^{k-1}}$ *and* $\mathbb{P}[e \in E(\partial P)] \leqslant \frac{1}{L}\frac{1}{n^{k-1}}$ *for all* $e \in E$*.*

Proof. The key idea of the proof is as follows. Constructing $\mathcal{P}$ by starting with $\mathcal{P} = \varnothing$ and iteratively adding suitable random walks in H to $\mathcal{P}$ yields a $\mathcal{P}$ as desired apart from (v) with probability at least $1/5$. This can be used to obtain an appropriate probability distribution on such collections $\mathcal{P}$.

More precisely, for $t_\star = n^{1/3}$, we will construct $\mathcal{P}$ in $s_\star = \vartheta^2 n/t_\star$ *stages*, where in stage $s \in [s_\star]$ we potentially add the L-walks generated by an (L, ω^{s-1})-random walk $X^s = X_1^s \dots X_{t_\star}^s$ in a random subgraph H^{s-1} of H for a balanced perfect fractional matching ω^{s-1} in H^{s-1}. To this end, for every subgraph $S \subseteq H$ with a balanced perfect fractional matching, fix one such perfect fractional matching in S to which we refer as the perfect fractional matching *assigned* to S. We inductively define the random k-graphs $H = H^0 \supseteq \dots \supseteq H^{s_\star}$ and random walks $X^1, \dots, X^{s_\star}$ as follows.

(1) Let $H^0 = H$;

(2) for $s \in [s_\star]$, define the random walk $X^s = X_1^s \dots X_{t_\star}^s$ in H^{s-1} and $H^s \subseteq H^{s-1}$ as follows.

  (2a) If there is a balanced perfect fractional matching in H^{s-1}, let $X^s = X_1^s \dots X_{t_\star}^s$ be an (L, ω^{s-1})-random walk in H^{s-1}, where ω^{s-1} denotes the perfect fractional matching assigned to H^{s-1}; otherwise let $X^s = X^{s-1}$;

  (2b) if X^s is self-avoiding, let $H^s = H^{s-1} - \{X_t^s \colon t \in [t_\star]\}$; otherwise let $H^s = H^{s-1}$.

Note that Lemma 4.3.8 guarantees a balanced perfect fractional matching in $H_0 = H$ and that for $s \in [s_\star]$, if $X^s = X^{s-1}$, then $H^s = H^{s-1}$ by construction of H^{s-1} even if X^s is self-avoiding.

We may think of stages $s \in [s_\star]$ where there is no balanced perfect fractional matching in H^{s-1} and stages $s \in [s_\star]$ where X^s is not self-avoiding as *failed* stages as these stages fail to generate L-paths disjoint from each other and all L-paths previously added to $\mathcal{P}$. While stages that fail due to the absence of a balanced perfect fractional matching in H^{s-1} are *fatal* in the sense that they entail failure in all subsequent stages, we can otherwise recover from failure in the sense that subsequent stages may still be successful. We repeat previously generated paths in the case of fatal failure simply for convenience.

Later we show that with probability at least 4/5 no fatal failure occurs; that is, there is a balanced perfect fractional matching in H^{s-1} for all $s \in [s_\star]$. We also show that with probability at least 4/5, there are at most $n^{1/2}$ non-fatal failures, that is, there are at most $n^{1/2}$ stages $s \in [s_\star]$ such that X^s is not self-avoiding.

For $s \in [s_\star]_0$, let $V^s = V(H^s)$ and let $V = V^0 = V(H)$. We use $p_\star = t_\star/L$ to refer to the number of L-walks generated in every stage. For $s \in [s_\star]$ and $p \in [p_\star]$, let $P^s_p = X^s_{L(p-1)+1} \ldots X^s_{Lp}$ denote the p-th walk generated in stage s and let $\mathcal{P}^s = \{P^s_p \colon p \in [p_\star]\}$ denote the set of all walks generated in stage s. Let

$$\mathcal{P}' = \bigcup_{s\in[s_\star]} \mathcal{P}^s, \qquad \mathcal{B}(\mathcal{P}') = \bigcup_{P\in\mathcal{P}'} \mathcal{B}^{a,\ell}(P),$$
$$\mathcal{P} = \bigcup_{s\in[s_\star]\colon X^s \text{ is self-avoiding}} \mathcal{P}^s \qquad \text{and} \qquad \mathcal{B}(\mathcal{P}) = \bigcup_{P\in\mathcal{P}} \mathcal{B}^{a,\ell}(P).$$

In accordance with (iv), we say that an (a,ℓ)-block B is *good* if there are at most $\vartheta^4 n$ vertices $x \in V_+$ such that $\mathcal{A}^\ell(B)$ does not contain an x-absorber in H_+ and we call B *bad* otherwise. We define events $\mathcal{E}_1$, $\mathcal{E}_2$, $\mathcal{E}_3$, and $\mathcal{E}_4$ as follows.

- $\mathcal{E}_1$: For all $x \in V_+$, there are at least $5\vartheta^4 n$ triples (s,p,i) with $s \in [s_\star]$, $p \in [p_\star]$, $i \in [L/(a(2k+\ell))]$, and $\mathcal{A}^\ell(B^{a,\ell}_i(P^s_p)) \cap \mathcal{A}_x \neq \varnothing$;
- $\mathcal{E}_2$: there are at most $\vartheta^4 n$ bad blocks in $\mathcal{B}(\mathcal{P}')$;
- $\mathcal{E}_3$: there are at most $n^{1/2}$ stages $s \in [s_\star]$ such that X^s is not self-avoiding;
- $\mathcal{E}_4$: $H^{s_\star}$ is $(\varrho + n^{-1/12})$-almost regular.

We claim the following.

If $\mathcal{E} = \mathcal{E}_1 \cap \ldots \cap \mathcal{E}_4$ occurs, then $\mathcal{P}$ with $\mathcal{B} = \{B \in \mathcal{B}(\mathcal{P}) : B$ is good$\}$ satisfies (i)–(iv). (4.5.7)

To see that this is true, we argue as follows. First observe that since $H - X$ is $(\eta/2)$-intersecting for all $X \subseteq V$ of size at most $\eta n/2$, the random k-graphs H^s with $s \in [s_\star]$ are $(\eta/2)$-intersecting. Thus if $\mathcal{E}_4$ occurs, there exists a balanced perfect fractional matching in $H^{s_\star}$ by Lemma 4.3.8. This implies that there was a balanced perfect fractional matching in H^s for all $s \in [s_\star]_0$ and thus no fatal failure occured; otherwise $H^{s_\star}$ would be equal to the first such k-graph without a perfect fractional matching by construction.

Next note that if $\mathcal{E}_1 \cap \mathcal{E}_3$ occurs, the number of triples (s, p, i) with $s \in [s_\star]$, $p \in [p_\star]$, $i \in [L/(a(2k+\ell))]$, and $\mathcal{A}^\ell(B_i^{a,\ell}(P_p^s)) \cap \mathcal{A}_x \neq \varnothing$ such that X^s is self-avoiding is at least $4\vartheta^4 n$ for all $x \in V_+$. This together with the previous observation shows that whenever $\mathcal{E}_1 \cap \mathcal{E}_3 \cap \mathcal{E}_4$ occurs we have that for all $x \in V_+$, there are at least $4\vartheta^4 n$ blocks $B \in \mathcal{B}(\mathcal{P})$ with $\mathcal{A}^\ell(B) \cap \mathcal{A}_x \neq \varnothing$. If $\mathcal{E}_2$ now also occurs in addition to $\mathcal{E}_1 \cap \mathcal{E}_3 \cap \mathcal{E}_4$, we lose at most $\vartheta^4 n$ blocks by dropping bad blocks which shows that (iii) holds.

Finally, since we only consider good blocks, (iv) holds, $s_\star \cdot p_\star = \vartheta^2 n/L$ implies that (i) holds by construction of $\mathcal{P}$, and if $\mathcal{E}_4$ occurs, (ii) holds because $H - V(\mathcal{P}) = H^{s_\star}$. This proves (4.5.7).

Let us finish the proof assuming $\mathbb{P}[\mathcal{E}] \geqslant 1/5$. By (4.5.7), this implies that choosing $\hat{\mathcal{P}}$ from all possible realisations of $\mathcal{P}$ such that $\mathbb{P}[\hat{\mathcal{P}} = \mathcal{Q}] = \mathbb{P}[\mathcal{P} = \mathcal{Q} \mid \mathcal{E}]$ for all possible realisations $\mathcal{Q}$ of $\mathcal{P}$ is a probabilistic construction as desired because for all $e \in E$ and $i \in [L-k+1]$, Lemma 4.5.3 entails

$$\begin{aligned}
\mathbb{P}[\exists (X_t)_{t\in[L]} \in \hat{\mathcal{P}} \colon \{X_i, \ldots, X_{i+k-1}\} = e] &= \mathbb{P}[\exists (X_t)_{t\in[L]} \in \mathcal{P} \colon \{X_i, \ldots, X_{i+k-1}\} = e \mid \mathcal{E}] \\
&\leqslant 5 \sum_{s\in[s_\star]} \sum_{(X_t)_{t\in[L]}\in\mathcal{P}^s} \mathbb{P}[\{X_i, \ldots, X_{i+k-1}\} = e] \\
&= 5 \sum_{s\in[s_\star]} \sum_{(X_t)_{t\in[L]}\in\mathcal{P}^s} k! \frac{\omega^{s-1}(e)}{k!\omega^{s-1}(\varnothing)} \\
&\leqslant 5 \cdot s_\star \cdot p_\star \cdot k! \cdot 2 \frac{1}{\frac{\eta}{2} n^k} \leqslant \frac{\vartheta}{L} \frac{1}{n^{k-1}},
\end{aligned}$$

which implies (v).

It remains to prove $\mathbb{P}[\mathcal{E}] \geqslant 1/5$. This easily follows if $\mathbb{P}[\mathcal{E}_i] \geqslant 4/5$ for all $i \in [4]$, which is what we prove next. For $x \in V_+$ and $s \in [s_\star]$, let $Y_{x,s}$ denote the random variable counting the pairs (p, i) with $p \in [p_\star]$, $i \in [L/(2k+\ell)]$, and $A_i^\ell(P_p^s) \in \mathcal{A}_x$. Note that $\mathcal{E}_1$ occurs if $Y_{x,s} \geqslant 5\vartheta^3 n/s_\star$ for all $x \in V_+$ and $s \in [s_\star]$. Observation 4.5.1 in conjunction with Observation 4.5.4 shows that $Y_{x,s}$ stochastically dominates a binomial random variable with parameters $p_\star \frac{L}{2k+\ell}$ and $\vartheta^{1/2}$. Hence Chernoff's inequality (Lemma 4.3.1) implies that

$$\mathbb{P}\left[Y_{x,s} < \frac{5\vartheta^3 n}{s_\star}\right] \leqslant \mathbb{P}\left[Y_{x,s} \leqslant \frac{1}{2} \cdot \vartheta^{1/2} \cdot p_\star \frac{L}{2k+\ell}\right] \leqslant \exp(-n^{1/4})$$

and the union bound yields $\mathbb{P}[\mathcal{E}_1] \geqslant 4/5$.

For all $x \in V_+$ and $B \in \mathcal{B}(\mathcal{P}')$, Observation 4.5.4 implies

$$\mathbb{P}[\mathcal{A}^\ell(B) \cap \mathcal{A}_x = \varnothing] \leqslant (1 - \vartheta^{1/2})^a \leqslant \vartheta^5/2.$$

This entails

$$\mathrm{ex}[|\{x \in V_+ \colon \mathcal{A}^\ell(B) \cap \mathcal{A}_x = \varnothing\}|] \leqslant \vartheta^5 n.$$

Using Markov's inequality we obtain $\mathbb{P}[B \text{ is bad}] \leqslant \vartheta$ for all $B \in \mathcal{B}(\mathcal{P}')$. This shows that the expected value of the random variable counting the number of bad blocks in $\mathcal{B}(\mathcal{P}')$ is at most

$$\vartheta \cdot s_\star \cdot p_\star \cdot \frac{L}{a(2k+\ell)} = \frac{\vartheta^4}{2k+\ell} n \leqslant \vartheta^4 n/5.$$

Again using Markov's inequality, this yields $\mathbb{P}[\mathcal{E}_2] \geqslant 4/5$.

For all $s \in [s_\star]$, Observation 4.5.2 implies that X^s is self-avoiding with probability at least $1 - 2n^{-1/4}$. From this, we obtain that the expected value of the random variable counting the stages $s \in [s_\star]$ where X^s is not self-avoiding is at most $2n^{-1/4} \cdot s_\star = 2\vartheta^2 n^{5/12} \leqslant n^{1/2}/5$. Using Markov's inequality we conclude that $\mathbb{P}[\mathcal{E}_3] \geqslant 4/5$.

To see that $\mathbb{P}[\mathcal{E}_4] \geqslant 4/5$ holds, we show that all random k-graphs H^s with $s \in [s_\star]_0$ are almost regular with high probability. More precisely, for $s \in [s_\star]_0$, let $\mathcal{E}_4^s$ denote the event that H^s is $(\varrho + sn^{-3/4})$-almost regular. Our goal is to show

$$\mathbb{P}\Big[\bigcap_{s'=0}^{s} \mathcal{E}_4^{s'}\Big] \geqslant 1 - s\exp(n^{-1/15}) \tag{4.5.8}$$

for all $s \in [s_\star]_0$. This suffices because $\mathcal{E}_4^{s_\star} \subseteq \mathcal{E}_4$. We proceed by induction on s. First note that $\mathbb{P}[\mathcal{E}_4^0] = 1$. Next, assume that (4.5.8) holds for some $s \in [s_\star - 1]_0$. Then, Lemma 4.3.8 guarantees that there is no fatal failure in stage $s+1$. For $U = V^s \smallsetminus \{X_t^{s+1} \colon t \in [t_\star]\}$

and $\mathbf{x} \in \binom{V^s}{k-1}$, Lemma 4.5.5 entails

$$\mathbb{P}\left[d_{H^s[U\cup\mathbf{x}]}(\mathbf{x}) = (1 \pm n^{-31/40})\left(1 - \frac{t_\star}{|V^s|}\right) d_{H^s}(\mathbf{x}) \;\middle|\; \bigcap_{s'=0}^{s} \mathcal{E}_4^{s'}\right] \geqslant 1 - \exp(-n^{1/14}).$$

Lemma 4.3.10 and the union bound yield $\mathbb{P}[\mathcal{E}_4^{s+1} \mid \bigcap_{s'=0}^{s} \mathcal{E}_4^{s'}] \geqslant 1 - \exp(-n^{1/15})$ and thus by induction hypothesis $\mathbb{P}[\bigcap_{s'=0}^{s+1} \mathcal{E}_4^{s'}] \geqslant 1 - (s+1)\exp(-n^{1/15})$. □

In the following Proposition 4.5.8 we employ the absorbers provided by Lemma 4.5.6 to absorb sets of vertices X in the sense that for all suitable X, we find a selection of these absorbers such that the selected absorbers can simultaneously absorb all vertices in X. As this requires matching all vertices $x \in X$ to an x-absorber, we make use of the following observation which follows easily by iteratively using Hall's theorem and removing perfect matchings.

Observation 4.5.7. *Suppose G is a bipartite graph on $2n$ vertices with bipartition $\{V_1, V_2\}$ with $|V_1| = |V_2| = n$. For $i \in [2]$, let $\delta_i = \min\{d_G(v)\colon v \in V_i\}$. Then, there are at least $(\delta_1 + \delta_2 - n)/2$ disjoint perfect matchings in G.*

Proposition 4.5.8. *Let $1/n \ll \varrho \ll 1/L \ll \vartheta \ll \nu_+ \ll \eta, 1/k$. Suppose H_+ is an η-intersecting k-graph on at most $(1+\nu_+)n$ vertices with vertex set V_+ and edge set E_+ and H is a ϱ-almost regular induced subgraph of H_+ on n vertices. Then, there is a probabilistic construction of a collection $\mathcal{P}$ of L-paths in H together with a function $\sigma\colon \mathcal{P} \to [L]_0$ such that the following holds.*

(i) *$|\mathcal{P}| \leqslant \vartheta^2 n/L$;*

(ii) *$c = \sum_{P\in\mathcal{P}} \sigma(P) \geqslant \vartheta^4 n$;*

(iii) *$H - V(\mathcal{P})$ is 2ϱ-almost regular;*

(iv) *for all probabilistic constructions of a set $X \subseteq V_+$ with $X \cap V(\mathcal{P}) = \varnothing$ and $|X| = c$, there is a probabilistic construction of a collection $\mathcal{P}'$ of paths in H_+ such that the following holds.*

  (iv.i) *There is a bijection $\varphi : \mathcal{P} \to \mathcal{P}'$ such that for all $P \in \mathcal{P}$, the path $\varphi(P)$ is an $(L + \sigma(P))$-path with $V(\varphi(P)) \subseteq V(P) \cup X$ that has the same ordered end-edges as P;*

(iv.ii) *for all $e \in E_+$, we have*

$$\mathbb{P}[e \in E(\mathcal{P}')] \leqslant \frac{1}{\vartheta^4 n^{k-1}} \quad \textit{and} \quad \mathbb{P}[e \in E(\partial\mathcal{P}')] \leqslant \frac{1}{L n^{k-1}}.$$

Proof. For the probabilistic construction of the collection $\mathcal{P}$ together with a function σ, we will employ Lemma 4.5.6. Then, for the probabilistic construction of $\mathcal{P}'$ we will randomly absorb all vertices $x \in X$ into the paths in $\mathcal{P}$; that is, we will randomly place every vertex $x \in X$ in the center of an x-absorber in H_+ that is a subgraph of a path in $\mathcal{P}$.

In more detail, we argue as follows. Let $E = E(H)$. For $x \in V_+$, let $\mathcal{A}_x$ denote the set of x-absorbers in H_+. Choose ℓ such that $\vartheta \ll \nu_+, 1/\ell \ll \eta, 1/k$. Let $a = 1/\vartheta$. Lemma 4.5.6 provides a probabilistic construction of a collection $\mathcal{P}$ of L-paths in H together with a set $\mathcal{B} \subseteq \bigcup_{P\in\mathcal{P}} \mathcal{B}^{a,\ell}(P)$ such that the following holds.

(1) $|\mathcal{P}| \leqslant \vartheta^2 n/L$;

(2) $H - V(\mathcal{P})$ is 2ϱ-almost regular;

(3) $|\{B \in \mathcal{B} \colon \mathcal{A}^\ell(B) \cap \mathcal{A}_x \neq \varnothing\}| \geqslant 3\vartheta^4 n$ for all $x \in V_+$;

(4) $|\{x \in V_+ \colon \mathcal{A}^\ell(B) \cap \mathcal{A}_x = \varnothing\}| \leqslant \vartheta^4 n$ for all $B \in \mathcal{B}$;

(5) $\mathbb{P}[e \in E(\mathcal{P})] \leqslant \vartheta \frac{1}{n^{k-1}}$ and $\mathbb{P}[e \in E(\partial P)] \leqslant \frac{1}{L}\frac{1}{n^{k-1}}$ for all $e \in E$.

Let $\sigma \colon \mathcal{P} \to [L]_0$ with $\sigma(P) = |\mathcal{B} \cap \mathcal{B}^{a,\ell}(P)|$ for all $P \in \mathcal{P}$ and let $c = |\mathcal{B}|$. Then (i)–(iii) hold.

For all (a,ℓ)-blocks B in H and all $x \in V_+$ with $\mathcal{A}_x \cap \mathcal{A}^\ell(B) \neq \varnothing$, we fix one x-absorber $A_x(B) \in \mathcal{A}^\ell(B)$ in H_+ for later use. For $\mathcal{P}$ as above and a probabilistic construction of a set $X \subseteq V_+$ with $X \cap V(\mathcal{P}) = \varnothing$ and $|X| = c$, we obtain $\mathcal{P}'$ through random absorption of all elements of X into the paths in $\mathcal{P}$ as follows. Consider the auxiliary bipartite graph G with bipartition $\{X, \mathcal{B}\}$ and an edge between $x \in X$ and $B \in \mathcal{B}$ if and only if $\mathcal{A}_x \cap \mathcal{A}^\ell(B) \neq \varnothing$. Intuitively, edges in this graph represent possible absorptions of elements of X into the adjacent blocks. Thus, we may think of perfect matchings in G as representations of valid stategies for the absorption of all elements of X. Due to (3) and (4), we have $d_G(x) \geqslant 3\vartheta^4 n$ for all $x \in X$ and $d_G(B) \geqslant c - \vartheta^4 n$ for all $B \in \mathcal{B}$. Therefore, Observation 4.5.7 guarantees the existence of a set of $\vartheta^4 n$ edge-disjoint perfect matchings in G. Pick one matching uniformly at random from this set and interpret this matching as a bijection $\mu \colon X \to \mathcal{B}$. Let φ denote the bijection defined on $\mathcal{P}$ that maps $P \in \mathcal{P}$ to the path obtained from P when placing $\mu^{-1}(B)$ in the center of $A_{\mu^{-1}(B)}(B)$ for all $B \in \mathcal{B} \cap \mathcal{B}^{a,\ell}(P)$. Let $\mathcal{P}' = \mathrm{Im}(\varphi)$.

It remains to check that this defines a probabilistic construction of a collection $\mathcal{P}'$ of paths satisfying (iv.ii). For all $\mathbf{x} \in \binom{V_+}{k-1}$ and all possible realisations $\mathcal{Q}$ of $\mathcal{P}$, there is at most one block $B \in \bigcup_{P \in \mathcal{Q}} \mathcal{B}^{a,\ell}(P)$ with $d_B(\mathbf{x}) \geqslant 1$. Hence, for all $x \in V_+$ and $\mathbf{x} \in \binom{V_+}{k-1}$, we have

$$\mathbb{P}[x \in X \wedge d_{\mu(x)}(\mathbf{x}) \geqslant 1 \mid \mathcal{P}, X] \leqslant \frac{1}{\vartheta^4 n} \tag{4.5.9}$$

by construction of μ. Thus, for all $e \in E_+$, (5) yields

$$\begin{aligned}
\mathbb{P}[e \in E(\mathcal{P}')] &\leqslant \mathbb{P}[\mathbf{x} \cap X = \varnothing \wedge e \in E(\mathcal{P})] + \sum_{x \in e} \mathbb{P}[x \in X \wedge d_{\mu(x)}(e \smallsetminus \{x\}) \geqslant 1] \\
&\overset{(5)}{\leqslant} \frac{\vartheta}{n^{k-1}} + \sum_{x \in e} \mathbb{P}[x \in X \wedge d_{\mu(x)}(e \smallsetminus \{x\}) \geqslant 1 \wedge d_{\mathcal{P}}(e \smallsetminus \{x\}) \geqslant 1] \\
&\overset{(4.5.9)}{\leqslant} \frac{\vartheta}{n^{k-1}} + \sum_{x \in e} \frac{1}{\vartheta^4 n} \mathbb{P}[d_{\mathcal{P}}(e \smallsetminus \{x\}) \geqslant 1] \\
&\leqslant \frac{\vartheta}{n^{k-1}} + \sum_{x \in e} \frac{1}{\vartheta^4 n} \sum_{v \in V_+} \mathbb{P}[(e \smallsetminus \{x\}) \cup \{v\} \in E(\mathcal{P})] \\
&\overset{(5)}{\leqslant} \frac{\vartheta}{n^{k-1}} + j \cdot \frac{1}{\vartheta^4 n} \cdot (1 + \nu_+) n \cdot \frac{\vartheta}{n^{k-1}} \leqslant \frac{1}{\vartheta^4 n^{k-1}}.
\end{aligned}$$

Furthermore, we obtain

$$\mathbb{P}[e \in E(\partial \mathcal{P}')] = \mathbb{P}[e \in E(\partial \mathcal{P})] \leqslant \frac{1}{L} \frac{1}{n^{k-1}},$$

which completes the proof. □

4.6 From paths to cycles

In this section, we perform the step from the yield of Proposition 4.4.1 to the decomposition into cycle factors as in our main theorem. We do this by describing a random process which converts one almost spanning path collection into a cycle factor and subsequently using another random process which repeatedly applies the first one to transform path collection after path collection. The first step is done in Lemma 4.6.3 and the second in Proposition 4.6.4.

We begin with the following somewhat standard lemma, which states that in any η-intersecting k-graph, there is a small (reservoir) set such that between any two ordered edges, there are many short paths with all "inner" vertices in this set. Complementing the notation ∂P, for a path $P = v_1, \ldots, v_\ell$, we define the *interior* of P as (the subpath) $P^\circ =$

$v_{k+1}, \dots, v_{\ell-k}$ if $\ell \geqslant 2k+1$ and $P^\circ = P[\varnothing]$ otherwise. The vertex set of P° is the set of *inner vertices* of P. For a collection $\mathcal{P}$ of paths, we define $\mathcal{P}^\circ = \bigcup_{P \in \mathcal{P}} P^\circ$.

Lemma 4.6.1. *Suppose $1/n \ll \beta, \varrho \ll 1/\ell_0, 1/\ell_1 \ll \eta, 1/k$, where $\ell_0 \leqslant \ell_1$. Suppose H is an η-intersecting ϱ-almost regular k-graph on n vertices with vertex set V and edge set E. Then there is a set $R \subseteq V$ such that the following holds.*

(i) *$\beta n/2 \leqslant |R| \leqslant \beta n$;*

(ii) *for all $\vec{s}, \vec{t} \in \overrightarrow{E}$ with $s \cap t = \varnothing$ and all integers $\ell \in [\ell_0, \ell_1]$, the number of $\vec{s}$–$\vec{t}$-paths P in H with $|V(P^\circ)| = \ell$ and $V(P^\circ) \subseteq R$ is at least $\beta |R|^\ell$;*

(iii) *$H - R$ is 2ϱ-almost regular.*

This follows easily from Lemma 4.3.11 by considering a random set of vertices in which each vertex is included independently at random (for instance, with probability $3\beta/4$) and using Chernoff's inequality, McDiarmid's inequality, and the union bound to show that the random set has the desired properties with high probability. We omit the calculations, which are standard by now.

The next lemma ensures that under the right conditions, many tuples can be connected in a probabilistically well behaved way. It will later be useful when building the cycle factor in Lemma 4.6.3.

Lemma 4.6.2. *Suppose $1/n \ll \zeta \ll \beta, 1/\ell_1, 1/\ell_0 \ll 1/k$, where $\ell_0 \leqslant \ell_1$. Suppose H is a k-graph on n vertices with vertex set V and edge set E. Suppose that $\mathcal{Q} = \{\vec{s}_1, \vec{t}_1, \dots, \vec{s}_m, \vec{t}_m\} \subseteq \overrightarrow{E}$ is a random set with $m \leqslant \zeta n$, $e \cap e' = \varnothing$ for all distinct $\vec{e}, \vec{e}' \in \mathcal{Q}$, and $\mathbb{P}[\vec{e} \in \mathcal{Q}] \leqslant \frac{\zeta}{n^{k-1}}$ for every $\vec{e} \in \overrightarrow{E}$. Further, for all $i \in [m]$, let $\lambda_i \in [\ell_0, \ell_1]$ and suppose that $\mathcal{P}_i$ is a set of at least βn^{λ_i} $\vec{s}_i$–$\vec{t}_i$-paths with λ_i inner vertices. Then, there is a probabilistic construction of a collection $\mathcal{W} \subseteq \bigcup_{i \in [m]} \mathcal{P}_i$ of paths with $|\mathcal{W} \cap \mathcal{P}_i| = 1$ for all $i \in [m]$ and $\mathbb{P}[e \in E(\mathcal{W})] \leqslant \frac{\zeta^{1/2}}{n^{k-1}}$ for every $e \in E$.*

Proof. We aim to connect the ordered edges $\vec{s}_i$ and $\vec{t}_i$ by choosing one of the paths in $\mathcal{P}_i$ uniformly at random. However, since the paths shall be vertex-disjoint, we employ an iterative procedure where in each step we only consider those paths in $\mathcal{P}_i$ which are vertex-disjoint from all previously chosen ones and all ordered edges in $\mathcal{Q}$.

Suppose we have already chosen paths $W_1, \dots, W_{i-1}$ for some $i \in [m-1]$, where W_j is an $\vec{s}_j$–$\vec{t}_j$-path with λ_j inner vertices for all $j \in [i-1]$. Then

$$\Big| \bigcup_{j \in [i-1]} V(W_j) \cup \bigcup_{\vec{e} \in \mathcal{Q}} e \Big| \leqslant \ell_1 \cdot m + 2(k-1) \cdot m \leqslant \frac{\beta}{2} n \,.$$

Thus, since $|\mathcal{P}_i| \geqslant \beta n^{\lambda_i}$, there are still at least $\beta n^{\lambda_i} - \frac{\beta}{2}n^{\lambda_i} = \frac{\beta}{2}n^{\lambda_i}$ paths in $P \in \mathcal{P}_i$ with $V(P) \cap (\bigcup_{j\in[i-1]} V(W_j) \cup \bigcup_{\vec{e}\in\mathcal{Q}} e) = \varnothing$. Choose one of these uniformly at random as W_i.

Now, set $\mathcal{W} = \{W_1, \dots, W_m\}$ and let us analyse the probabilities. For all $j \in [k]_0$, $\mathbf{x} \in \binom{V}{j}$ and $i \in [m]$, the number of $\vec{s}_i$–$\vec{t}_i$-paths W in H with $|V(W^\circ)| = \lambda_i$ and $\mathbf{x} \subseteq V(W^\circ)$ is at most $\ell_1^k \cdot n^{\lambda_i - j}$. Therefore, by the choice of $\mathcal{W}$, this implies $\mathbb{P}[\mathbf{x} \subseteq V(W_i^\circ) \mid \vec{s}_1, \dots, \vec{s}_m, \vec{t}_1, \dots, \vec{t}_m] \leqslant \frac{2\ell_1^k}{\beta n^j}$. Furthermore, for all $j \in [k]$ and $\mathbf{x} \in \binom{V}{j}$, the number of ordered edges $\vec{e} \in \vec{E}$ with $\mathbf{x} \subseteq e$ is at most $k^k \cdot n^{k-j}$ and hence $\mathbb{P}[\exists \vec{e} \in \mathcal{Q} \colon \mathbf{x} \subseteq e] \leqslant \frac{\zeta k^k}{n^{j-1}}$ holds. Thus, for all $e \in E$, we obtain

$$\begin{aligned}
\mathbb{P}[e \in E(\mathcal{W})] &\leqslant \sum_{i\in[m]} \mathbb{P}[e \subseteq s_i \cup V(W_i^\circ) \vee e \subseteq V(W_i^\circ) \cup t_i] \\
&\leqslant \sum_{i\in[m]} \sum_{\mathbf{y}\subseteq e} \sum_{\vec{e}'\in\{\vec{s}_i,\vec{t}_i\}} \mathbb{P}[e \smallsetminus \mathbf{y} \subseteq V(W_i^\circ) \mid \mathbf{y} \subseteq e']\mathbb{P}[\mathbf{y} \subseteq e'] \\
&\leqslant \sum_{\mathbf{y}\subseteq e} \frac{2\ell_1^k}{\beta n^{|e\smallsetminus\mathbf{y}|}} \sum_{i\in[m]} \sum_{\vec{e}'\in\{\vec{s}_i,\vec{t}_i\}} \mathbb{P}[\mathbf{y} \subseteq e'] \\
&= \frac{4\ell_1^k m}{\beta n^k} + \sum_{\mathbf{y}\subseteq e\colon \mathbf{y}\neq\varnothing} \frac{2\ell_1^k}{\beta n^{|e\smallsetminus\mathbf{y}|}} \mathbb{P}[\exists \vec{e}' \in \mathcal{Q} \colon \mathbf{y} \subseteq e'] \\
&\leqslant \frac{4\zeta\ell_1^k}{\beta n^{k-1}} + 2^k \cdot \frac{\zeta k^k \cdot 2\ell_1^k}{\beta n^{k-1}} \leqslant \frac{\zeta^{1/2}}{n^{k-1}},
\end{aligned}$$

which completes the proof. □

In the following lemma, we transform a path collection as yielded by Proposition 4.4.1 into a cycle factor. It resembles the usual final step in a proof by absorption, in particular, this is where we will use Proposition 4.5.8. However, in order to subsequently apply this construction process multiple times in the proof of Proposition 4.6.4 without ruining certain quasirandom properties, we need to construct the cycle factor probabilistically and take care that the probability for any edge to occur in the constructed cycle factor is small enough. Figure 4.6.1 illustrates this process.

