

EXTREMAL GRAPH THEORY VIA STRUCTURAL ANALYSIS

by

FREDERIK GARBE

A thesis submitted to
The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

School of Mathematics
The University of Birmingham
April, 2018

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

ABSTRACT

In this thesis we discuss two extremal problems in graph theory which can be solved by investigating the structure of extremal constructions and minimal counterexamples. In the first part we establish a precise characterisation of 4-uniform hypergraphs with minimum codegree close to $n/2$ which contain a Hamilton 2-cycle. As a corollary we determine the exact Dirac threshold for Hamilton 2-cycles in 4-uniform hypergraphs, and we provide a polynomial-time algorithm which, given a 4-uniform hypergraph H with minimum codegree close to $n/2$, either finds a Hamilton 2-cycle in H or provides a certificate that no such cycle exists. In contrast we also show that the corresponding decision problem for tight Hamilton cycles in k -graphs with minimum degree close to $n/2$ is NP -complete.

In the second part we study the following bootstrap percolation process: given a connected graph G , we *infect* an initial set $A \subseteq V(G)$, and in each step a vertex v becomes infected if at least a ρ -proportion of its neighbours are infected. Once infected, a vertex remains infected forever. A set A which infects the whole graph is called a *contagious set*. We denote the minimal size of a contagious set by $h_\rho(G)$. Our main result states that for every $\rho \in (0, 1]$, every connected graph G on n vertices has $h_\rho(G) < 2\rho n$ or $h_\rho(G) = 1$. This improves previous results of Chang, Chang and Lyuu, and Gentner and Rautenbach and is best-possible. We also provide a stronger bound in the case of graphs of girth at least five. Both proofs exploit the structure of a minimal counterexample.

ACKNOWLEDGEMENTS

First of all I am indebted to Richard Mycroft for his always truly kind supervision during my PhD programme. Furthermore, I want to thank the Combinatorics group in Birmingham for the cordial working environment. Special thanks also to my fellow PhD students and in particular my office mates for the stimulating coffee breaks.

CONTENTS

1	Introduction	1
1.1	Characterising hypergraphs admitting spanning structures	2
1.1.1	Perfect matchings	5
1.1.2	Hamilton cycles	7
1.1.3	New results	9
1.1.4	Open questions	13
1.1.5	Proof overview	15
1.1.6	Further related results	17
1.2	Bootstrap Percolation	18
1.2.1	r -neighbourhood percolation	19
1.2.2	Degree proportional percolation	19
1.2.3	Critical probability	20
1.2.4	Contagious sets	21
1.2.5	New results	24
1.2.6	Open questions	27
2	Hamilton Cycles below the Dirac Threshold	30
2.1	Organisation of the chapter	30
2.2	A characterisation of dense 4-graphs with no Hamilton 2-cycle	31
2.2.1	Notation	31
2.2.2	Odd-good and even-good bipartitions of 4-graphs	33
2.2.3	The exact Dirac threshold for Hamilton 2-cycles in 4-graphs.	35
2.3	Proof of Theorem 1.11	36
2.3.1	Non-extremal 4-graphs	38
2.3.2	Extremal 4-graphs	40
2.4	Hamilton 2-cycles in 4-graphs: non-extremal case	41
2.4.1	A connecting lemma	41
2.4.2	A reservoir lemma	46
2.4.3	Proof of the absorbing lemma (Lemma 2.6)	52
2.4.4	Proof of the long cycle lemma (Lemma 2.7)	64
2.5	Hamilton 2-cycles in 4-graphs: even-extremal case	70
2.5.1	Hamilton paths of split pairs	71
2.5.2	Proof of Lemma 2.9	76

2.6	Hamilton 2-cycles in 4-graphs: odd-extremal case	85
2.7	An algorithm to find a Hamilton 2-cycle in a dense 4-graph	99
2.8	Tight Hamilton cycles in k -graphs	101
2.9	Algorithmic details for Sections 2.4 and 2.7	104
2.10	Full proofs for the reductions claimed in Section 2.8	113
3	Contagious Sets in Bootstrap Percolation	122
3.1	Proof of Theorem 1.15	122
3.2	Proof of Theorem 1.17.	131
	List of References	134

CHAPTER 1

INTRODUCTION

Since the second half of the 20th century, extremal graph theory has been a thriving field of combinatorics that has often addressed a variant of the following problem. We would like to maximise (or minimise) a graph parameter under the condition that some graph property is fulfilled. For example the historical question, first investigated by Mantel [71], which gave birth to extremal graph theory is the following. What is the maximum number of edges in a triangle-free graph?

Usually, the aim of such an investigation is not just to determine the maximum or minimum value of the graph parameter in question, but to understand the structure of the extremal graph(s) which do have this property. Therefore Turán's Theorem [85] does not just state that a K_{r+1} -free graph can have at most $(r-1)n^2/(2r)$ edges, but also classifies the unique extremal construction which achieves this bound, i.e. the complete r -partite graph with partite classes of size n/r , known since then as the *Turán graph*.

Keeping such an emphasis in mind, in the first part of this thesis we will be interested in minimising a certain notion of degree whilst ensuring the existence of a certain spanning subgraph in a hypergraph. Moreover, we will attempt not just to determine the extremal constructions, but to classify all graphs with a weaker minimum degree condition which do admit the spanning subgraph. In the second part of this thesis we will be interested in

minimising the size of a certain contagious set which has the property that it can infect a graph in a certain bootstrap percolation process. In this investigation the main idea of our approach will be to determine the structure of a minimal counterexample which we then can exploit to find a small contagious set.

The new results in the first part are due to joint work with Richard Mycroft, and Section 1.1 and Chapter 2 are based on [40, 41]. The second part, which consists of Section 1.2 and Chapter 3, is based on [39] which is joint work with Andrew McDowell and Richard Mycroft.

1.1 Characterising hypergraphs admitting spanning structures

One of the fundamental motivations in graph theory is the quest for necessary and sufficient conditions which guarantee the existence of certain spanning subgraphs. Traditional subgraphs of particular interest are perfect matchings and Hamilton cycles. The classic theorem of Tutte gives a precise characterisation of all graphs which contain a perfect matching whilst Dirac's theorem [32] gives probably the best-known sufficient condition to ensure the existence of a Hamilton cycle, i.e. if a graph has minimum degree at least $n/2$. A fundamental difference between both problems is that for the corresponding decision problem – given a graph G , determine whether it contains a Hamilton cycle or a perfect matching – exists Edmond's polynomial time algorithm [34] in the case of perfect matchings, but in the case of Hamilton cycles this decision problem is one of Karp's famous 21 NP-complete problems [61]. Moreover, even if restricted to graphs of minimum degree at least $n/2 - \varepsilon n$, for $\varepsilon > 0$, the Hamilton cycle problem is NP-complete [30]. Therefore, assuming that $P \neq NP$, it is not likely that we can find a 'nice' characterisation of Hamiltonian graphs like Tutte found for perfect matchings, not even in the degree-restricted setting. Here we would like to discuss the situation in the case of hypergraphs which has

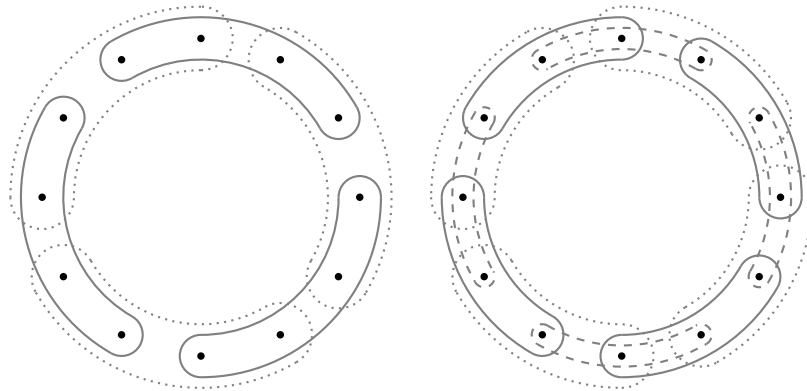


Figure 1.1: A 2-cycle 4-graph and a 2-cycle 3-graph.

received a great amount of attention in recent years. To discuss this work, we first need to find suitable generalisations for those concepts in the context of hypergraphs.

A k -uniform hypergraph, or k -graph H consists of a set of vertices $V(H)$ and a set of edges $E(H)$, where each edge consists of exactly k vertices. This generalises the notion of an ordinary graph, which coincides with the case $k = 2$. Given any integer $1 \leq \ell \leq k - 1$, we say that a k -graph C is an ℓ -cycle if C has at least one edge and the vertices of C may be cyclically ordered in such a way that every edge of C consists of k consecutive vertices and each edge intersects the subsequent edge (in the natural ordering of the edges) in precisely ℓ vertices (see Figure 1.1). Note that the latter condition implies that the number of vertices of an ℓ -cycle k -graph C is divisible by $k - \ell$, as each edge of C contains exactly $k - \ell$ vertices which are not contained in the previous edge. We say that a k -graph H on n vertices is ℓ -Hamiltonian or contains a Hamilton ℓ -cycle if it contains an n -vertex ℓ -cycle as a subgraph; as above, a necessary condition for this is that $k - \ell$ divides n , and we assume this implicitly throughout the following discussion. It is common to refer to $(k - 1)$ -cycles as *tight cycles* and to speak of *tight Hamilton cycles* accordingly. This is the most commonly used definition of a cycle in a uniform hypergraph, but more general definitions, such as a Berge cycle [17], have also been considered. A *perfect matching* of H is a set of n/k vertex disjoint edges. Given a k -graph H and a set $S \subseteq V(H)$, the

degree of S , denoted $d_H(S)$ (or $d(S)$ when H is clear from the context), is the number of edges of H which contain S as a subset. The *minimum codegree* of H , denoted $\delta(H)$, is the minimum of $d(S)$ taken over all sets of $k - 1$ vertices of H , and the *maximum codegree* of H , denoted $\Delta(H)$, is the maximum of $d(S)$ taken over all sets of $k - 1$ vertices of H . Note that for graphs the maximum and minimum codegree are simply the maximum and minimum degree respectively.

A major focus has been to find analogues of Dirac's theorem for hypergraphs and different kinds of spanning subgraphs. This is to determine the best-possible minimum codegree condition which guarantees that a k -graph on n vertices contains the desired subgraph. We refer to this as the corresponding *Dirac threshold*. In this chapter we will investigate the Dirac threshold but we will also go significantly further and ask the following related extremal question. We say that a decision problem is *tractable* if there exists a polynomial time algorithm to decide it.