Lemma 4.6.3. *Suppose $1/n \ll \mu, 1/L \ll \delta \ll \varrho \ll \eta, 1/k$. Suppose H is a k-graph on n vertices with vertex set V and F is an η-intersecting ϱ-almost regular spanning subgraph of H. Let $\mathcal{P} \subseteq H - F$ be a collection of L-paths with $|V(\mathcal{P})| \geqslant (1-\mu)n$ and let $\mathcal{C}$ be a cycle factor on n vertices of girth at least L^3.*

Then, there is a probabilistic construction of a copy $\mathcal{C}'$ of $\mathcal{C}$ in H such that $\mathcal{C}' \subseteq \mathcal{P} \cup F$ and for every $e \in E(F)$ of type $\tau \in \mathcal{T} \smallsetminus \{\text{k-end}\}$, we have $\mathbb{P}[e \in E(F \cap \mathcal{C}')] \leqslant p_\tau$, *where* $p_{1\text{-con}} = \frac{\delta^{1/2}}{n^{k-1}}$*,* $p_{j\text{-end}} = \frac{1}{\delta^k n^{k-j}}$ *for all* $j \in [k-1]$*, and* $p_{j\text{-con}} = p_{\text{lo}} = \frac{1}{\delta^k n^{k-1}}$ *for all* $j \in [k] \smallsetminus \{1\}$*.*

Proof. We use the absorption method to transform the almost spanning collection of paths $\mathcal{P}$ into the desired cycle factor $\mathcal{C}'$. To this end, we perform the usual steps of a proof via absorption in a k-graph similar to $F[V \smallsetminus V(\mathcal{P})]$. That is, we set aside the absorbing structure (via Proposition 4.5.8), cover almost everything (via Proposition 4.4.1), connect all the paths in the approximate covering, in the absorbing structure, and, in this case, paths in $\mathcal{P}$ (via Lemma 4.3.11), and finally, we absorb the remaining vertices. Note however, that here we actually want to perform all these steps as a probabilistic construction and analyse the probabilities for edges of F to occur in $F \cap \mathcal{C}'$.

Suppose

$$1/n \ll \mu, 1/L \ll \delta \ll \varrho \ll 1/L' \ll \mu' \ll \beta \ll \vartheta \ll \ell_1 \ll \ell_0 \ll \eta, 1/k\,. \tag{4.6.1}$$

We say a set $\hat{\mathcal{P}} \subseteq \mathcal{P}$ is *good* if for $\hat{V} = V \smallsetminus V(\hat{\mathcal{P}})$, we have $|\hat{V}| \in [\frac{\delta}{2}n, \frac{3}{2}\delta n]$, for all $\mathbf{x}, \mathbf{y} \in \binom{V}{k-1}$, we have $|N_F(\mathbf{x}) \cap N_F(\mathbf{y}) \cap \hat{V}| \geqslant \frac{\eta}{2}|\hat{V}|$, and $F[\hat{V}]$ is 2ϱ-almost regular. Consider a random selection $\mathcal{P}_{\text{rand}}$ of paths in $\mathcal{P}$ which includes every path in $\mathcal{P}$ independently with probability $1-\delta$. Then McDiarmid's inequality (Lemma 4.3.2) guarantees that $\mathbb{P}[\mathcal{P}_{\text{rand}} \text{ is good}] \geqslant \frac{99}{100}$, say. Denote the set of good sets by $\mathscr{P}$ and pick a set $\mathcal{P}' = \{P_1, \dots, P_m\}$ at random from $\mathscr{P}$ such that

$$\mathbb{P}[\mathcal{P}' = \mathcal{Q}] = \mathbb{P}[\mathcal{P}_{\text{rand}} = \mathcal{Q} \mid \mathcal{P}_{\text{rand}} \in \mathscr{P}]$$

for all $\mathcal{Q} \in \mathscr{P}$. Then $\mathcal{P}'$ is good and we set $V_1 = V \smallsetminus V(\mathcal{P}')$. For $P \in \mathcal{P}$, we have $V(P) \subseteq V_1$ if and only if $P \notin \mathcal{P}'$, and each path in $\mathcal{P}$ is included in $\mathcal{P}'$ independently with probability $1-\delta$. Thus, for distinct $\widetilde{P}_1, \dots, \widetilde{P}_k \in \mathcal{P}$, we have

$$\begin{aligned}\mathbb{P}\Big[\bigcup_{i \in [k]} V(\widetilde{P}_i) \subseteq V_1\Big] &= \frac{\mathbb{P}[\mathcal{P}_{\text{rand}} \in \mathscr{P} \wedge \forall i \in [k] : \widetilde{P}_i \notin \mathcal{P}_{\text{rand}}]}{\mathbb{P}[\mathcal{P}_{\text{rand}} \in \mathscr{P}]} \\ &\leqslant \mathbb{P}[\forall i \in [k] : \widetilde{P}_i \notin \mathcal{P}_{\text{rand}}] \cdot \frac{100}{99} \leqslant \frac{100\delta^k}{99}\,. \end{aligned} \tag{4.6.2}$$

Next, we describe the construction of a copy $\mathcal{C}'$ of $\mathcal{C}$ and analyse $\mathbb{P}[e \in E(F \cap \mathcal{C}') \mid \mathcal{P}']$ for all $e \in E$ (note that fixing $\mathcal{P}'$ in particular fixes V_1). In the end, we use this to deduce the upper bound on the probabilities for different types of edges of F to lie in $E(F \cap \mathcal{C}')$. Since we perform several probabilistic constructions sequentially, in principle, we could

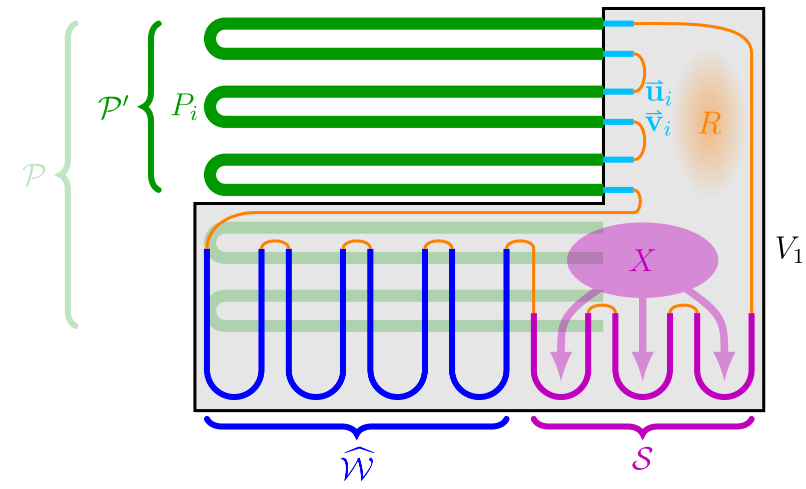


Figure 4.6.1: Transforming a path collection into a Hamiltonian cycle.

obtain all the probabilities conditioned on all previous steps. However, for our analysis it suffices to condition only on the choice of $\mathcal{P}'$.

Set aside a set $R \subseteq V_1$ provided by applying Lemma 4.6.1 with $F[V_1]$, β, ℓ_0, ℓ_1, 2ϱ, $\eta/2$ here taking the roles of H, β, ℓ_0, ℓ_1, ϱ, η in Lemma 4.6.1. Then $\frac{\beta}{2}|V_1| \leqslant |R| \leqslant \beta|V_1|$ and for all $\vec{s}, \vec{t} \in \vec{E}(F)$ with $s \cap t = \varnothing$ and all integers $\ell \in [\ell_0, \ell_1]$, the number of $\vec{s}$–$\vec{t}$-paths P in H with $|V(P^\circ)| = \ell$ and $V(P^\circ) \subseteq R$ is at least $\beta|R|^\ell$. Further, Lemma 4.6.1 guarantees that $F[V_1 \setminus R]$ is 4ϱ-almost regular.

Since we later only want to deal with edges of $F[V_1]$, we extend all paths in $\mathcal{P}'$ by an edge of F at each end. More precisely, for $i \in [m]$, let $\vec{a}_i$ be the ordered starting edge and $\vec{b}_i$ be the ordered ending edge of $P_i \in \mathcal{P}'$ and inductively choose $\vec{u}_i = (u_i^1, \ldots, u_i^k) \in \vec{E}(F)$ and $\vec{v}_i = (v_i^1, \ldots, v_i^k) \in \vec{E}(F)$ for each $i \in [m]$ as follows.

Suppose that for some $i \in [m]$, we already have defined ordered edges $\vec{u}_{i'}, \vec{v}_{i'} \in \vec{E}(F)$ for all $i' \in [i-1]$ such that $\vec{u}_{i'} P_{i'} \vec{v}_{i'}$ is a path in H. Set

$$U_i = V_1 \setminus \Big(R \cup \bigcup_{i' \in [i-1]} u_{i'} \cup v_{i'} \Big)$$

and note that since $m \leqslant \frac{n}{L}$, we know that $|U_i| \geqslant (1 - \frac{3}{2}\beta)|V_1|$ holds. Moreover, since $\mathcal{P}'$ is good, we know that for all $\mathbf{x}, \mathbf{y} \in \binom{V}{k-1}$ and $U \subseteq V_1$ with $|U| \geqslant (1-2\beta)|V_1|$, we have

$$|N_F(\mathbf{x}) \cap N_F(\mathbf{y}) \cap U| \geqslant \frac{\eta}{3}|V_1| \quad \text{and} \quad |N_F(\mathbf{x}) \cap U| \geqslant \frac{\eta}{3}|V_1|\,. \tag{4.6.3}$$

Suppose for some $j \in [k]$, vertices $u_i^{j'}$ are given for all $j' \in \{j+1, \ldots, k\}$. Then choose $u_i^j \in N_F(u_i^{j+1}, \ldots, u_i^k, a_i^1, \ldots, a_i^{j-1}) \cap U_i$ uniformly at random (note that due to (4.6.3), this means that u_i^j is chosen uniformly at random from a set of size at least $\frac{\eta}{3}|V_1|$). Subsequently, suppose that for some $j \in [k]$, vertices $v_i^{j'}$ are given for all $j' \in [j-1]$. Then choose

$$v_i^j \in N_F(b_i^{j+1}, \ldots, b_i^k, v_i^1, \ldots, v_i^{j-1}) \cap (U_i \setminus u_i)$$

uniformly at random. By this definition, the edges $u_1, v_1, \ldots, u_m, v_m$ are pairwise disjoint. Furthermore, observe that the definition of $\vec{u}_i$ and $\vec{v}_i$ together with (4.6.3) yield the following. For all $j \in [k]$, $i \in [m]$, and $\mathbf{x} \in \binom{V}{j}$,

$$\mathbb{P}[\mathbf{x} \subseteq u_i \vee \mathbf{x} \subseteq v_i \mid \mathcal{P}'] \leqslant 2k^j \left(\frac{3}{\eta|V_1|}\right)^j \leqslant \frac{\ell_0}{|V_1|^j}. \tag{4.6.4}$$

Now set $V_2 = V_1 \setminus (R \cup \bigcup_{i \in [m]} u_i \cup v_i)$ and note that by (4.6.3), $F[V_2]$ is $\eta/3$-intersecting. Furthermore, since $F[V_1 \setminus R]$ is $\eta/3$-intersecting and 4ϱ-almost regular and $|\bigcup_{i \in [m]} u_i \cup v_i| \leqslant \frac{2kn}{L}$, we know that $F[V_2]$ is 5ϱ-almost regular.

Now choose ν_+ such that $\vartheta \ll \nu_+ \ll \eta, 1/k$ and apply Proposition 4.5.8 with $F[V_1]$, $F[V_2]$, 5ϱ, L', ϑ, ν_+, $\eta/2$ here in place of H_+, H, ϱ, L, ϑ, ν_+, η there.

This engenders a probabilistic construction of a pair $(\mathcal{S}, \sigma)$ where $\mathcal{S}$ is a collection of L'-paths in $F[V_2]$ and $\sigma \colon \mathcal{S} \to [L']_0$ is a function such that the following holds.

(S1) $|\mathcal{S}| \leqslant \vartheta^2|V_2|/L'$;

(S2) $c = \sum_{S \in \mathcal{S}} \sigma(S) \geqslant \vartheta^4|V_2|$;

(S3) setting $V_3 = V_2 \setminus V(\mathcal{S})$, we have that $F[V_3]$ is 10ϱ-almost regular;

(S4) for all probabilistic constructions of a set $X \subseteq V_1$ with $X \cap V(\mathcal{S}) = \varnothing$ and $|X| = c$, there is a probabilistic construction of a collection $\mathcal{S}'$ of paths in $F[V_1]$ such that the following holds.

- (S4.1) There is a bijection $\varphi : \mathcal{S} \to \mathcal{S}'$ such that for all $S \in \mathcal{S}$, the path $\varphi(S)$ is an $(L' + \sigma(S))$-path with $V(\varphi(S)) \subseteq V(S) \cup X$ that has the same ordered end-edges as S;
- (S4.2) for all $e \in E(F)$, we have

$$\mathbb{P}[e \in E(\mathcal{S}') \mid \mathcal{P}'] \leqslant \frac{1}{\vartheta^4|V_2|^{k-1}} \quad \text{and} \quad \mathbb{P}[e \in E(\partial\mathcal{S}') \mid \mathcal{P}'] \leqslant \frac{1}{L'|V_2|^{k-1}}.$$

The next step is to cover almost all vertices of $F[V_3]$ by long paths. As in the other steps, we need to do this in a probabilistic way. This will be achieved by utilising a weak version of Proposition 4.4.1 followed by some random selections.

By (S1), $F[V_3]$ is $\eta/4$-intersecting and due to property (S3), it is 10ϱ-almost r-regular for some $r \geqslant \frac{\eta}{5}\binom{|V_3|}{k-1}$. Thus, setting $r' = (1-10\varrho)\frac{r}{k}$ yields $kr' \leqslant d_{F[V_3]}(v) \leqslant (1+21\varrho)kr'$ for all $v \in V_3$. Hence, we can apply Proposition 4.4.1 with $F[V_3]$, 21ϱ, L', $\eta/4$, μ', r' here instead of H, ϱ, L, η, μ, r there. As remarked after the statement of Proposition 4.4.1, the proof also provides edge-disjoint collections $\mathcal{W}_1, \ldots, \mathcal{W}_{r'}$ of L'-cycles in $F[V_3]$ with $V(\mathcal{W}_i) \geqslant (1-\mu')|V_3|$ for all $i \in [r']$. Let $\widehat{\mathcal{W}}$ be a random collection of L'-paths obtained by choosing one of $\mathcal{W}_1, \ldots, \mathcal{W}_{r'}$ uniformly at random and then independently deleting $k-1$ consecutive edges in each cycle. Note that $|V(\widehat{\mathcal{W}})| \geqslant (1-\mu')|V_3|$ and that for each $e \in E(F)$, we have

$$\mathbb{P}[e \in E(\widehat{\mathcal{W}}) \mid \mathcal{P}'] \leqslant \frac{1}{r'} \leqslant \frac{\ell_0}{|V_1|^{k-1}} \tag{4.6.5}$$

and

$$\mathbb{P}[e \in E(\partial\widehat{\mathcal{W}}) \mid \mathcal{P}'] \leqslant \frac{2}{L'r'} \leqslant \frac{1}{L'^{2/3}|V_1|^{k-1}}\,. \tag{4.6.6}$$

Next, we will utilise Lemma 4.6.2 applied to the ordered end-edges of the paths in $\widehat{\mathcal{W}}$, the ordered end-edges of the paths in $\mathcal{S}$, and the ordered edges $\vec{u}_1, \vec{v}_1, \ldots, \vec{u}_m, \vec{v}_m$ to connect the respective paths to cycles. First, we say which paths we aim to put into one cycle and afterwards we prepare for the application of Lemma 4.6.2.

Let $C_1, \ldots, C_h$ be the cycles in $\mathcal{C}$ and for $i \in [h]$, set $L_i = |V(C_i)| \geqslant L^3$. We now inductively define collections of paths $\mathcal{Z}_1, \ldots, \mathcal{Z}_h$ which we use to construct copies of the cycles $C_1, \ldots, C_h$. Suppose that for $i \in [h]$, we already have chosen collections of paths $\mathcal{Z}_1, \ldots, \mathcal{Z}_{i-1}$. Next, we pick a set $\mathcal{Z}_i$ of previously unused paths. More precisely, for $i' \in [m]$, write $P'_{i'} = \vec{u}_{i'}P_{i'}\vec{v}_{i'}$ and choose

$$\mathcal{Z}_i \subseteq \big(\{P'_1, \ldots, P'_m\} \cup \widehat{\mathcal{W}} \cup \mathcal{S}\big) \setminus \bigcup_{j \in [i-1]} \mathcal{Z}_j$$

such that $|\mathcal{Z}_i \cap \mathcal{S}|$ is maximal with $|\mathcal{Z}_i| \cdot \ell_0 + |V(\mathcal{Z}_i)| + \sum_{S \in \mathcal{S} \cap \mathcal{Z}_i} \sigma(S) \leqslant L_i$ and such that subject to this $|\mathcal{Z}_i|$ is maximal. Observe that since $|V(\{P'_1, \ldots, P'_m\} \cup \widehat{\mathcal{W}} \cup \mathcal{S})| \geqslant n - \vartheta^4|V_2|$, since each path in $\{P'_1, \ldots, P'_m\} \cup \widehat{\mathcal{W}} \cup \mathcal{S}$ has at most $2L$ vertices (and for a path $S \in \mathcal{S}$, we in fact have $|V(S)| + \sigma(S) \leqslant 2L$), and by (S2), this definition implies

$$L_i - 2L - \ell_0 < |\mathcal{Z}_i| \cdot \ell_0 + |V(\mathcal{Z}_i)| + \sum_{S \in \mathcal{S} \cap \mathcal{Z}_i} \sigma(S) \leqslant L_i\,. \tag{4.6.7}$$

Therefore, $L_i \geqslant L^3$ entails $|\mathcal{Z}_i| \geqslant L$ and so $|\mathcal{Z}_i|(\ell_1 - \ell_0) \geqslant 2L - \ell_0$. Together with (4.6.7), this implies that there are integers $\lambda_i^1, \ldots, \lambda_i^{|\mathcal{Z}_i|} \in [\ell_0, \ell_1]$ with

$$\sum_{j \in [|\mathcal{Z}_i|]} \lambda_i^j = L_i - \Big(|V(\mathcal{Z}_i)| + \sum_{S \in \mathcal{S} \cap \mathcal{Z}_i} \sigma(S)\Big)\,.$$

Next, we aim to connect the paths in $\mathcal{Z}_i$ by means of Lemma 4.6.2 to a cycle Z_i for every $i \in [h]$, using paths with $\lambda_i^1, \ldots, \lambda_i^{|\mathcal{Z}_i|}$ inner vertices. To this end, list the paths in $\mathcal{Z}_i$ arbitrarily as $A_i^1, \ldots, A_i^{|\mathcal{Z}_i|}$ for all $i \in [h]$, and for all $i \in [h]$ and $g \in [|\mathcal{Z}_i|]$, let $\vec{s}_i^{\,g}$ and $\vec{t}_i^{\,g}$ be the starting and ending edge of A_i^g, respectively. Further, write $p_\star = \sum_{i \in [h]} |\mathcal{Z}_i|$ and for $p \in [p_\star]$, let $\iota_1(p), \iota_2(p)$ be such that $(\iota_1(p), \iota_2(p))$ is the p-th element in the lexicographic ordering of the tuples in $\{(i,g) : i \in [h], g \in [|\mathcal{Z}_i|]\}$. Now, for every $p \in [p_\star]$, set $\vec{s}_p = \vec{s}_{\iota_1(p)}^{\,\iota_2(p)+1}$ (where we view the upper index modulo $\iota_2(p)$) and $\vec{t}_p = \vec{t}_{\iota_1(p)}^{\,\iota_2(p)}$. Note that when for every $p \in [p_\star]$, we connect $\vec{s}_p$ and $\vec{t}_p$ by a path with $\lambda_{\iota_1(p)}^{\iota_2(p)}$ inner vertices, we obtain cycles $Z_1, \ldots, Z_h$ such that $|V(Z_i)| = L_i - \sum_{S \in \mathcal{S} \cap \mathcal{Z}_i} \sigma(S)$ for all $i \in [h]$.

Set $\mathcal{Q} = \{\vec{s}_1, \vec{t}_1, \ldots, \vec{s}_{p_\star}, \vec{t}_{p_\star}\}$ and note that

$$|\mathcal{Q}| \leqslant 2\big(m + |\widehat{\mathcal{W}}| + |\mathcal{S}|\big) \leqslant \frac{2n}{L} + \frac{2|V_1|}{L'} + \frac{2\vartheta^2|V_2|}{L'} \leqslant \frac{|V_1|}{L'^{1/2}}\,.$$

Further, by (S4.2), by (4.6.4) (together with the union bound), and by (4.6.6) we know that for $e \in E(F)$,

$$\mathbb{P}[\vec{e} \in \mathcal{Q} \mid \mathcal{P}'] \leqslant \frac{\vartheta}{L'|V_2|^{k-1}} + \frac{1}{L^{1/2}|V_1|^{k-1}} + \frac{1}{L'^{2/3}|V_1|^{k-1}} \leqslant \frac{1}{L'^{1/2}|V_1|^{k-1}}\,. \tag{4.6.8}$$

For $i \in [p_\star]$, let $\mathcal{P}_i$ be the set of all $\vec{s}_i$–$\vec{t}_i$-paths P with $|V(P^\circ)| = \lambda_{\iota_1(i)}^{\iota_2(i)}$ and $V(P^\circ) \subseteq R$. Recall that the properties of R guarantee that $|\mathcal{P}_i| \geqslant \frac{\beta}{2}|R|^{\lambda_{\iota_1(i)}^{\iota_2(i)}} \geqslant \beta^{\ell_1+2}|V_1|^{\lambda_{\iota_1(i)}^{\iota_2(i)}}$.

Now we apply Lemma 4.6.2 with $F[V_1]$, $\mathcal{Q}$, $\big(\lambda_{\iota_1(i)}^{\iota_2(i)}\big)_{i \in [p_\star]}$, $(\mathcal{P}_i)_{i \in [p_\star]}$, $\frac{1}{L'^{1/2}}$, β^{ℓ_1+2}, ℓ_1, ℓ_0 here instead of H, $\mathcal{Q}$, $(\lambda_i)_{i \in [m]}$, $(\mathcal{P}_i)_{i \in [m]}$, ζ, β, ℓ_1, ℓ_0 there.

Lemma 4.6.2 then yields a probabilistic construction of a collection of paths $\mathcal{W}' \subseteq$

$\bigcup_{i\in[p_\star]}\mathcal{P}_i$ with $|\mathcal{W}'\cap\mathcal{P}_i|=1$ for all $i\in[p_\star]$ such that for every $e\in E(F)$, we have

$$\mathcal{P}[e\in E(\mathcal{W}')\mid\mathcal{P}']\leqslant\frac{1}{L'^{1/4}|V_1|^{k-1}}\,. \tag{4.6.9}$$

This leaves us with cycles $Z_1,\ldots,Z_h$ such that for all $i\in[h]$, the cycle Z_i contains the paths in $\mathcal{Z}_i$ as subpaths and $|V(Z_i)|=L_i-\sum_{S\in\mathcal{S}\cap\mathcal{Z}_i}\sigma(S)$. Observe that in particular, every element of $\mathcal{S}$ is a subpath in one of the cycles $Z_1,\ldots,Z_h$.

We aim to absorb the set $X=V_1\smallsetminus\bigcup_{i\in[h]}V(Z_i)$ of not yet covered vertices into the paths in $\mathcal{S}$. For this, note that since $|V(Z_i)|=L_i-\sum_{S\in\mathcal{S}\cap\mathcal{Z}_i}\sigma(S)$ for all $i\in[h]$, we have $|X|=\sum_{S\in\mathcal{S}}\sigma(S)=c$ because $\sum_{i\in[h]}L_i=n$. So property (S4) indeed allows us to absorb X into the cycles Z_i. More precisely, there is a probabilistic construction of a set $\mathcal{S}'$ of vertex-disjoint paths in $F[V_1]$ such that

- there is a bijection $\varphi:\mathcal{S}\to\mathcal{S}'$ such that for all $S\in\mathcal{S}$, the path $\varphi(S)$ is an $(L'+\sigma(S))$-path with $V(\varphi(S))\subseteq V(S)\cup X$ that has the same end-tuples as S;
- for all $e\in E(F)$, we have

$$\mathbb{P}[e\in E(\mathcal{S}')\mid\mathcal{P}']\leqslant\frac{1}{\vartheta^4|V_2|^{k-1}}\quad\text{and}\quad\mathbb{P}[e\in E(\partial\mathcal{S}')\mid\mathcal{P}']\leqslant\frac{1}{L'|V_2|^{k-1}}. \tag{4.6.10}$$

Due to the properties of $\mathcal{S}'$, replacing the every path $S\in\mathcal{S}\cap\mathcal{Z}_i$ in the cycle Z_i by the path $\varphi(S)$ for all $i\in[h]$ leaves us with vertex-disjoint cycles $C_1',\ldots,C_h'$ with $|C_i'|=L_i$, that is, C_i' is a copy of C_i for all $i\in[h]$. Thus we have constructed a copy of the cycle factor $\mathcal{C}$.

Finally, let us collect the upper bounds for the probabilities for the edges of different types to occur in the constructed cycle factor $\mathcal{C}'=\bigcup_{i\in[h]}C_i'$. First note that for $i'\in[m]$ and $e\in E(F)$ which is not ending (with respect to $\mathcal{P}$), we have that $e\in E(\vec{u}_{i'}\vec{a}_{i'})$ can only happen for some $i'\in[m]$ if $e=u_{i'}$ and the analogous remark holds for $\vec{b}_{i'}\vec{v}_{i'}$. Therefore, given $e\in E(F)$ which is not ending, (4.6.4) (together with the union bound) yields

$$\mathbb{P}\Big[e\in\bigcup_{i'\in[m]}E(\vec{u}_{i'}\vec{a}_{i'})\cup E(\vec{b}_{i'}\vec{v}_{i'})\mid\mathcal{P}'\Big]\leqslant\frac{1}{L^{1/2}|V_1|^{k-1}}\,. \tag{4.6.11}$$

For $j\in[k-1]$ and a j-ending edge $e\in E(F)$, we consider all partitions $\{\mathbf{x}_1,\mathbf{x}_2\}$ of e for which there exists a $P\in\mathcal{P}$ such that $\mathbf{x}_1$ is an end-set of P. The event $e\in E(\vec{u}_{i'}\vec{a}_{i'})$ (respectively $e\in E(\vec{b}_{i'}\vec{v}_{i'})$) can only happen for some $i'\in[m]$, if for one of these partitions, $\mathbf{x}_1$

is an end-set of $P_{i'} \in \mathcal{P}'$ and $\mathbf{x}_2 \subseteq u_{i'}$ (respectively $\mathbf{x}_2 \subseteq v_{i'}$). Note that for a fixed realisation of $\mathcal{P}'$, this can happen for at most one partition and one $i' \in [m]$. Further, since e is j-ending, for each such partition, we have $|\mathbf{x}_1| \leqslant j$. Thus, (4.6.4) implies that

$$\mathbb{P}\Big[e \in \bigcup_{i' \in [m]} E(\vec{u}_{i'}\vec{a}_{i'}) \cup E(\vec{b}_{i'}\vec{v}_{i'}) \mid \mathcal{P}'\Big] \leqslant \frac{\ell_0}{|V_1|^{k-j}} . \tag{4.6.12}$$

Note that the probabilities analysed after (4.6.2) were the probabilities for edges of F to occur in some subpath conditioned on the choice of $\mathcal{P}'$. While we do not need to make use of this for most types, we will do so for 1-concentrated edges. First note that for $e \in E(F)$, if $e \in E(F \cap \mathcal{C}')$, then

$$e \in E(\widehat{\mathcal{W}}) \cup E(\mathcal{W}') \cup E(\mathcal{S}') \cup \bigcup_{i' \in [m]} E(\vec{u}_{i'}\vec{a}_{i'}) \cup E(\vec{b}_{i'}\vec{v}_{i'}) , \tag{4.6.13}$$

and if e is not ending, then in addition $e \subseteq V_1$ holds. If $e \in E(F)$ is 1-concentrated with respect to $\mathcal{P}$, for the event $e \subseteq V_1$ to occur, $V(P) \subseteq V_1$ has to hold for each of the k paths in $\mathcal{P}$ which contain a vertex of e. Therefore, using (4.6.13) together with the bounds on the respective probabilities in (4.6.5), (4.6.9), (4.6.10), (4.6.11), and (4.6.2), entails

$$\mathbb{P}[e \in E(F \cap \mathcal{C}')] = \mathbb{P}\big[e \in E(F \cap \mathcal{C}') \mid e \subseteq V_1\big]\mathbb{P}[e \subseteq V_1] \leqslant \frac{L'}{|V_1|^{k-1}} \frac{100\delta^k}{99} \leqslant \frac{\delta^{1/2}}{n^{k-1}} .$$

If $e \in E(F)$ is j-ending (with respect to $\mathcal{P}$) for a $j \in [k-1]$, (4.6.13) together with the bounds in (4.6.5), (4.6.9), (4.6.10), and (4.6.12) give that

$$\begin{aligned}\mathbb{P}[e \in E(F \cap \mathcal{C}')] &\leqslant \mathbb{P}\big[e \in E(\widehat{\mathcal{W}}) \cup E(\mathcal{W}') \cup E(\mathcal{S}')\big] + \mathbb{P}\Big[e \in \bigcup_{i' \in [m]} E(\vec{u}_{i'}\vec{a}_{i'}) \cup E(\vec{b}_{i'}\vec{v}_{i'})\Big] \\ &\leqslant \frac{L'}{|V_1|^{k-1}} + \frac{\ell_0}{|V_1|^{k-j}} \leqslant \frac{1}{\delta^k n^{k-j}} .\end{aligned}$$

Lastly, if $e \in E(F)$ is neither 1-concentrated nor ending, (4.6.13) together with (4.6.5), (4.6.9), (4.6.10), and (4.6.11) entail that

$$\mathbb{P}[e \in E(F \cap \mathcal{C}')] \leqslant \frac{L'}{|V_1|^{k-1}} \leqslant \frac{1}{\delta^k n^{k-1}} .$$

Summarised, we provided a probabilistic construction of a cycle factor $\mathcal{C}'$ which is a copy of the cycle factor $\mathcal{C}$ such that for every $e \in E(F)$, the probability $\mathbb{P}[e \in E(F \cap \mathcal{C}')]$ is bounded as claimed in the statement. $\square$

The next proposition says that given an approximate decomposition of the edge set into approximate partitions of the vertex set into long paths, that is, given the setup after applying Proposition 4.4.1, we can indeed obtain an approximate decomposition of the edge set into given cycle factors of not too small girth.

Proposition 4.6.4. *Suppose $1/n \ll 1/L \ll \mu \ll \varrho \ll \eta, 1/k$. Suppose H is a k-graph on n vertices with vertex set V and F is an η-intersecting ϱ-almost regular spanning subgraph of H. Suppose $\mathcal{P}_1, \ldots, \mathcal{P}_r$ are edge-disjoint collections of L-paths in $H - F$ with $|V(\mathcal{P}_i)| \geqslant (1-\mu)n$ for all $i \in [r]$ such that for all $\mathbf{e} \in \binom{V}{k}$, we have*

(i) $|\mathcal{I}_{\mathrm{lo}}(\mathbf{e})| \leqslant \mu r$;

(ii) $|\mathcal{I}_{j\text{-con}}(\mathbf{e})| \leqslant 2^{L^2} n^{k-j}$ *for all* $j \in [k-1]$ *and* $|\mathcal{I}_{k\text{-con}}(\mathbf{e})| \leqslant 2^{L^2} n$;

(iii) $|\mathcal{I}_{j\text{-end}}(\mathbf{e})| \leqslant \frac{n^{k-j}}{L^{1/2}}$ *for all* $j \in [k-1]$.

Then, for all cycle factors $\mathcal{C}_1, \ldots, \mathcal{C}_r$ on n vertices of girth at least L^3, there are edge-disjoint copies of $\mathcal{C}_1, \ldots, \mathcal{C}_r$ in H.

Proof. In the following, we will analyse a random process in which we utilise Lemma 4.6.3 to transform each collection of paths $\mathcal{P}_i$ in a fairly uniform way into a cycle factor $\mathcal{C}'_i$ that is a copy of $\mathcal{C}_i$.

Suppose $\mu \ll \delta \ll \varrho$ and define a random process inductively as follows. Suppose that for some $i \in [r]$ and all $j \in [i-1]$ and $\mathbf{x} \in \binom{V}{k-1}$, we have defined a cycle factor $\mathcal{C}'_j$ in H and a $\{0,1\}$-valued random variable $Y_j^{\mathbf{x}}$. Further, set $\mathcal{C}'_{<i} = \bigcup_{j \in [i-1]} \mathcal{C}'_j$.

Let $\mathcal{E}_i$ be the event that for all $\mathbf{x} \in \binom{V}{k-1}$, we have that $d_{F \cap \mathcal{C}'_{<i}}(\mathbf{x}) \leqslant \delta^{1/4} n$. If $\mathcal{E}_i$ does not occur, we set $\mathcal{C}'_i = \varnothing$ and choose ${Y_i^{\mathbf{x}}}_{|\mathcal{E}_i^c}$ such that $\mathbb{P}[Y_i^{\mathbf{x}} = 1 \mid (Y_j^{\mathbf{x}})_{j \in [i-1]}, \mathcal{E}_i^c] = \sum_{\mathbf{e} \in \binom{V}{k} \colon \mathbf{x} \subseteq \mathbf{e}} p_{\mathcal{T}(\mathbf{e}, i)}$. In the end, we will show that with high probability, $\mathcal{E}_i$ occurs for all $i \in [r]$.

If $\mathcal{E}_i$ occurs, it follows that for all $\mathbf{x}, \mathbf{y} \in \binom{V}{k-1}$, we have

$$|N_{F - \mathcal{C}'_{<i}}(\mathbf{x}) \cap N_{F - \mathcal{C}'_{<i}}(\mathbf{y})| \geqslant |N_F(\mathbf{x}) \cap N_F(\mathbf{y})| - 2\delta^{1/4} n \geqslant \frac{\eta}{2} n$$

and that $F - \mathcal{C}'_{<i}$ is 2ϱ-almost regular. This means that if $\mathcal{E}_i$ occurs, we may apply Lemma 4.6.3, with H, $F - \mathcal{C}'_{<i}$, L, $\mathcal{C}_i$, $\mathcal{P}_i$, $\eta/2$, 2ϱ, δ, μ here instead of H, F, L, $\mathcal{C}$, $\mathcal{P}$, η, ϱ, δ, μ there, to obtain a probabilistic construction of a cycle factor $\mathcal{C}'_i$ in H such that

- $\mathcal{C}'_i$ is a copy of the cycle factor $\mathcal{C}_i$;

- $E(\mathcal{C}'_i) \subseteq E(\mathcal{P}_i) \cup E(F - \mathcal{C}'_{<i})$ (in particular, $\mathcal{C}'_i$ is edge-disjoint from every cycle factor $\mathcal{C}'_j$ for all $j \in [i-1]$);

- $\mathbb{P}[\mathbf{e} \in E(F \cap \mathcal{C}'_i) \mid (Y^{\mathbf{x}}_j)_{j\in[i-1]}, \mathcal{E}_i] \leqslant p_{\tau(\mathbf{e},i)}$ for every $\mathbf{e} \in \binom{V}{k}$ where $p_{1\text{-con}} = \frac{\delta^{1/2}}{n^{k-1}}$, $p_{j\text{-end}} = \frac{1}{\delta^k n^{k-j}}$ for all $j \in [k-1]$, $p_{k\text{-end}} = 0$, and $p_{j\text{-con}} = p_{\mathrm{lo}} = \frac{1}{\delta^k n^{k-1}}$ for all $j \in [k] \setminus \{1\}$ (here $p_{k\text{-end}} = 0$ holds since if $\mathbf{e} \in \binom{V}{k}$ is k-ending with respect to $\mathcal{P}_i$, then $\mathbf{e} \notin E(F)$).

Denoting the indicator variable of the event $d_{F\cap\mathcal{C}'_i}(\mathbf{x}) \geqslant 1$ by $I^{\mathbf{x}}_i$ for every $\mathbf{x} \in \binom{V}{k-1}$, this definition of $\mathcal{C}'_i$ implies $\mathbb{P}[I^{\mathbf{x}}_i = 1 \mid (Y^{\mathbf{x}}_j)_{j\in[i-1]}, \mathcal{E}_i] \leqslant \sum_{\mathbf{e}\in\binom{V}{k}:\, \mathbf{x}\subseteq\mathbf{e}} p_{\tau(\mathbf{e},i)}$ and we set ${Y^{\mathbf{x}}_i}_{|\mathcal{E}_i} = {I^{\mathbf{x}}_i}_{|\mathcal{E}_i}$. Thus, for all $\mathbf{x} \in \binom{V}{k-1}$, we have

$$\big|\{i \in [r] \colon d_{F\cap\mathcal{C}'_i}(\mathbf{x}) \geqslant 1\}\big| \leqslant \sum_{i\in[r]} Y^{\mathbf{x}}_i\,. \tag{4.6.14}$$

By definition, we have that $\mathbb{P}[Y^{\mathbf{x}}_i = 1 \mid (Y^{\mathbf{x}}_j)_{j\in[i-1]}] \leqslant \sum_{\mathbf{e}\in\binom{V}{k}:\, \mathbf{x}\subseteq\mathbf{e}} p_{\tau(\mathbf{e},i)}$ holds for all $i \in [r]$ and this entails that for all $\mathbf{x} \in \binom{V}{k-1}$, we have

$$\begin{aligned}
\sum_{i\in[r]} \mathbb{E}\big[Y^{\mathbf{x}}_i = 1 \mid (Y^{\mathbf{x}}_j)_{j\in[i-1]}\big] &\leqslant \sum_{\mathbf{e}\in\binom{V}{k}:\, \mathbf{x}\subseteq\mathbf{e}} \sum_{\tau\in\mathcal{T}} p_\tau |\mathcal{I}_\tau(\mathbf{e})| \\
&\overset{\text{(i)–(iii)}}{\leqslant} n\Big(\mu r \cdot \frac{1}{\delta^k n^{k-1}} + r \cdot \frac{\delta^{1/2}}{n^{k-1}} + 2 \sum_{j\in[k-1]\setminus\{1\}} 2^{L^2} n^{k-j} \frac{1}{\delta^k n^{k-1}} + \sum_{j\in[k-1]} \frac{n^{k-j}}{L^{1/2}} \frac{1}{\delta^k n^{k-j}}\Big) \\
&\leqslant \frac{\delta^{1/3} n}{2}\,.
\end{aligned}$$

Thus, we obtain by Freedman's inequality (Lemma 4.3.3) that

$$\mathbb{P}\Big[\sum_{i\in[r]} Y^{\mathbf{x}}_i \geqslant \delta^{1/3} n\Big] \leqslant \exp(-n^{1/2})\,. \tag{4.6.15}$$

Suppose now that $\sum_{i\in[r]} Y^{\mathbf{x}}_i \leqslant \delta^{1/3} n$ holds for all $\mathbf{x} \in \binom{V}{k-1}$. Then by (4.6.14), we conclude that $d_{F\cap\mathcal{C}'_{<i}}(\mathbf{x}) \leqslant \delta^{1/4} n$ holds for all $i \in [r]$ and $\mathbf{x} \in \binom{V}{k-1}$, meaning that the event $\mathcal{E}_i$ occurs for all $i \in [r]$. Consequently, by (4.6.15) and the union bound we conclude that

$$\mathbb{P}\Big[\bigcap_{i\in[r]} \mathcal{E}_i\Big] \geqslant \mathbb{P}\Big[\sum_{i\in[r]} Y^{\mathbf{x}}_i \leqslant \delta^{1/3} n \text{ for all } \mathbf{x} \in \binom{V}{k-1}\Big] > 0\,.$$

Thus, with positive probability it happens that $\mathcal{E}_i$ occurs for all $i \in [r]$, meaning that in each step i of this random process we construct a copy of $\mathcal{C}_i$ that is edge-disjoint from all previously constructed cycle factors. This yields edge-disjoint copies of $\mathcal{C}_1, \ldots, \mathcal{C}_r$, as desired. $\square$

4.7 Proof of Theorem 4.1.3

We now prove our main theorem by combining Propositions 4.4.1 and 4.6.4.

Proof of Theorem 4.1.3. Suppose $1/n \ll \varrho, 1/L \ll \mu \ll \varrho_F \ll \varepsilon \ll \eta, 1/k$. Suppose H is an η-intersecting ϱ-almost r-regular k-graph on n vertices. Let $r' = (1-\varepsilon)r/k$ and suppose $\mathcal{C}_1, \ldots, \mathcal{C}_{r'}$ are cycle factors, whose girth is at least L. We show that there is a suitable spanning subgraph F of H such that Proposition 4.4.1 guarantees the existence of collections $\mathcal{P}_1, \ldots, \mathcal{P}_{r'}$ of paths in $H - F$ that we can connect to cycles forming a copy of $\mathcal{C}_i$ for all $i \in [r']$ via Proposition 4.6.4.

Let $V = V(H)$ and $E = E(H)$. Let F be a random spanning subgraph of H for which every edge $e \in E$ is included in $E(F)$ independently at random with probability $p = \varepsilon - 2\varrho$. Then Chernoff's inequality (Lemma 4.3.1) and the union bound show that with positive probability F is an $\varepsilon^2\eta/2$-intersecting 2ϱ-almost pr-regular spanning subgraph of H such that $H - F$ is $\eta/2$-intersecting. From now on, let F denote such a subgraph.

For all $v \in V$, we have

$$d_{H'}(v) \geqslant ((1-\varrho) - (1+2\varrho)(\varepsilon - 2\varrho))r \geqslant kr'$$

and

$$d_{H'}(v) \leqslant ((1+\varrho) - (1-2\varrho)(\varepsilon - 2\varrho))r \leqslant (1+4\varrho)kr'.$$

Thus with 4ϱ, $L^{1/3}$, $\eta/2$, μ, H' playing the roles of ϱ, L, η, μ, H, Proposition 4.4.1 yields edge-disjoint collections $\mathcal{P}_1, \ldots, \mathcal{P}_{r'}$ of $L^{1/3}$-paths in H' with $|V(\mathcal{P}_i)| \geqslant (1-\mu)n$ for all $i \in [r']$ such that the following holds for all $\mathbf{e} \in \binom{V}{k}$.

- $|\mathcal{I}_{\text{lo}}(\mathbf{e})| \leqslant \mu r'$;
- $|\mathcal{I}_{j\text{-con}}(\mathbf{e})| \leqslant n^{k-j}/(\eta/2)^{2L^{1/3}}$ for all $j \in [k-1]$ and $|\mathcal{I}_{k\text{-con}}(\mathbf{e})| \leqslant n/(\eta/2)^{2L^{1/3}}$;
- $|\mathcal{I}_{j\text{-end}}(\mathbf{e})| \leqslant n^{k-j}/L^{1/6}$ for all $j \in [k-1]$.

Since F is 2ϱ-almost regular, F is in particular ϱ_F-almost regular. Consequently, with $L^{1/3}$, μ, ϱ_F, $\varepsilon^2\eta/2$, H, F playing the roles of L, μ, ϱ, η, H, F, Proposition 4.6.4 yields copies of the given cycle factors as desired. $\square$

4.8 Concluding remarks

In this paper, we prove a strong generalization of both the well-known Dirac-type result for the existence of one Hamiltonian cycle in large k-graphs due to Rödl, Ruciński, and Szemerédi [98–100] and a result concerning asymptotically optimal packings of Hamiltonian cycles in graphs by Ferber, Krivelevich, and Sudakov [36]. In fact, our result even applies to cycle factors of large girth.

It was recently proved independently by Lang and Sanhueza-Matamala [77] and by Polcyn, Reiher, Rödl, and Schülke [93] that every large k-graph on n vertices with $\delta_{k-2}(H) \geqslant (5/9 + o(1))\binom{n}{2}$ contains a Hamiltonian cycle. We wonder whether such k-graphs actually contain $(1 - o(1))\,\mathrm{reg}_k(H)/k$ edge-disjoint Hamiltonian cycles.

There are yet other sufficient conditions for Hamiltonicity in hypergraphs, see for instance [102]. It would be interesting to know which other sufficient conditions imply a packing result similar to Theorem 4.1.1. Another tempting question in this direction is as follows. Call a k-graph H *robustly Hamiltonian* if we can delete $o(n)$ edges incident to each $(k-1)$-set and H still contains a Hamiltonian cycle. Does every large robustly Hamiltonian k-graph H contain $(1 - o(1))\,\mathrm{reg}_k(H)/k$ edge-disjoint Hamiltonian cycles?

Lastly, it would of course be desirable to obtain a real decomposition of a k-graph into Hamiltonian cycles, or even stronger, to show that any k-graph satisfying certain conditions contains $\mathrm{reg}_k(H)/k$ edge-disjoint Hamiltonian cycles. However, even decompositions of cliques into Hamiltonian cycles are in general not known to exist. Further, it was recently shown by Piga and Sanhueza-Matamala [88] that there are arbitrarily large 3-graphs H with $\delta_2(H) \geqslant (\frac{2}{3} - o(1))|V(H)|$ which do not contain $\mathrm{reg}_3(H)/3$ edge-disjoint Hamiltonian cycles. Therefore, our results cannot be improved to exact decompositions without increasing the lower bound on the minimum degree significantly.

5. Covering 3-edge-coloured random graphs with monochromatic trees

5.1 Introduction

Given a graph G and a positive integer r, let $\mathrm{tc}_r(G)$ denote the minimum number k such that in any r-edge-colouring of G, there are k monochromatic trees $T_1, \ldots, T_k$ such that the union of their vertex sets covers $V(G)$, i.e.,

$$V(G) = V(T_1) \cup \cdots \cup V(T_k).$$

We define $\mathrm{tp}_r(G)$ analogously by requiring the union above to be disjoint.

It is easy to see that $\mathrm{tp}_2(K_n) = 1$ for all $n \geqslant 1$, and Erdős, Gyárfás, and Pyber [33] proved that $\mathrm{tp}_3(K_n) = 2$ for all $n \geqslant 1$, and conjectured that $\mathrm{tp}_r(K_n) = r - 1$ for every n and r. Haxell and Kohayakawa [59] showed that $\mathrm{tp}_r(K_n) \leqslant r$ for all sufficiently large $n \geqslant n_0(r)$. We remark that it is easy to see that $\mathrm{tc}_r(K_n) \leqslant r$ (just pick any vertex $v \in V(K_n)$ and let T_i, for $i \in [r]$, be a maximal monochromatic tree of colour i containing v), but it is not even known whether or not $\mathrm{tc}_r(K_n) \leqslant r - 1$ for every n and r (as would be implied by the conjecture of Erdős, Gyárfás, and Pyber).

Concerning general graphs instead of complete graphs, Gyárfás [53] noted that a well-known conjecture due to Ryser [61] on matchings and transversal sets in hypergraphs is equivalent to the statement that for every graph G and integer $r \geqslant 2$, we have $\mathrm{tc}_r(G) \leqslant (r-1)\alpha(G)$. In particular, Ryser's conjecture, if true, would imply that $\mathrm{tc}_r(K_n) \leqslant r - 1$, for every $n \geqslant 1$ and $r \geqslant 2$. Ryser's conjecture was proved in the case $r = 3$ by Aharoni [1], but for $r \geqslant 4$ very little is known. For example, Haxell and Scott [60] proved (in the context of Ryser's original conjecture) that there exists $\varepsilon > 0$ such that for $r \in \{4, 5\}$, we have $\mathrm{tc}_r(G) \leqslant (r - \varepsilon)\alpha(G)$, for any graph G.

Bal and DeBiasio [6] initiated the study of covering and partitioning random graphs by monochromatic trees. They proved that if $p \ll \left(\frac{\log n}{n}\right)^{1/r}$, then with high probability[1] we have $\mathrm{tc}_r(G(n,p)) \to \infty$. They conjectured that for any $r \geqslant 2$, this was the correct threshold for the event $\mathrm{tp}_r(G(n,p)) \leqslant r$. Kohayakawa, Mota, and Schacht [72] proved that this conjecture holds for $r = 2$, while Ebsen, Mota, and Schnitzer[2] showed that it does not hold for more than two colours.

Bucić, Korándi, and Sudakov [15] proved that if $p \ll \left(\frac{\log n}{n}\right)^{\sqrt{r}/2^{r-2}}$, then w.h.p. we have $\mathrm{tc}_r(G(n,p)) \geqslant r+1$, which implies that the threshold for the event $\mathrm{tc}_r(G(n,p)) \leqslant r$ is in fact significantly larger than the one conjectured by Bal and DeBiasio when r is large. Bucić, Korándi, and Sudakov also proved that w.h.p. we have $\mathrm{tc}_r(G(n,p)) \leqslant r$ for $p \gg \left(\frac{\log n}{n}\right)^{1/2^r}$. They were also able to roughly determine the typical behaviour of $\mathrm{tc}_r(G(n,p))$ in terms of the range where p lies in (see [15, Theorems 1.3 and 1.4]).

Considering colourings with three colours, the general results from [15], as stated, imply that if $p \gg \left(\frac{\log n}{n}\right)^{1/8}$, then w.h.p. we have $\mathrm{tc}_3(G(n,p)) \leqslant 3$, and if $p \gg \left(\frac{\log n}{n}\right)^{1/6}$, then w.h.p. $\mathrm{tc}_3(G(n,p)) \leqslant 88$ (the methods from [15] may actually give a somewhat better upper bound than 88, if one optimizes their calculations). Our main result improves these bounds.

Theorem 5.1.1. *If $p = p(n)$ satisfies $p \gg \left(\frac{\log n}{n}\right)^{1/6}$, then with high probability we have*

$$\mathrm{tc}_3\big(G(n,p)\big) \leqslant 3.$$

It is easy to see that if $p = 1 - \omega(n^{-1})$, then w.h.p. there is a 3-edge-colouring of $G(n,p)$ for which three monochromatic trees are needed to cover all vertices — it suffices to consider three non-adjacent vertices x_1, x_2, and x_3, and colour the edges incident to x_i with colour i and colour all the remaining edges with any colour. Therefore, the bound for $\mathrm{tc}_3(G(n,p))$ in Theorem 5.1.1 is best possible as long as p is not too close to 1.

We remark that, from the example described in [72], we know that for $p \ll \left(\frac{\log n}{n}\right)^{1/4}$, we have w.h.p. $\mathrm{tc}_3(G(n,p)) \geqslant 4$. It would be very interesting to describe the behaviour of $\mathrm{tc}_3(G(n,p))$ when $\left(\frac{\log n}{n}\right)^{1/4} \ll p \ll \left(\frac{\log n}{n}\right)^{1/6}$.

This paper is organized as follows. In Section 5.2 we present some definitions and auxiliary results that we will use in the proof of Theorem 5.1.1, which is outlined in Section 5.3. The details of the proof of Theorem 5.1.1 are given in Section 5.4.

[1] We will write shortly *w.h.p.* for *with high probability*.
[2] A description of this construction can be found in [72].

5.2 Preliminaries

Most of our notation is standard (see [8, 10, 26] and [9, 62]). However, we will mention in the following few definitions regarding hypergraphs that will play a major role in our proofs just for completeness.

We say that a set A of vertices in a hypergraph $\mathcal{H}$ is a *vertex cover* if every hyperedge of $\mathcal{H}$ contains at least one element of A. The *covering number* of $\mathcal{H}$, denoted by $\tau(\mathcal{H})$, is the smallest size of a vertex cover in $\mathcal{H}$. A *matching* in $\mathcal{H}$ is a collection of disjoint hyperedges in $\mathcal{H}$. The *matching number* of $\mathcal{H}$, denoted by $\nu(\mathcal{H})$, is the largest size of a matching in $\mathcal{H}$. An immediate relationship between $\tau(\mathcal{H})$ and $\nu(\mathcal{H})$ is the inequality $\nu(\mathcal{H}) \leqslant \tau(\mathcal{H})$. If additionally $\mathcal{H}$ is r-uniform, then we have $\tau(\mathcal{H}) \leqslant r\nu(\mathcal{H})$. A conjecture due to Ryser (which first appeared in the thesis of his Ph.D. student, Henderson [61]) states that for every r-uniform r-partite hypergraph $\mathcal{H}$, we have $\tau(\mathcal{H}) \leqslant (r-1)\nu(\mathcal{H})$. Note that the König–Egerváry theorem corresponds to Ryser's conjecture for $r = 2$. Aharoni [1] proved that Ryser's conjecture holds for $r = 3$, but the conjecture remains open for $r \geqslant 4$.

Given a vertex v in a 3-uniform hypergraph $\mathcal{H}$, the *link graph* of $\mathcal{H}$ with respect to v is the graph $L_v = (V, E)$ with vertex set $V = V(\mathcal{H})$ and edge set $E = \{xy : \{x, y, v\} \subseteq \mathcal{H}\}$.

We will use the following theorem due to Erdős, Gyárfás and Pyber [33] in the proof of our main result.

Theorem 5.2.1 (Erdős, Gyárfás and Pyber)**.** *For any 3-edge-colouring of a complete graph K_n, there exists a partition of $V(K_n)$ into 2 monochromatic trees.*

We will also use the following lemma, which is a simple application of Chernoff's inequality. For a proof of the first item see [73, Lemma 3.8]. The second item is an immediate corollary of [73, Lemma 3.10].

Lemma 5.2.2. *Let $\varepsilon > 0$. If $p = p(n) \gg \left(\frac{\log n}{n}\right)^{1/6}$, then w.h.p. $G \in G(n,p)$ has the following properties.*

(i) For any disjoint sets $X, Y \subseteq V(G)$ with $|X|, |Y| \gg \frac{\log n}{p}$, we have

$$|E_G(X, Y)| = (1 \pm \varepsilon)p|X||Y|.$$

(ii) Every vertex $v \in V(G)$ has degree $d_G(v) = (1 \pm \varepsilon)pn$ and every set of $i \leqslant 6$ vertices has $(1 \pm \varepsilon)p^i n$ common neighbours.

5.3 A sketch of the proof

In this section we will give an overview of the proof of 5.1.1. Let $G = G(n, p)$, with $p \gg \left(\frac{\log n}{n}\right)^{1/6}$, and let $\phi : E(G) \to \{\text{red}, \text{ green}, \text{ blue}\}$ be any 3-edge-colouring of G. We consider an auxiliary graph F, with $V(F) = V(G)$ and $ij \in E(F)$ if and only if there is, in the colouring ϕ, a monochromatic path in G connecting i and j. Then we define a 3-edge-colouring ϕ' of F with $\phi'(ij)$ being the colour of any monochromatic path in G connecting i and j. Note that any covering of F with monochromatic trees with respect to the colouring ϕ' corresponds to a covering of G with monochromatic trees with respect to the colouring ϕ with the same number of trees.

Next, we consider different cases depending on the value of $\alpha(F)$. If $\alpha(F) = 1$, then F is a complete 3-edge-coloured graph and by a theorem of Erdős, Gyárfás and Pyber (see 5.2.1), there exists a partition of $V(F)$ into 2 monochromatic trees. The remaining proof now is divided into the cases $\alpha(F) \geqslant 3$ and $\alpha(F) = 2$.

Case $\alpha(F) \geqslant 3$. From the condition on the independence number of G, there exist three vertices $r, b, g \in V(G)$ that pairwise do not have any monochromatic path connecting them. With high probability, they have a common neighbourhood in G of size at least $np^3/2$. Let X_{rbg} be the largest subset of this common neighbourhood such that for each $i \in \{r, b, g\}$, the edges from i to X_{rbg} in G are all coloured with one colour. Then, since there are no monochromatic paths between any two of r, b, g, we have $|X_{rbg}| \geqslant np^3/12$ and moreover we may assume that all edges between r and X_{rbg} are red, all between b and X_{rbg} are blue and those between g and X_{rbg} are green. Now we notice that all vertices that have a neighbour in X_{rbg} are covered by the union of the spanning trees of the red component of r, the blue component of b and the green component of g.

We are done in the case where every vertex has a neighbour in X_{rbg}, as the vertices in $X_{rbg} \cup N_G(X_{rbg})$ are covered by the red, blue and green component containing r, b and g, respectively. Otherwise, w.h.p. any vertex $y \in V \setminus (X_{rbg} \cup N_G(X_{rbg}))$ has many common neighbours with r, b and g in G that are also neighbours of some vertex in X_{rbg}. An analysis of the possible colourings of the edges between X_{rbg} and the common neighbourhood of the vertices r, b, g and y yields the following: for some $i \in \{r, b, g\}$, let us say $i = r$, every vertex $y \in X_{rbg}$ can be connected to r by a monochromatic path in colour red or either to g or b by a monochromatic path in the colour blue or green, respectively.

This already gives us that all vertices in G can be covered by 5 monochromatic trees, since all the vertices in $N_G(X_{rbg})$ lie in the red component of r, or the green component of g, or in the blue component of b and every vertex in $V \setminus N_G(X_{rbg})$ lies in the red component of r, in the blue component of g or in the green component of b. By analysing the colours of edges to the common neighbourhood of carefully chosen vertices, we are able to show that actually three of those five trees already cover all the vertices of G.

Case $\alpha(F) = 2$. Let us consider a 3-uniform hypergraph $\mathcal{H}$ defined as follows (this definition is inspired by a construction of Gyárfás [53] and also appears in [15]). The vertices of $\mathcal{H}$ are the monochromatic components of F and three vertices form a hyperedge if the corresponding three components have a vertex in common (in particular, those three monochromatic components must be of different colours). Hence, $\mathcal{H}$ is a 3-uniform 3-partite hypergraph.

We observe that if A is a vertex cover of $\mathcal{H}$, then the monochromatic components associated with the vertices in A cover all the vertices of G. This implies that $\mathrm{tc}_3(G) \leqslant \tau(\mathcal{H})$. Also, it is easy to see that $\nu(\mathcal{H}) \leqslant \alpha(F) = 2$. Now, recall that Aharoni's result [1] (which corresponds to Ryser's conjecture for $r = 3$) states that for every 3-uniform 3-partite hypergraph $\mathcal{H}$ we have $\tau(\mathcal{H}) \leqslant 2\nu(\mathcal{H})$. Together with the previous observation, this implies $\mathrm{tc}_3(G) \leqslant 4$. But our goal is to prove that $\mathrm{tc}_3(G) \leqslant 3$. To this aim, we analyse the hypergraph $\mathcal{H}$ more carefully, reducing the situation to a few possible settings of components covering all vertices. In each of those cases, we can again analyse the possible colouring of edges of common neighbours of specific vertices, inferring that indeed there are 3 monochromatic components which cover all vertices.

5.4 Proof of 5.1.1

Instead of analysing the colouring of the graph $G = G(n, p)$, it will be helpful to analyse the following auxiliary graph.

Definition 5.4.1 (Shortcut graph)**.** *Let G be a graph and φ be a 3-edge-colouring of G. The* shortcut graph *of G (with respect to φ) is the graph $F = F(G, \phi)$ that has $V(G)$ as the vertex set and the following edge set:*

$$\{uv : u, v \in V(G) \text{ and } u \text{ and } v \text{ are connected in } G \text{ by a path monochromatic under } \phi\}.$$

Let us consider an edge-multicolouring ϕ' of $F = F(G, \phi)$ which assigns to an edge $uv \in E(F(G, \phi))$ the list of all the colours of monochromatic paths connecting u and v in G

under the colouring ϕ. We will say that ϕ' is the *inherited colouring*[3] of $F(G, \phi)$. We say that an edge $e \in F(G, \phi)$ *has colour* ϱ (or *is coloured with* ϱ) if ϱ belongs to the list of colours assigned to e by ϕ'. We say that a subgraph H of $F(G, \phi)$ is *monochromatic under* ϕ' if all the edges of H are coloured with a common colour. Let $\mathrm{tc}(F, \phi')$ be the minimum number k such that there are k trees $T_1, \ldots, T_k$ which are monochromatic under ϕ' such that $V(F) = V(T_1) \cup \cdots \cup V(T_k)$. Note that any covering of $F(G, \phi)$ with monochromatic trees under ϕ' corresponds to a covering of G with monochromatic trees under the colouring ϕ. In particular, if we show that for every 3-edge-colouring ϕ of G, we have $\mathrm{tc}(F, \phi') \leqslant 3$, where $F = F(G, \phi)$ is the shortcut graph of G with respect to ϕ, and ϕ' is the inherited colouring of F, then we have shown that $\mathrm{tc}_3(G) \leqslant 3$. Therefore, 5.1.1 follows from the following lemma.

Lemma 5.4.2. *Let $p \gg \left(\frac{\log n}{n}\right)^{1/6}$ and let $G = G(n, p)$. The following holds with high probability. For any 3-edge-colouring ϕ of G, we have $\mathrm{tc}(F, \phi') \leqslant 3$, where F is the shortcut graph $F = F(G, \phi)$ and ϕ' is the inherited colouring of F.*

The proof of 5.4.2 is divided into two different cases, depending on the independence number of F. Subsections 5.4.1 and 5.4.2 are devoted, respectively, to the proof of 5.4.2 when $\alpha(F) \geqslant 3$ and $\alpha(F) \leqslant 2$.