Question 1.1. *Let S be a k -graph. What is the best-possible minimum codegree condition $f : \mathbb{N} \rightarrow \mathbb{N}$ which makes the decision problem, whether a k -graph H with $\delta(H) \geq f(|V(H)|)$ contains S as a subgraph, tractable?*

We refer to this as the corresponding *tractability threshold*. For the discussion of this threshold we assume that $P \neq NP$ and therefore that a decision problem is not tractable, if it is NP -complete. So if on the one hand we consider a characterisation of a graph property to be *useful*, if it gives rise to a polynomial time algorithm for the corresponding decision problem, and on the other hand assume that $NP \neq P$ and therefore an NP -complete problem does not allow a polynomial time algorithm, then this question asks essentially for the best-possible minimum codegree condition which ensures the existence of a useful characterisation of graphs containing a certain subgraph. In the following we will discuss the situation for the two prime examples of spanning subgraphs, i.e. perfect matchings and Hamilton ℓ -cycles. We will usually assume that n is sufficiently

large. Furthermore, for a given function $f : \mathbb{N} \rightarrow \mathbb{R}$ we denote by $\text{PM}(k, f)$ the decision problem of determining whether a k -graph H with $\delta(H) \geq f(|V(H)|)$ contains a perfect matching and by $\text{HC}_\ell(k, f)$ the decision problem of determining whether a k -graph H with $\delta(H) \geq f(|V(H)|)$ contains a Hamilton ℓ -cycle.

1.1.1 Perfect matchings

Unlike in the graph case, the perfect matching problem in k -uniform hypergraphs, for $k \geq 3$, is known to be NP-complete (see [43]). Furthermore Szymańska [84] showed that for any $\varepsilon > 0$ the problem remains NP-complete even when restricted to k -graphs H with $\delta(H) \geq n/k - \varepsilon n$.

Theorem 1.2 ([84]). *For $k \geq 3$ and $\alpha < 1/k$ the problem $\text{PM}(k, \alpha n)$ is NP-complete.*

On the other hand the Dirac threshold for perfect matchings was determined to be $n/2 - o(1)$, first asymptotically by Kühn and Osthus [70] and later exactly by Rödl, Ruciński and Szemerédi [82].

Theorem 1.3 ([82]). *Suppose that n is sufficiently large and $k \mid n$. If H is a k -graph on n vertices with*

$$\delta(H) \geq n/2 - k + C ,$$

where C is an explicit constant $C \in \{3/2, 2, 5/2, 3\}$ only dependent on the parity of n and k , then H contains a perfect matching. This bound is best-possible.

It is not difficult to find an extremal example with minimum codegree at least $n/2 - k$ which does not contain a perfect matching. We construct a k -graph H whose edges are all k -tuples that have an even size intersection with some set S such that $n/2 - 1 \leq |S| \leq (n + 1)/2$ and $|S|$ is odd. Hence H cannot contain a perfect matching as every matching of H can only cover an even size subset of S . Note that this construction satisfies $\delta(H) \geq n/2 - k + C - 1$ and therefore shows that Theorem 1.3 is best-possible

in the case of k being odd. A very similar construction gives the lower bound in the case of k being even.

Theorem 1.2 and Theorem 1.3 imply that it seems to be unlikely to find a useful characterisation for graphs containing a perfect matching with minimum codegree at least αn , where $\alpha \in [0, 1/k)$, whilst on the other hand every k -graph with minimum codegree at least $n/2$ does contain a perfect matching which leads to a trivial characterisation for those graphs. Hence the natural question is whether we can characterise k -graphs with minimum codegree at least n/k and can close the ‘hardness gap’ of $\text{PM}(k, \alpha n)$ when $\alpha \in [1/k, 1/2)$.

Indeed, based on work of Keevash and Mycroft in [66], Knox, Keevash and Mycroft [64] and Han [52] could give a precise characterisation of such graphs. The algorithmic consequence thus is that the perfect matching problem can be solved in polynomial time in k -graphs H with $\delta(H) \geq n/k$.

Theorem 1.4 ([64, 52]). *For $k \geq 3$, there exists a polynomial time algorithm to decide $\text{PM}(k, n/k)$.*

Note that Knox, Keevash and Mycroft [64] even formulated an algorithm which can find such a perfect matching if it exists and which can be applied to graphs H with $\delta(H) \geq \alpha n$, where $\alpha \in (1/k, 1]$. The underlying characterisation which leads to this algorithm can be described the following way. H contains a perfect matching if and only if it is not a subgraph of an element of a certain class of extremal constructions. We give the precise statement for the 3-uniform case.

Theorem 1.5 ([64]). *For any $\varepsilon > 0$ and n large enough with $3 \mid n$ the following holds. If H is a 3-graph on n vertices and $\delta(H) \geq (1/3 + \varepsilon)n$, then H does not contain a perfect matching if and only if there is a subset $A \subseteq V(H)$ such that $|A|$ is odd but every edge of H intersects A in an even number of vertices.*

This implies that a hypergraph with minimum codegree at least $n/3 + \varepsilon n$ does not contain a perfect matching if and only if it is a subgraph of a construction which is very similar to the one we used above to show that Theorem 1.3 is best-possible. For higher uniformity the characterisation becomes more involved, as more so called *divisibility barriers* can occur. Since we would like to continue with the case of Hamilton cycles, we refer the interested reader to [64] for further details. We can conclude that the hardness of the restricted perfect matching problem $PM(k, \alpha n)$ is known for every $k \geq 3$ and $\alpha \in [0, 1]$ and that the tractability threshold for perfect matchings is considerably lower than the Dirac threshold.

1.1.2 Hamilton cycles

Since the Hamilton cycle problem is already NP -complete for ordinary graphs, an elementary reduction from the graph case can show that it is also NP -complete for k -graphs with $k \geq 3$. Moreover, similarly to the case of perfect matchings Karpiński, Ruciński and Szymańska [62] showed that also the restricted problem $HM_\ell(k, \alpha n)$, for $\alpha \in [0, 1/(\lceil k/(k-\ell) \rceil(k-\ell))]$, is NP -complete. Actually, their statement is formulated only for tight cycles, but the same construction gives the result for ℓ -cycles for any $1 \leq \ell \leq k-1$.

Theorem 1.6 ([62]). *For any $1 \leq \ell < k$ and any $\varepsilon > 0$ the following problem is NP -complete. Given a k -graph H with $\delta(H) \geq (\frac{1}{\lceil \frac{k}{k-\ell} \rceil(k-\ell)} - \varepsilon)|V(H)|$, determine whether H contains a Hamilton ℓ -cycle.*

Especially in recent years the research about the Dirac threshold for Hamilton cycles in hypergraphs has received a great amount of attention. First non-trivial upper and lower bounds on the Dirac threshold were established by Katona and Kierstead [63] for tight cycles in k -graphs with $k \geq 3$. A considerable breakthrough was achieved by Rödl, Ruciński and Szemerédi, who determined the asymptotic Dirac threshold for tight cycles in

3-graphs [80]. In their proof they introduced the since then well known ‘absorbing method’ which had an immense impact on research in extremal graph theory. All subsequent results about the Dirac threshold of Hamilton cycles in hypergraphs are based on this method. They later extended their result and showed that $n/2$ is the asymptotic Dirac threshold for Hamilton ℓ -cycles, for $1 \leq \ell \leq k - 1$ in the case of $k - \ell$ dividing k . The cases in which $k - \ell$ does not divide k were settled in a series of papers by Kühn and Osthus [69], Keevash, Kühn, Mycroft and Osthus [65], Hàn and Schacht [51] and Kühn, Mycroft and Osthus [68]. These results can all be collectively described by the following theorem, whose statement gives the asymptotic Dirac threshold for any $k \geq 3$ and $1 \leq \ell \leq k - 1$.

Theorem 1.7 ([51, 65, 68, 69, 80, 81]). *For any $k \geq 3$, $1 \leq \ell < k$ and $\varepsilon > 0$, there exists n_0 such that if $n \geq n_0$ is divisible by $k - \ell$ and H is a k -graph on n vertices with*

$$\delta(H) \geq \begin{cases} \left(\frac{1}{2} + \varepsilon\right) n & \text{if } k - \ell \text{ divides } k, \\ \left(\frac{1}{\lceil \frac{k}{k-\ell} \rceil (k-\ell)} + \varepsilon\right) n & \text{otherwise,} \end{cases}$$

then H contains a Hamilton ℓ -cycle.

Similar to the case of perfect matchings it is not difficult to see that this result is asymptotically best-possible. In the case that $k - \ell$ divides k the exact same construction like in the case of perfect matchings gives rise to an extremal example, as a Hamilton ℓ -cycle also contains a perfect matching. In the case that $k - \ell$ does not divide k let H be the k -graph whose edges are all k -sets which contain at least one vertex of a set S of size $\lceil n/(\lceil k/(k - \ell) \rceil (k - \ell)) \rceil - 1$. H does not contain a Hamilton ℓ -cycle as such a cycle would contain $n/(k - \ell)$ edges, but a vertex lies in at most $\lceil k/(k - \ell) \rceil$ edges and $n/(k - \ell) > |S| \lceil k/(k - \ell) \rceil$.

More recently, the exact Dirac threshold has been identified in some cases, namely for $k = 3, \ell = 2$ by Rödl, Ruciński and Szemerédi [83], for $k = 3, \ell = 1$ by Czygrinow and

Molla [29], and for any $k \geq 3$ and $\ell < k/2$ by Han and Zhao [54]. Moreover, for tight cycles Karpiński, Ruciński and Szymańska [62] derandomised the proof of Theorem 1.7 to describe a polynomial-time algorithm which actually finds a tight Hamilton cycle in such a k -graph. This raises the question whether the threshold for tractability could lie substantially below the Dirac threshold.

Note that in the case of $(k - \ell)$ not dividing k Theorem 1.6 and Theorem 1.7 imply that the hardness of $\text{HC}_\ell(k, \alpha n)$ is known for every $\alpha \in [0, 1] \setminus \{1/(\lceil k/(k - \ell) \rceil (k - \ell))\}$, as by Theorem 1.6 $\text{HC}_\ell(k, \alpha n)$ is *NP*-complete for $\alpha < 1/(\lceil k/(k - \ell) \rceil (k - \ell))$ and by Theorem 1.7 it is trivial for $\alpha > 1/(\lceil k/(k - \ell) \rceil (k - \ell))$. However, in the case of $(k - \ell)$ dividing k the tractability of $\text{HC}_\ell(k, \alpha n)$ is not known for $\alpha \in [1/k, 1/2]$. For tight cycles Karpiński, Ruciński and Szymańska [62] noticed this ‘hardness gap’ and stressed the similarity to the case of perfect matchings. Can we expect a similar behaviour for tight Hamilton cycles?