From now on, we fix $\varepsilon > 0$ and assume that $p \gg \left(\frac{\log n}{n}\right)^{1/6}$ and n is sufficiently large. Then, by 5.2.2, we may assume that the following holds w.h.p.:

1. There is an edge between any two sets of size $\omega\left((\log n)/p\right)$.

2. Every vertex $v \in V(G)$ has degree $d_G(v) = (1 \pm \varepsilon)pn$.

3. Every set of $i \leqslant 6$ vertices has $(1 \pm \varepsilon)p^i n$ common neighbours.

5.4.1 Shortcut graphs with independence number at least three

Proof of 5.4.2 for $\alpha(F) \geqslant 3$. Since $\alpha(F) \geqslant 3$, there exist three vertices $r, b, g \in V(G)$ that pairwise do not have any monochromatic path connecting them in G. In particular, if v is a common neighbour of r, b and g in G, then the edges vr, vb and vg have all different colours. The common neighbourhood of r, b and g in G has size at least $np^3/2$. Let X_{rbg} be the largest subset of this common neighbourhood such that for each $i \in \{r, b, g\}$, the edges between i and the vertices of X_{rbg} are all coloured with the same colour in G.

[3] Although ϕ' is a multicolouring, in the sense that we assigned several colours to each edge, we will refer to it as colouring, for simplicity.

Then $|X_{rbg}| \geqslant np^3/12$. Without loss of generality, assume that all edges between r and the vertices of X_{rbg} are red, between b and the vertices of X_{rbg} are blue and those between g and the vertices of X_{rbg} are green. Let $C_{\text{red}}(r)$, $C_{\text{blue}}(b)$ and $C_{\text{green}}(g)$ be respectively the red, blue and green components in G containing r, g and b.

Notice that all vertices of F that have a neighbour in X_{rbg} are covered by $C_{\text{red}}(r)$, $C_{\text{blue}}(b)$ or $C_{\text{green}}(g)$. Therefore, the proof would be finished if every vertex had a neighbour in X_{rbg}. If this is not the case, we fix an arbitrary vertex $y \in V \setminus (X_{rbg} \cup N_G(X_{rbg}))$. By our choice of p, there are at least $np^4/2$ common neighbours of y, r, b and g. Let X_{yrbg} be the largest subset of the common neighbourhood of y, r, b and g such that for each $i \in \{r, b, g\}$, the edges between i and X_{yrbg} are all coloured the same. Then $|X_{yrbg}| \geqslant np^4/12$. Note that since $y \notin N_G(X_{rbg})$, the sets X_{yrbg} and X_{rbg} are disjoint. Furthermore, since $|X_{yrbg}|, |X_{rbg}| \gg \frac{\log n}{p}$, we have

$$|E_G(X_{yrbg}, X_{rbg})| \geqslant 1.$$

We now analyse the colours between r, b, g and the set X_{yrbg}. Again, since there is no monochromatic path connecting any two of r, b and g, all $i \in \{r, b, g\}$ have to connect to X_{yrbg} in different colours. Since X_{yrbg} is disjoint from X_{rbg}, by the maximality of X_{rbg} we cannot have r, b and g being simultaneously connected to X_{yrbg} by red, blue and green edges, respectively. Assume first that for each $i \in \{r, b, g\}$, the edges between i and X_{yrbg} have different colours from the edges between i and X_{rbg}. Then let uv be an edge between X_{yrbg} and X_{rbg} and notice that whatever the colour of uv is, we will have a monochromatic path connecting two of the vertices in $\{r, g, b\}$. Therefore, we can assume that for some $i \in \{r, g, b\}$, we have that all the edges between i and X_{rbg} and all the edges between i and X_{yrbg} coloured the same. Without loss of generality, we may say that such i is r. In this case, the edges between b and X_{yrbg} are green and the edges between g and X_{yrbg} are blue. Finally, all the edges between X_{yrbg} and X_{rbg} are red, otherwise we would be able to connect b and g by some monochromatic path. Figure 5.4.1 shows the colouring of the edges that we have analysed so far.

Let us now consider any further vertex $x \in V \setminus (X_{rbg} \cup N_G(X_{rbg}))$ with $x \neq y$, if such a vertex exists. We define X_{xrbg} analogously to X_{yrbg} and observe that the colour pattern from r, b, g to X_{xrbg} must be the same as the one to X_{yrbg}. Indeed, if this is not the case, then a similar analysis of the colours of the edges between $\{r, b, g\}$ and X_{xrbg} yields that for some $i \in \{b, g\}$, we know that the edges connecting i to X_{xrbg} are of the same colour as the edges connecting i to X_{rbg}. Without loss of generality, let us say that i is g.

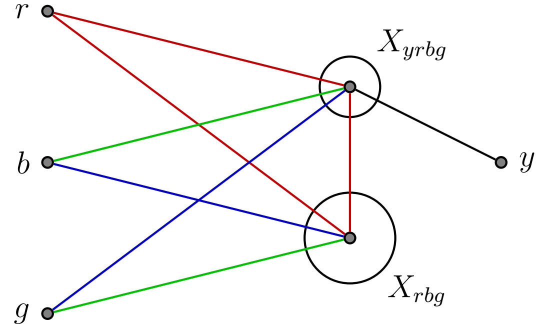


Figure 5.4.1: Analysis of the colouring of the edges incident on X_{rbg} and on X_{yrbg}.

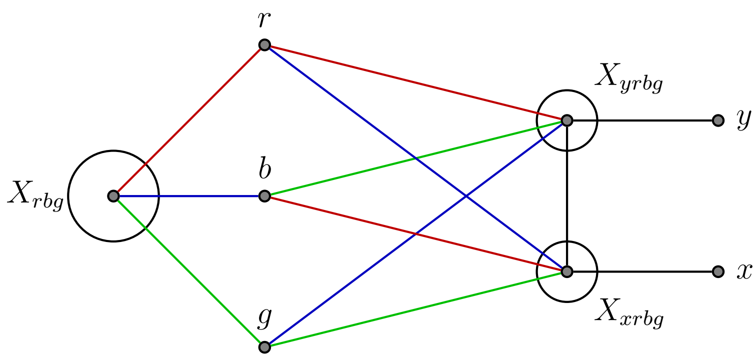


Figure 5.4.2: Analysis of the colour of the edges incident on X_{yrbg} and on X_{xrbg}.

Then the edges between b and X_{xrbg} are red and the edges between r and X_{xrbg} are green, otherwise X_{xrbg} and X_{rbg} would not be disjoints sets. Figure 5.4.2 shows the colouring of the edges incident to X_{yrbg} and X_{xrbg}. Since $|X_{yrbg}|, |X_{xrbg}| \gg \frac{\log n}{p}$, we have that there is some edge uv between X_{yrbg} and X_{xrbg}. But then however we colour uv, we will get a monochromatic path connecting two vertices in $\{r, b, g\}$, which is a contradiction. Thus, the colour pattern of edges between $\{r, b, g\}$ and X_{xrbg} is the same as the colour pattern of the edges between $\{r, b, g\}$ and X_{yrbg}.

Therefore, we have that each vertex in $X_{rbg} \cup N_G(X_{rbg})$ belongs to one of the monochromatic components $C_{\text{red}}(r)$, $C_{\text{blue}}(b)$ or $C_{\text{green}}(g)$, while a vertex in $V(G) \setminus (X_{rbg} \cup N_G(X_{rbg}))$ belongs to one of the monochromatic components $C_{\text{red}}(r)$, $C_{\text{green}}(b)$ or $C_{\text{blue}}(g)$ where the latter two are the green component containing b and the blue component containing g, respectively. This gives a covering of G with five monochromatic trees. Next we will show that actually three of those trees already cover all the vertices.

Suppose that at least four among the components $C_{\text{red}}(r)$, $C_{\text{blue}}(b)$, $C_{\text{green}}(b)$, $C_{\text{green}}(g)$, and $C_{\text{blue}}(g)$ are needed to cover all vertices. Since there does not exist any monochromatic path between any two of r, b, g, we know that for each $i \in \{r, b, g\}$, any monochromatic component containing i does not intersect $\{r, g, b\} \setminus \{i\}$. Hence, for each $i \in \{r, b, g\}$, one of these components contains i. Also, one element in $\{r, b, g\}$ belongs to two of these components. Without loss of generality, let us say that b belongs to two of these components. Therefore, $C_{\text{red}}(r)$, $C_{\text{blue}}(b)$ and $C_{\text{green}}(b)$ are three of these at least four components needed to cover all the vertices. Now there are two cases regarding the fourth component: we need $C_{\text{green}}(g)$ as the fourth component or we need $C_{\text{blue}}(g)$ (those two cases might intersect).

We begin with the first case, where we need the components $C_{\text{red}}(r)$, $C_{\text{blue}}(b)$, $C_{\text{green}}(b)$ and $C_{\text{green}}(g)$ to cover all the vertices of G. Let

$$\tilde{b} \in C_{\text{blue}}(b) \setminus (C_{\text{red}}(r) \cup C_{\text{green}}(b) \cup C_{\text{green}}(g))$$

and let

$$\tilde{g} \in C_{\text{green}}(b) \setminus (C_{\text{red}}(r) \cup C_{\text{blue}}(b) \cup C_{green}(g)) \,.$$

Then let $X_{\tilde{b}\tilde{g}rbg}$ be the maximum set of common neighbours of $\tilde{b}, \tilde{g}, r, g, b$ such that for each $i \in \{\tilde{b}, \tilde{g}, r, b, g\}$, the edges from i to $X_{\tilde{b}\tilde{g}rbg}$ are all coloured the same. Since $|X_{\tilde{b}\tilde{g}rbg}| \geqslant np^5/240 \gg \frac{\log n}{p}$, we have

$$|E_G(X_{\tilde{b}\tilde{g}rbg}, X_{yrbg})| \geqslant 1 \quad \text{and} \quad |E_G(X_{\tilde{b}\tilde{g}rbg}, X_{rbg})| \geqslant 1.$$

We will analyse the possible colours of the edges between the specified vertices and $X_{\tilde{b}\tilde{g}rbg}$. If for each of r, b, g, the colour it sends to $X_{\tilde{b}\tilde{g}rbg}$ is different from the colour it sends to X_{rbg}, then any edge between $X_{\tilde{b}\tilde{g}rbg}$ and X_{rbg} ensures a monochromatic path between two of r, b, g (in the colour of that edge). Similarly, it cannot happen that for each of r, b, g, the colour it sends to $X_{\tilde{b}\tilde{g}rbg}$ is different from the colour it sends to X_{yrbg}. Thus, since r sends red to both X_{rbg} and X_{yrbg} while the colours from b (and g) to X_{rbg} and X_{yrbg} are switched, the colour of the edges between r and $X_{\tilde{b}\tilde{g}rbg}$ is red.

Now note that, by the choice of $\tilde{b}$ and $\tilde{g}$, the edges between each of them and $X_{\tilde{b}\tilde{g}rbg}$ can not be red. Further, the choice implies that an edge between $\tilde{b}$ and $X_{\tilde{b}\tilde{g}rbg}$ can not be of the same colour (green or blue) as an edge between $\tilde{g}$ and $X_{\tilde{b}\tilde{g}rbg}$. If g would send blue (and hence b would send green) edges to $X_{\tilde{b}\tilde{g}rbg}$, there would either be a blue path between b and g (if the edges between $\tilde{b}$ and $X_{\tilde{b}\tilde{g}rbg}$ are blue) or $\tilde{b}$ would lie in $C_{\text{green}}(b)$ (if the edges between $\tilde{b}$ and $X_{\tilde{b}\tilde{g}rbg}$ are green). Since both those situations would mean a contradiction, we may assume that each of r, b, g sends edges with that colour to $X_{\tilde{b}\tilde{g}rbg}$ as it does to X_{rbg}. But then $X_{\tilde{b}\tilde{g}rbg}$ is actually a subset of X_{rbg} and since $\tilde{g}$ has an edge to X_{rbg}, it lies in one of $C_{\text{red}}(r)$, $C_{\text{blue}}(b)$, or $C_{\text{green}}(g)$; a contradiction.

In the case where the forth component that we need is $C_{\text{blue}}(g)$, we repeat the construction of $X_{\tilde{b}\tilde{g}rbg}$ similarly as before by letting

$$\tilde{b} \in C_{\text{blue}}(b) \setminus (C_{\text{red}}(r) \cup C_{\text{green}}(b) \cup C_{\text{blue}}(g))$$

and

$$\tilde{g} \in C_{\text{green}}(b) \setminus (C_{\text{red}}(r) \cup C_{\text{blue}}(b) \cup C_{\text{blue}}(g)).$$

Also as before, we end up with $X_{\tilde{b}\tilde{g}rbg}$ being part of X_{rbg}. From the choice of $\tilde{g}$, the edges it sends to $X_{\tilde{b}\tilde{g}rbg}$ have to be green, since otherwise it would be in $C_{\text{red}}(r)$ or $C_{\text{blue}}(b)$. But that gives a green path between b and g, a contradiction.

Summarising, we infer that three components among $C_{\text{red}}(r)$, $C_{\text{blue}}(b)$, $C_{\text{green}}(b)$, $C_{\text{green}}(g)$ and $C_{\text{blue}}(g)$ cover the vertex set of G. □

5.4.2 Shortcut graphs with independence number at most two

Proof of 5.4.2 for $\alpha(F) \leqslant 2$. We start by noticing that if $\alpha(F) = 1$, then the graph F together with the colouring φ' is a complete 3-coloured graph and therefore, by 5.2.1, there exists a partition of $V(F)$ into 2 monochromatic trees. Thus, we may assume that $\alpha(F) = 2$.

Let $\mathcal{H}$ be the 3-uniform hypergraph with $V(\mathcal{H})$ being the collection of all the monochromatic components of F under the colouring φ' and three monochromatic components form a hyperedge in $\mathcal{H}$ if they share a vertex. Notice that $\mathcal{H}$ is 3-partite, since distinct monochromatic components of the same colour do not have a common vertex and therefore they can not belong to the same hyperedge. In other words, the colour of each component give us a 3-partition of the vertex set of $\mathcal{H}$. We denote by V_{red},V_{blue} and V_{green} the set of vertices of $V(\mathcal{H})$ that correspond to, respectively, red, blue and green components. Such construction was inspired by a construction due to Gyárfás [53] and it was also used in [15].

Note that every vertex v of F is contained in a monochromatic component for each one of the colours (a monochromatic component could consist only of v). Therefore, any vertex cover of $\mathcal{H}$ corresponds to a covering of the vertices of F with monochromatic trees. Indeed, if A is a vertex cover of $\mathcal{H}$, then consider the monochromatic components corresponding to each vertex in A. If any vertex v of F is not covered by those components, then the vertices in $\mathcal{H}$ corresponding to the red, green and blue components in F containing v do not belong to A and they form an hyperedge. But this contradicts the fact that A is a vertex cover of $\mathcal{H}$. Therefore,

$$\mathrm{tc}(F, \phi') \leqslant \tau(\mathcal{H}). \tag{5.4.1}$$

The inequality (5.4.1) corresponds to Proposition 4.1 in [15] in our setting.

Let $L = \bigcup_{s \in V_{\text{red}}} L_s$ be the union of the link graphs L_s of all vertices $s \in V_{\text{red}}$. Any vertex cover of this bipartite graph L corresponds to a vertex cover of $\mathcal{H}$ of the same size. Therefore,

$$\tau(\mathcal{H}) \leqslant \tau(L). \tag{5.4.2}$$

Furthermore, by the König–Egerváry theorem we know that $\tau(L) = \nu(L)$. Thus, if $\nu(L) \leqslant 3$, then by (5.4.1) and (5.4.2), we have

$$\mathrm{tc}(F, \phi') \leqslant \tau(\mathcal{H}) \leqslant \tau(L) = \nu(L) \leqslant 3.$$

Therefore, we may assume that $\nu(L) \geqslant 4$, and fix a matching M_L of size at least four in L. Let us say that M_L consists of the edges G_1B_1, G_2B_2, G_3B_3, and G_4B_4, where $\{G_1, G_2, G_3, G_4\} \subseteq V_{\text{green}}$ and $\{B_1, B_2, B_3, B_4\} \subseteq V_{\text{blue}}$.

Now we give an upper bound for $\nu(\mathcal{H})$. Note that any matching $M_{\mathcal{H}}$ in $\mathcal{H}$ gives us an independent set I in F. Indeed, for each hyperedge $e \in M_{\mathcal{H}}$, let $v_e \in V(F)$ be any vertex in the intersection of those monochromatic components associated to the vertices in e and let $I = \{v_e : e \in M_{\mathcal{H}}\}$. We claim that I is an independent set in F. Indeed, if v_e and v_f were adjacent vertices in I, then e and f intersect, as the edge connecting v_e to v_f in F will connect the monochromatic components containing v_e and v_f of that colour that is given to the edge v_ev_f. Therefore, since $\alpha(F) = 2$, we have

$$\nu(\mathcal{H}) \leqslant \alpha(F) = 2. \tag{5.4.3}$$

Now, if there are three different edges in M_L that are edges in the link graphs of three different vertices of V_{red}, then there would be a matching of size 3 in $\mathcal{H}$, contradicting (5.4.3). Therefore, we may assume that M_L is contained in the union of at most two link graphs, say L_{R_1} and L_{R_2}, of vertices $R_1, R_2 \in V_{\text{red}}$. Now we are left with three cases: (Case 1) two edges of M_L belong to L_{R_1} and two belong to L_{R_2}; (Case 2) three edges of M_L belong to L_{R_1} and one to L_{R_2}; (Case 3) the four edges of M_L belong to L_{R_1}. Without loss of generality, we can describe each of those three cases as follows (see Figures 5.4.3, 5.4.4

and 5.4.5):

Case 1: The edges G_1B_1 and G_2B_2 belong to L_{R_1} and the edges G_3B_3 and G_4B_4 belong to L_{R_2}. That means that all the following four sets are non-empty:

$$J_1 := R_1 \cap G_1 \cap B_1,$$
$$J_2 := R_1 \cap G_2 \cap B_2,$$
$$J_3 := R_2 \cap G_3 \cap B_3,$$
$$J_4 := R_2 \cap G_4 \cap B_4.$$

Case 2: The edges G_1B_1, G_2B_2 and G_3B_3 belong to L_{R_1} and the edge G_4B_4 belongs to L_{R_2}. That means that all the following four sets are non-empty:

$$J_1 := R_1 \cap G_1 \cap B_1,$$
$$J_2 := R_1 \cap G_2 \cap B_2,$$
$$J_3 := R_1 \cap G_3 \cap B_3,$$
$$J_4 := R_2 \cap G_4 \cap B_4.$$

Case 3: The edges G_1B_1, G_2B_2, G_3B_3 and G_4B_4 belong to L_{R_1}. That means that all the following four sets are non-empty:

$$J_1 := R_1 \cap G_1 \cap B_1,$$
$$J_2 := R_1 \cap G_2 \cap B_2,$$
$$J_3 := R_1 \cap G_3 \cap B_3,$$
$$J_4 := R_1 \cap G_4 \cap B_4.$$

In this case, let R_2 be any other red component different from R_1 and let B and G be, respectively, a blue and a green component with $R_2 \cap B \cap G \neq \varnothing$. Suppose that $G \notin \{G_1, G_2, G_3, G_4\}$. Then the three of the edges G_1B_1, G_2B_2, G_3B_3 and G_4B_4 are not incident to GB (because B must be different from at least three of the sets B_1, B_2, B_3 and B_4) and these three edges together with GB may be analysed just as in Case 2. Therefore, we may suppose that $G \in \{G_1, G_2, G_3, G_4\}$. Let us say, without loss of generality, that $G = G_4$. If $B \notin \{B_1, B_2, B_3\}$, then the edges G_1B_1, G_2B_2 and G_3B_3 belong to L_{R_1}, the edge GB belongs to L_{R_2} and this case may be analysed, again, just as in Case 2. Therefore, we may assume that $B \in \{B_1, B_2, B_3\}$. Let us say, without loss of generality that $B = B_3$.

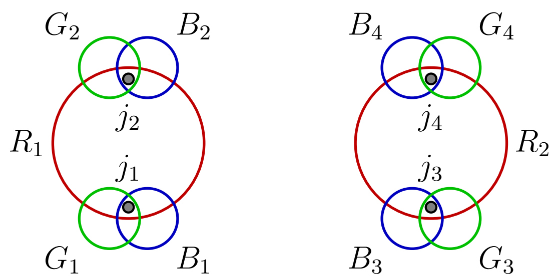


Figure 5.4.3: Case 1

Then let J_5 be the following non-empty set:

$$J_5 := R_2 \cap G_4 \cap B_3. \tag{5.4.4}$$

Let us further remark that, since $\nu(\mathcal{H}) \leqslant 2$, in each of the three cases above, we have

$$V(F) = R_1 \cup R_2 \cup G_1 \cup G_2 \cup G_3 \cup G_4 \cup B_1 \cup B_2 \cup B_3 \cup B_4.$$

Otherwise, for any uncovered vertex $v \in V(F)$, the hyperedge given by the red, blue and green components containing v together with the hyperedges $R_1B_1G_1$ and $R_2B_3G_3$ (in Cases 1 and 2) or $R_2B_3G_4$ (in Case 3) is a matching of size 3 in $\mathcal{H}$.

Let us start with Case 1.

Proof in Case 1: We will prove that R_1 and R_2 together with possibly one further monochromatic component cover $V(F)$. For each $i \in \{1, 2, 3, 4\}$, let $\tilde{B}_i = B_i \setminus (R_1 \cup R_2)$ and $\tilde{G}_i = G_i \setminus (R_1 \cup R_2)$.

Pick vertices $j_i \in J_i$, with $i \in \{1, 2, 3, 4\}$, arbitrarily. Consider a vertex $o \in \tilde{B}_1$ (if such a vertex exists). Since $\alpha(F) = 2$, there is an edge connecting two of o, j_2, j_3. Because j_2 and j_3 belong to different components of each colour, such an edge must be incident to o. So let us say that such edge is oj_i, for some $i \in \{2, 3\}$. Since $o \notin R_1 \cup R_2$, the edge oj_i cannot be red. And since $o \in B_1$, oj_i cannot be blue either, otherwise we would connect the blue components B_1 and B_i. Now assume that o and j_2 are not adjacent. Then oj_3 is a green edge in F. By analogously analysing the edge between o, j_2 and j_4 together with the supposition that oj_2 is not an edge in F, we get that oj_4 must be a green edge in F. But then we have a green path j_3oj_4 connecting j_3 to j_4, a contradiction. Therefore oj_2 is an edge in F and it is green. That implies that $o \in G_2$. Therefore $\tilde{B}_1 \subseteq G_2$. Analogously, we can conclude the following:

$$\begin{aligned}
\tilde{B}_1 \subseteq G_2, \quad \tilde{G}_1 \subseteq B_2,\\
\tilde{B}_2 \subseteq G_1, \quad \tilde{G}_2 \subseteq B_1,\\
\tilde{B}_3 \subseteq G_4, \quad \tilde{G}_3 \subseteq B_4,\\
\tilde{B}_4 \subseteq G_3, \quad \tilde{G}_4 \subseteq B_3.
\end{aligned} \tag{5.4.5}$$

Claim 5.4.3. *We have* $\tilde{B}_1 \cup \tilde{G}_1 \cup \tilde{B}_2 \cup \tilde{G}_2 = \varnothing$ *or* $\tilde{B}_3 \cup \tilde{G}_3 \cup \tilde{B}_4 \cup \tilde{G}_4 = \varnothing$.

Proof. Suppose for a contradiction that there exist $o_1 \in \tilde{B}_1 \cup \tilde{G}_1 \cup \tilde{B}_2 \cup \tilde{G}_2$ and $o_2 \in \tilde{B}_3 \cup \tilde{G}_3 \cup \tilde{B}_4 \cup \tilde{G}_4$. Recall that from our choice of p, there is some $z \in N(j_1, j_2, j_3, j_4, o_1, o_2)$. Two of the edges zj_i,for $i \in \{1, 2, 3, 4\}$, have the same colour. Since each j_i belongs to different green and blue components, those two edges are red. Since $\{j_1, j_2\} \in R_1$ and $\{j_3, j_4\} \in R_2$, those two red edges are either zj_1 and zj_2 or zj_3 and zj_4. Let us say that zj_1 and zj_2 are red (the other case is similar). Then one of the edges zj_3 and zj_4 has to be green and the other blue. Now, since $o_1 \notin R_1$, the edge zo_1 is either green or blue. Then one of the paths o_1zj_3 or o_1zj_4 is green or blue. This implies that $o_1 \in B_3 \cup G_3 \cup B_4 \cup G_4$. On the other hand, (5.4.5) implies that $o_1 \in (B_1 \cup B_2) \cap (G_1 \cup G_2)$. But then we reached a contradiction, since that would mean that o_1 belongs to two different components of the same colour. □

We may assume without loss of generality that $\tilde{B}_3 \cup \tilde{G}_3 \cup \tilde{B}_4 \cup \tilde{G}_4$ is empty. Then, recalling that $\nu(\mathcal{H}) \leqslant 2$ and in view of (5.4.5), the union of the components R_1, B_1, G_1 and R_2 covers every vertex of F. If we show that $B_1 \subseteq G_1 \cup R_1 \cup R_2$ or that $G_1 \subseteq B_1 \cup R_1 \cup R_2$, then we get three monochromatic components covering the vertices of F. Our next claim states precisely that.

Claim 5.4.4. *We have* $\tilde{B}_1 \smallsetminus G_1 = \varnothing$ *or* $\tilde{G}_1 \smallsetminus B_1 = \varnothing$.

Proof. Suppose that there exist two distinct vertices $b \in \tilde{B}_1 \smallsetminus G_1$ and $g \in \tilde{G}_1 \smallsetminus B_1$. Let $z \in N(j_1, j_2, j_3, j_4, b, g)$. As before, either zj_1 and zj_2 or zj_3 and zj_4 are red edges. First assume that zj_1 and zj_2 are red. Then one of the edges zj_3 and zj_4 has to be green and the other blue. Now, since $b \notin R_1$, the edge zb is either green or blue. Then one of the paths bzj_3 or bzj_4 is green or blue. This implies that $b \in B_3 \cup G_3 \cup B_4 \cup G_4$. On the other hand, (5.4.5) implies that $b \in B_1 \cap G_2$. Then we reached a contradiction, since that would mean that b belongs to two different components of the same colour.

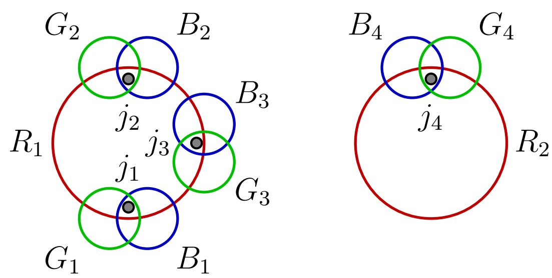


Figure 5.4.4: Case 2

Therefore, the edges zj_3 and zj_4 are red and one of the edges zj_1 and zj_2 is green and the other is blue. First let us say that zj_1 is green and zj_2 is blue. Since $b \notin (R_1 \cup R_2)$, the edge zb cannot be red. Also the edge zb cannot be blue otherwise the path bzj_2 would connect the components B_1 and B_2. Finally, zb cannot be green, otherwise the path bzj_1 would gives us that $b \in G_1$. Therefore, zj_1 is blue and zj_2 is green. But this case analogously leads to a contradiction (with g and G_i instead of b and B_i and green and blue switched). $\square$

We proceed to the proof of Case 2.

Proof in Case 2: As in Case 1, pick vertices $j_i \in J_i$, with $i \in \{1, 2, 3, 4\}$ arbitrarily. We claim that $V(F) \subseteq R_1 \cup R_2 \cup B_4 \cup G_4$. Indeed, let $o \in V(F) \setminus (R_1 \cup R_2)$. Notice that since $\alpha(F) = 2$, there is an edge in each of the following sets of three vertices: $\{o, j_4, j_1\}$, $\{o, j_4, j_2\}$, and $\{o, j_4, j_3\}$. We claim that oj_4 is an edge of F. Indeed, if this was not the case, then since there cannot be an edge between j_4 and j_i for $i = 1, 2, 3$, we would have the edges oj_1, oj_2 and oj_3 and all of them would be coloured green or blue. Thus, two of them would be coloured the same, connecting two distinct components of one colour in this colour, a contradiction. So $oj_4 \in E(F)$ and since oj_4 cannot be red, we conclude that $o \in (B_4 \cup G_4)$. Therefore, R_1, R_2, B_4 and G_4 cover all vertices of F.

If $B_4 \setminus (R_1 \cup R_2 \cup G_4) = \varnothing$ or $G_4 \setminus (R_1 \cup R_2 \cup B_4) = \varnothing$, then we get three monochromatic components covering $V(F)$. So let us assume that there exist $b \in B_4 \setminus (R_1 \cup R_2 \cup G_4)$ and $g \in G_4 \setminus (R_1 \cup R_2 \cup B_4)$. If b and g are not adjacent, then since each of the sets $\{b, g, j_i\}$, for $i = 1, 2, 3$, has to induce at least one edge, there are two edges between b and $\{j_1, j_2, j_3\}$ or two edges between g and $\{j_1, j_2, j_3\}$. However, from the choice of b, we know that all the edges between b and $\{j_1, j_2, j_3\}$ are green, and therefore, two of such edges would give us a green connection between two different green components, a contradiction. Similarly, from the choice of g, we know that all the edges between b and $\{j_1, j_2, j_3\}$ are blue, and two of such edges would give us a blue connection between two different blue components,

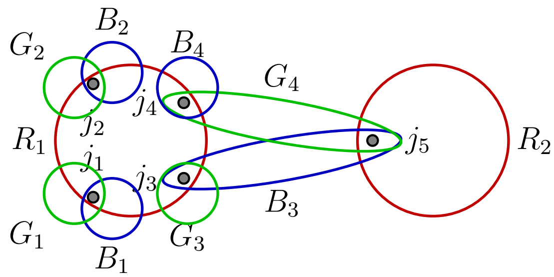


Figure 5.4.5: Case 3

again a contradiction.

Hence, we conclude that $bg \in F$ for any $b \in B_4 \setminus (R_1 \cup R_2 \cup G_4)$ and any $g \in G_4 \setminus (R_1 \cup R_2 \cup B_4)$ and any such edge bg is red. Therefore, there is a red component R_3 covering $(B_4 \triangle G_4) \setminus (R_1 \cup R_2)$, where $B_4 \triangle G_4$ denotes the symmetric difference. If $(B_4 \cap G_4) \setminus (R_1 \cup R_2) = \varnothing$, then R_1, R_2 and R_3 cover $V(F)$ and we are done. Therefore, suppose there is a vertex $x \in (B_4 \cap G_4) \setminus (R_1 \cup R_2)$. If $R_2 \setminus (B_4 \cup G_4) = \varnothing$, then R_1, B_4, G_4 cover $V(F)$ and we are done. Therefore, suppose there is a vertex $y \in R_2 \setminus (B_4 \cup G_4)$. Note that $xy \notin E(F)$, since x and y belong to different components in each of the colours. Also, $xj_i \notin E(F)$, for $i \in \{1, 2, 3\}$, since otherwise two different components of the same colour would be connected in that colour by the edge xj_i. Now $\alpha(F) = 2$ implies that $yj_i \in E(F)$, for $i \in \{1, 2, 3\}$ (otherwise, $\{x, y, j_i\}$ would be an independent set). But these edges must all be green or blue, hence two of them are of the same colour, connecting two different components of one colour in that colour, a contradiction.

We arrived at the last case, Case 3.

Proof in Case 3: Similarly to the previous cases, let us pick vertices $j_i \in J_i$, with $i \in \{1, 2, 3, 4, 5\}$ arbitrarily. We will show first that we can cover all vertices of F with four monochromatic components. Let $o_1, o_2 \in V(F) \setminus (R_1 \cup B_3 \cup G_4)$ and let z be a vertex in $N(j_1, j_2, j_3, o_1, o_2, j_5)$. At least one of the edges zj_1, zj_2 and zj_3 is red, as otherwise we would connect two distinct components of one colour in that colour. Therefore, $z \in R_1$. Since $o_1, o_2, j_5 \notin R_1$, the edges zo_1, zo_2 and zj_5 cannot be red. Furthermore, $o_1 z$ and $o_2 z$ are coloured with a colour different from the colour of the edge $j_5 z$, as otherwise they would belong to B_3 or G_4. Thus, o_1 and o_2 are connected by a monochromatic path in green or blue. Hence, we showed that any two vertices of $V(F) \setminus (R_1 \cup B_3 \cup G_4)$ are connected by a monochromatic path in green or blue. We infer that there is a green or blue component covering $V(F) \setminus (R_1 \cup B_3 \cup G_4)$. Therefore, R_1, B_3, G_4 and one further blue or green component C cover all vertices of G. Let us assume that C is a green component; the case where C is a blue component is analogous.

We claim that $R_1 \cup B_3 \cup C$, or $R_1 \cup G_4 \cup C$, or $R_1 \cup B_3 \cup G_4$ covers $V(F)$. Indeed, suppose for the sake of contradiction that there exist vertices $g \in G_4 \smallsetminus (R_1 \cup B_3 \cup C)$, $b \in B_3 \smallsetminus (R_1 \cup G_4 \cup C)$ and $c \in C \smallsetminus (R_1 \cup B_3 \cup G_4)$. Let $z \in N(j_1, j_2, j_3, g, b, c)$ and note that one of zj_1, zj_2 and zj_3 is red. Consequently gz, cz and bz are not red. Notice, however, that gz and bz can not be both green and neither both blue. Now let us say cz is green. Since $c \notin G_4$ and $g \in G_4$, we would have gz blue in this case. But then bz must be green and since $c \in C$ and C is a green component, we have $b \in C$, which is a contradiction. Therefore, cz must be blue. Then, since $c \notin B_3$ and $b \in B_3$, the edge bz should be green. Thus the edge gz is blue. Since this argument holds for any $g \in G_4 \smallsetminus (R_1 \cup B_3 \cup C)$ and $c \in C \smallsetminus (R_1 \cup B_3 \cup G_4)$, we conclude that $V(F) \smallsetminus (R_1 \cup B_3)$ can be covered by one blue tree. Hence, G can be covered by the three monochromatic trees. This finishes the last case and thereby the proof of 5.4.2. $\square$

6. Convex graphon parameters and graph norms

6.1 Introduction

In extremal combinatorics, quantifying quasirandomness by using a suitable norm has been an extremely useful strategy. For instance, the main idea in the proof of the celebrated Szemerédi regularity lemma is to use an L^2-norm increment, the Gowers norms play a central role in additive combinatorics, and the cut-norm is the key concept in the theory of dense graph limits [84].

It is a natural question to ask what norms can be defined on the space of two-variable real symmetric functions on $[0,1]^2$, which appear as the limit objects of sequences of (weighted) large graphs. To formalise, a *graphon* (resp. *signed graphon*) W is a two-variable symmetric measurable function from $[0,1]^2$ to $[0,1]$ (resp. $[-1,1]$). We consider the vector space $\mathcal{W}$ of two-variable symmetric bounded measurable functions on $[0,1]^2$, which contains the set of (signed) graphons as a convex subset. Given a graph H and $W \in \mathcal{W}$, the *homomorphism density* of H is defined by the functional

$$t_H(W) = \int \prod_{ij \in E(H)} W(x_i, x_j) d\mu^{v(H)},$$

where μ is the Lebesgue measure on $[0,1]$.

Let $\|W\|_H := |t_H(W)|^{1/e(H)}$ and let $\|W\|_{r(H)} := t_H(|W|)^{1/e(H)}$. We then say that a graph H is *(semi-)norming* if $\|\cdot\|_H$ defines a (semi-)norm on $\mathcal{W}$, and *weakly norming* if $\|\cdot\|_{r(H)}$ is a norm on $\mathcal{W}$. With this notation, we now state the following central question in the area, asked by Lovász [83] and Hatami [58]:

Question 6.1.1 ([83], Problem 24)**.** What graphs H are (weakly) norming?

A moment's thought will prove the fact that a weakly norming graph H must be bipartite and that, as the name suggests, every (semi-)norming graph is weakly norming.

The particular example $\|\cdot\|_{C_{2k}}$, where C_{2k} is the even cycle of length $2k$, is already interesting, as it corresponds to the Schatten–von Neumann norms in operator theory.

Perhaps one of the most important applications of weakly norming graphs is to Sidorenko's conjecture, a major open problem in extremal graph theory also proposed by Erdős and Simonovits [35] in a slightly different form.

Conjecture 6.1.2 (Sidorenko's conjecture [105])**.** *Let H be a bipartite graph and let W be a graphon. Then*

$$t_H(W) \geqslant t_{K_2}(W)^{e(H)}. \tag{6.1.1}$$

If a graph H satisfies (6.1.1) for every graphon W, then we say that H is *Sidorenko*. Szegedy observed[1] that every weakly norming graph is Sidorenko. Moreover, Conlon and the first author [22] proved that weakly norming graphs can be used as 'building blocks' to construct a Sidorenko graph. On the other hand, there are Sidorenko graphs that are verified to be not weakly norming. For instance, a bipartite graph that has a vertex adjacent to all the vertices on the other side, proven to be Sidorenko by Conlon, Fox, and Sudakov [20], is not weakly norming unless it is a complete bipartite graph. Moreover, Král', Martins, Pach, and Wrochna [74] recently proved that there exists an edge-transitive Sidorenko graph that is not weakly norming.

Although the weakly norming property is strictly stronger than being Sidorenko, partial answers to Question 6.1.1 have also made significant progress towards Sidorenko's conjecture. Hatami [58], who firstly studied Question 6.1.1, showed that even cycles C_{2k} are norming, and complete bipartite graphs $K_{m,n}$ and hypercubes Q_d are weakly norming. Lovász [81] later proved that $K_{n,n}$ minus a perfect matching is weakly norming. Before their work, Q_d and $K_{n,n}$ minus a perfect matching were unknown to be Sidorenko. Recently, Conlon and the first author [22] obtained a much larger class of weakly norming graphs, which also added many new examples to the class of Sidorenko graphs that played a crucial rôle in their subsequent work [23].

Despite a fair amount of recent progress [20–23, 58, 70, 80, 110], Sidorenko's conjecture remains open. In particular, none of the partial results succeeded in determining whether the notorious *Möbius ladder* graph $K_{5,5} \setminus C_{10}$, suggested by Sidorenko [104, 105], is Sidorenko or not, although Conlon and the first author [23] proved that its 'square' is Sidorenko. We make some progress in understanding this mysterious graph, by proving that it is not weakly norming.

[1] It appeared in [58].

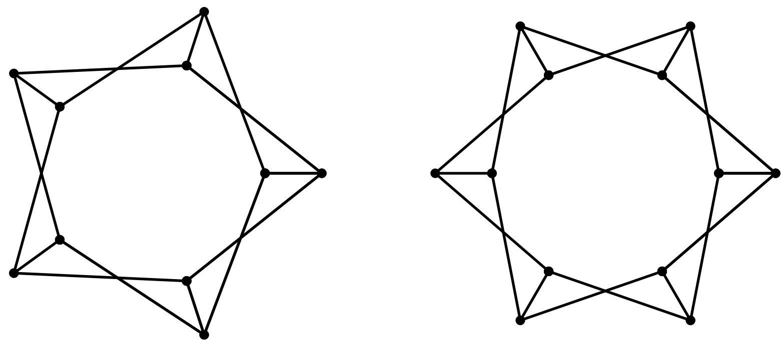


Figure 6.1.1: $C_5^{\bowtie}$ (the Möbius ladder) and $C_6^{\bowtie}$.

Theorem 6.1.3. *The Möbius ladder graph $K_{5,5} \setminus C_{10}$ is not weakly norming.*

For a graph H, let $H^{\bowtie}$ be the graph obtained by blowing up every vertex v of H by an edge v_1v_2 and putting two edges u_2v_1 and u_1v_2 between each pair of blown-up edges u_1u_2 and v_1v_2 whenever $uv \in E(H)$. The resulting graph $H^{\bowtie}$ is always a bipartite graph whose bipartite adjacency matrix is the (symmetric) adjacency matrix of H plus the identity. This blow-up was considered by Kim, Lee, and the first author [70]. They observed (see Figure 6.1.1) that $C_5^{\bowtie}$ is isomorphic to the Möbius ladder and, if H is bipartite, $H^{\bowtie}$ is isomorphic to $H \,\square\, K_2$, where $\square$ denotes the *Cartesian product* of graphs. In particular, $C_4^{\bowtie}$ is the 3-cube graph, proven to be weakly norming by Hatami. We prove a more general result that implies Theorem 6.1.3.

Theorem 6.1.4. *For every $k > 4$, $C_k^{\bowtie}$ is not weakly norming.*

In [58], Hatami asked whether two particular graphs, the Möbius strip and $C_{2k} \,\square\, K_2$, are weakly norming. Theorem 6.1.4 hence answers both questions at once. We remark that every $C_{2k}^{\bowtie}$ is known to be Sidorenko by [70], but except the case $C_3^{\bowtie} \cong K_{3,3}$ it is still an open question whether every $C_{2k+1}^{\bowtie}$ is Sidorenko or not.

Our proof relies on determining an equivalent condition of the (weakly) norming property. A function f defined on the set of graphons is a *(signed) graphon parameter* if $f(W) = f(W')$ for (signed) graphons W and W' for which there exists a measure-preserving bijection $\varphi : [0,1] \to [0,1]$ satisfying $W(\varphi(x), \varphi(y)) = W'(x, y)$. In particular, $t_H(W)$ is always a graphon parameter for any graph H.

Theorem 6.1.5. *Let H be a graph. Then*

(i) H is weakly norming if and only if $t_H(\cdot)$ is a convex graphon parameter.

(ii) H is norming if and only if $t_H(\cdot)$ is a strictly convex signed graphon parameter.

By using Theorem 6.1.5(ii), we also prove that $K_{t,t}$ minus a perfect matching, proven to be weakly norming by Lovász, is not norming if $t > 3$.

Theorem 6.1.6. *For every $t > 3$, $K_{t,t}$ minus a perfect matching is not norming.*

As observed by Hatami [58, Observation 2.5(ii)], every norming graph must be *eulerian*, i.e., every vertex has even degree. Thus, we only prove Theorem 6.1.6 for odd integers t, which gives the first examples of weakly norming graphs that are eulerian but not norming.

6.2 Preliminaries

Given an $n \times n$ symmetric matrix $A = (a_{ij})$, let U_A be the two-variable symmetric step function on $[0,1]^2$ defined by

$$U_A(x,y) = a_{ij}, \text{ if } (i-1)/n \leqslant x < i/n \text{ and } (j-1)/n \leqslant y < j/n$$

and $U_A = 0$ on the measure-zero set $x = 1$ or $y = 1$ for simplicity. Trivially, $A \mapsto U_A$ is a linear map and U_A satisfies the identity

$$t_H(U_A) = n^{-v(H)} \sum_{\phi : V(H) \to [n]} \prod_{uv \in E(H)} a_{\phi(u)\phi(v)}.$$

In other words, $t_H(U_A)$ is $n^{-v(H)}$ times a homogeneous $\binom{n+1}{2}$-variable polynomial of degree $e(H)$. We call the polynomial $P_{H,n}(A)$ for $A \in \mathrm{Sym}_n$, where Sym_n denotes the $\binom{n+1}{2}$-dimensional vector space of $n \times n$ real symmetric matrices.

The *cut norm* $\|\cdot\|_\square$ on $\mathcal{W}$ is defined by

$$\|W\|_\square := \sup_{S,T \subseteq [0,1]} \left| \int_{S \times T} W(x,y)dxdy \right|.$$

Then the corresponding counting lemma is stated as follows:

Lemma 6.2.1 ([81], Exercise 10.28)**.** *Let U and W be signed graphons and let H be a graph. Then*

$$|t_H(U) - t_H(W)| \leqslant 4e(H) \|U - W\|_\square .$$

The following lemma, which connects a (signed) graphon W to a step function of the form U_A, is an easy consequence of the fact $\|W\|_\square \leqslant \|W\|_1$ and the dominated convergence theorem.

Lemma 6.2.2. *Let W be a signed graphon. For every $\varepsilon > 0$, there exists a symmetric matrix A such that $\|W - U_A\|_\square < \varepsilon$.*

To prove Theorem 6.1.5(ii), we shall use some facts about norming graphs, appeared in [81].

Lemma 6.2.3 ([81], Exercise 14.8)**.** *Let H be a norming graph. Then $t_H(W)$ is always positive for a nonzero signed graphon W. In particular, $e(H)$ is even, since $t_H(-W) = (-1)^{e(H)} t_H(W)$.*

We follow the standard notion of convexity and related definitions. A *convex set* is a subset C of a vector space such that $\lambda x + (1-\lambda)y \in C$ whenver $x, y \in C$ and $\lambda \in (0,1)$. A function $f : C \to \mathbb{R}$ is said to be *convex* if, for each $0 < \lambda < 1$,

$$f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda) f(y).$$

We say that a function f is *strictly convex* if the inequality above is strict whenever x and y are distinct. We shall use a simple fact about convexity repeatedly in what follows:

Lemma 6.2.4. *Let U be a convex subset of a vector space and let f be a convex nonnegative function on U. If $g : \mathbb{R}_{\geqslant 0} \to \mathbb{R}$ is an increasing convex function, then $g \circ f$ is also convex.*

Proof. Let $u, v \in U$. Then for each $\lambda \in (0,1)$.

$$g(f(\lambda u + (1-\lambda)v)) \leqslant g(\lambda f(u) + (1-\lambda) f(v)) \leqslant \lambda g(f(u)) + (1-\lambda) g(f(v)),$$

where the first inequality uses convexity of f and monotonicity of g and the second uses convexity of g. $\square$

For a real-valued function $f(x_1, \cdots, x_n)$ that is twice differentiable on an open set $U \subseteq \mathbb{R}^n$, the *Hessian* of f, denoted by $\nabla^2 f$, is the $n \times n$ matrix $H = (h_{ij})$, where $h_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$. We will only consider polynomials f, so its Hessian $\nabla^2 f$ is always a symmetric matrix with polynomial-valued entries. Standard results in convex analysis, e.g., Section 3.1.4 in [13], imply the following equivalence.

Lemma 6.2.5. *Every n-variable polynomial P is convex on a convex set $C \subseteq \mathbb{R}^n$ if and only if its Hessian $\nabla^2 P$ is positive semidefinite on the interior of C.*

We also recall a basic fact in functional analysis.

Lemma 6.2.6. *Let f be a nonnegative function on a vector space V such that $f(x) = 0$ if and only if $x = 0$, and $f(\lambda x) = |\lambda| f(x)$. If $B := \{x \in V : f(x) \leqslant 1\}$ is convex, then f defines a norm on V.*

Proof. It is enough to prove the triangle inequality $f(x+y) \leqslant f(x) + f(y)$. For nonzero x and y, both $x_1 := x/f(x)$ and $y_1 := y/f(y)$ lie in the convex set B. Set $\lambda = \frac{f(x)}{f(x)+f(y)}$. Then by convexity, $\lambda x_1 + (1-\lambda)y_1 \in B$, and thus,

$$f\left(\frac{x+y}{f(x)+f(y)}\right) = f(\lambda x_1 + (1-\lambda)y_1) \leqslant 1.$$

This proves subadditivity of f. □

6.3 Convexity and weakly norming graphs

Theorem 6.1.5(i) is a consequence of the following equivalence.

Theorem 6.3.1. *Let H be a graph. Then the following are equivalent:*

(i) H is weakly norming.

(ii) $t_H(\cdot)$ is a convex graphon parameter.

(iii) $P_{H,n}(\cdot)$ is a convex polynomial on the positive orthant for every $n \in \mathbb{N}$.

Proof. (i)⇒ (ii). If $\|\cdot\|_{r(H)}$ is convex, then by Lemma 6.2.4, $t_H(W) = \|W\|_{r(H)}^{e(H)}$ is also convex on the set of graphons.

(ii)⇒(i). Convexity of $t_H(\cdot)$ for graphons naturally extends to all $U, W \in \mathcal{W}$ with nonnegative values. Thus, for all $U, W \in \mathcal{W}$ and $\lambda \in (0,1)$,

$$t_H(|\lambda W + (1-\lambda)U|) \leqslant t_H(\lambda|W| + (1-\lambda)|U|) \leqslant \lambda t_H(|W|) + (1-\lambda)t_H(|U|).$$

Indeed, $0 \leqslant W' \leqslant W$ pointwise implies $0 \leqslant t_H(W') \leqslant t_H(W)$, which gives the first inequality, and the second follows from convexity of $t_H(\cdot)$. Therefore, the set

$$B := \{W \in \mathcal{W} : t_H(|W|) \leqslant 1\} = \{W \in \mathcal{W} : t_H(|W|)^{1/e(H)} \leqslant 1\}$$

is convex. Lemma 6.2.6 now proves the triangle inequality for $\|\cdot\|_{r(H)}$.

(ii)⇒(iii). Let $A = (a_{ij})$ and $B = (b_{ij})$ be two $n \times n$ symmetric matrices with positive entries. We may assume that $\max a_{ij} \leqslant 1$ and $\max b_{ij} \leqslant 1$. Then convexity of $P_{H,n}$ immediately follows from linearity of the map $A \mapsto U_A$ and convexity of $t_H(\cdot)$ for graphons.

(iii)⇒(ii). Let W_1 and W_2 be two graphons. By Lemma 6.2.2, there exist $n \times n$ symmetric matrices $A_{1,n}$ and $A_{2,n}$ such that $\left\| W_i - U_{A_{i,n}} \right\|_{\square} \to 0$ as $n \to \infty$ for each $i = 1, 2$. Convexity of $P_{H,n}$ gives

$$t_H(\lambda U_{A_{1,n}} + (1-\lambda) U_{A_{2,n}}) \leqslant \lambda t_H(U_{A_{1,n}}) + (1-\lambda) t_H(U_{A_{2,n}}).$$

Letting $n \to \infty$ finishes the proof, as $t_H(W_n) \to t_H(W)$ if $\|W_n - W\|_{\square} \to 0$ by Lemma 6.2.1. □

Remark 6.3.2. After proving the statement, we found that the equivalence between (i) and (ii) in fact implicitly appeared in Doležal et al. [28] by a different approach using weak* limits. We include our shorter proof for the sake of completeness.

In particular, (iii) enables a computational way of verifying weakly norming property, by using Lemma 6.2.5.

Corollary 6.3.3. *A graph H is weakly norming if and only if the Hessian $\nabla^2 P_{H,n}(A)$ is positive semidefinite for every $A \in \mathrm{Sym}_n$ with positive entries and $n \in \mathbb{N}$.*

To prove Theorem 6.1.4, we need some auxiliary facts about $C_k^{\bowtie}$. For a vertex subset $X \subseteq V(H)$, let $N^*(X) := N(X) \setminus X$, where $N(X)$ denotes the *union* of all neighbours of $x \in X$.

Lemma 6.3.4. *Let $H = C_k^{\bowtie}$ for $k > 4$. Then*

(i) there is an edge e in H such that $N^*(e)$ induces exactly one edge, i.e., e is contained in exactly one 4-cycle, and

(ii) if X spans exactly two edges, then $N^*(X)$ contains an edge.

Note that $C_3^{\bowtie} \cong K_{3,3}$ and $C_4^{\bowtie} \cong Q_3$ violate (i). We omit the proof, as it is seen by a straightforward case analysis.

Proof of Theorem 6.1.4. Let H be the graph $C_k^{\bowtie}$. Since $\nabla^2 P_{H,n}(A)$ is a matrix with polynomial entries, its positive semidefiniteness for $A \in \mathrm{Sym}_n$ with positive entries extends to those $A \in \mathrm{Sym}_n$ with nonnegative entries. We analyse a 2×2 submatrix of the Hessian $\nabla^2 P_{H,3}(A)$, where

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix},$$

with respect to the two variables a_{13} and a_{33}. Namely, we write $h(x,y) := P_{H,3}(A_{x,y})$, where

$$A_{x,y} := \begin{bmatrix} 1 & 1 & y \\ 1 & 0 & 1 \\ y & 1 & x \end{bmatrix},$$

and claim that $\nabla^2 h(x,y)$ is not positive semidefinite at $x = y = 0$.

We may decompose h into $h(x,y) = q(x,y) + \ell(x,y) + r(y)$, where $q(x,y)$ is the sum of all monomials with x-degree at least two, $\ell(x,y)$ is the sum of all monomials with x-degree one, and $r(y)$ is the rest only depending on y. Then the Hessian $\nabla^2 h(0,0)$ is the matrix

$$\begin{bmatrix} q_{xx}(0,0) & \ell_{xy}(0,0) \\ \ell_{xy}(0,0) & r_{yy}(0) \end{bmatrix}$$

and we claim that $q_{xx}(0,0) = 0$ and that $\ell_{xy}(0,0) > 0$. We regard $A_{x,y}$ as a looped, simple, and edge-weighted graph on $\{1,2,3\}$ with the weight a_{ij} for each edge ij. Then $q(x,y)$ counts the weight on the homomorphisms from H to $A_{x,y}$ that use the x-edge at least twice.

If a homomorphism uses the x-edge more than twice, then the corresponding monomial is divisible by x^3 and vanishes in $q_{xx}(0,0)$. Thus, to compute $q_{xx}(0,0)$, we only count those H-homomorphisms which use the x-edge exactly twice. Suppose that $e_1, e_2 \in E(H)$ are mapped to the vertex 3 with the looped x-edge. If a vertex in $N^*(e_1 \cup e_2)$ is mapped to the vertex 1, the homomorphism uses the y-edge and the corresponding monomial vanishes in $q_{xx}(0,0)$. Otherwise, if all the vertices in $N^*(e_1 \cup e_2)$ are mapped to the vertex 2, an edge contained in $N^*(e_1 \cup e_2)$, which exists by Lemma 6.3.4(ii), receives the loop weight 0. Thus, $q_{xx}(0,0) = 0$.

It remains to prove $\ell_{xy}(0,0) > 0$. By Lemma 6.3.4(i), there is an edge e contained in at most one 4-cycle. Let $e' = uv$ be the edge disjoint from e in the 4-cycle that contains

e. Consider the homomorphism that maps an edge e to the x-edge, i.e., both ends of e to 3, exactly one end u of e' to 1, all vertices in $N^*(e) \setminus \{u\}$ to 2, and the other vertices to 1. Since $N^*(e) \setminus \{u\}$ is an independent set by the uniqueness of the 4-cycle containing e, this is a homomorphism that uses both the x- and the y-edge exactly once. Thus, the corresponding monomial is xy, which proves that $\ell_{xy}(0,0) \geqslant 1$. □

6.4 Strict convexity and norming graphs

Theorem 6.1.5(ii) follows from a result analogous to Theorem 6.3.1.

Theorem 6.4.1. *Let H be a graph. Then the following are equivalent:*

(i) H is norming.

(ii) $t_H(\cdot)$ is a strictly convex parameter for signed graphons.

(iii) $P_{H,n}(\cdot)$ is a strictly convex polynomial on Sym_n for every $n \in \mathbb{N}$.

(iv) $P_{H,n}(\cdot)^{1/e(H)}$ is a norm on Sym_n.

Proof. (i)⇒(ii). Let $U, W \in \mathcal{W}$. Hatami proved the following inequality (see (34) in [58]):

$$t_H(U+W) + t_H(U-W) \leqslant 2^{e(H)-1}(t_H(U) + t_H(W)).$$

Since H is norming, $t_H(U-W) > 0$ unless $U = W$ almost everywhere by Lemma 6.2.3. This implies strict convexity of $t_H(\cdot)$.

(ii)⇒(iii). This immediately follows from the linearity of the map $A \mapsto U_A$.

(iii)⇒(iv). If $e(H)$ is odd, then $P_{H,n}(A) + P_{H,n}(-A) = 0$ for every $A \in \mathrm{Sym}_n$, which contradicts strict convexity. Thus, $e(H)$ is even. Again by strict convexity, $2P_{H,n}(A) = P_{H,n}(A) + P_{H,n}(-A) > 0$ whenever $A \neq 0$. Hence, $P_{H,n}(A)^{1/e(H)}$ is well-defined and positive for every nonzero A. Furthermore, $P_{H,n}(\lambda A)^{1/e(H)} = |\lambda| P_{H,n}(A)^{1/e(H)}$. Since

$$B := \{A \in \mathrm{Sym}_n : P_{H,n}(A) \leqslant 1\} = \{A \in \mathrm{Sym}_n : P_{H,n}(A)^{1/e(H)} \leqslant 1\}$$

is a convex set, we may apply Lemma 6.2.6 and conclude that $P_{H,n}(A)$ is a norm on Sym_n.

(iv)⇒(i). The proof is the same as the part (iii)⇒(ii) of Theorem 6.3.1. □

Indeed, positive definiteness of the Hessian implies strict convexity of a polynomial, but the converse is not true in general. Thus, the naive analogue of Corollary 6.3.3 obtained by replacing weakly norming and positive semidefinite by norming and positive definite, respectively, does not hold. One might still hope to prove that a graph H is norming by showing that the Hessian $\nabla^2 P_{H,n}(A)$ is positive definite at each nonzero $A \in \mathrm{Sym}_n$, using the one-sided implication. However, we show that this is impossible by proving that every norming graph has a singular Hessian $\nabla^2 P_{H,n}(A)$ at some $A \neq 0$ whenever n is even.

Proposition 6.4.2. *For every n, there exists a nonzero $2n \times 2n$ symmetric matrix A such that $\nabla^2 P_{H,2n}(A)$ is singular for every norming graph H.*

Proof. Let $A = \begin{bmatrix} J_n & -J_n \\ -J_n & J_n \end{bmatrix}$, where J_n denotes the $n \times n$ matrix with all entries equal to 1. We claim that $\nabla^2 P_{H,2n}(A)$ has eigenvalue 0 with the eigenvector $1_n = (1, 1, \cdots, 1)^T \in \mathbb{R}^{n(2n+1)}$. Recall the folklore fact [81, Example 5.14] that $t_F(U_A)$ is the indicator function that F is eulerian. In particular, H is eulerian and $e(H)$ is even. Thus,

$$t_H(U_A - \varepsilon) + t_H(U_A + \varepsilon) = 2t_H(U_A) + 2\sum_J t_J(U_A)\varepsilon^{e(H)-e(J)} = 2 + 2\varepsilon^2 \sum_F t_F(U_A) + O(\varepsilon^4),$$

where the first sum is taken over all proper subgraphs J of H with even number of edges and the second is taken over all subgraphs $F \subseteq H$ with $e(F) = e(H) - 2$. Since one always obtains a non-eulerian subgraph F by deleting two edges from an eulerian graph H, $t_F(U_A) = 0$. Thus, $t_H(U_A - \varepsilon) + t_H(U_A + \varepsilon) = 2 + O(\varepsilon^4)$. On the other hand, by the Taylor expansions of $P_{H,2n}$ at A,

$$P_{H,2n}(A + \varepsilon J_{2n}) + P_{H,2n}(A - \varepsilon J_{2n}) = P_{H,2n}(A) + 2\varepsilon^2 1_n^T \nabla^2 P_{H,2n}(A) 1_n + O(\varepsilon^3).$$

Since $P_{H,2n}(A + \varepsilon J_{2n}) + P_{H,2n}(A - \varepsilon J_{2n}) = (2n)^{v(H)}(t_H(U_A - \varepsilon) + t_H(U_A + \varepsilon))$, it follows that $1_n^T \nabla^2 P_{H,2n}(A) 1_n = 0$. Since $\nabla^2 P_{H,2n}(A)$ is positive semidefinite, $\nabla^2 P_{H,2n}(A) 1_n$ must be zero. This completes the proof of the claim. $\square$

As already used in the last line of the proof, we are only able to obtain a weaker analogue of Corollary 6.3.3.

Corollary 6.4.3. *For a norming graph H, $\nabla^2 P_{H,n}(A)$ is positive semidefinite for every $A \in \mathrm{Sym}_n$.*

It is still enough to find $A \in \mathrm{Sym}_n$ such that $\nabla^2 P_{H,n}(A)$ is not positive semidefinite to prove that H is not norming. This is exactly what we do in the proof of Theorem 6.1.6.

Proof of Theorem 6.1.6. Let H_t be the graph $K_{2t+1,2t+1} \setminus (2t+1) \cdot K_2$. As mentioned before, it is enough to prove that H_t is not norming, as $K_{2t,2t}$ minus a perfect matching is not eulerian and thus not norming. Let

$$A = \begin{bmatrix} x & y & \varepsilon \\ y & 1 & 1 \\ \varepsilon & 1 & -1 \end{bmatrix}$$

and let $h(x,y) := P_{H,3}(A)$. Here we suppress the dependency on $0 < \varepsilon < 1$, since ε is a small constant to be chosen later. We analyse the 2×2 Hessian matrix $\nabla^2 h$ at $(0,0)$. As in the proof of Theorem 6.1.4, we decompose $h(x,y)$ into three parts, i.e.,

$$h(x,y) = q(x,y) + \ell(x,y) + r(y),$$

where $q(x,y)$ is the sum of monomials divisible by x^2, ℓ is the sum of monomials whose x-degree is 1, and r is the remaining terms. Then the Hessian $\nabla^2 h(0,0)$ is the matrix

$$\begin{bmatrix} q_{xx}(0,0) & \ell_{xy}(0,0) \\ \ell_{xy}(0,0) & r_{yy}(0) \end{bmatrix}.$$

We regard A as a looped, simple, and edge-weighted graph on $\{1,2,3\}$ with the weight a_{ij} for each edge ij. For the same reason as in the proof of Theorem 6.1.4, $q_{xx}(0,0)$ is equal to the number of homomorphisms that use the x-edge exactly twice without using the y-edge. Such a homomorphism ϕ maps at least three vertices V_1 in H_t that induce exactly two edges to the vertex 1 and never maps their neighbour to the vertex 2. Thus, $N^*(V_1)$ must be embedded to the vertex 3. Since H_t is $2t$-regular and V_1 contains exactly two edges, $e(V_1, N^*(V_1)) \geqslant 6t-4$ and thus, ϕ uses the ε-edge at least $6t-4$ times.

Analogously, $\ell_{xy}(0,0)$ counts the number of homomorphisms that use both the x- and y-edges exactly once and hence, use the ε-edge at least $4t-3$ times. The homomorphisms using the y-edge exactly twice and avoiding the x-edge must use ε-edge at least $2t-2$ times. Therefore,

$$\nabla^2 h(0,0) = \begin{bmatrix} q_{xx}(0,0) & \ell_{xy}(0,0) \\ \ell_{xy}(0,0) & r_{yy}(0) \end{bmatrix} = \begin{bmatrix} O(\varepsilon^{6t-4}) & O(\varepsilon^{4t-3}) \\ O(\varepsilon^{4t-3}) & O(\varepsilon^{2t-2}) \end{bmatrix}.$$

Here $O(\cdot)$ notation includes implicit multiplicative constants depending only on t.