Question 1.8. *For $\alpha \in [1/k, 1/2)$, can $\text{HC}_{k-1}(k, \alpha n)$ be decided in polynomial time?*

1.1.3 New results

Our first result closes the aforementioned ‘hardness gap’ identified by Karpiński, Ruciński and Szymańska [62] by answering Question 1.8 negatively.

Theorem 1.9 ([40, 41]). *For any $k \geq 3$ there exists a constant $C \in \mathbb{N}$ such that the problem to determine whether a k -graph H with $\delta(H) \geq |V(H)|/2 - C$ admits a tight Hamilton cycle is *NP*-complete.*

This result can be seen as a hypergraph analogue of the result of Dahlhaus, Hajnal, and Karpiński [30] for ordinary graphs. Furthermore, Theorem 1.9 together with Theorem 1.7 shows that, in contrast to the situation for perfect matchings described in Section 1.1.1, the tractability threshold and the Dirac threshold for tight Hamilton cycles are asymptotically

equal. Since the Dirac and the tractability threshold are asymptotically determined in the case of $k - \ell$ not dividing k and with the above result also in the case of tight cycles, this leaves those values of k and ℓ such that $1 \leq \ell \leq k/2$ and $k - \ell$ divides k . Hence the next interesting case is $\ell = k/2$. We therefore investigated Hamilton 2-cycles in 4-graphs (that is the case $k = 4$ and $\ell = 2$). One result is the previously unknown exact Dirac threshold Hamilton 2-cycles in 4-graphs, the asymptotic bound of Theorem 1.7 is the best previously known result.

Theorem 1.10 ([40, 41]). *There exists n_1 such that if $n \geq n_1$ is even and H is a 4-graph on n vertices with*

$$\delta(H) \geq \begin{cases} \frac{n}{2} - 2 & \text{if } n \text{ is divisible by } 8, \\ \frac{n}{2} - 1 & \text{otherwise,} \end{cases}$$

then H contains a Hamilton 2-cycle. Moreover, this minimum codegree condition is best-possible in each case.

To see that Theorem 1.10 is best-possible, for an even integer n , let A and B be disjoint sets with $|A \cup B| = n$ such that $|A| = n/2 - 1$, if 8 divides n , and $|A| = n/2$, otherwise. We define H to be the 4-graph with vertex set $A \cup B$ whose edges are all 4-sets $e \subseteq A \cup B$ with odd sized intersection with A . We call this the *odd-extremal example*. It is easy to see that $\delta(H) = n/2 - 3$, if 8 divides n , and $n/2 - 2$ otherwise. Furthermore, assume that $C = (v_1, v_2, \dots, v_n)$ is a Hamilton 2-cycle in H . Write $P_i = \{v_{2i-1}, v_{2i}\}$ for each $1 \leq i \leq n/2$, so the edges of C are $e_i := P_i \cup P_{i+1}$ for $1 \leq i \leq n/2$ (with addition taken modulo $n/2$). Note that for exactly $n/4$ pairs P_i we must have that $|P_i \cap A| \in \{0, 2\}$ and for all other P_i we have $|P_i \cap A| = |P_i \cap B| = 1$. Hence we must have $4 \mid n$ and that $|A| - |B| \equiv 2 \cdot n/4 \pmod{4}$. In the case that $8 \mid n$ this leads to the contradiction that $|A| - |B| \equiv 0 \pmod{4}$ and if 8 does not divide n , then we get the contradiction that $|A| - |B| \equiv 2 \pmod{4}$.

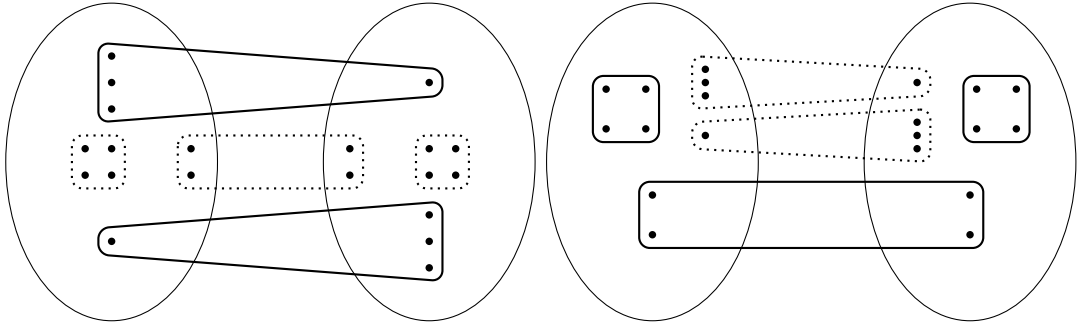


Figure 1.2: A visualisation of odd- and even-extremality.

It is also worth to mention a second construction, which we call the *even-extremal example* and which we have already used to show the lower bound for the Dirac threshold of perfect matchings. Let n be an even integer and H be a 4-graph of order n with vertex set $V(H) = A \cup B$ such that $|A| \neq |B|$ and A is odd. Let the edge set of H consist of all 4-sets of $V(H)$ with an even sized intersection with A . Then H cannot contain a Hamilton 2-cycle, as such a Hamilton 2-cycle can only cover an even number of vertices of A . Note that by choosing $|A|$ to be the largest odd integer such that $|A| < n/2$ we cannot quite, but almost, achieve the bound of Theorem 1.10, i.e. $\delta(H) \geq n/2 - 5$.

However, as mentioned before we are not merely interested in determining the Dirac threshold, but our actual goal is to find a useful characterisation for 2-Hamiltonian 4-graphs with minimum codegree below the Dirac threshold. Surprisingly, and in contrast to the case of tight cycles, we can indeed give a precise characterisation of all 4-graphs with minimum codegree close to the Dirac threshold according to whether or not they contain a Hamilton 2-cycle. This is expressed in the following theorem. Note for this that a bipartition of a set V simply means a partition of V into two sets. The precise definitions of the terms ‘even-good’ and ‘odd-good’ are somewhat technical, so we defer them to Section 2.2.2. For now, loosely speaking, we can interpret a graph H to be ‘even-good’, if by deleting only very few edges it cannot become a subgraph of a construction

which is very similar to the even extremal example. Similarly, a graph H is ‘odd-good’, if by deleting only very few edges it cannot become a subgraph of a construction which is very similar to the odd extremal example. This is depicted in Figure 1.2. The depicted graph only containing edges of the solid type and none of the dotted ones results into an odd-extremal (respectively even-extremal) example. Adding just two (or three) disjoint edges of the dotted type results into the graph being odd-good (or even-good). Recall also that a 4-graph can only contain a Hamilton 2-cycle if it has even order.

Theorem 1.11 ([40, 41]). *There exist $\varepsilon, n_0 > 0$ such that the following statement holds for any even $n \geq n_0$. Let H be a 4-graph on n vertices with $\delta(H) \geq n/2 - \varepsilon n$. Then H admits a Hamilton 2-cycle if and only if every bipartition of $V(H)$ is both even-good and odd-good.*

Theorem 1.10 then actually follows by a straightforward deduction from the stronger Theorem 1.11 (which is given in Section 2.2.3). Furthermore, note that the spirit of Theorem 1.11 is very similar to the one of Theorem 1.5, as both state that H either admits the spanning structure or H is (almost) a subgraph of a graph from a certain class of extremal constructions. Concerning the tractability threshold it is not immediately apparent that the criterion of Theorem 1.11 can be tested in polynomial time, but in Section 2.2.2 we explain why, due to the minimum codegree of H , this is in fact the case. Consequently, we can determine in polynomial time whether or not a 4-graph H whose codegree is close to the Dirac threshold admits a Hamilton 2-cycle. Moreover, by derandomising the proof of Theorem 1.11 we can actually find such a cycle, should one exist, giving the following theorem.

Theorem 1.12 ([41]). *There exist a constant $\varepsilon > 0$ and an algorithm which, given a 4-graph H on n vertices with $\delta(H) \geq n/2 - \varepsilon n$, runs in time $O(n^{32})$ and returns either a Hamilton 2-cycle in H or a certificate that no such cycle exists (that is, a bipartition of $V(H)$ which is either not even-good or not odd-good).*

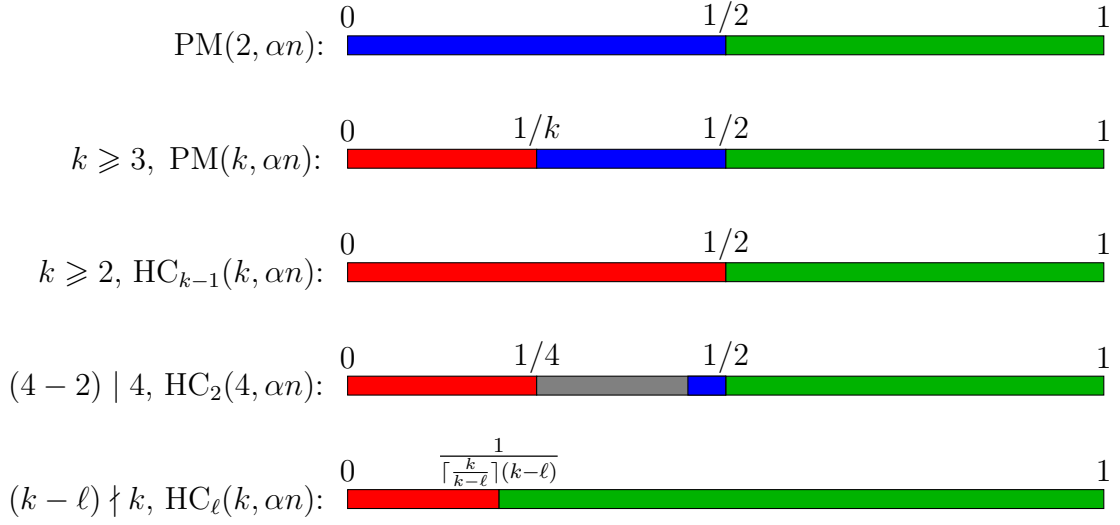


Figure 1.3: The complexity of $\text{PM}(k, \alpha n)$ and $\text{HC}_{\ell}(k, \alpha n)$ for $\alpha \in [0, 1]$. Red means that the problem is NP -complete, blue means that it has a non-trivial polynomial time algorithm and green means that it is above the Dirac threshold.

Therefore Theorem 1.12 and Theorem 1.10 prove the existence of a linear-size gap between the tractability threshold and the Dirac threshold for 2-Hamiltonicity. This provides a surprising contrast to the graph setting and to the case of tight cycles in hypergraphs.

1.1.4 Open questions

Figure 1.3 summarises the known results about the complexity of the decision problems $\text{PM}(k, \alpha n)$ and $\text{HC}_{\ell}(k, \alpha n)$ for $\alpha \in [0, 1]$. We can observe an interesting contrast between the case of tight cycles and Hamilton 2-cycles in 4-graphs. Whilst $\text{HC}_{k-1}(k, \alpha n)$ is NP -complete for $\alpha \in [0, 1/2)$ until it suddenly becomes trivial for $\alpha > 1/2$ we can observe that $\text{HC}_2(4, \alpha n)$ opens up a linear sized gap for a non-trivial polynomial time algorithm for $\alpha \in [1/2 - \varepsilon, 1/2)$ which is related to a precise characterisation of 2-Hamiltonian 4-graphs with minimum codegree at least $n/2 - \varepsilon n$. Compared to the case of perfect matchings this means that tight cycles behave very differently whilst some looser cycles seem to exhibit a

similar behaviour. Furthermore, in the case that $k - \ell$ does not divide k we have that the tractability threshold and the Dirac threshold for $\text{HC}_\ell(k, \alpha n)$ are asymptotically equal; similar to tight cycles. It would be interesting to know, for other values of k and ℓ such that $k - \ell$ does divide k , whether the tractability threshold for a Hamilton ℓ -cycle in a k -graph is essentially equal to the Dirac threshold (as Theorem 1.9 shows is the case for tight cycles), or whether it is significantly different (as Theorem 1.12 shows is the case for 2-cycles in 4-graphs).

Furthermore, we made no attempt to quantify the constant ε in Theorem 1.11 which arises from our proof. However, the question remains how large ε can be chosen such that the characterisation of Theorem 1.11 still holds. We conjecture that the characterisation is still true for $\varepsilon = 1/6$.

Conjecture 1.13 ([41]). *There exists n_0 such that the following statement holds. Let H be a 4-graph on $n \geq n_0$ vertices with $\delta(H) \geq n/3$. Then H admits a Hamilton 2-cycle if and only if every bipartition of $V(H)$ is both even-good and odd-good.*

If true, the minimum codegree condition of Conjecture 1.13 would be essentially best-possible. To see this, fix any n which is divisible by 4 and take disjoint sets X, Y and Z each of size $\frac{n}{3} \pm 1$ such that $|X| \neq |Y|$ and $|X \cup Y \cup Z| = n$. Define H to be the 4-graph on vertex set $V := X \cup Y \cup Z$ whose edges are all sets $S \in \binom{V}{4}$ with $|S \cap X| \equiv |S \cap Y| \pmod{3}$. It is easily checked that we then have $\delta(H) \geq n/3 - 4$ and that every bipartition of $V(H)$ is both even-good and odd-good. However, there is no Hamilton 2-cycle in H . Indeed, since 4 divides n , taking every other edge of such a cycle would give a perfect matching M in H . Since each edge of M covers equally many vertices of X and Y (modulo 3), the same is true of M as a whole, contradicting the fact that $|X| \not\equiv |Y| \pmod{3}$.

If Conjecture 1.13 holds, then by the same argument used to establish Corollary 2.3 we may determine in polynomial time whether a 4-graph H on n vertices with $\delta(H) \geq n/3$ admits a Hamilton 2-cycle. Moreover, we speculate that under the weaker assumption

that $\delta(H) \geq n/4$ it may be possible to prove a similar statement to Conjecture 1.13 which considers partitions of $V(H)$ into three parts as well as into two parts. If so, this would allow us to determine in polynomial time whether H contains a Hamilton 2-cycle under this weaker assumption. Such a result would neatly complement Theorem 1.6, which shows that for any $c < 1/4$ it is NP-complete to determine whether a 4-graph H on n vertices with $\delta(H) \geq cn$ contains a Hamilton 2-cycle. However, our proof of Theorem 1.11 relies extensively on ε being small; it seems that significant new ideas and techniques would be needed to prove Conjecture 1.13 or this proposed extension.

1.1.5 Proof overview

The proof of Theorem 1.11 consists of two directions. The easier one is to show that every 2-Hamiltonian 4-graph is even-good and odd-good. This is shown using elementary arguments in Section 2.3. The main difficulty is to prove the backwards direction which requires most of Chapter 2. The core idea is to split the proof into three cases; two ‘extremal’ and one ‘general’ case. The extremal cases are an ‘even extremal’ case in which the graph is very close to the even-extremal example and an ‘odd extremal’ case in which the graph is very close to the odd-extremal example which we have already seen before. In Section 2.3 we will specify this notion of closeness. In such an extremal case we have, due to the high codegree, a lot of information about the structure of the graph, for example that we have a bipartition such that almost all edges have an even sized intersection with a partite set and that almost all edges of this type are present, and can therefore by some ad hoc methods use the exceptional edges of different type, which are guaranteed by being even-good and odd-good, to find a Hamilton 2-cycle. The precise definition of odd-good and even-good comes into play here. In the general case the proof will follow the so called ‘absorbing method’ whose locus classicus is the article of Rödl, Ruciński and Szemerédi [80] in which they encapsulated the idea of absorbing for

the first time in a standardised methodological approach. Implicitly the idea was already used before, so for example by Erdős, Gyárfás, and Pyber in [36]. This method was so impactful that nowadays a huge number of articles discussing packing problems in graphs and hypergraphs is based on it. The core observation is that by not being even-extremal we can connect almost all pairs via short paths which results into a ‘connecting lemma’ which we show in Section 2.4.1. Furthermore, by not being odd-extremal almost all pairs p of vertices will lie in a short so called ‘absorbing structure’ S , which is a tight path such that $V(S) \cup p$ also forms a tight path in H with the same ends as S . We will show this ‘absorbing lemma’ in Section 2.4.3. This means that, if S is a section of a longer path P , we can simply replace the section S by a path consisting of $V(S) \cup p$ to gain a path P' with vertex set $V(P) \cup p$. We then say that P *absorbs* p . The outline of the general case then is as follows. We first find an ‘absorbing path’ P , this is a path which is not too long and which contains many pairwise disjoint absorbing structures which can absorb any not too large set of $V(H) \setminus V(P)$. We then construct an ‘almost spanning cycle’ C which contains P as a segment and covers almost all vertices. Finally, the absorbing path P can absorb the leftover vertices which results in the desired Hamilton cycle. Obviously, we will need the absorbing lemma to find the absorbing path and we will make use of the connecting lemma repeatedly to connect the absorbing structures or while building the almost spanning cycle. The precise statements of the absorbing lemma and the long cycle lemma are in Section 2.3.1. The straightforward deduction of the Dirac statement Theorem 1.10 from the characterisation Theorem 1.11 can be found in Section 2.2.3. The NP-completeness result of Theorem 1.9 is shown by a series of polynomial time reductions from the Hamilton cycle problem in subcubic graphs, which is known to be NP-complete due to Garey, Johnson and Stockmeyer [44]. We give an overview of the reductions in Section 2.8 while Section 2.10 contains all the details.

1.1.6 Further related results

We would like to note that the Dirac threshold has also been investigated for other types of degree conditions for uniform hypergraphs. Specifically, for a k -graph H and $1 \leq s < k$ we say the *minimum s -degree of H* is the minimum of $d(S)$ taken over all sets of s vertices of H . For $s < k - 1$ much less is known compared to the case of $s = k - 1$ which we discussed before. The exact $(k - 2)$ -degree Dirac threshold for a Hamilton ℓ -cycle in a k -graph for $1 \leq \ell < k/2$ was determined very recently by Bastos, Mota, Schacht, Schnitzer and Schulenburg [14, 15]. This generalises previous results of Han and Zhao [55] and Buß, Hàn and Schacht [19] for 3-graphs. One significant unresolved problem is to determine the asymptotic 1-degree Dirac threshold for a tight Hamilton cycle in a k -graph. For $k = 3$ this problem was solved recently by Reiher, Rödl, Ruciński, Schacht and Szemerédi [76] (partial results were previously given by Rödl and Ruciński [78] and Rödl, Ruciński, Schacht and Szemerédi [79]). For general $k \geq 4$ the problem remains open; upper bounds were established by Glebov, Person and Weps [46]. Also note that Rödl and Ruciński conjectured in [77] that for every s -degree the Dirac threshold for tight Hamilton cycles and the Dirac threshold for perfect matchings are asymptotically equal. Although this conjecture holds in the case of codegree ($s = k - 1$) it was shown to be false in general by Han and Zhao [56]. Lastly, we would like to mention that the tractability threshold was investigated for F -packings in hypergraphs by Han and Treglown [53]. A perfect F -packing of H is a collection of vertex disjoint copies of F in H which together cover all vertices of H . They give a general tool for investigating the complexity of such a packing problem in hypergraphs. In particular, they determine the tractability threshold for the problem of packing k -partite k -graphs and therefore generalise the result of Keevash, Knox and Mycroft [64] and Han [52] concerning perfect matchings, which is the special case of F being a single edge.

1.2 Bootstrap Percolation

Many phenomena in biology, physics and sociology can be described using a model which is called a *cellular automaton*. Those automata are dynamical systems on graphs (often regular grids) whose vertices have one of a finite number of states at a discrete time t . The rules according to which the states of the vertices change from one time step to the next one are local and homogeneous, meaning that the rule only depends on a finite neighbourhood of the considered vertex and that the same rule applies to all vertices. Those systems were first considered by Ulam [86] and von Neumann [87]. The simplest case is the one in which there are only two states for each vertex. A famous example modelling the behaviour of living cells is Conway's 'Game of life' [42]. The underlying graph is the infinite 2-dimensional grid and each vertex has two states, alive and dead. The update rule allows vertices to die or to become alive depending on the number of living neighbours. Another important inspiration comes from statistical physics. The idea is to use a cellular automaton to model a simplification of the Glauber dynamics of the Ising model of ferromagnetism. This was first attempted by Chalupa, Leath, and Reich [24] who introduced a model called *bootstrap percolation*. In this model there are again only two states for each vertex, we call them healthy and infected, but this time only healthy vertices can change their state and become infected while infected vertices remain infected forever. We would like to discuss a related model on finite graphs.

Formally, we define the following more general framework, which is not necessarily homogeneous, for a *bootstrap percolation process*. Let G be a finite graph and $\varphi : V(G) \rightarrow \mathbb{R}_{\geq 0}$ be a function. For a set $A \subseteq V(G)$ we define a set $H_{\varphi, G}(A) \subseteq V(G)$ iteratively in the following way. We set $A_0 := A$ and $A_{i+1} := A_i \cup \{v \in V(G) : |N_G(v) \cap A_i| \geq \varphi(v)\}$. We call $(A_i)_{i \in \mathbb{N}}$ the *infecting sequence* in G induced by A with respect to φ . Note that $A_i \subseteq A_{i+1}$ for every $i \in \mathbb{N}$. Furthermore, if $A_i = A_{i+1}$ for some $i \in \mathbb{N}$, then $A_{i+1} = A_{i+2}$.