Unfortunately, the product of the diagonal entries and the product of the off-diagonal entries are in the same order $O(\varepsilon^{8t-6})$. However, we claim that $r_{yy}(0)$ is asymptotically smaller than $O(\varepsilon^{2t-2})$ and also that $|\ell_{xy}(0,0)| = \Omega(\varepsilon^{4t-3})$, which implies that $\nabla^2 h(0,0)$ is not positive semidefinite for a sufficiently small $\varepsilon > 0$.

Let $A \cup B$ be the bipartition of H_t and let $A = \{a_1, \cdots, a_{2t+1}\}$ and $B = \{b_1, \cdots, b_{2t+1}\}$ such that $a_i b_i$, $1 \leqslant i \leqslant 2t+1$, is the missing perfect matching in H_t. Firstly, let Φ_{yy} be the set of homomorphisms that use the y-edge twice and the ε-edge exactly $2t-2$ times while avoiding the x-edge. Each $\varphi \in \Phi_{yy}$ must map one vertex, say a_1, to 1, two neighbours of a_1 to 2, and the other $2t-2$ neighbours of a_1 to 3. That is, once we choose the vertex a_1 and two of its neighbours to be mapped to 2, all the embeddings of the neighbours of a_1 are fixed. Consider these vertices as pre-embedded. Let V_3 be the set of $2t-2$ vertices mapped to 3 and let U be the vertices that are not yet embedded. Then $U = \{a_2, \cdots, a_{2t+1}\} \cup \{b_1\}$. In particular, U induces a star centred at b_1 with $2t$ edges. Also note that by the definition of Φ_{yy}, the homomorphisms in Φ_{yy} do not map any other vertex than a_1 to 1. For each $\varphi \in \Phi_{yy}$, denote by U_φ the subset of U mapped to the vertex 3. Then the coefficient of the term $\varepsilon^{2t-2} y^2$ in $r(y)$ is determined by

$$\sum_{\varphi \in \Phi_{yy}} (-1)^{e(V_3, U_\varphi) + e(V_3) + e(U_\varphi)}. \tag{6.4.1}$$

Suppose $b_1 \in U_\varphi$. For each φ, let b_i and b_j, $i < j$, be the two vertices mapped to the vertex 2. Then both a_i and a_j have all their $2t-1$ other neighbours than b_i and b_j mapped to 3. Thus, by switching the image of a_i under φ between 2 and 3, we produce another homomorphism $\bar{\varphi}$ whose weight $(-1)^{e(V_3, U_{\bar{\varphi}}) + e(V_3) + e(U_{\bar{\varphi}})}$ has exactly the opposite sign of that of φ. This switching is an involution, and thus, the two terms pair up and cancel each other in (6.4.1). If $b_1 \notin U_\varphi$, then one may do an analogous switching with the minimum indexed vertex amongst $a_2, \cdots, a_{2t+1}$ that has an odd degree to those vertices mapped to 3. Thus, (6.4.1) evaluates to zero.

To prove $|\ell_{xy}(0,0)| = \Omega(\varepsilon^{4t-3})$, let Ψ_{xy} be the set of homomorphisms that use each of the x- and y-edge exactly once. Suppose that, under $\psi \in \Psi_{xy}$, a_i and b_j, $i, j > 1$ and $i \neq j$, are mapped to 1 and b_k, $i \neq k > 1$, is mapped to 2. To avoid using the y-edge more than once, ψ must map $(B \setminus \{b_i, b_j, b_k\}) \cup (A \setminus \{a_i, a_j\})$ to the vertex 3. Thus, there are only two vertices a_j and b_i whose embedding is not yet determined. Note that a_j and b_i have $2t-2$ and $2t-1$ neighbours mapped to 3, respectively, and they are adjacent. Let α_ψ and β_ψ be the indicator function that a_j and b_i are mapped to 3 by ψ, respectively. Then the coefficient of the term $\varepsilon^{4t-3}xy$ in $\ell(x,y)$ is a nonzero constant times

$$\sum_{\psi\in\Psi_{xy}}(-1)^{(2t-2)\alpha_\psi+(2t-1)\beta_\psi+\alpha_\psi\beta_\psi} = \sum_{\psi\in\Psi_{xy}}(-1)^{\beta_\psi+\alpha_\psi\beta_\psi}.$$

Since each choice $(\alpha_\psi,\beta_\psi)\in\{0,1\}^2$ determines a homomorphism $\psi\in\Psi_{xy}$, (α_ψ,β_ψ) is uniformly distributed on $\{0,1\}^2$. Hence, the sum above evaluates to a nonzero constant, which proves the claim. □

6.5 Concluding remarks

Our method using the Hessian matrix $\nabla^2 P_{H,n}$ is reminiscent of [74] in the sense that both rely on determining positive semidefiniteness of matrices given by homomorphism counts. More precisely, in [74] they looked at two edges e and e' in a graph G sharing a vertex and used non-positive semidefiniteness of the 2×2 matrix

$$A_{e,e'} = \begin{bmatrix} h_{e,e} & h_{e,e'} \\ h_{e,e'} & h_{e',e'} \end{bmatrix},$$

where h_{e_1,e_2} counts the number of those homomorphisms from H to G which map a $K_{1,2}$ in H to the homomorphic copy of $K_{1,2}$'s formed by e_1 and e_2, to prove that a certain H is not weakly norming.

This is somewhat analogous to the Hessian matrix obtained by evaluating the corresponding weights of e and e' to be zero. However, the Hessian does not take the particular $K_{1,2}$-structure into account, so it has larger entries than $A_{e,e'}$ above. We did not attempt to reprove their result using our language, but we remark that there are non-weakly norming graphs that satisfy their positive semidefiniteness condition. For instance, take a vertex-disjoint union of two non-isomorphic connected weakly norming graphs. This is proven to be not weakly norming in [47], but the corresponding 2×2 matrix in [74] is positive semidefinite, since it is a positive linear combination of the respective matrices of the components. It would be interesting to see if the two distinct positive semidefiniteness conditions are equivalent for connected graphs H.

7. On extremal problems concerning the traces of sets

7.1 Introduction

A hypergraph $\mathcal{H}$ is a pair $(V, \mathcal{F})$ where V is the set of vertices and $\mathcal{F} \subseteq 2^V = \mathscr{P}(V)$ is the set of edges. In the literature, the problems we consider in this article are often presented in the context of families rather than hypergraphs. If not necessary, it is then not distinguished between the family $\mathcal{F} \subseteq 2^V$ and the hypergraph $(V, \mathcal{F})$. We will follow this notational path and also use the "family" and "hypergraph" essentially synonymously.

Let V be an n-element set and let $\mathcal{F}$ be a family of subsets of V. For a subset T of V define the *trace* of $\mathcal{F}$ on T by $\mathcal{F}_{|T} = \{F \cap T : F \in \mathcal{F}\}$. For integers n, m, a, and b, we write

$$(n, m) \to (a, b)$$

if for every family $\mathcal{F} \subseteq 2^V$ with $|\mathcal{F}| \geqslant m$ and $|V| = n$ there is an a-element set $T \subseteq V$ such that $|\mathcal{F}_{|T}| \geqslant b$ (we also say that (n, m) *arrows* (a, b)).

The first type of question that was asked for this arrowing notation is similar to the spirit of the classic Turán problem: For a fixed number of vertices n, how many edges are needed such that there is a subset of s vertices such that all its subsets lie in the trace. The following result on this question was conjectured by Erdős [46] and was proved independently by Sauer [101], Shelah and Perles [103], and Vapnik and Červonenkis [114]. It states that for a large family $\mathcal{F}$ on n vertices, there is an s-set of vertices such that all its subsets lie in the trace of $\mathcal{F}$. More precisely, they showed that $(n, m) \to (s, 2^s)$ whenever $m > \sum_{0 \leqslant i < s} \binom{n}{i}$.

Another fundamental question that was raised in the area is how large a family can be at most so that there will still be a vertex v such that the trace on $V \setminus \{v\}$ is not much smaller than the original family. More precisely, the following problem was posed by Füredi and Pach [46] and, more recently, by Frankl and Tokushige as Problem 3.8 in their monograph [42]:[1]

Problem 7.1.1. *Given non-negative integers n and s, what is the maximum integer $m(n, s)$ such that for every integer $m \leqslant m(n, s)$, we have*

$$(n, m) \to (n - 1, m - s).$$

A family $\mathcal{F}$ is *hereditary* if for every $F' \subseteq F \in \mathcal{F}$, we have that $F' \in \mathcal{F}$. In [41], Frankl proves that among families with a fixed number of edges and vertices, the trace is minimised by hereditary families. Thus, the problems considered here, and in particular Problem 7.1.1, can be reduced to hereditary families (see Lemma 7.2.1). Hence, Problem 7.1.1 is asking for the maximum integer $m(n, s)$ such that every hereditary hypergraph on n vertices with at most $m(n, s)$ edges contains a vertex of degree at most s.[2] Formulated differently, $m(n, s)$ is the minimal integer such that every hereditary hypergraph on n vertices with minimum degree at least $s + 1$ has at least $m(n, s) + 1$ edges.

The investigation of this problem started with Bondy [11] and Bollobás [82] determining $m(n, 0)$ and $m(n, 1)$, respectively. Later Frankl [41] and Frankl and Watanabe [43] proved part (1) and (2), respectively, of the following theorem.

Theorem 7.1.2. *For $d, n \in \mathbb{N}$ and $d|n$, we have*

1. $m(n, 2^{d-1} - 1) = \frac{n}{d}(2^d - 1)$,
2. $m(n, 2^{d-1} - 2) = \frac{n}{d}(2^d - 2)$.

Consider a family consisting of a set of size d and all possible subsets, and take n/d vertex disjoint copies of it. The resulting family has minimum degree 2^{d-1} and $\frac{n}{d}(2^d - 1) + 1$ edges. Thus, this family is an extremal construction for (1). By taking out all sets of size d, we obtain an extremal construction for (2).

Our main result makes further progress on Problem 7.1.1, solving it for general $s = 2^{d-1} - c$ as long as c is linearly small in d.

Theorem 7.1.3 (Main theorem). *Let $d, c, n \in \mathbb{N}$ with $d \geqslant 4c$ and $d|n$. Then*

$$m(n, 2^{d-1} - c) = \frac{n}{d}(2^d - c).$$

[1] There have been slightly different versions in use for the arrowing notation and for what we denote by $m(n, s)$. In this work, we follow the notation in [42].

[2] As usual, we define the degree of a vertex v in a hypergraph $\mathcal{H} = (V, \mathcal{F})$ as $\deg(v) = |\{F \in \mathcal{F} : v \in F\}|$ and the minimum degree of $\mathcal{H}$ is $\delta(\mathcal{H}) = \min_{v \in V} \deg(v)$.

Remark 7.1.4. In fact, our proof of Theorem 7.1.3 yields that for $d \geqslant 4c$ and $m \leqslant \frac{n}{d}(2^d - c)$ we have $(n, m) \to (n-1, m-(2^{d-1}-c))$ without any divisibility conditions on n. The assumption $d|n$ is only necessary for the extremal constructions showing the maximality of $\frac{n}{d}(2^d - c)$. Analogous remarks hold for Theorem 7.1.2 above and Theorem 7.1.5 below. In Section 7.5 we provide a construction showing that the equality in Theorem 7.1.3 does not hold for $d = c$ (see Construction 1).

One might also try to solve Problem 7.1.1 for small values of s. Apart from the aforementioned results by Bondy and Bollobás, progress was made by Frankl [41], Watanabe [115,116], and by Frankl and Watanabe [43]. In [43], they conjectured that $m(n, 12) = (28/5 + o(1))n$. Theorem 7.1.3 does not consider cases for which d is very small in terms of c. The following results extend Theorem 7.1.2 to $c = 3$ and 4 and every $d \geqslant 5$ (for smaller d the respective $m(n, s)$ is not defined or has been determined previously). In particular, it proves the conjecture of Frankl and Watanabe for $s = 12$ in a strong sense.

Theorem 7.1.5. *Let $d, n \in \mathbb{N}$ with $d \geqslant 5$ and $d|n$. Then*

1. $m(n, 2^{d-1} - 3) = \frac{n}{d}(2^d - 3)$ *and*
2. $m(n, 2^{d-1} - 4) = \frac{n}{d}(2^d - 4)$. *In particular,* $m(n, 12) = \frac{28}{5}n$.

Note that for larger d, this theorem is of course a special case of Theorem 7.1.3.

7.2 Preliminaries

In this work we consider the set of natural numbers $\mathbb{N}$ to start with 1 and the logarithms considered are to the base 2. Further, for $i \in \mathbb{N}$ we set as usual $[i] = \{1, \dots, i\}$, and it is also convenient to define $[i]_0 = \{0, \dots, i\}$. Given a set $F \subseteq \mathbb{N}$ and some $i \in \mathbb{N}$, we denote by $F + i$ the set $\{j + i : j \in F\}$. For our considerations isolated vertices, i.e., vertices that are contained in the vertex set of a hypergraph but do not lie in any edges, usually do not play an important rôle. This will lead to a few easy peculiarities in notation. For two hypergraphs $\mathcal{H}$ and $\mathcal{H}'$ we write $\mathcal{H} \cong \mathcal{H}'$ if they are isomorphic up to isolated vertices, more precisely, if there are vertex sets V disjoint to $V(\mathcal{H})$ and V' disjoint to $V(\mathcal{H}')$ such that the hypergraph $(V(\mathcal{H}) \cup V, E(\mathcal{H}))$ is isomorphic to $(V(\mathcal{H}') \cup V', E(\mathcal{H}'))$.

For a hypergraph $\mathcal{H} = (V, E)$ and $v \in V$ we define the link L_v of v to be the hypergraph on V with edge set $\{F \setminus \{v\} : v \in F \in E\}$. Further, we write

$$V_v = \{w \in V : \text{there is an } e \text{ with } \{v, w\} \subseteq e \in E\},$$

note that if v is not an isolated vertex, then $v \in V_v$. This notation will be useful in the proof of Theorem 7.1.3 when defining the clusters mentioned in the overview of the proof.

The following lemma due to Frankl [41] provides the aforementioned reduction of Problem 7.1.1 to hereditary families.

Lemma 7.2.1. *For $n, m, a, b \in \mathbb{N}$ the following statements are equivalent.*

1. *For every n-set V and every hereditary family $\mathcal{F} \subseteq 2^V$ with $|\mathcal{F}| \geqslant m$, there exists a set $T \subseteq V$ with $|T| = a$ such that $|\mathcal{F}_{|T}| \geqslant b$.*

2. *$(n, m) \to (a, b)$.*

In particular, this means that in the proof of our results we only need to consider hereditary families.

Let $n \in \mathbb{N}$, for $A, B \in 2^{[n]}$ we say that $A \prec_{col} B$ or A precedes B in the *colexicographic order* if $\max(A \triangle B) \in B$. Let $m \in \mathbb{N}$ with $m \leqslant 2^n$ and define $\mathcal{R}_n(m)$ to be the family on n vertices containing the first m sets of $2^{[n]}$ according to the colexicographic order. Note that for $n \leqslant n'$ and $m \leqslant 2^n$, we have $\mathcal{R}_n(m) \cong \mathcal{R}_{n'}(m)$ and hence, we will not distinguish between $\mathcal{R}_n(m)$ and $\mathcal{R}_{n'}(m)$ and we will omit the subscript. The following theorem due to Katona [66] is a generalisation of the well-known Kruskal-Katona theorem.

Theorem 7.2.2. *Let $f : \mathbb{N}_0 \to \mathbb{R}$ be a monotone non-increasing function and let $\mathcal{F}$ be a hereditary family with $|\mathcal{F}| = m$. Then*

$$\sum_{F \in \mathcal{F}} f(|F|) \geqslant \sum_{R \in \mathcal{R}(m)} f(|R|).$$

Observe that for a hypergraph $\mathcal{H} = (V, E)$ a double counting argument yields

$$\sum_{x \in V} \sum_{H \in L_x} \frac{1}{|H| + 1} = |E \setminus \varnothing|.$$

For the proofs of Theorems 7.1.3 and 7.1.5 we generalise this argument by considering weights $w_{\mathcal{H}}(v)$ for all vertices v. We will refer to $\sum_{H \in L_x} \frac{1}{|H|+1}$ as *uniform weight* since it can be imagined as uniformly distributing the unit weight among its vertices.

Accordingly, Theorem 7.2.2 will normally be applied with $\mathcal{F}$ being the link of a vertex and, as we usually consider the uniform weight, the function f will often be $f(k) = \frac{1}{k+1}$. The weight of $\mathcal{R}(m)$ with respect to this f will come up repeatedly and hence, for brevity we set $W(m) := \sum_{R \in \mathcal{R}(m)} \frac{1}{|R|+1}$. Note that we have $W(2^{d-1}) = \frac{2^d - 1}{d}$ and further the following estimate[3] for $W(2^{d-1}-c)$ for a $c \in [2^{d-2}]$:

$$W(2^{d-1}-c) \geqslant \frac{2^d-1}{d} - \frac{c}{d-\log c} \tag{7.2.1}$$

Indeed, if $A \in 2^{[d-1]} \setminus \mathcal{R}(2^{d-1}-c)$, then there are at least $2^{d-1-|A|}$ sets in $2^{[d-1]} \setminus \mathcal{R}(2^{d-1}-c)$. Thus, it follows that for every $A \in 2^{[d-1]} \setminus \mathcal{R}(2^{d-1}-c)$ we have $|A| \geqslant d-1-\log c$. This gives that $W(2^{d-1}) - W(2^{d-1}-c) \leqslant \frac{c}{d-\log c}$ and thereby (7.2.1).

7.3 Proof of Theorem 7.1.3

For proving Theorem 7.1.3 we introduce two "local" lemmas. The first lemma says that if a family deviates enough from $\mathcal{R}(m)$, the weight of this family will have a surplus with respect to $W(m)$.[3]

Lemma 7.3.1. *Let $d \geqslant 3$ and $c \leqslant 2^d$ be integers. For a hereditary family $\mathcal{H}$, with $|\mathcal{H}| \geqslant 2^d - c$ the following holds.*

1. $\sum_{H\in\mathcal{H}} \frac{1}{|H|+1} \geqslant W(2^d-c)$.
2. *If there are at least $d+1$ non isolated vertices in $\mathcal{H}$, then*
$$\sum_{H\in\mathcal{H}} \frac{1}{|H|+1} \geqslant W(2^d-c) + \frac{1}{6}\,.$$
3. *If $c \in \{2,3\}$ and $\mathcal{H} \not\cong \mathcal{R}(2^d-c)$, then we have*
$$\sum_{H\in\mathcal{H}} \frac{1}{|H|+1} \geqslant W(2^d-c) + \min\left(\frac{1}{6}, \frac{1}{d}\right)\,.$$

Proof. Let d, n, c, and $\mathcal{H}$ be given as in the statement. The first part follows by applying Theorem 7.2.2 with $f(k) = \frac{1}{k+1}$.

In order to prove part (2) and (3) we need some preparation. Denote by h_i and r_i the number of i-sets in $\mathcal{H}$ and $\mathcal{R}(2^d-c)$, respectively. Given $s \in [d]_0$ set $g(k) = 1$ for $k \leqslant s$ and $g(k) = 0$ for $k > s$. Then Theorem 7.2.2 applied with $f = g$ yields

$$\sum_{i\in[s]_0} h_i \geqslant \sum_{i\in[s]_0} r_i. \tag{7.3.1}$$

[3]To have a clearer presentation of our main results and their proofs, we refrained from striving for optimal bounds.

Next, let $H_1, \dots, H_{|\mathcal{H}|}$ be an enumeration of the elements of $\mathcal{H}$ such that $|H_j| \leqslant |H_{j+1}|$. Given $i \in [d-1]$ let $\varphi(i)$ be the number of edges of size at most i in the family $\mathcal{R}(2^d - c)$, i.e., $\varphi(i) = \sum_{j \in [i]_0} r_j$. Let $\mathcal{H}_0 = \{H_1\} = \{\varnothing\}$ and for $i \in [d-1]$ consider the following set of edges $\mathcal{H}_i = \{H_{\varphi(i-1)+1}, \dots, H_{\varphi(i)}\}$ and observe that its size is r_i. Inequality (7.3.1) implies that for $H \in \mathcal{H}_i$, where $i \in [d-1]_0$, we have $|H| \leqslant i$. Thus,

$$\sum_{i \in [d-1]_0} \sum_{H \in \mathcal{H}_i} \frac{1}{|H|+1} \geqslant \sum_{i \in [d-1]_0} \frac{r_i}{i+1} = W(2^d - c)\,. \tag{7.3.2}$$

If now at least $d+1$ vertices are contained in edges of $\mathcal{H}$, then even for $H_{d+2} \in \mathcal{H}_2$ it holds that $|H_{d+2}| = 1$. Hence, (7.3.2) now becomes $\sum_{H \in \mathcal{H}} \frac{1}{|H|+1} \geqslant \frac{1}{2} - \frac{1}{3} + W(2^d - c)$ and (2) is proved.

For proving (3), let $c \in \{2, 3\}$ and note that if there are at least $d+1$ non isolated vertices in $\mathcal{H}$, then the result follows from (2). Thus, assume that there are only d non isolated vertices in $\mathcal{H}$. Observe that $r_i = \binom{d}{i}$ for $i \in [d-2]$, $r_d = 0$ and $r_{d-1} = d - (c-1)$. Hence, due to (7.3.1) we have $h_i = \binom{d}{i}$ for $i \in [d-2]$ and because of $\mathcal{H}$ being hereditary and the size of $\mathcal{H}$, further $h_{d-1} \geqslant d - (c-1)$. In fact, $h_{d-1} > r_{d-1} = d - (c-1)$ has to hold since $\mathcal{H} \ncong \mathcal{R}(2^d - c)$. Together with (7.3.2) the result follows. □

The following is the second local lemma. Part (2) states that a hereditary family on d vertices with high minimum degree contains many edges and therefore, considering Lemma 7.2.1, this is a local version of Theorem 7.1.3. Moreover, Part (1) states that if a hereditary family has not enough edges, then there are several vertices of low degree.

Lemma 7.3.2. *Let $d, c \in \mathbb{N}$ such that $d > c$, let V be a d-set and let $\mathcal{H} \subseteq 2^V$ be hereditary.*

1. *If $|\mathcal{H}| \leqslant 2^d - c - 1$, then $\deg(v) \leqslant 2^{d-1} - c - 1$ for at least $d - c$ vertices v.*
2. *If $\delta(\mathcal{H}) \geqslant 2^{d-1} - c$, then $|\mathcal{H}| \geqslant 2^d - c$.*

Proof. (1): By $\bar{\mathcal{H}}$ denote the family $\{V \smallsetminus F : F \in 2^V \smallsetminus \mathcal{H}\}$. The bound on $|\mathcal{H}|$ implies that $c + 1 \leqslant |\bar{\mathcal{H}}|$ and observe that since $\mathcal{H}$ is hereditary, $\bar{\mathcal{H}}$ is hereditary. Consider some ordering $\bar{\mathcal{H}} = \{H_1, \dots, H_{|\bar{\mathcal{H}}|}\}$ with $|H_i| \leqslant |H_{i+1}|$. Note that because $\bar{\mathcal{H}}$ is hereditary, we know that if some vertex $v \in V$ is contained in one of the edges $H_1, \dots, H_j$, then in fact $\{v\} = H_i$ for some $i \in [j]$. Thus, there are $d - c$ vertices that do not lie in any of $H_1, \dots, H_{c+1}$. Note that these vertices lie in at least $c + 1$ sets of $2^V \smallsetminus \mathcal{H}$ and therefore, for each such v we have $\deg_{\mathcal{H}}(v) \leqslant 2^{d-1} - c - 1$.

(2): Assume for contradiction that $|\mathcal{H}| \leqslant 2^d - c - 1$. Then (1) gives the contradiction. □

Now we are ready to prove Theorem 7.1.3.

Proof of Theorem 7.1.3. Let n, d, and c be given as in the theorem. First note that

$$\mathcal{F}_0 = \left\{F + (i-1)d : F \in \mathcal{R}(2^d - (c-1)) \text{ and } i \in \left[\frac{n}{d}\right]\right\} \subseteq 2^{[n]}$$

shows that for $m = \frac{2^d - c}{d} n + 1$, we have $(n, m) \nrightarrow \big(n-1, m - (2^{d-1} - c)\big)$.

In a hereditary family on n vertices with m edges the existence of a set of size $n-1$ on which the trace of the family has size at least $m - (2^{d-1} - c)$ is equivalent to the existence of a vertex with degree at most $2^{d-1} - c$. Therefore, Lemma 7.2.1 implies that it is sufficient to show that for every hereditary family $\mathcal{F}$ on n vertices with minimum degree at least $2^{d-1} - c + 1$ we have $|\mathcal{F}| \geqslant \frac{2^d - c}{d} n + 1$. Let now $\mathcal{F} \subseteq 2^V$ be such a hereditary family on some n-set V in which every vertex has degree at least $2^{d-1} - c + 1$.

To prove the lower bound on the number of edges, we will define a weight function w on V with the property that $1 + \sum_{v \in V} w_{\mathcal{F}}(v) \leqslant |\mathcal{F}|$. Subsequently, it will be enough to show that $\sum_{v \in V} w_{\mathcal{F}}(v) \geqslant \frac{2^d - c}{d} n$. Indeed, for $c = 1$ from Lemma 7.3.1 (1) for $\mathcal{H} = L_v$, it follows that the weight function $\sum_{H \in L_v} \frac{1}{|H|+1}$ satisfies the desired inequality. Hence, from now on we assume $c \geqslant 2$. Note however, that for this uniform weight and c large, in $\mathcal{F}_0$ there are vertices with weight below and above $\frac{2^d - c}{d}$. As mentioned in the overview, we overcome this difficulty by using non-uniform weights and by bounding the average weight of sets of vertices instead of bounding the weight of every single vertex.

To that aim, we will in the following consider a partition of V. Let us call a vertex $v \in V$ *light* if $|V_v| = d$. Further, let $\mathcal{L}$ be a maximum set of light vertices such that $V_v \cap V_{v'} = \varnothing$ for all $v, v' \in \mathcal{L}$ and call the sets V_v with $v \in \mathcal{L}$ *clusters*. Later, the weight of a vertex will be defined depending on how it relates to these clusters. Moreover, call the vertices $u \in V \setminus \bigcup_{v \in \mathcal{L}} V_v$ with $|V_u| > d$ *heavy* vertices and let $\mathfrak{H}$ be the set of all heavy vertices. The vertices in $\mathcal{L}$ will be distinguished further into two different types $\mathcal{L}_1$ and $\mathcal{L}_2$ as follows. Let $\mathcal{L}_1$ be the set of those vertices $v \in \mathcal{L}$ for which every vertex in V_v is only contained in edges of 2^{V_v}, that is

$$\mathcal{L}_1 = \{v \in \mathcal{L} \colon \text{there is no } e \in \mathcal{F} \setminus 2^{V_v} \text{ with } e \cap V_v \neq \varnothing\}.$$

Furthermore, let $\mathcal{L}_2$ be the set of those vertices $v \in \mathcal{L}$ for which there exists an $x \in V_v$ that is contained in an edge of $\mathcal{F} \setminus 2^{V_v}$, in other words,

$$\mathcal{L}_2 = \{v \in \mathcal{L} \colon \text{there is an } e \in \mathcal{F} \setminus 2^{V_v} \text{ with } e \cap V_v \neq \varnothing\}. \tag{7.3.3}$$

Note that we have $\mathcal{L} = \mathcal{L}_1 \cup \mathcal{L}_2$. Lastly, we collect the remaining vertices in the set $\bar{\mathcal{L}} = V \setminus (\mathfrak{H} \cup \bigcup_{v \in \mathcal{L}} V_v)$. Thus, we have $V = \mathfrak{H} \cup \bigcup_{v \in \mathcal{L}_1} V_v \cup \bigcup_{v \in \mathcal{L}_2} V_v \cup \bar{\mathcal{L}}$.

Next, for each of the partition classes $\mathfrak{H}$, $\bigcup_{v \in \mathcal{L}_1} V_v$, $\bigcup_{v \in \mathcal{L}_2} V_v$, and $\bar{\mathcal{L}}$ the weights will be defined and we will show that the average weight in each partition class is bounded from below by $\frac{2^d - c}{d}$.

Assign the uniform weight $w_{\mathcal{F}}(u) = \sum_{H \in L_u} \frac{1}{|H|+1}$ to every heavy vertex $u \in \mathfrak{H}$. This definition and (2) from Lemma 7.3.1 give that every heavy vertex has weight at least

$$\frac{1}{6} + W(2^{d-1} - c + 1) \geqslant \frac{1}{6} + \frac{2^d - 1}{d} - \frac{c-1}{d - \log(c-1)} \geqslant \frac{2^d - c}{d}\,, \tag{7.3.4}$$

where we used the bound (7.2.1) for the first inequality and $d \geqslant 4c$ and $\log x \leqslant \frac{2}{3}x$ for $x \geqslant 1$ for the second (recall that we can assume $c \geqslant 2$).

Given $v \in \mathcal{L}_1$, we have that $\mathcal{F}[V_v]$ is a family on d vertices with minimum degree at least $2^{d-1} - c + 1$. Thus, from Lemma 7.3.2 (2) (with $c-1$ here in place of c there) it follows that $|\mathcal{F}[V_v]| \geqslant 2^d - c + 1$. Since summing the uniform vertex weights of all vertices of a family amounts to the number of non-empty edges in that family, assigning the uniform weight $w_{\mathcal{F}}(x) = \sum_{H \in L_x} \frac{1}{|H|+1}$ to every $x \in V_v$ yields

$$\frac{1}{d} \sum_{x \in V_v} w_{\mathcal{F}}(x) = \frac{|\mathcal{F}[V_v] \setminus \{\varnothing\}|}{d} \geqslant \frac{2^d - c}{d}\,. \tag{7.3.5}$$

Given $v \in \mathcal{L}_2$, the idea is that the vertices in V_v already have a relatively large uniform weight just taking into account the edges on V_v. Thus, they only need a smaller proportion of the weight of an edge that includes vertices outside of V_v. More precisely, we assign the weight

$$w_{\mathcal{F}}(x) = \sum_{H \in L_x} \frac{1}{|H|+1} - |V_x \setminus V_v| \left(\frac{1}{2} - \frac{c-1}{d-c} \right)$$

to every vertex $x \in V_v$. Since $\mathcal{F}$ is hereditary, the number of 2-uniform edges containing x and crossing from the inside of the cluster to the outside is exactly $|V_x \setminus V_v|$. Hence, this definition can be understood as vertices in V_v having basically the uniform weight but then renouncing part of their uniform share of those crossing edges. Later, these edges will contribute more than their uniform share to weight of the outside vertex.