So for a finite graph G there must exist $k \in \mathbb{N}$ such that $A_k = A_{k+1} = A_{k+2} = \dots$, and we set $H_{\varphi,G}(A) := A_k$ for this k . For simplicity we will write $H_{\varphi,G}(v)$ in place of $H_{\varphi,G}(\{v\})$ for a vertex $v \in V(G)$. We say that a vertex v *becomes infected at step i* , if $v \in A_i \setminus A_{i-1}$, and for a set $S \subseteq H_{\varphi,G}(A)$ we say that A *infects S* . We now call a set $A \subseteq V(G)$ a *contagious set* for G with respect to φ if $H_{\varphi,G}(A) = V(G)$. The literature also calls infected vertices *active* and it is said that A *percolates*, if A is a contagious set. Sometimes a contagious set is also called a *perfect target set* or an *irreversible dynamic monopoly*; the latter in contrast to the different process of *dynamic monopolies* in which, similarly to the ‘Game of life’, active vertices may become inactive again as investigated in [73] and [16].

1.2.1 r -neighbourhood percolation

The simplest and most widely studied variant is based on the already mentioned idea of Chalupa, Leath, and Reich [24]. For a fixed $r \in \mathbb{N}$, they considered the process in which vertices become infected when at least r of their neighbours are infected, independently of their degree. Since then this case has often been called *r -neighbourhood bootstrap percolation*. Besides its applications in statistical physics, this threshold model was also used to describe problems in sociology, epidemiology, economics and computer science such as the diffusion of information and opinion in social networks, the spread of diseases, and cascading failures in financial or computer networks (see for example [18, 33, 67] and the references therein). In our framework this model is the special case of the constant function $\varphi \equiv r$.

1.2.2 Degree proportional percolation

A related model was inspired by observations in sociology. Granovetter [48] expressed the opinion that it is a natural assumption that an individual may be more likely to be influenced, if a significant proportion of their friends hold a particular view, rather than

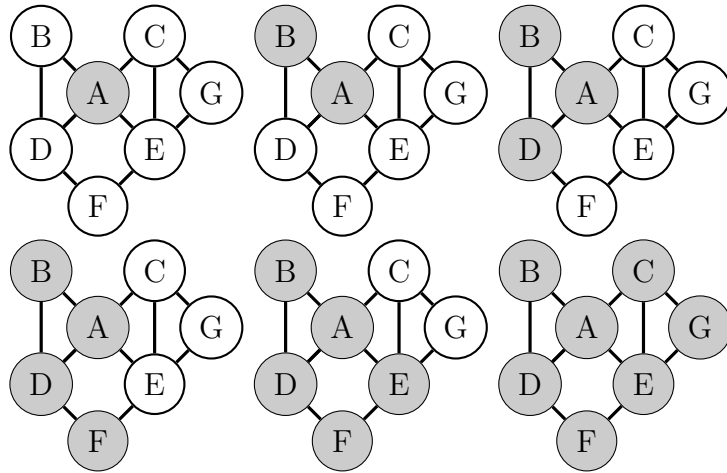


Figure 1.4: *Majority rule* $\varphi(v) = d(v)/2$ with initial set $A_0 = \{A\}$.

simply encountering a fixed number of people who have adopted that position. Sociologists call this process *complex propagation* in contrast to *simple propagation* in which one active neighbour is enough to influence a vertex. For studies investigating this complex model, see for example [21, 20, 88]. Therefore Granovetter suggested a percolation process in which a number of vertices are initially infected, and then, for a fixed constant $\rho \in [0, 1]$, after each time step an uninfected vertex will become infected if at least a ρ -proportion of its neighbours are already infected. Depending on the value of ρ and the choice of initially infected vertices, this process may result in the infection spreading to the entire graph. In our framework this corresponds to the function $\varphi(v) = \rho d_G(v)$. An example of such a process is shown in Figure 1.4.

1.2.3 Critical probability

A key problem in percolation theory, also inspired by statistical physics, is to analyse the behaviour of the bootstrap percolation model for various infinite families of graphs with an initial set A_0 chosen randomly according to a product measure with density p . The aim is to approximate the *critical probability* of the phase transition which is the infimal density p at which the initial random selection is likely to be contagious for a sequence

of graphs with size of the vertex set tending to infinity. In an abundant line of research this problem was investigated for multidimensional grids [2, 12, 22, 23, 57, 8, 10, 50, 7], the hypercube [6], the Hamming torus [49], which is the Hamming graph with vertex set $\{1, \dots, n\}^d$, and various distributions of random graphs [13, 59, 4, 60]. An example of a degree-proportional function φ which was investigated is the so-called *majority bootstrap percolation*, i.e. $\varphi(v) = d_G(v)/2$; for this function the critical probability was studied for the hypercube in [9].

1.2.4 Contagious sets

Another research focus, which we will discuss in the following, is the extremal problem of bounding the minimal size of a contagious set in a bootstrap percolation process. This problem originated from the following exercise in Hungarian folklore, which is taken from [12]. On an $n \times n$ chessboard consisting of white squares there are two states for each square, namely ‘clean’ and ‘weedy’. A clean square stays clean if and only if the majority of its neighbours are clean whilst a weedy square stays weedy forever. What is the minimum number of weedy squares necessary to make the whole chessboard weedy?

It is a nice exercise to show that the answer to this question is n and we notice that this question is actually asking for the minimal size of a contagious set in our framework, where G is the 2-dimensional grid and $\varphi \equiv 2$ is the 2-neighbourhood percolation. In general for a given function φ and a graph G we denote the minimal size of a contagious set by $h_\varphi(G)$. In the case of the r -neighbourhood bootstrap percolation this number was investigated for special families of graphs. Inspired by the above Hungarian folklore exercise Pete [74] formulated this problem for finite multidimensional grids and gave first bounds on the size of a minimal contagious set (see Balogh and Pete [12] for a summary of these results). The case $r = 2$ was settled by Balogh and Bollobás [6]. Recently, exact asymptotics for the hypercube were determined by Morrison and Noel [72]. This proved a conjecture of

Balogh and Bollobás [6] and improved earlier bounds of Balogh, Bollobás and Morris [10]. Additionally, Morrison and Noel also proved bounds for multidimensional rectangular grids. Furthermore, the minimal size of a contagious set has been investigated for expander graphs [27], very dense graphs [38], random graphs [59, 3, 37, 5], with additional Ore and Chvátal-type degree conditions [31], and in the setting of hypergraphs [11].

More generally, Ackermann, Ben-Zwi and Wolfowitz [1] found an elegant random selection argument which leads to an expression which bounds the minimal size of a contagious set for every function φ and every graph G . We choose an order σ of the vertices of $V(G)$ uniformly at random. Let A be the set of vertices $v \in V(G)$ for which fewer than $\lceil \varphi(v) \rceil$ neighbours of v precede v in the order σ . We claim that A is a contagious set for G with respect to φ . Suppose for a contradiction that this is not the case, and let $v \in V(G) \setminus H_{\varphi,G}(A)$ be the first vertex according to the order σ which is not infected by A . Then $v \notin A$, so v has at least $\lceil \varphi(v) \rceil$ neighbours which precede v . The minimality of v in σ implies that all of these neighbours are infected by A , but this contradicts the fact that v is not infected at any time step t . So A is indeed a contagious set for G with respect to φ . Moreover, for any $v \in V(G)$ we have $v \in A$ if and only if v is one of the first $\lceil \varphi(v) \rceil$ members of $\{v\} \cup N(v)$ in the order σ . It follows that $\mathbb{P}[v \in A] = \frac{\lceil \varphi(v) \rceil}{1+d_G(v)}$. Hence the size of a minimal contagious set is bounded by the expected size of A which is

$$h_\varphi(G) \leq E[|A|] = \sum_{v \in V(G)} \frac{\lceil \varphi(v) \rceil}{d_G(v) + 1}. \quad (1.2.1)$$

This bound was also independently derived by Reichmann [75] for the special case of r -neighbourhood bootstrap percolation. Very recently this bound was slightly improved by Cordasco, Gargano, Mecchia, Rescigno and Vaccaro in [28] to

$$h_\varphi(G) \leq \sum_{v \in V'} \min \left\{ 1, \frac{\lceil \varphi(v) \rceil}{d'(v) + 1} \right\}, \quad (1.2.2)$$

where $V' = \{v \in V(G) \mid d_G(v) \geq 2 \text{ or } \varphi(v) \neq 1\}$ and $d'(v) = |N_G(v) \cap V'|$.

In this thesis we focus on general upper bounds on the size of a smallest contagious set in a degree-proportional bootstrap percolation process, i.e. $\varphi_\rho(v) = \rho d_G(v)$, where $\rho \in [0, 1]$ is a fixed constant. Observe that if G is not connected, then a smallest contagious set of G is simply the union of smallest contagious sets, one from each connected component, so we need only consider connected graphs. Throughout this thesis we will often shorten the notation of φ_ρ to simply ρ , so for example we will write $h_\rho(G)$ instead of $h_{\varphi_\rho}(G)$.

For majority bootstrap percolation (where $\rho = 1/2$), expression (1.2.1) shows that $h_{1/2}(G) \leq |V(G)|/2$ for any graph G . This bound is best-possible, as shown by taking G to be a clique of even order. However, for small ρ it is possible to construct graphs with many vertices of small degree such that expression (1.2.1) and also expression (1.2.2) are far from the actual minimal size of a contagious set. For example, for $1/n < \rho \ll 1/3$, let $K_{1,n-1}$ be the star on n vertices with partite sets $\{v\} \cup B$. We add a perfect matching to B and call the resulting graph S . Note that according to the notation of expression (1.2.2) we have $V' = V(S)$ and $d'(v) = d_S(v)$. Hence

$$h_\rho(S) = 1 < \rho n \ll (n-1)/3 < \sum_{v \in V'} \frac{[\rho d_S(v)]}{d'(v) + 1}.$$

For this reason attention has recently focused on providing different kinds of upper bounds which might work better for small values of ρ ; in particular on upper bounds of the form $h_\rho(G) < C\rho n$ for some constant C . Observe that, since every contagious set has size at least one, such bounds must necessarily exempt the case when $h_\rho(G) = 1$ (some previous works have assumed a lower bound on the maximum degree of G or the order of G , but these assumptions are easily seen to be at least as restrictive as our exemption). This leads to the following question.

Question 1.14. *What is the smallest constant C such that, for every $\rho \in (0, 1]$, every*

connected graph G fulfils $h_\rho(G) = 1$ or $h_\rho(G) < C\rho n$?

Chang [25] gave the first bound of the form described above and showed that every connected graph G on n vertices satisfies $h_\rho(G) = 1$ or

$$h_\rho(G) \leq (2\sqrt{2} + 3)\rho n < 5.83\rho n . \quad (1.2.3)$$

Later, Chang and Lyuu [26] improved this bound by showing that every such G satisfies $h_\rho(G) = 1$ or

$$h_\rho(G) \leq 4.92\rho n . \quad (1.2.4)$$

This was the best general bound prior to this work. Very recently Gentner and Rautenbach [45] provided a stronger upper bound under the additional assumptions that G has girth at least five and ρ is sufficiently small. More precisely, they showed that for any $\varepsilon > 0$, if ρ is sufficiently small and G has girth at least five, then $h_\rho(G) = 1$ or

$$h_\rho(G) \leq (2 + \varepsilon)\rho n . \quad (1.2.5)$$

1.2.5 New results

The first main result of this chapter completely answers Question 1.14 and improves on each of the aforementioned results by giving the optimal bound of the described form.

Theorem 1.15 ([39]). *For every $\rho \in (0, 1]$, every connected graph G of order n has $h_\rho(G) = 1$ or*

$$h_\rho(G) < 2\rho n .$$

In particular, for connected graphs of order $n > 1/(2\rho)$ we always have $h_\rho(G) < 2\rho n$. Already Chang and Lyuu [26] observed that the complete graph provides a lower bound of order $O(\rho n)$ on $h_\rho(G)$. Indeed, for $n := \lfloor 1/\rho \rfloor + 2$, every contagious set for K_n has size

at least $2 \geq 2\rho n - 4\rho$, from which we derive the following observation which shows that Theorem 1.15 is best-possible.

Observation 1.16 ([39]). *For every $\varepsilon > 0$ there exists $\rho > 0$ and a connected graph G on n vertices such that $h_\rho(G) \neq 1$ and*

$$h_\rho(G) \geq (2 - \varepsilon)\rho n .$$

The second main result of this chapter is a stronger bound for graphs of girth at least five, obtained by combining our approach for Theorem 1.15 with the method used by Gentner and Rautenbach [45] to prove (1.2.5).

Theorem 1.17 ([39]). *For every $\varepsilon > 0$ there exists $\rho_0 > 0$ such that for every $\rho \in (0, \rho_0)$ and every connected graph G of order n and girth at least 5 we have $h_\rho(G) = 1$ or*

$$h_\rho(G) < (1 + \varepsilon)\rho n .$$

Theorem 1.17 is asymptotically best-possible, as shown by the following example of the balanced $(\lfloor 1/\rho \rfloor + 1)$ -regular tree.

Observation 1.18 ([39]). *For every $\varepsilon > 0$, every $\rho_0 > 0$ and every $N_0 \in \mathbb{N}$ there exists $\rho \in (0, \rho_0)$ and a connected graph G with girth at least 5 on $n > N_0$ vertices such that*

$$h_\rho(G) > (1 - \varepsilon)\rho n .$$

Proof. Given ε, ρ_0 and N_0 , choose $\rho \in (0, \rho_0)$ sufficiently small so that $d := \lfloor 1/\rho \rfloor + 1$ fulfils $d > \max\{2/\varepsilon, \sqrt{N_0 - 1}\}$. Let G be the balanced d -regular tree with root v of order $n := 1 + d^2 > N_0$, which consists of three ‘tiers’ $T_1 = \{v\}, T_2 = N_G(v)$ and $T_3 = V(G) \setminus (\{v\} \cup N_G(v))$ such that $|T_2| = d$ and $|T_3| = d(d - 1)$. We will show that

$h_\rho(G) = d$ and therefore that

$$h_\rho(G) = d = \frac{d}{\rho(1+d^2)}\rho n \geq \frac{(d-1)d}{1+d^2}\rho n = \left(1 - \frac{d+1}{1+d^2}\right)\rho n > (1-2/d)\rho n > (1-\varepsilon)\rho n .$$

First note that T_2 infects the whole graph G , so $h_\rho(G) \leq d$. On the other hand, observe that every vertex in T_2 needs at least two infected neighbours in order to become infected itself. For each $u \in T_2$ write $T_u := \{u\} \cup N_G(u) \setminus \{v\}$. Then, because of our previous observation, every contagious set $A \subseteq V(G)$ fulfils $|A \cap T_u| \geq 1$ for each $u \in T_2$. Since the sets T_u are pairwise disjoint it follows that $|A| \geq |T_2| = d$, so $h_\rho(G) \geq d$. \square

The proofs of Theorem 1.15 and Theorem 1.17 are presented in Chapter 3. The proofs of both results rely on the strategy to assume a minimal counterexample and to analyse its structure. In particular we bound the number of edges between vertices of low degree and vertices of high degree. We can then utilise this information about the degrees in a more refined version of the random selection of Ackermann, Ben-Zwi and Wolfowitz [1] which we discussed above. This way we can guarantee the existence of a small contagious set which contradicts the assumption that there exists a counterexample to our assertion. In the case of graphs of girth at least 5 we will instead combine our approach with the random selection of Gentner and Rautenbach [45]. Since the proofs are quite short, we refrain from giving a more detailed overview here and refer directly to Chapter 3.

Finally, we would like to remark that the extremal construction for Observation 1.16, which shows that Theorem 1.15 is best-possible, indeed relies on the fact that n is chosen close to $1/\rho$. It is therefore a natural question whether, for fixed $\rho \in (0, 1]$, stronger bounds could be given for sufficiently large n . Note that in our construction for Observation 1.18 the order n can be chosen to be arbitrarily large and therefore the bound of Theorem 1.17 cannot be replaced by a asymptotically stronger bound in the case of sufficiently large n . We give a detailed discussion of this aspect in the next Section 1.2.6.

1.2.6 Open questions

We conclude this chapter with the discussion of some related open problems. As mentioned above, although Observation 1.16 shows that Theorem 1.15 cannot be improved, there remains the question of whether, for fixed ρ , the optimal bound for graphs on at least n vertices is asymptotically ρn as n tends to infinity.

Question 1.19. *For $\rho \in (0, 1]$ fixed, is it true that*

$$\limsup_{n \rightarrow \infty} \max_{G \text{ connected, order } n} \frac{h_\rho(G)}{\rho n} = 1?$$

Note that Observation 1.18 and Theorem 1.17 show that the answer to Question 1.19 is positive, if we restrict it to graphs of girth at least 5. However, without this additional restriction the answer to this question is negative for uncountable many and arbitrarily small values of $\rho \in (0, 1]$, as demonstrated by the following construction.

Proposition 1.20 ([39]). *For every $\rho \in (0, 1/3)$ such that $1/\rho - \lfloor 1/\rho \rfloor > 1/2$ there exists $\varepsilon(\rho) > 0$ such that for every $N_0 \in \mathbb{N}$ there exists a connected graph G of order $n \geq N_0$ with*

$$h_\rho(G) > (1 + \varepsilon)\rho n .$$

Proof. Roughly speaking, we use the integrality gap to produce a number of graphs H_i each requiring two vertices to become infected, while containing slightly fewer than $2/\rho$ vertices. We connect these graphs together to a graph G in a way that ensures that even if the whole graph G except one H_i is infected, we then still need at least two additional infected vertices from H_i to infect the vertices in H_i . This ensures that a contagious set must contain at least two vertices from every H_i , while the size of these subgraphs bounds $h_\rho(G)$ away from ρn .

Formally, set $\gamma = 1/\rho - \lfloor 1/\rho \rfloor - 1/2$ and choose $k \in \mathbb{N}$ such that $k > \max\{1/(2\gamma), \rho N_0/2\}$

and $\varepsilon \in \mathbb{R}$ such that $0 < \varepsilon < \frac{\rho(\gamma-1/(2k))}{1-\rho(\gamma-1/(2k))}$. Let A and B be two disjoint sets of $\lfloor 1/\rho \rfloor$ vertices and u be an additional vertex. We define a graph H with vertex set $A \cup B \cup \{u\}$ and edge set $\{\{v, w\} \mid v, w \in B \text{ or } v \in A, w \in B \text{ or } v = u, w \in A\}$. In other words, B is a clique, and we add a complete bipartite graph from A to B and all edges from the vertices of A to u .

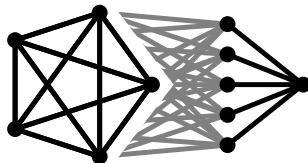


Figure 1.5: H when $\lfloor 1/\rho \rfloor = 5$

Let H_1, \dots, H_k be vertex disjoint copies of H and denote by u_1, \dots, u_k the vertices corresponding to u . Let s be another additional vertex. We define a graph G to be the union of the H_i and $\{s\}$ with the additional edges $\{\{s, u_i\} \mid i \in [k]\}$. Note that $n = |V(G)| = k|V(H)| + 1 \geq N_0$ and that a contagious set for G contains at least 2 vertices from each H_i . This follows since each vertex in H_i requires two infected neighbours to become infected. The vertex u may gain one infected neighbour if s is infected but to become infected, at least one vertex from A must also be infected. For a vertex from A to become infected, then either two vertices from B must be infected already, or one vertex from B and u must have been previously infected. Equally, no vertex from B can become infected without two infected neighbours from within $A \cup B$. We therefore see that only u can become infected without two vertices from H_i being in the contagious set, however the infection of u requires an infected vertex in A and will only lead to the infection of

H_i if there was also a vertex infected in B . Thus we have

$$\begin{aligned} h_\rho(G) &\geq 2k = \frac{2k}{\rho(k|V(H)| + 1)}\rho n = \frac{2}{\rho(2\lfloor \frac{1}{\rho} \rfloor + 1) + \rho/k}\rho n = \frac{1}{\rho\lfloor \frac{1}{\rho} \rfloor + \rho/2 + \rho/2k}\rho n \\ &= \frac{1}{1 - \rho(\gamma - 1/(2k))}\rho n = \left(1 + \frac{\rho(\gamma - 1/(2k))}{1 - \rho(\gamma - 1/(2k))}\right)\rho n > (1 + \varepsilon)\rho n . \end{aligned}$$

□

Although this demonstrates that for a given ρ , in general it is not possible to give a bound tending to ρn as n tends to infinity, it remains possible that for sufficiently small ρ , sufficiently large graphs have contagious sets whose size is close to ρn . We formulate this formally as the following question.