Of course, if $|\mathcal{F}[V_v]| \geqslant 2^d - c + 1$, then again the bound (7.3.5) follows for v directly by double counting and thus, we may assume that $|\mathcal{F}[V_v]| \leqslant 2^d - c$. Define the set C as the set of vertices $x \in V_v$ for which there exists some F_x with $x \in F_x \in \mathcal{F} \setminus 2^{V_v}$. Note that in fact, since $\mathcal{F}$ is hereditary, we may assume $|F_x| = 2$. Considering the minimum degree condition in $\mathcal{F}$ and applying Lemma 7.3.2 (1) to $\mathcal{F}[V_v]$ (with $c-1$ here instead of c there) it follows that

$$|C| \geqslant d - c + 1\,. \tag{7.3.6}$$

Moreover, the minimum degree of $\mathcal{F}$ implies $deg(v) \geqslant 2^{d-1} - c + 1$ and hence, $\mathcal{F}$ being hereditary gives that $|2^{V_v} \smallsetminus \mathcal{F}| \leqslant 2(c-1)$. Therefore, double counting the non-empty edges in $\mathcal{F}[V_v]$ yields

$$|\mathcal{F}[V_v] \smallsetminus \{\varnothing\}| = \sum_{x \in V_v} \sum_{H \in L_x \cap 2^{V_v}} \frac{1}{|H|+1} \geqslant 2^d - 2c + 1\,. \tag{7.3.7}$$

Now, for every vertex $x \in C$ there is at least one 2-uniform edge $F_x \in \mathcal{F} \smallsetminus F[V_v]$ which contributes $\frac{c-1}{d-c}$ to the sum of the weights. This, together with (7.3.7) and (7.3.6) give

$$\begin{aligned}
\frac{1}{d}\sum_{x \in V_v} w_{\mathcal{F}}(x) &\geqslant \frac{1}{d}\left(|\mathcal{F}[V_v] \smallsetminus \{\varnothing\}| + |C|\frac{c-1}{d-c}\right) \\
&\geqslant \frac{1}{d}\left(2^d - 2c + 1 + (d-c+1)\frac{c-1}{d-c}\right) \\
&\geqslant \frac{2^d - c}{d}\,.
\end{aligned} \tag{7.3.8}$$

Lastly consider vertices from $\bar{\mathcal{L}}$. Recall that in particular, these vertices are light and could potentially have a too low weight if the uniform weight would be used. Note that by the maximality of $\mathcal{L}$, for every vertex $a \in \bar{\mathcal{L}}$ we can pick a $v(a) \in \mathcal{L}_2$ such that there exists an edge containing a and a vertex of $V_{v(a)}$. Since the vertices in $\bigcup_{v \in \mathcal{L}_2} V_v$ renounced part of their share of some of those edges, the vertices in $\bar{\mathcal{L}}$ can be given a larger fraction. To be precise, the weight for $a \in \bar{\mathcal{L}}$ is defined as

$$w_{\mathcal{F}}(a) = \sum_{H \in L_a} \frac{1}{|H|+1} + \left|V_a \cap V_{v(a)}\right|\left(\frac{1}{2} - \frac{c-1}{d-c}\right)\,.$$

Lemma 7.3.1 (1) yields that

$$w_{\mathcal{F}}(a) \geqslant W(2^{d-1} - c + 1) + \frac{1}{2} - \frac{c-1}{d-c} \geqslant W(2^{d-1} - c + 1) + \frac{1}{6} \geqslant \frac{2^d - c}{d}\,, \tag{7.3.9}$$

where the second inequality follows from $d \geqslant 4c$ and the third follows as in (7.3.4). Observe that the definition of $w_{\mathcal{F}}$ implies $\sum_{x\in V} w_{\mathcal{F}}(x) \leqslant 1 + |\mathcal{F}|$ because the left-hand side counts every edge of $\mathcal{F}$ apart from the empty set at most once. Since (7.3.4), (7.3.5), (7.3.8), and (7.3.9) say that the average weight per vertex in $\mathcal{F}$ is at least $\frac{2^d-c}{d}$, the proof is complete. $\square$

7.4 Proof of Theorem 7.1.5

This section is dedicated to the proof of Theorem 7.1.5. The proof is very similar to the proof of the main theorem just with some adaptions to obtain more precise bounds at certain points. Hence, we will omit some details that already appeared in the last section.

Proof of Theorem 7.1.5. Let $c \in \{3,4\}$ and $d \geqslant 5$. Firstly, the family $\mathcal{F}_0$ from the proof of Theorem 7.1.3 shows that for $m = \frac{n}{d}(2^d - c) + 1$, we have $(n, m) \nrightarrow \big(n-1, m-(2^{d-1}-c)\big)$.

Let now $\mathcal{F} \subseteq 2^V$ be a hereditary family on some n-set V in which every vertex has degree at least $2^{d-1} - c + 1$. In the following we will show that $|\mathcal{F}| \geqslant \big(2^d - c\big)\,\frac{n}{d} + 1$.

To gain more precision later, this time we call a vertex $v \in V$ *light* if $L_v \cong \mathcal{R}(2^{d-1} - (c-1))$. Again, let $\mathcal{L}$ be a maximum set of light vertices such that $V_v \cap V_{v'} = \varnothing$ for all $v, v' \in \mathcal{L}$. Call the vertices $u \in V \smallsetminus \bigcup_{v\in\mathcal{L}} V_v$ with $L_u \ncong \mathcal{R}(2^{d-1} - (c-1))$ *heavy* vertices. The sets $\mathcal{L}_i$, $\mathfrak{H}$, $\bar{\mathcal{L}}$ are defined similarly as in the proof of Theorem 7.1.3, just according to the different definitions of light and heavy vertices here.

Again we assign the uniform weight to every heavy vertex of $\mathcal{F}$. Note that then, due to Lemma 7.3.1 (3) and the structure of $\mathcal{R}(2^{d-1} - (c-1))$ for $c \leqslant 4$, every heavy vertex has weight at least

$$\begin{aligned}&\min\left(\frac{1}{6}, \frac{1}{d-1}\right) + W(2^{d-1} - (c-1))\\ \geqslant &\min\left(\frac{1}{6}, \frac{1}{d-1}\right) + \frac{2^d-1}{d} - \frac{(c-1)d-1}{(d-1)d} \geqslant \frac{2^d-c}{d}\,. \end{aligned} \tag{7.4.1}$$

For $v \in \mathcal{L}_1$ and $x \in V_v$ the weight is again defined as the uniform weight and as in the proof of Theorem 7.1.3, we obtain

$$\frac{1}{d}\sum_{x\in V_v} w_{\mathcal{F}}(x) \geqslant \frac{2^d-c}{d}\,. \tag{7.4.2}$$

To write the next weight definitions in a compact way, we define the following set

$$\mathcal{S} = \Big\{H \in \mathcal{F} : |H| = 3 \text{ and } H \cap \bigcup_{v \in \mathcal{L}_2} V_v \neq \varnothing, H \cap \bar{\mathcal{L}} \neq \varnothing\Big\}$$

Note that $\mathcal{S}$ is the set of those edges of size 3 in $\mathcal{F}$ crossing from the inside of some V_v with $v \in \mathcal{L}_2$ to its outside and contain a vertex from $\bar{\mathcal{L}}$. For $v \in \mathcal{L}_2$ and a vertex $x \in V_v$, assign the weight $w_{\mathcal{F}}(x) = \sum_{H \in L_x} \frac{1}{|H|+1} - \frac{1}{9}|\{H \in L_x : H \cup \{x\} \in \mathcal{S}\}|$.

Claim 7.4.1. *For $v \in \mathcal{L}_2$ we have $\frac{1}{d}\sum_{x \in V_v} w_{\mathcal{F}}(x) \geqslant \frac{2^d - c}{d}$.*

We postpone the proof of this claim to the end of the section and first finish the proof of Theorem 7.1.5 using the claim.

For a vertex define the weight $a \in \bar{\mathcal{L}}$ as $w_{\mathcal{F}}(a) = \sum_{H \in L_a} \frac{1}{|H|+1} + \frac{1}{18}|\{H \in L_a : H \cup \{a\} \in \mathcal{S}\}|$. Note that by the maximality of $\mathcal{L}$, there exists a $v(a) \in \mathcal{L}_2$ such that there are an edge F and a vertex $x_a \in V_{v(a)}$ with $a, x_a \in F$. In fact, it is easy to check that since $L_a \cong \mathcal{R}(2^{d-1} - (c-1))$, the number of 2-sets in L_a that contain x_a is at least $d-2 \geqslant 3$. Thus, Lemma 7.3.1 (1) and the definition of the weight yield

$$w_{\mathcal{F}}(a) \geqslant W(2^{d-1} - (c-1)) + \frac{d-2}{18} \geqslant W(2^{d-1} - (c-1)) + \frac{1}{6} \geqslant \frac{2^d - c}{d}\,, \tag{7.4.3}$$

where the last inequality follows similarly as in (7.4.1).

Now observe that the definition of $w_{\mathcal{F}}$ implies $\sum_{x \in V} w_{\mathcal{F}}(x) \leqslant 1 + |\mathcal{F}|$ because the left-hand side counts every edge of $\mathcal{F}$ apart from the empty set at most once. In particular, for $H \in \mathcal{S}$ there are at least one $x \in H \cap \bigcup_{v \in \mathcal{L}_2} V_v$ and at most two $a, a' \in H \cap \bar{\mathcal{L}}$. Thus, H contributes at most 1 to $\sum_{x \in V} w_{\mathcal{F}}(x)$.

Since (7.4.1), (7.4.2), Claim 7.4.1, and (7.4.3) say that the average weight per vertex in $\mathcal{F}$ is at least $\frac{2^d - c}{d}$, the proof is complete. $\square$

Proof of Claim 7.4.1. Here, we will differ slightly depending on the value of c.

Case $c = 3$: If $\delta(\mathcal{F}[V_v]) \geqslant 2^{d-1} - 2$, then (7.4.2) holds for v as well and so we may assume $\delta(\mathcal{F}[V_v]) < 2^{d-1} - 2$ and thereby $|2^{V_v} \smallsetminus \mathcal{F}| \geqslant 3$. On the other hand, since $\deg(v) \geqslant 2^{d-1} - 2$ and $\mathcal{F}$ is hereditary, $|2^{V_v} \smallsetminus \mathcal{F}| \leqslant 4$. So we can assume that $|2^{V_v} \smallsetminus \mathcal{F}| \in \{3, 4\}$. If $|2^{V_v} \smallsetminus \mathcal{F}| = 3$, then $\deg(v) \geqslant 2^{d-1} - 2$ and $\mathcal{F}$ being hereditary imply that the sets in $2^{V_v} \smallsetminus \mathcal{F}$ are V_v, $V_v \smallsetminus \{v\}$, and some $A \in V_v^{(d-1)}$ with $v \in A$. Thus, each vertex $x \in A \smallsetminus \{v\}$ lies in all three sets of $2^{V_v} \smallsetminus \mathcal{F}$, and so there has to be an $F_x \in L_x \cap (V \smallsetminus V_v)^{(1)}$ because of the minimum degree of $\mathcal{F}$. Thus, the definition of the weight and double counting the non-empty edges in $\mathcal{F}[V_v]$ implies

$$\sum_{x\in V_v} w_{\mathcal{F}}(x) \geqslant |\mathcal{F}[V_v] \setminus \{\varnothing\}| + \frac{|A \setminus \{v\}|}{2} \geqslant 2^d - 4 + \frac{d-2}{2} \geqslant 2^d - 3.$$

Similarly, if $|2^{V_v} \setminus \mathcal{F}| = 4$, then the sets in $2^{V_v} \setminus \mathcal{F}$ are V_v, $V_v \setminus \{v\}$, some $A \in V_v^{(d-1)}$ with $v \in A$, and $A \setminus \{v\}$. Hence, there are $d-2$ vertices x (namely, the vertices in $A \setminus \{v\}$) for which there has to be an $F_x \in L_x \cap (V \setminus V_v)^{(1)}$ and at least one further $F'_x \in L_x$ with $F'_x \cap (V \setminus V_v) \neq \varnothing$ and $|F'_x| \leqslant 2$. Noting that each F_x contributes $1/2$ to $\sum_{x\in V_v} w_{\mathcal{F}}(x)$ and each F'_x at least $1/3 - 1/9 = 2/9$, we obtain in the usual way

$$\sum_{x\in V_v} w_{\mathcal{F}}(x) \geqslant 2^d - 5 + \frac{d-2}{2} + \frac{2(d-2)}{9} \geqslant 2^d - 3$$

and thereby the claim if $c = 3$.

Case $c = 4$: In a similar way as in the beginning of the case $c = 3$, we observe that we may assume $|2^{V_v} \setminus \mathcal{F}| \in \{4, 5, 6\}$. Further observe that if $|2^{V_v} \setminus \mathcal{F}| = 4$, then since $\deg(v) = 2^{d-1} - 3$, the sets in $2^{V_v} \setminus \mathcal{F}$ are V_v, $V_v \setminus \{v\}$, A, and B for some distinct $A, B \in V_v^{(d-1)}$ which both contain v. Thus, there are at least $d-3$ vertices (namely those in $A \cap B \setminus \{v\}$) that lie in four sets of $2^{V_v} \setminus \mathcal{F}$. Since for any such vertex x there has to be an $F_x \in L_x \cap (V \setminus V_v)^{(1)}$, we get $\sum_{x\in V_v} w_{\mathcal{F}}(x) \geqslant 2^d - 5 + \frac{d-3}{2} \geqslant 2^d - 4$.

Similarly, if $|2^{V_v} \setminus \mathcal{F}| = 5$, the sets in $2^{V_v} \setminus \mathcal{F}$ are V_v, $V_v \setminus \{v\}$, A, B, and $A \setminus \{v\}$ for some distinct $A, B \in V_v^{(d-1)}$ which both contain v. Hence, for the $d-3$ vertices $x \in A \cap B \setminus \{v\}$ there have to be an $F_x \in L_x \cap (V \setminus V_v)^{(1)}$ and at least one further $F'_x \in L_x$ with $F'_x \cap (V \setminus V_v) \neq \varnothing$ and $|F'_x| \leqslant 2$. In addition, for the one vertex $x \in A \setminus B$ there has to be an $F_x \in L_x \cap (V \setminus V_v)^{(1)}$. For a vertex $x \in A \cap B \setminus \{v\}$ we observe the following. If $F'_x \cup \{x\} \notin \mathcal{S}$, then F'_x contributes at least $1/3$ to $\sum_{x\in V_v} w_{\mathcal{F}}(x)$. On the other hand, if $F'_x \cup \{x\} \in \mathcal{S}$, then there is some $a \in \bar{\mathcal{L}}$ with $a \in F'_x$. Since for any $a \in \bar{\mathcal{L}}$ we have $L_a \cong \mathcal{R}(2^{d-1} - 3)$ (and $d \geqslant 5$), the number of 2-sets in L_a which contain x is at least $d - 2 \geqslant 3$. So in this case the edges in $\{H \in L_x : H \cup \{x\} \in \mathcal{S}\}$ contribute at least $\frac{2}{9} \cdot 3 = 2/3$. In either case, we derive

$$\sum_{x\in V_v} w_{\mathcal{F}}(x) \geqslant 2^d - 6 + \frac{d-2}{2} + \frac{d-3}{3} \geqslant 2^d - 4\,.$$

Lastly, if $|2^{V_v} \setminus \mathcal{F}| = 6$, then the sets in $2^{V_v} \setminus \mathcal{F}$ are V_v, $V_v \setminus \{v\}$, A, B, $A \setminus \{v\}$, and $B \setminus \{v\}$ for some distinct $A, B \in V_v^{(d-1)}$ which both contain v. Thus, for the $d - 3$

vertices $x \in A \cap B \setminus \{v\}$ there is an $F_x \in L_x \cap (V \setminus V_v)^{(1)}$ and at least two further $F_x^i \in L_x$ with $F_x^i \cap (V \setminus V_v) \neq \varnothing$ and $|F_x^i| \leqslant 2$, $i \in [2]$. In addition, there are two further vertices $x \in A \triangle B$ for which there is at least one $F_x \in L_x \cap (V \setminus V_v)^{(1)}$. For a vertex $x \in A \cap B \setminus \{v\}$ we observe the following. If $F_x^i \cup \{x\} \notin \mathcal{S}$ for $i = 1, 2$, then these two edges together contribute at least 2/3 to $\sum_{x \in V_v} w_{\mathcal{F}}(x)$. If $F_x^i \cup \{x\} \in \mathcal{S}$ for some $i \in \{1, 2\}$, then the edges in $\{H \in L_x : H \cup \{x\} \in \mathcal{S}\}$ contribute at least 2/3 as noted above. Therefore the definition of the weight entails

$$\sum_{x \in V_v} w_{\mathcal{F}}(x) \geqslant 2^d - 7 + \frac{d-1}{2} + \frac{2(d-3)}{3} \geqslant 2^d - 4$$

and thereby the claim is proved if $c = 4$. □

7.5 Further remarks and open problems

Consider $m(s)$ to be the following limit introduced in [43]

$$m(s) := \lim_{n \to \infty} \frac{m(n, s)}{n}.$$

It is not difficult to check that $m(s)$ is well-defined (see [43]). Rephrased by means of this definition, Theorem 7.1.3 implies that for $c \leqslant \frac{d}{4}$ we have that

$$m(2^{d-1} - c) = \frac{2^d - c}{d}. \tag{7.5.1}$$

The first open problem we would like to mention concerns finding a sharp relation between d and c such that (7.5.1) holds. More precisely, finding the maximum integer $c_0(d)$ such that the equality (7.5.1) holds for every $c \leqslant c_0$. In view of Theorem 7.1.3 we have that $c_0(d) \geqslant \lfloor \frac{d}{4} \rfloor$, and below we will give a construction that proves that $c_0(d) \leqslant d$ for $d \geqslant 5$.

Let $\mathcal{F} \subseteq 2^V$ with $|V| = n$ and d be a positive integer such that $d|n$. We say that $\mathcal{F}$ is *d-local* if there exists a partition of V into sets of size d such that every $F \in \mathcal{F}$ is a subset of one of the sets of the partition. Observe that the extremal construction presented in the proof of Theorem 7.1.3 is a d-local hypergraph with minimum degree $2^{d-1} - c + 1$ and with $m(n, 2^{d-1} - c) + 1$ edges. That construction can be generalised in the following way.

Take $d \geqslant 5$ and $c \in [2^{d-2}]$ and set $s = 2^{d-1} - c$, for simplicity let $d|n$. By definition of $m(\cdot, \cdot)$, there is a family $\mathcal{F}$ on a d-set V with $m(d, s) + 1$ edges such that for every $v \in V$ we have that $|\mathcal{F}_{|V \setminus \{v\}}| \leqslant |\mathcal{F}| - s - 1$. Note that we may assume that $\varnothing \in \mathcal{F}$ and take n/d vertex disjoint copies of $\mathcal{F}$. It is easy to see that for the resulting family $\mathcal{F}'$ we have $|\mathcal{F}'_{|V'\smallsetminus\{v\}}| \leqslant |\mathcal{F}'| - s - 1$ for every $v \in V'$ (where V' is the vertex set of $\mathcal{F}'$) and that further $|\mathcal{F}'| = m(d,s)\frac{n}{d} + 1$. This gives the following general upper bound on $m(2^{d-1} - c)$

$$m(2^{d-1} - c) \leqslant \frac{m(d, 2^{d-1} - c)}{d}\,. \tag{7.5.2}$$

Moreover, we observe that for $c = d+1$ we have that $m(d, 2^{d-1} - (d+1)) < 2^d - (d+1)$. To see this, consider the family $\mathcal{F} \subseteq 2^{[d]}$ containing all sets with at most $d-2$ vertices. Then $\mathcal{F}$ has $2^d - (d+1)$ edges and minimum degree $2^{d-1} - d > 2^{d-1} - c$. Thus, from (7.5.2) it follows that

$$m(2^{d-1} - (d+1)) \leqslant \frac{2^d - (d+2)}{d}\,.$$

This means (7.5.1) does not hold for $c = d+1$, and hence $c_0(d) \leqslant d$.

Note that this construction is also d-local. An interesting problem is to find the values of c for which there are no d-local extremal families.

Problem 7.5.1. *Given a positive integer $d \geqslant 2$, find the minimal $c_\star(d) \in [1, 2^{d-2}]$ such that for all $c \geqslant c_\star$ we have*

$$m(2^{d-1} - c) < \frac{m(d, 2^{d-1} - c)}{d}\,.$$

A solution to this problem would give an insight into the structural behaviour of the extremal families: For $c \geqslant c_\star$ and large n (possibly satisfying certain divisibility conditions) there is no d-local extremal family for $m(n, 2^{d-1} - c)$. Note that the results in [41, 43, 115] solved Problem 7.5.1 for $d \leqslant 4$.

In the following, given a vertex set of size n we describe a non d-local family that has less edges than any possible d-local hereditary family with the same minimum degree. More precisely, the construction below yields that, given $d \geqslant 5$ and $c = d$, we have

$$m(2^{d-1} - d) \leqslant \frac{2^d - d - \frac{1}{2}}{d} < \frac{m(d, 2^{d-1} - d)}{d}\,. \tag{7.5.3}$$

Construction 1. Let $d \geqslant 5$ and k a positive integer, set $n = 2dk$. Take V to be a set of n vertices. Consider $U_1, \ldots, U_{2k}$ to be a partition of V into sets of size d, and for every set U_i arbitrarily pick a vertex $x_i \in U_i$. Define

$$\mathcal{G} = \{S \subseteq V\colon \text{ there is an } i \text{ such that } S \subseteq U_i \text{ and } |S| \leqslant d-2\}$$

$$\mathcal{H} = \{U_i \smallsetminus \{x_i\} \colon \text{for } i \in \{1, 2, \dots, 2k\}\}$$
$$\mathcal{I} = \{\{x_i, x_{i+1}\} \colon \text{ for } i \in \{1, 3, 5, \dots, 2k-1\}\}.$$

One can check that the number of edges of the family $\mathcal{F} = \mathcal{G} \cup \mathcal{H} \cup \mathcal{I}$ is given by

$$|\mathcal{G}| + |\mathcal{H}| + |\mathcal{I}| = \frac{2^d - d - 2}{d} n + 1 + \frac{n}{d} + \frac{n}{2d} = \frac{2^d - d - \frac{1}{2}}{d} n + 1.$$

Moreover, every vertex in V has degree $s = 2^{d-1} - d + 1$. This implies the first inequality of (7.5.3). Taking $d = c + 1$ in Lemma 7.3.2 (1) yields

$$2^d - d \leqslant m(d, 2^{d-1} - d),$$

and thereby the second inequality in (7.5.3).

For $s \leqslant 16$ (that is $d \leqslant 5$), considering the results from [41, 43, 116] and Theorem 7.1.5 all values of $m(s)$ are found, except $m(11)$. We recall the conjecture of Frankl and Watanabe [43], which states that Construction 1 is extremal for $d = 5$.

Conjecture 7.5.2 ([43])**.** $m(11) = 5.3$.

A complementary approach than the one taken in this paper could be as follows.

Problem 7.5.3. *Given a positive integer d and an integer $c \in [0, 2^{d-1})$, find the value of $m(2^{d-1} + c)$.*

Naturally, for $c \geqslant 2^{d-1} - \frac{d}{4}$ Problem 7.5.3 is solved by Theorem 7.1.3. For $c \leqslant 2^{d-2}$, the only general result is given in [43], where it is shown that $m(2^{d-1}) = \frac{2^d - 1}{d} + \frac{1}{2}$. For other values of c Problem 7.5.3 is still open.

Observe that Theorems 7.1.3 and 7.1.5 and the results presented in [41, 43, 115] concern cases in which s is close to 2^d for some value of d. In general, there are still large intervals between powers of 2 for which the only bounds on $m(s)$ that are known are those that follow directly from the previously mentioned results. Finding a solution for Problem 7.5.1 might shed light on this problem by possibly providing a first understanding of the structural behaviour in those intervals.

Bibliography

[1] R. Aharoni, *Ryser's conjecture for tripartite 3-graphs*, Combinatorica **21** (2001), no. 1, 1–4. ↑1.1.2, 5.1, 5.2, 5.3

[2] N. Alon and I. Z. Ruzsa, *Non-averaging subsets and non-vanishing transversals*, J. Combin. Theory Ser. A **86** (1999), no. 1, 1–13. ↑2.2.2, 2.2.2

[3] M. Anthony and P. L. Bartlett, *Neural network learning: theoretical foundations*, Cambridge University Press, Cambridge, 1999. ↑1.3

[4] P. Araújo, S. Piga, and M. Schacht, *Localised codegree conditions for tight Hamilton cycles in 3-uniform hypergraphs* (2020), available at arXiv:2005.11942. Submitted. ↑2.3

[5] D. Bal and A. Frieze, *Packing tight Hamilton cycles in uniform hypergraphs*, SIAM J. Discrete Math. **26** (2012), 435–451. ↑1.1.1, 4.1

[6] D. Bal and L. DeBiasio, *Partitioning random graphs into monochromatic components*, Electron. J. Combin. **24** (2017), no. 1, Paper No. 1.18, 25. ↑1.1.2, 5.1

[7] G. R. Blakley and P. Roy, *A Hölder type inequality for symmetric matrices with nonnegative entries*, Proc. Amer. Math. Soc. **16** (1965), 1244–1245. ↑2.2.2

[8] B. Bollobás, *Modern graph theory*, Graduate Texts in Mathematics, vol. 184, Springer-Verlag, New York, 1998. ↑5.2

[9] ———, *Random graphs*, Second, Cambridge Studies in Advanced Mathematics, vol. 73, Cambridge University Press, Cambridge, 2001. ↑5.2

[10] J. A. Bondy and U. S. R. Murty, *Graph theory*, Graduate Texts in Mathematics, vol. 244, Springer, New York, 2008. ↑5.2

[11] J. A. Bondy, *Induced subsets*, J. Combin. Theory Ser. B **12** (1972), no. 2, 201–202. ↑1.3, 7.1

[12] C. Bowtell and J. Hyde, *A degree sequence strengthening of the vertex degree threshold for a perfect matching in 3-uniform hypergraphs* (2020), available at arXiv:2008.12222. ↑3.1

[13] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University Press, 2004. ↑6.2

[14] W. G. Brown, P. Erdős, and V. T. Sós, *Some extremal problems on r-graphs*, New directions in the theory of graphs (Proc. Third Ann Arbor Conf., Univ. Michigan, Ann Arbor, Mich, 1971), 1973, pp. 53–63. ↑1.3

[15] M. Bucić, D. Korándi, and B. Sudakov, *Covering graphs by monochromatic trees and helly-type results for hypergraphs*, Combinatorica (2021). ↑1.1.2, 12, 5.1, 5.3, 5.4.2, 5.4.2, A.1, A.2

[16] E. Buß, H. Hàn, and M. Schacht, *Minimum vertex degree conditions for loose Hamilton cycles in 3-uniform hypergraphs*, J. Combin. Theory Ser. B **103** (2013), no. 6, 658–678. ↑3.1

[17] F. R. K. Chung, *Regularity lemmas for hypergraphs and quasi-randomness*, Random Structures Algorithms **2** (1991), no. 2, 241–252. ↑3.6

[18] V. Chvátal, *On Hamilton's ideals*, J. Combinatorial Theory Ser. B **12** (1972), 163–168. ↑1.1.1, 2.1.1, 3.1

[19] P. Condon, J. Kim, D. Kühn, and D. Osthus, *A bandwidth theorem for approximate decompositions*, Proc. Lond. Math. Soc. **118** (2019), 1393–1449. ↑4.1

[20] D. Conlon, J. Fox, and B. Sudakov, *An approximate version of Sidorenko's conjecture*, Geom. Funct. Anal. **20** (2010), no. 6, 1354–1366. ↑1.2, 6.1

[21] D. Conlon, J. H. Kim, C. Lee, and J. Lee, *Some advances on Sidorenko's conjecture*, J. London Math. Soc. **98** (2018), no. 3, 593–608. ↑1.2, 6.1

[22] D. Conlon and J. Lee, *Finite reflection groups and graph norms*, Adv. Math. **315** (2017), 130–165. ↑1.2, 6.1

[23] ———, *Sidorenko's conjecture for blow-ups*, Discrete Anal. (2021), Paper No. 2, 13. ↑1.2, 6.1

[24] B. Csaba, D. Kühn, A. Lo, D. Osthus, and A. Treglown, *Proof of the* 1*-factorization and Hamilton decomposition conjectures*, Mem. Amer. Math. Soc. **244** (2016), monograph 1154, 164 pages. ↑1.1.1, 4.1

[25] A. Czygrinow and T. Molla, *Tight codegree condition for the existence of loose Hamilton cycles in 3-graphs*, SIAM J. Discrete Math. **28** (2014), no. 1, 67–76. ↑3.1

[26] R. Diestel, *Graph theory*, Fifth, Graduate Texts in Mathematics, vol. 173, Springer, Berlin, 2017. ↑1, 5.2

[27] G. A. Dirac, *Some theorems on abstract graphs*, Proc. London Math. Soc. (3) **2** (1952), 69–81. ↑1.1.1, 2.1, 3.1

[28] M. Doležal, J. Grebík, J. Hladký, I. Rocha, and V. Rozhoň, *Relating the cut distance and the weak* topology for graphons*, J. Combin. Theory Ser. B **147** (2021), 252–298. ↑6.3.2

[29] S. Ehard, S. Glock, and F. Joos, *Pseudorandom hypergraph matchings*, Combin. Probab. Comput. **29** (2020), 868–885. ↑1.1.1, 4.2, 4.3.5

[30] S. Ehard and F. Joos, *Decompositions of quasirandom hypergraphs into hypergraphs of bounded degree* (2020), available at arXiv:2011.05359. Submitted. ↑1.1.1

[31] P. Erdős, *On extremal problems of graphs and generalized graphs*, Israel J. Math. **2** (1964), 183–190. ↑1.1.1, 2.5.2

[32] P. Erdős, P. Frankl, and V. Rödl, *The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent*, Graphs Combin. **2** (1986), no. 2, 113–121. ↑1.3

[33] P. Erdős, A. Gyárfás, and L. Pyber, *Vertex coverings by monochromatic cycles and trees*, J. Combin. Theory Ser. B **51** (1991), no. 1, 90–95. ↑1.1.2, 1.1.2, 5.1, 5.2

[34] P. Erdős, *Problems and results in combinatorial number theory*, Journees Arithmétiques de Bordeaux (Conf., Univ. Bordeaux, Bordeaux, 1974), 1975, pp. 295–310. Astérisque, Nos. 24–25. ↑1.3

[35] P. Erdős and M. Simonovits, *Supersaturated graphs and hypergraphs*, Combinatorica **3** (1983), no. 2, 181–192. ↑13, 2.5.2, 6.1

[36] A. Ferber, M. Krivelevich, and B. Sudakov, *Counting and packing Hamilton cycles in dense graphs and oriented graphs*, J. Combin. Theory Ser. B **122** (2017), 196–220. ↑4.1, 4.1, 4.8, A.1

[37] A. Ferber, M. Krivelevich, and B. Sudakov, *Counting and packing Hamilton ℓ-cycles in dense hypergraphs*, J. Comb. **7** (2016), no. 1, 135–157. ↑1.1.1, 4.1, 4.1, 4.3.5, 4.3.9

[38] M. Fiedler, *Theory of graphs and its applications*, Proceedings of the Symposium held in Smolenice in June 1963, Publishing House of the Czechoslovak Academy of Sciences, Prague, 1964. ↑2.1.1

[39] M. Fitch, *Rational exponents for hypergraph Turán problems*, J. Comb. **10** (2019), no. 1, 61–86.↑2.2.2

[40] P. Frankl and V. Rödl, *The uniformity lemma for hypergraphs*, Graphs Combin. **8** (1992), no. 4, 309–312. ↑3.6

[41] P. Frankl, *On the trace of finite sets*, J. Combin. Theory Ser. A **34** (1983), no. 1, 41–45. ↑1.3, 7.1, 7.1, 7.2, 7.5, 7.5, 7.5, A.1, A.2

[42] P. Frankl and N. Tokushige, *Extremal problems for finite sets*, Vol. 86, American Mathematical Soc., 2018. ↑1.3, 1.3, 16, 7.1, 1