Question 1.21 ([39]). *Is it true that*

$$\limsup_{\rho \rightarrow 0} \limsup_{n \rightarrow \infty} \max_{G \text{ connected, order } n} \frac{h_\rho(G)}{\rho n} = 1?$$

Lastly, it is also possible that Theorem 1.17 can be improved in several ways. Since the example of the balanced $(\lfloor 1/\rho \rfloor + 1)$ -regular tree in Observation 1.18 only shows that Theorem 1.17 is asymptotically best-possible, there remains the obvious question whether it is possible to drop the ε in the statement of Theorem 1.17. Note that the assumption about ρ being sufficiently small is necessary for Theorem 1.17. To see this, choose any $\rho \in (1/2, 3/5)$, and observe that the 5-cycle C_5 then fulfils $h_\rho(C_5) = 3 > \rho n(C_5)$. This illustrates that graphs of girth at least 5 behave different than trees for which Gentner and Rautenbach [45] showed that for every $\rho \in (0, 1]$ and every tree T of order at least $1/\rho$ we have $h_\rho(T) \leq \rho n(T)$. It would also be interesting to know whether Theorem 1.17 remains valid under the weaker assumption that G has girth at least four (note that Observation 1.16 and Proposition 1.20 each show that the girth assumption cannot be dropped altogether).

CHAPTER 2

HAMILTON CYCLES BELOW THE DIRAC THRESHOLD

2.1 Organisation of the chapter

This chapter is organised as follows. In Section 2.2 we establish the definitions and notation we need, including the definition of even-good and odd-good bipartitions needed for our characterisation (Theorem 1.11), before giving the brief deduction of the exact Dirac threshold for Hamilton 2-cycles in 4-graphs (Corollary 1.10) and discussing the complexity aspects of Theorem 1.11. Next, in Section 2.3 we give the proof of Theorem 1.11, postponing a number of key lemmas to future sections. Specifically we distinguish a non-extremal case, an even-extremal case and an odd-extremal case, the necessary lemmas for which are postponed to Sections 2.4, 2.5 and 2.6 respectively. In parallel with the proofs of these lemmas we establish analogous algorithmic results, and in Section 2.7 we combine these results to formulate a polynomial-time algorithm as claimed in Theorem 1.12. We then consider tight Hamilton cycles in Section 2.8, proving Theorem 1.9 via a sequence of polynomial-time reductions.

2.2 A characterisation of dense 4-graphs with no Hamilton 2-cycle

2.2.1 Notation

Let H be a k -graph. We write $e(H)$ for the number of edges of H . Also, for any set $S \subseteq V(H)$ we define the *neighbourhood* of S to be $N_H(S) := \{S' \subseteq V(H) \setminus S : S \cup S' \in E(H)\}$. That is, $N_H(S)$ is the collection of all sets which together with S form an edge of H . So $N_H(S)$ is a collection of $(k - |S|)$ -sets and $|N_H(S)| = d_H(S)$. In particular, if $|S| = k - 1$ then $N_H(S)$ is a set of singleton sets of vertices, in which case we identify $N_H(S)$ with the corresponding set of vertices for notational simplicity, for example writing $v \in N_H(S)$ instead of $\{v\} \in N_H(S)$. Furthermore, when H is clear from the context we write simply $N(S)$ and to avoid clutter we frequently omit braces around sets, for example writing $N(x, y, z)$ instead of $N(\{x, y, z\})$. Given a set $X \subseteq V(H)$, we write $H[X]$ to denote the subgraph of H induced by X , that is, the k -graph with vertex set X and whose edges are all edges $e \in E(H)$ with $e \subseteq X$.

We define ℓ -paths in k -graphs in a similar way to ℓ -cycles. Indeed, a k -graph P is an ℓ -path if P has no isolated vertices and, moreover, the vertices of P can be linearly ordered in such a way that every edge of P consists of k consecutive vertices and each edge intersects the subsequent edge in precisely ℓ vertices. So the number of vertices in an ℓ -path must be congruent to k modulo $k - \ell$. As for cycles we refer to $(k - 1)$ -paths as *tight paths*. The *length* of an ℓ -path or ℓ -cycle is the number of edges it contains. A *segment* of an ℓ -path P or ℓ -cycle C is an ℓ -path P' which is a subgraph of P or C respectively.

Now suppose that H is a 4-graph. Given a bipartition $\{A, B\}$ of $V(H)$, we say that an edge $e \in E(H)$ is *odd* if $|e \cap A|$ is odd (or equivalently, if $|e \cap B|$ is odd) and *even* otherwise. We denote the subgraph of H consisting only of the even edges of H by H_{even}

and the subgraph of H consisting only of the odd edges of H by H_{odd} . Also, we say that a pair p of distinct vertices of H is a *split* pair if $|p \cap A| = 1$, and a *connate* pair otherwise. These terms are all dependent on the bipartition $\{A, B\}$ and 4-graph H in question, but these will always be clear from the context.

We use various ways of describing a 2-path or 2-cycle in a 4-graph. One is to list a sequence of vertices, that is, (v_1, \dots, v_m) for some even $m \geq 4$; the edges of C are then $\{v_i, v_{i+1}, v_{i+2}, v_{i+3}\}$ for each even i . Another is to give a sequence of pairs of vertices, that is, $p_1 p_2 p_3 \dots p_m$ for some integer $m \geq 2$; the edges of C are then $p_i \cup p_{i+1}$ for each i . The *ends* of a 2-path in a 4-graph are the initial pair and final pair, that is $\{v_1, v_2\}$ and $\{v_{m-1}, v_m\}$ in the first style of notation, and p_1 and p_m in the second style of notation. We concatenate 2-paths in the natural way, for example, if P is a 2-path with ends p and p' , and Q is a 2-path with ends p' and q , and P and Q have no vertices in common outside p' , then PQ is a 2-path with ends p and q . We sometimes say that P is a path *from* p *to* q to mean that P has ends p and q , however, note that a path from p to q has the same meaning as a path from q to p .

Given a 4-graph H , the *total 2-pathlength* of H is the maximum sum of lengths of vertex-disjoint 2-paths in H . For example, H having total 2-pathlength 3 indicates the presence in H of three disjoint edges (*i.e.* three 2-paths of length 1), or of a 2-path of length 3, or of two vertex-disjoint 2-paths, one of length 1 and one of length 2.

For an integer k we write $[k]$ for the set of integers from 1 to k and, given a set V , we write $\binom{V}{k}$ for the set of subsets of V of size k . Also, we write $x \ll y$ (“ x is sufficiently smaller than y ”) to mean that for any $y > 0$ there exists $x_0 > 0$ such that for any $x \leq x_0$ the subsequent statement holds. Similar statements with more variables are defined accordingly. We omit floors and ceilings throughout this chapter where they do not affect the argument.

2.2.2 Odd-good and even-good bipartitions of 4-graphs

Using the definitions introduced in the previous subsection, we can now give the central definition of our characterisation.

Definition 2.1. *Let H be a 4-graph on n vertices, where n is even, and let $\{A, B\}$ be a bipartition of $V(H)$. We say that $\{A, B\}$ is even-good if at least one of the following statements holds.*

- (i) $|A|$ is even or $|A| = |B|$.
- (ii) H_{odd} contains edges e and e' such that either $e \cap e' = \emptyset$ or $e \cap e'$ is a split pair.
- (iii) $|A| = |B| + 2$ and H_{odd} contains edges e and e' with $e \cap e' \in \binom{A}{2}$.
- (iv) $|B| = |A| + 2$ and H_{odd} contains edges e and e' with $e \cap e' \in \binom{B}{2}$.

Now let $m \in \{0, 2, 4, 6\}$ and $d \in \{0, 2\}$ be such that $m \equiv n \pmod{8}$ and $d \equiv |A| - |B| \pmod{4}$. Then we say that $\{A, B\}$ is odd-good if at least one of the following statements holds.

- (v) $(m, d) \in \{(0, 0), (4, 2)\}$.
- (vi) $(m, d) \in \{(2, 2), (6, 0)\}$ and H_{even} contains an edge.
- (vii) $(m, d) \in \{(4, 0), (0, 2)\}$ and H_{even} has total 2-pathlength at least two.
- (viii) $(m, d) \in \{(6, 2), (2, 0)\}$ and either there is an edge $e \in E(H_{\text{even}})$ with $|e \cap A| = |e \cap B| = 2$ or H_{even} has total 2-pathlength at least three.

We have now introduced all notation and definitions needed to understand and make use of Theorem 1.11.

Remark 2.2. *Suppose that H has even order. If $\{A, B\}$ is a bipartition of $V(H)$ which is not even-good, then, since H_{odd} must not contain two disjoint edges, there exists a set X of at most four vertices of H such that each edge of H_{odd} meets X . Similarly, if $\{A, B\}$*

is a bipartition of $V(H)$ which is not odd-good, then there exists a set X of at most eight vertices of H such that every edge of H_{even} meets X .

Using Remark 2.2 we can test the criterion of Theorem 1.11 in polynomial time in graphs of high minimum codegree. A special case of a result of Keevash, Knox and Mycroft [64, Lemma 2.2] states that, given a k -graph H on n vertices with minimum codegree $\delta(H) \geq n/3$, we can list in time $O(n^5)$ all bipartitions $\{A, B\}$ of $V(H)$ with no odd edge and all bipartitions $\{A, B\}$ of $V(H)$ with no even edge, and that there are at most a constant number of such bipartitions. Hence we can first check whether the order of H is even and, if so, we can establish all candidates for a bipartition $\{A, B\}$ which is not even-good or not odd-good in the following way: for each set X of eight vertices of H , delete the vertices of X from H to form H' , and list all bipartitions $\{A', B'\}$ of $V(H')$ with no even edge or no odd edge, then for each such $\{A', B'\}$ list each of the 2^8 possible extensions to a bipartition of $V(H)$. Clearly we can check in polynomial time whether a given bipartition of $V(H)$ is even-good and odd-good, so we can test the criterion of Theorem 1.11 by checking this for all listed bipartitions. Together with Theorem 1.11 this proves that we can determine in polynomial time whether a 4-graph H satisfying the minimum degree condition of Theorem 1.11 contains a Hamilton 2-cycle. More precisely, we have the following corollary (see [40] for a more detailed description of this algorithm and proof of its correctness under the assumption that Theorem 1.11 holds).

Corollary 2.3. *There exist a constant $\varepsilon > 0$ and an algorithm which, given a 4-graph H on n vertices with $\delta(H) \geq n/2 - \varepsilon n$, determines in time $O(n^{25})$ whether H contains a Hamilton 2-cycle. Furthermore, if H does not contain a Hamilton 2-cycle, then the algorithm returns a bipartition $\{A, B\}$ of $V(H)$ which is not even-good or odd-good.*

Theorem 1.12 supersedes Corollary 2.3 by showing that we can actually find a Hamilton 2-cycle in such a 4-graph in polynomial time (if such a cycle exists). This does not follow

from Theorem 1.11 directly, but instead follows by giving algorithms for each step involved in the proof of Theorem 1.11. We do this in parallel with the results needed for the proof of Theorem 1.11 in Sections 2.4, 2.5 and 2.6, before presenting the algorithm claimed in Theorem 1.12 in Section 2.7.