[43] P. Frankl and M. Watanabe, *Some best possible bounds concerning the traces of finite sets*, Graphs Combin. **10** (1994), no. 2-4, 283–292. ↑1.3, 1.3, 7.1, 7.1, 7.5, 7.5, 7.5, 7.5.2, 7.5, A.1, A.2

[44] D. A. Freedman, *On Tail Probabilities for Martingales*, The Annals of Probability **3** (1975), 100 –118. ↑4.3.3

[45] A. Frieze and M. Krivelevich, *Packing Hamilton cycles in random and pseudo-random hypergraphs*, Random Structures Algorithms **41** (2012), no. 1, 1–22. ↑4.1

[46] Z. Füredi and J. Pach, *Traces of finite sets: extremal problems and geometric applications*, Extremal problems for finite sets **3** (1991), 255–282. ↑1.3, 7.1

[47] F. Garbe, J. Hladký, and J. Lee, *Two remarks on graph norms*, Discrete Comput. Geom. (2021). ↑6.5

[48] D. Gerbner and B. Patkós, *Extremal finite set theory*, Discrete Mathematics and its Applications (Boca Raton), CRC Press, Boca Raton, FL, 2019. ↑1.3

[49] S. Glock, F. Joos, J. Kim, D. Kühn, and D. Osthus, *Resolution of the Oberwolfach problem*, J. Eur. Math. Soc. To appear. ↑4.1

[50] S. Glock, D. Kühn, A. Lo, and D. Osthus, *The existence of designs via iterative absorption: hypergraph F-designs for arbitrary F*, Mem. Amer. Math. Soc. To appear. ↑4.1

[51] S. Glock, D. Kühn, and D. Osthus, *Extremal aspects of graph and hypergraph decomposition problems*, Surveys in combinatorics. To appear. ↑4.1, 4.1

[52] S. Glock, S. Gould, F. Joos, D. Kühn, and D. Osthus, *Counting Hamilton cycles in Dirac hypergraphs*, Combin. Probab. Comput. (2020), 1–23. ↑4.3.4, 4.3.7

[53] A. Gyárfás, *Partícióféfedések és lefogóhalmazok hipergráfokban*, Tanulmányok—MTA Számitástech. Automat. Kutató Int. Budapest **71** (1977), 62 (Hungarian). ↑1.1.2, 1.1.2, 5.1, 5.3, 5.4.2

[54] ______, *Vertex covers by monochromatic pieces—a survey of results and problems*, Discrete Math. **339** (2016), no. 7, 1970–1977. ↑1.1.2

[55] A. Hajnal and E. Szemerédi, *Proof of a conjecture of P. Erdős*, Combinatorial theory and its applications, II (Proc. Colloq., Balatonfüred, 1969), 1970, pp. 601–623. ↑2.1.1

[56] J. Han and Y. Zhao, *Minimum vertex degree threshold for loose Hamilton cycles in 3-uniform hypergraphs*, J. Combin. Theory Ser. B **114** (2015), 70–96. ↑3.1

[57] ______, *Forbidding Hamilton cycles in uniform hypergraphs*, J. Combin. Theory Ser. A **143** (2016), 107–115. ↑1.1.1, 1.1.1, 2.1.1

[58] H. Hatami, *Graph norms and Sidorenko's conjecture*, Israel J. Math. **175** (2010), no. 1, 125–150. ↑1.2, 6.1, 6.1, 1, 6.1, 6.1, 6.4, A.1, A.2

[59] P. E. Haxell and Y. Kohayakawa, *Partitioning by monochromatic trees*, J. Combin. Theory Ser. B **68** (1996), no. 2, 218–222. ↑1.1.2, 5.1

[60] P. E. Haxell and A. D. Scott, *On Ryser's conjecture*, Electron. J. Combin. **19** (2012), no. 1, Paper 23, 10. ↑5.1

[61] J. R. Henderson, *Permutation decompositions of (0,1)-matrices and decomposition transversals*, ProQuest LLC, Ann Arbor, MI, 1971. Thesis (Ph.D.)–California Institute of Technology. ↑1.1.2, 5.1, 5.2

[62] S. Janson, T. Łuczak, and A. Rucinski, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000. ↑3.4.1, 3.4.2, 5.2

[63] F. Joos and M. Kühn, *Fractional cycle decompositions in hypergraphs* (2021), available at arXiv:2101.05526. Submitted. ↑1.1.1, 4.2, 4.3.3, 4.3.4, 4.3.7, 4.3.11

[64] F. Joos, M. Kühn, and B. Schülke, *Decomposing hypergraphs into cycle factors* (2021), available at arXiv:2104.06333. Submitted. ↑1.1.1, 1.3, A.2, A.4

[65] R. M. Karp, *Reducibility among combinatorial problems*, Complexity of computer computations (Proc. Sympos., IBM Thomas J. Watson Res. Center, Yorktown Heights, N.Y., 1972), 1972, pp. 85–103. ↑1.1.1

[66] G. O. H. Katona, *Optimization for order ideals under a weight assignment*, Problèmes Comb, et Théorie des Graphes (1978), 257–258. ↑7.2

[67] G. Y. Katona and H. A. Kierstead, *Hamiltonian chains in hypergraphs*, J. Graph Theory **30** (1999), no. 3, 205–212. ↑1.1.1, 2.1.1, 3.1, A.1, A.2

[68] P. Keevash, *The existence of designs* (2014), available at arXiv:1401.3665. Submitted. ↑4.1

[69] P. Keevash and K. Staden, *Ringel's tree packing conjecture in quasirandom graphs* (2020), available at arXiv:2004.09947. Submitted. ↑4.1, 4.3.4

[70] J. H. Kim, C. Lee, and J. Lee, *Two approaches to Sidorenko's conjecture*, Trans. Amer. Math. Soc. **368** (2016), no. 7, 5057—5074. ↑1.2, 6.1, 6.1, 6.1

[71] Y. Kohayakawa, W. Mendonça, G. O. Mota, and B. Schülke, *Covering 3-edge-colored random graphs with monochromatic trees*, SIAM J. Discrete Math. **35** (2021), no. 2, 1447–1459. ↑1.1.2, 1.3, A.2, A.4

[72] Y. Kohayakawa, G. O. Mota, and M. Schacht, *Monochromatic trees in random graphs*, Math. Proc. Cambridge Philos. Soc. **166** (2019), no. 1, 191–208. ↑1.1.2, 5.1, 5.1, 2, A.4

[73] D. Korándi, F. Mousset, R. Nenadov, N. Škorić, and B. Sudakov, *Monochromatic cycle covers in random graphs*, Random Structures Algorithms **53** (2018), no. 4, 667–691. ↑5.2

[74] D. Král', T. Martins, P. P. Pach, and M. Wrochna, *The step Sidorenko property and non-norming edge-transitive graphs*, J. Combin. Theory Ser. A **162** (2019), 34–54. ↑1.2, 6.1, 6.5

[75] D. Kühn, J. Lapinskas, and D. Osthus, *Optimal packings of Hamilton cycles in graphs of high minimum degree*, Combin. Probab. Comput. **22** (2013), 394–416. ↑4.1

[76] D. Kühn and D. Osthus, *Loose Hamilton cycles in 3-uniform hypergraphs of high minimum degree*, J. Combin. Theory Ser. B **96** (2006), no. 6, 767–821. ↑3.1

[77] R. Lang and N. Sanhueza-Matamala, *Minimum degree conditions for tight Hamilton cycles* (2020), available at arXiv:2005.05291. Submitted. ↑1.1.1, 1.1.1, 2.1.1, 3.1, 4.1, 4.8

[78] J. Lee, *On some graph densities in locally dense graphs*, Random Structures Algorithms **58** (2021), no. 2, 322–344. ↑2.2.2

[79] J. Lee and B. Schülke, *Convex graphon parameters and graph norms*, Isr. J. Math. To appear. ↑1.2, 1.3, A.2, A.4

[80] J. X. Li and B. Szegedy, *On the logarithmic calculus and Sidorenko's conjecture*, Combinatorica. To appear. ↑1.2, 6.1

[81] L. Lovász, *Large networks and graph limits*, American Mathematical Society colloquium publications, American Mathematical Society, 2012. ↑1.2, 1.2, 6.1, 6.2.1, 6.2, 6.2.3, 6.4, A.1, A.2

[82] L. Lovász, *Combinatorial problems and exercises*, second, AMS Chelsea Publishing, Providence, RI, 2007. ↑1.3, 7.1

[83] L. Lovász, *Graph homomorphisms: Open problems*, 2008. manuscript available at http://www. cs. elte. hu/lovasz/problems. pdf. ↑6.1, 6.1.1

[84] L. Lovász and B. Szegedy, *Limits of dense graph sequences*, J. Combin. Theory Ser. B **96** (2006), no. 6, 933 –957. ↑6.1

[85] C. McDiarmid, *On the method of bounded differences*, Surveys in combinatorics, 1989 (Norwich, 1989), 1989, pp. 148–188. ↑3.4.2

[86] R. Montgomery, A. Pokrovskiy, and B. Sudakov, *A proof of Ringel's Conjecture* (2020), available at arXiv:2001.02665. Submitted. ↑4.1

[87] C. St. J. A. Nash-Williams, *Edge-disjoint Hamiltonian circuits in graphs with vertices of large valency*, Studies in Pure Mathematics (Presented to Richard Rado), 1971, pp. 157–183. ↑4.1

[88] S. Piga and N. Sanhueza-Matamala, *Cycle decompositions in* 3*-uniform hypergraphs* (2021), available at arXiv:2101.12205. Submitted. ↑4.8

[89] S. Piga and B. Schülke, *On extremal problems concerning the traces of sets*, J. Combin. Theory Ser. A **182** (2021), 105447, 15. ↑1.3, 1.3, A.1, A.2, A.4

[90] N. Pippenger and J. Spencer, *Asymptotic behaviour of the chromatic index for hypergraphs*, J. Combin. Theory Ser. A **51** (1989), 24–42. ↑4.3.4

[91] J. Polcyn and Chr. Reiher, 2018. Personal communication. ↑1.1.1

[92] J. Polcyn, Chr. Reiher, V. Rödl, A. Ruciński, M. Schacht, and B. Schülke, *Minimum pair degree condition for tight Hamiltonian cycles in 4-uniform hypergraphs*, Acta Math. Hungar. **161** (2020), no. 2, 647–699. ↑1.1.1, 1.1.1, 2.1.1, 2.2.1, 2.2.1, 2.2.3, 2.2.5, 2.2.3, 2.2.3, 2.4, 2.5.1, 2.6, 2.6

[93] J. Polcyn, Chr. Reiher, V. Rödl, and B. Schülke, *On Hamiltonian cycles in hypergraphs with dense link graphs*, J. Combin. Theory Ser. B **150** (2021), 17–75.↑1.1.1, 1.1.1, 1.3, 3.1, 3.3, 4.1, 4.8, A.1, A.2, A.4

[94] L. Pósa, *A theorem concerning Hamilton lines*, Magyar Tud. Akad. Mat. Kutató Int. Közl. **7** (1962), 225–226 (English, with Russian summary). ↑1.1.1, 2.1.1, 3.1, A.1, A.2

[95] Chr. Reiher, V. Rödl, A. Ruciński, M. Schacht, and E. Szemerédi, *Minimum vertex degree condition for tight Hamiltonian cycles in 3-uniform hypergraphs*, Proc. Lond. Math. Soc. (3) **119** (2019), no. 2, 409–439. ↑1.1.1, 1.1.1, 2.1.1, 2.2.1, 2.2.1, 2.2.3, 2.2.3, 2.3, 2.3, 3.1, 3.1, 3.3, 3.3.1, 3.4, 3.7, 4.1

[96] Chr. Reiher, V. Rödl, and M. Schacht, *On a Turán problem in weakly quasirandom 3-uniform hypergraphs*, J. Eur. Math. Soc. (JEMS) **20** (2018), no. 5, 1139–1159. ↑2, 3.6

[97] V. Rödl, *On a packing and covering problem*, European J. Combin. **6** (1985), no. 1, 69–78. ↑4.3.4

[98] V. Rödl, A. Ruciński, and E. Szemerédi, *A Dirac-type theorem for 3-uniform hypergraphs*, Combin. Probab. Comput. **15** (2006), no. 1-2, 229–251. ↑1.1, 1.1.1, 1.1.1, 3.1, 3.1, 3.2, 4.1, 4.2, 4.8, A.1, A.2

[99] ———, *An approximate Dirac-type theorem for k-uniform hypergraphs*, Combinatorica **28** (2008), no. 2, 229–260. ↑1.1.1, 2.1.1, 4.1, 4.2, 4.8, A.1, A.2

[100] ———, *Dirac-type conditions for Hamiltonian paths and cycles in 3-uniform hypergraphs*, Adv. Math. **227** (2011), no. 3, 1225–1299. ↑3.1, 3.7, 4.1, 4.2, 4.8, A.1, A.2

[101] N. Sauer, *On the density of families of sets*, J. Combin. Theory Ser. A **13** (1972), no. 1, 145–147. ↑1.3, 7.1

[102] B. Schülke, *A pair degree condition for Hamiltonian cycles in* 3*-uniform hypergraphs* (2021), available at arXiv:1910.02691. Submitted. ↑1.3, 2.1.1, 4.8, A.1, A.2, A.4

[103] S. Shelah, *A combinatorial problem; stability and order for models and theories in infinitary languages*, Pacific Journal of Mathematics **41** (1972), no. 1, 247–261. ↑1.3, 7.1

[104] A. Sidorenko, *A correlation inequality for bipartite graphs*, Graphs Combin. **9** (1993), no. 2, 201–204. ↑1.2, 1.2, 2.2.2, 6.1

[105] ———, *Inequalities for functionals generated by bipartite graphs*, Discrete Math. Appl. **2** (1993), 489—504. ↑1.2, 1.2, 6.1.2, 6.1

[106] M. Simonovits, *Extremal graph problems, degenerate extremal problems, and supersaturated graphs*, Progress in graph theory (Waterloo, Ont., 1982) (1984), 419–437. ↑2.2.2

[107] V. T. Sós, P. Erdős, and W. G. Brown, *On the existence of triangulated spheres in* 3*-graphs, and related problems*, Period. Math. Hungar. **3** (1973), no. 3-4, 221–228. ↑1.3

[108] K. Staden and A. Treglown, *On degree sequences forcing the square of a Hamilton cycle*, SIAM J. Discrete Math. **31** (2017), no. 1, 383–437. ↑2.1.1, 3.1

[109] A. Steger, *Die Kleitman-Rothschild Methode*, Ph.D. thesis, Forschungsinstitut für Diskrete Mathematik, Rheinische Friedrich-Wilhelms-Universität Bonn **114** (1990), no. 1, 70–96. ↑3.6

[110] B. Szegedy, *An information theoretic approach to Sidorenko's conjecture* (2014), available at arXiv:1406.6738. arXiv:1406.6738. ↑1.2, 6.1

[111] E. Szemerédi, *Is laziness paying off? ("Absorbing" method)* (2013), 17–34. ↑1.1, 1.1.1, 2.1.2

[112] A. Treglown, *A degree sequence Hajnal-Szemerédi theorem*, J. Combin. Theory Ser. B **118** (2016), 13–43. ↑2.1.1, 3.1, 3.6

[113] V. N. Vapnik and A. Ja. Červonenkis, *The uniform convergence of frequencies of the appearance of events to their probabilities*, Dokl. Akad. Nauk SSSR **181** (1968), 781–783. ↑1.3

[114] V. N. Vapnik and A. Y. Chervonenkis, *On the uniform convergence of relative frequencies of events to their probabilities*, Measures of complexity, 2015, pp. 11–30. ↑7.1

[115] M. Watanabe, *Arrow relations on families of finite sets*, Discrete Math. **94** (1991), no. 1, 53–64. ↑1.3, 7.1, 7.5, 7.5

[116] ———, *Some best possible bounds concerning the traces of finite sets II*, Graphs Combin. **11** (1995), no. 3, 293–303. ↑1.3, 7.1, 7.5

[117] R. Wilson and J. J. Watkins (eds.), *Combinatorics: ancient and modern*, Oxford University Press, Oxford, 2013. ↑1

Appendix

English summary

This thesis deals with several problems in extremal combinatorics. In extremal combinatorics, thresholds and extremes in the behaviour of discrete structures like graphs and hypergraphs are studied.

The first part is about Hamiltonicity in hypergraphs. Here, the goal is to generalise results on Hamiltonian cycles in graphs to hypergraphs. In Chapter 2, we follow up on the investigation of minimum degree conditions for Hamiltonian cycles in hypergraphs initiated by Katona and Kierstead [67] and continued in particular by Rödl, Ruciński, and Szemerédi [98–100], who generalised Dirac's theorem to k-uniform hypergraphs and the minimum $(k-1)$-degree. In joint work with Polcyn, Reiher, and Rödl [93], the author proved an analogue of Dirac's theorem for k-uniform hypergraphs and the $(k-2)$-degree, that is, an asymptotically tight condition on $\delta_{k-2}(H)$ guaranteeing the existence of a Hamiltonian cycle.

We then continue by proving a 3-uniform analogue of a result by Pósa [94] in Chapter 3, based on the author's work in [102]. Pósa's theorem states that graphs with certain degree sequences contain a Hamiltonian cycle and it is stronger than Dirac's result since it allows vertices of small degree. Similarly, our result strengthens the result by Rödl, Ruciński, and Szemerédi in [98] by also allowing pairs of vertices of small degree (and even vertices of small degree).

As it turns out, the (asymptotically tight) minimum $(k-1)$-degree condition assumed in [98] and [99] is already enough to guarantee not only one but as many disjoint Hamiltonian cycles in H as there "possibly could be" (due to degree regularity reasons). We proof this and even stronger decomposition type results in Chapter 4, some of which generalise results in graphs due to Ferber, Krivelevich, and Sudakov [36] to hypergraphs.

After this, we turn to a different problem in Chapter 5, in which we mix the search for spanning substructures with a Ramsey-type setting. The problem is to cover edge-coloured random graphs $G(n,p)$ with as few trees as possible and for as a small p as possible. More precisely, we prove that if $p(n) \gg \left(\frac{\log n}{n}\right)^{1/6}$, then with high probability we have that for every 3-colouring of the edges, $G(n,p)$ can be covered with three monochromatic trees. This improves some previous results on this problem due to Bucić, Kórandi, and Sudakov [15].

In Chapter 6, we solve a problem in the theory of graphons, which are limit objects of sequences of large weighted graphs and can be seen as a continuous generalisation of graphs. Lovász [81] and Hatami [58] asked for which graphs H one can define a norm via the homomorphism density; a problem that is closely connected with Sidorenko's conjecture. We show that for certain graphs this is not possible, answering two questions of Hatami [58]. One of these graphs, $K_{5,5} \setminus C_{10}$, is of particular interest for Sidorenko's conjecture.

This thesis ends with a result in extremal set theory. Given a non-negative integer s, we investigate the maximum real $m(s)$ such that every hereditary hypergraph (also called abstract simplicial complex) with "edge density" at most $m(s)$ has minimum (vertex) degree at most s. While previously $m(s)$ was only known for some small values of s and, due to work by Frankl [41] and Frankl and Watanabe [43], for $s = 2^d - c$ with $d \in \mathbb{N}$ and $c \in \{0, 1, 2\}$, Piga and the author [89] determined $m(2^d - c)$ for all $c, d \in \mathbb{N}$ with $c \leqslant d/4$. In addition, we determined $m(s)$ for more small values of s, in particular proving a conjecture by Frankl and Watanabe [43].

German summary (Deutsche Zusammenfassung)

Diese Arbeit behandelt verschiedene Probleme in extremaler Kombinatorik. Extremale Kombinatorik beschäftigt sich mit Schranken und Extremen im Verhalten diskreter Strukturen wie Graphen und Hypergraphen.

Im ersten Teil geht es um Hamiltonkreise in Hypergraphen. Ziel ist es dabei, Resultate über Hamiltonkreise in Graphen zu verallgemeinern. In Kapitel 2 setzen wir Untersuchungen zu minimalen Gradbedingungen für Hamiltonkreise in Hypergraphen fort, die von Katona und Kierstead [67] begonnen und insbesondere von Rödl, Ruciński und Szemerédi [98–100] weitergeführt wurden, indem sie den Satz von Dirac auf k-uniforme Hypergraphen und den minimalen $(k-1)$-Grad verallgemeinerten. In Zusammenarbeit mit Polcyn, Reiher und Rödl [93] bewies der Autor ein Analogon des Satzes von Dirac für k-uniforme Hypergraphen und den $(k-2)$-Grad, das heißt eine asymptotisch scharfe Bedingung an $\delta_{k-2}(H)$, die die Existenz eines Hamiltonkreises garantiert.

In Kapitel 3 zeigen wir basierend auf dem Artikel [102] des Autors eine 3-uniforme Entsprechung des Satzes von Pósa [94], der besagt, dass Graphen mit gewissen Gradsequenzen einen Hamiltonkreis besitzen. Ebenso wie Pósa's Satz den von Dirac verallgemeinert, indem auch Ecken kleinen Grades zugelassen sind, verallgemeinert unser Resultat in Kapitel 3 das von Rödl, Ruciński und Szemerédi in [98], indem wir Eckenpaare (und sogar Ecken) kleinen Grades zulassen.

Wie sich herausstellt, ist die (asymptotisch scharfe) Bedingung an den minimalen $(k-1)$-Grad in [98] und [99] schon genug, um so viele Hamiltonkreise in H zu garantieren, wie es (aus Regularitätsgründen) überhaupt nur geben kann. Dies und noch stärkere Zerlegungsresultate zeigten Joos, Kühn und der Autor in der Arbeit [64], die die Grundlage für Kapitel 4 bildet.

Anschließend wenden wir uns in Kapitel 5 einem etwas anderen Problem zu. Wir kombinieren hier die Suche nach aufspannenden Substrukturen mit Ramseytheorie, indem wir versuchen, kantengefärbte Zufallsgraphen mit so wenig monochromatischen Bäumen wie möglich zu überdecken. Genauer gesagt zeigten Kohayakawa, Mendonça, Mota und der Autor [71], dass wenn $p(n) \gg \left(\frac{\log n}{n}\right)^{1/6}$ gilt, $G(n,p)$ mit hoher Wahrscheinlichkeit für jede 3-Färbung der Kanten mit drei monochromatischen Bäumen überdeckt werden kann. Dies verbessert einige frühere Schranken von Bucić, Kórandi und Sudakov [15].

In Kapitel 6 lösen wir ein Problem in der Theorie der Graphone, die Grenzwertobjekte von gewichteten Graphen sind. Lovász [81] und Hatami [58] fragten für welche Graphen H

man mithilfe der Homomorphismendichte eine Norm definieren kann. Dieses Problem ist eng mit Sidorenko's Vermutung verknüpft. In Zusammenarbeit mit Lee [79] zeigte der Autor, dass dies für bestimmte Graphen nicht der Fall ist, was zwei Fragen von Hatami [58] beantwortet. Einer dieser Graphen, der $K_{5,5} \setminus C_{10}$, steht in Bezug auf Sidorenko's Vermutung besonders im Fokus.

Wir schließen diese Arbeit mit einem Resultat in der extremalen Mengenlehre ab. Gegeben eine nichtnegative ganze Zahl s untersuchen wir das maximale $m(s)$, sodass jeder abstrakte Simplizialkomplex mit "Kantendichte" höchstens $m(s)$ eine Ecke von Grad höchstens s enthält. Während $m(s)$ zuvor nur für einige kleine Werte von s und, durch Arbeiten von Frankl [41] und Frankl und Watanabe [43], für $s = 2^d - c$ mit $d \in \mathbb{N}$ und $c \in \{0, 1, 2\}$ bekannt war, bestimmten Piga und der Autor [89] $m(2^{d-1} - c)$ für alle $c, d \in \mathbb{N}$ mit $c \leqslant d/4$. Außerdem bestimmten wir $m(s)$ für weitere kleine Werte von s, wodurch wir insbesondere eine Vermutung von Frankl und Watanabe lösten.

Publications related to this dissertation

Articles

[1] F. Joos, M. Kühn, and B. Schülke, *Decomposing hypergraphs into cycle factors* (2021), available at arXiv:2104.06333. Submitted. ↑

[2] P. Gupta, Y. Mogge, S. Piga, and B. Schülke, *r-cross t-intersecting families via necessary intersection points* (2020), available at arXiv:2010.11928. Submitted. ↑

[3] B. Schülke, *A pair degree condition for Hamiltonian cycles in* 3*-uniform hypergraphs* (2021), available at arXiv:1910.02691. To appear in Combin. Probab. Comput. ↑

[4] Y. Kohayakawa, W. Mendonça, G. O. Mota, and B. Schülke, *Covering 3-edge-colored random graphs with monochromatic trees*, SIAM J. Discrete Math. **35** (2021), no. 2, 1447–1459, DOI 10.1137/20M137464X. ↑

[5] J. Lee and B. Schülke, *Convex graphon parameters and graph norms*, Israel J. Math. **242** (2021), no. 2, 549–563, DOI 10.1007/s11856-021-2112-6. ↑

[6] J. Polcyn, Chr. Reiher, V. Rödl, and B. Schülke, *On Hamiltonian cycles in hypergraphs with dense link graphs*, J. Combin. Theory Ser. B **150** (2021), 17–75, DOI 10.1016/j.jctb.2021.04.001. ↑

[7] S. Piga and B. Schülke, *On extremal problems concerning the traces of sets*, J. Combin. Theory Ser. A **182** (2021), 105447, 15, DOI 10.1016/j.jcta.2021.105447. ↑

[8] J. Bellmann and B. Schülke, *Short proof that Kneser graphs are Hamiltonian for $n \geqslant 4k$*, Discrete Math. **344** (2021), no. 7, 112430, 2, DOI 10.1016/j.disc.2021.112430. ↑

[9] O. Ebsen, G. S. Maesaka, Chr. Reiher, M. Schacht, and B. Schülke, *Embedding spanning subgraphs in uniformly dense and inseparable graphs*, Random Structures Algorithms **57** (2020), no. 4, 1077–1096, DOI 10.1002/rsa.20957. ↑

[10] J. Polcyn, Chr. Reiher, V. Rödl, A. Ruciński, M. Schacht, and B. Schülke, *Minimum pair degree condition for tight Hamiltonian cycles in 4-uniform hypergraphs*, Acta Math. Hungar. **161** (2020), no. 2, 647–699, DOI 10.1007/s10474-020-01078-7. ↑

Extended abstracts

[1] P. Gupta, Y. Mogge, S. Piga, and B. Schülke, *Maximum size of r-cross t-intersecting families*, Procedia Computer Science **195** (2021), 453-458, DOI 10.1016/j.procs.2021.11.055. Proceedings of the XI Latin and American Algorithms, Graphs and Optimization Symposium. ↑

[2] S. Piga and B. Schülke, *On extremal problems concerning the traces of sets*, Extended Abstracts EuroComb 2021, Trends in Mathematics, 2021, DOI 10.1007/978-3-030-83823-2_104. ↑

[3] Y. Kohayakawa, W. Mendonça, G. Mota, and B. Schülke, *Covering 3-coloured random graphs with monochromatic trees*, Acta Math. Univ. Comenian. (N.S.) **88** (2019), no. 3, 871–875. ↑

[4] O. Ebsen, G. S. Maesaka, Chr. Reiher, M. Schacht, and B. Schülke, *Powers of Hamiltonian cycles in μ-inseparable graphs*, Acta Math. Univ. Comenian. (N.S.) **88** (2019), no. 3, 637–641. ↑

[5] Chr. Reiher, V. Rödl, A. Ruciński, M. Schacht, and B. Schülke, *Minimum pair-degree for tight Hamiltonian cycles in 4-uniform hypergraphs*, Acta Math. Univ. Comenian. (N.S.) **88** (2019), no. 3, 1023–1027. ↑

Declaration of contributions

This thesis is based on work I did with several co-authors and all chapters have benefited from the discussions with them.

Chapter 2 is essentially the article [93], which is joint work with Joanna Polcyn, Christian Reiher, and Vojtěch Rödl. Vojtěch Rödl has been a driving force behind the progress on hypergraph generalisations of Dirac's theorem for more than 15 years. Christian Reiher gave us notes on the the proof of the main result and Joanna Polcyn and I worked out the details. All of us were involved in proofreading the article.

Chapter 3 is essentially the article [102], which is a continuation of my master thesis, for which Christian Reiher suggested me to work towards a hypergraph generalisation of Chvátal's theorem on Hamiltonian cycles in graphs.

Chapter 4 is essentially the article [64], which is joint work with Felix Joos and Marcus Kühn. Felix Joos suggested the basic problem and the general proof strategy. Subsequently, all of us discussed various strengthenings which are now our main theorems and their proofs as well as occurring problems and their fixes. Then Marcus Kühn and I worked out the details in discussion with each other and Felix Joos. All of us were involved in proofreading.

Chapter 5 is essentially the article [71], which is joint work with Yoshiharu Kohayakawa, Walner Mendonça, and Guilherme Mota. The problem was suggested by Yoshiharu Kohayakawa and he and Guilherme Mota told us about previous work on a related problem, namely [72]. Walner Mendonça and Guilherme Mota turned a draft of the proof into an article. All of us were involved in finalising and proofreading the article.

Chapter 6 is essentially the article [79] which is joint work with Joonkyung Lee. He introduced me to graphons, suggested the problem and that it might be possible to use computer calculations to find an example of a graph "at which convexity fails". After the example was found, we simplified it and worked out a proof that does not rely on computer calculations in discussion. Joonkyung Lee had a larger share of writing up the work, both of us were involved in finalising and proofreading the article.

Chapter 7 is essentially the article [89] which is joint work with Simón Piga.

For the introduction, I have in parts also used extended abstracts of some of the aforementioned articles.

Acknowledgements

First of all, I would like to thank my advisor Christian Reiher. He is not only an outstanding researcher but also a strikingly good teacher, who can convey both deep understanding of and a scintillating fascination for whichever topic at hand. I deeply appreciate all of his advice, on mathematics and beyond.

Further, I would like to thank Mathias Schacht, who always took extraordinarily great care of me in many respects during the years of my studies.

I am also grateful to all of my co-authors, from whom and with whom I could enjoy learning and thinking about mathematics and life in general. In addition, I would like to thank the whole discrete mathematics group at the University of Hamburg for creating a nice and friendly environment to learn and think. I also thank the German-Israeli Foundation for Scientific Research and Development, who supported me financially.

Moreover, I am thankful to all of my wonderful friends, who are like a family to me and always bring me peace, joy, and excitement.

I am grateful to my parents Karin Steinhardt and Michael Schülke, who raised me with care and effort, made me curious for the world, and left me all the freedom to explore it.

Lastly, I would like to thank my brother Rasmus Schülke, who has always protected, taught, and inspired me in more ways than I can comprehend.

Declaration of academic honesty

I hereby declare on oath, that I have written the present thesis on my own and have not used other than the acknowledged resources and aids - especially no uncited internet sources - and that I have not handed in this thesis in another examination procedure. The submitted written version corresponds to the version on the electronic storage medium.
Die vorliegende Arbeit habe ich selbständig verfasst und keine anderen als die angegebenen Hilfsmittel - insbesondere keine im Quellenverzeichnis nicht benannten Internet-Quellen - benutzt. Die Arbeit habe ich vorher nicht in einem anderen Prüfungsverfahren eingereicht. Die eingereichte schriftliche Fassung entspricht genau der auf dem elektronischen Speichermedium.

Bjarne Schülke
Hamburg, Oktober 20, 2021