2.2.3 The exact Dirac threshold for Hamilton 2-cycles in 4-graphs.

We now make use of our characterisation to deduce Corollary 1.10. First, we give a construction to show that the degree bound of Corollary 1.10 is best-possible. So fix an even integer $n \geq 6$ and let A and B be disjoint sets with $|A \cup B| = n$ such that $|A| = \frac{n}{2} - 1$ if 8 divides n and $|A| = \frac{n}{2}$ otherwise. We define H^* to be the 4-graph with vertex set $A \cup B$ whose edges are all 4-sets $e \subseteq A \cup B$ such that $|e \cap A|$ is odd. Then it is easy to see that $\delta(H^*) = \frac{n}{2} - 3$ if 8 divides n and $\frac{n}{2} - 2$ otherwise. Furthermore, the size of A implies that the bipartition $\{A, B\}$ of $V(H^*)$ is not odd-good, as H^* has no even edges. Hence by Theorem 1.11 there is no Hamilton 2-cycle in H^* .

Proof of Theorem 1.10. Choose ε, n_0 as in Theorem 1.11, and set $n_1 := \max\{n_0, \frac{2}{\varepsilon}, 100\}$. Let $n \geq n_1$ be even and let H be a 4-graph on n vertices which satisfies the minimum codegree condition of Corollary 1.10. Also let $\{A, B\}$ be a bipartition of $V(H)$, and assume without loss of generality that $|A| \leq \frac{n}{2}$. By Theorem 1.11 it suffices to prove that $\{A, B\}$ is even-good and odd-good. This follows immediately from Definition 2.1 if $|A| < 18$, so we may assume that $18 \leq |A| \leq |B|$ (this ensures that we may choose distinct vertices as required in what follows). Note that if 8 divides n and $|A| = \frac{n}{2}$, then $\{A, B\}$ is even-good by Definition 2.1(i) and odd-good by Definition 2.1(v). So we may assume that if 8 divides n , then $|A| \leq \frac{n}{2} - 1$ and $\delta(H) \geq \frac{n}{2} - 2$, whilst otherwise we have $|A| \leq \frac{n}{2}$ and $\delta(H) \geq \frac{n}{2} - 1$. Either way, we must have $\delta(H) \geq |A| - 1$.

To see that $\{A, B\}$ must be even-good, arbitrarily choose vertices $x_1, x_2, y_1, y_2, z_1, z_2 \in$

A. Then $|N(x_1, y_1, z_1) \cap B|, |N(x_2, y_2, z_2) \cap B| \geq \delta(H) - (|A| - 3) \geq 2$, so we may choose distinct $w_1, w_2 \in B$ with $w_1 \in N(x_1, y_1, z_1) \cap B$ and $w_2 \in N(x_2, y_2, z_2) \cap B$. The sets $\{x_1, y_1, z_1, w_1\}$ and $\{x_2, y_2, z_2, w_2\}$ are then disjoint odd edges of H , so $\{A, B\}$ is even-good by Definition 2.1(ii).

We next show that $\{A, B\}$ is also odd-good. For this, arbitrarily choose distinct vertices $a_1, a_2, \dots, a_9, a'_1, \dots, a'_9 \in A$ and $b_1, \dots, b_9 \in B$. For any $1 \leq i, j \leq 9$ we have $|N(a_i, a'_i, b_j) \cap B| \geq \delta(H) - (|A| - 2) \geq 1$, so there must be $b_j^i \in B$ such that $\{a_i, a'_i, b_j, b_j^i\}$ is an (even) edge of H . If for each $1 \leq j \leq 9$ the vertices b_j^i for $1 \leq i \leq 9$ are all distinct, then H_{even} contains a set of nine disjoint edges, so there is no set $X \subseteq V(H)$ with $|X| \leq 8$ which intersects every even edge of H . However, by Remark 2.2 such a set X must exist if $\{A, B\}$ is not odd-good. So we may assume that $b_j^{i'} = b_j^i$ for some $1 \leq i, i', j \leq 9$ with $i \neq i'$. It follows that $\{a_i, a'_i, b_j, b_j^i\}$ is an even edge of H with exactly two vertices in A , whilst $(a_i, a'_i, b_j, b_j^i, a_{i'}, a'_{i'})$ is a 2-path of length 2 in H_{even} . So $\{A, B\}$ is odd-good by Definition 2.1(v), (vi), (vii) or (viii), according to the value of n . \square

2.3 Proof of Theorem 1.11

In this section we prove Theorem 1.11, our characterisation of 4-graphs with minimum codegree close to the Dirac threshold which contain a Hamilton 2-cycle, although the proofs of several lemmas are deferred to subsequent sections. We begin with the following proposition, which establishes the forward implication of Theorem 1.11; note that the minimum codegree assumption is not needed for this direction.

Proposition 2.4. *Let H be a 4-graph. If H contains a Hamilton 2-cycle, then every bipartition of $V(H)$ is both even-good and odd-good.*

Proof. Let n be the order of H , let $C = (v_1, v_2, \dots, v_n)$ be a Hamilton 2-cycle in H and let $\{A, B\}$ be a bipartition of $V(H)$. Write $P_i = \{v_{2i-1}, v_{2i}\}$ for each $1 \leq i \leq \frac{n}{2}$, so the edges of C are $e_i := P_i \cup P_{i+1}$ for $1 \leq i \leq \frac{n}{2}$ (with addition taken modulo $\frac{n}{2}$). The key

observation is that e_i is even if P_i and P_{i+1} are both split pairs or both connate pairs, and odd otherwise.

We first show that $\{A, B\}$ is even-good. This holds by Definition 2.1(ii) if H contains two disjoint odd edges, so we may assume without loss of generality that all edges of H other than e_1 and $e_{n/2}$ are even. It follows that the pairs $P_2, P_3, \dots, P_{n/2}$ are either all split pairs or all connate pairs. In the former case, if P_1 is a split pair then $|A| = |B|$, so Definition 2.1(i) holds, whilst if $P_1 \subseteq A$ then Definition 2.1(iii) holds, and if $P_1 \subseteq B$ then Definition 2.1(iv) holds. In the latter case, if P_1 is a connate pair then $|A|$ is even, so Definition 2.1(i) holds, whilst if P_1 is a split pair then Definition 2.1(ii) holds. So in all cases we find that $\{A, B\}$ is even-good.

To show that $\{A, B\}$ is odd-good, suppose first that 4 does not divide n , and note that by our key observation the number of even edges in C must then be odd. If C contains three or more even edges or an edge with precisely two vertices in A , then $\{A, B\}$ is odd-good by Definition 2.1(vi) and (viii), so we may assume without loss of generality that $e_{n/2}$ is the unique even edge in C and that $e_{n/2} \subseteq A$ or $e_{n/2} \subseteq B$. It follows that $P_1, P_3, \dots, P_{n/2}$ are connate pairs (of which there are $\lceil \frac{n}{4} \rceil$ in total) and the remaining pairs are split, so $|A| - |B| \equiv 2\lceil \frac{n}{4} \rceil \pmod{4}$. We must therefore have $(m, d) \in \{(2, 2), (6, 0)\}$, and $\{A, B\}$ is odd-good by Definition 2.1(vi). On the other hand, if 4 divides n , then by our key observation the number of even edges in C is even. If this number is at least two then $\{A, B\}$ is odd-good by Definition 2.1(v) and (vii). If instead every edge of C is odd, then exactly $\frac{n}{4}$ of the pairs P_i are connate pairs, so $|A| - |B| \equiv 2 \cdot \frac{n}{4} \pmod{4}$, and C is odd-good by Definition 2.1 (v). \square

The main difficulty in proving Theorem 1.11 is therefore to establish the backwards implication. For this note that if H is a 4-graph and $\{A, B\}$ is a bipartition of $V(H)$ which is not odd-good, then H has very few even edges, as it does not contain three disjoint even edges. Similarly, if $\{A, B\}$ is not even-good, then H has very few odd edges.

This motivates the following definition.

Definition 2.5. Fix $c > 0$, a 4-graph H on n vertices, and a bipartition $\{A, B\}$ of $V(H)$.

- (a) We say that $\{A, B\}$ is c -even-extremal if $\frac{n}{2} - cn \leq |A| \leq \frac{n}{2} + cn$ and H contains at most $c \binom{n}{4}$ odd edges.
- (b) We say that $\{A, B\}$ is c -odd-extremal if $\frac{n}{2} - cn \leq |A| \leq \frac{n}{2} + cn$ and H contains at most $c \binom{n}{4}$ even edges.
- (c) We say that H is c -even-extremal if $V(H)$ admits a c -even-extremal bipartition, and likewise that H is c -odd-extremal if $V(H)$ admits a c -odd-extremal bipartition.

In our proof of Theorem 1.11 we distinguish between the non-extremal case, in which H is neither even-extremal nor odd-extremal, and the two extremal cases.

2.3.1 Non-extremal 4-graphs

Suppose first that H is neither even-extremal nor odd-extremal. In this case we proceed by the so-called ‘absorbing’ method, introduced by Rödl, Ruciński and Szemerédi [83], to construct a Hamilton 2-cycle in H . More specifically, we adapt the approach used by Karpiński, Ruciński and Szymańska [62], proving an ‘absorbing lemma’ and a ‘long cycle lemma’. Loosely speaking, our ‘absorbing lemma’ allows us to find a short 2-path in H which can ‘absorb’ most small collections of pairs of H .

Lemma 2.6 (Absorbing lemma). *Suppose that $1/n \ll \varepsilon \ll \gamma \ll \lambda \ll c, \mu$. If H is a 4-graph of order n with $\delta(H) \geq n/2 - \varepsilon n$ which is neither c -even-extremal nor c -odd-extremal, then there is a 2-path P in H and a 2-graph G on $V(H)$ with the following properties.*

- (i) P has at most μn vertices.
- (ii) Every vertex of $V(H) \setminus V(P)$ is contained in at least $(1 - \lambda)n$ edges of G .

(iii) For any $s \leq \gamma n$ and any s disjoint edges e_1, \dots, e_s of G which do not intersect P , there is a 2-path P^* in H with the same ends as P such that $V(P^*) = V(P) \cup \bigcup_{j=1}^s e_j$.

Next, our ‘long cycle lemma’ states that, having found an absorbing-path P_0 in H , we can cover almost all vertices of H by a long 2-cycle C of which P_0 is a segment such that the vertices not covered by C form a collection of pairs which can be absorbed by P_0 .

Lemma 2.7 (Long cycle lemma). *Suppose that $1/n \ll \varepsilon \ll \gamma \ll \lambda \leq \mu \ll c$ and that n is even. Let $H = (V, E)$ be a 4-graph of order n with $\delta(H) \geq n/2 - \varepsilon n$ which is not c -even-extremal. Also let P_0 be a 2-path in H on at most μn vertices, and let G be a 2-graph on V such that each vertex $v \in V \setminus V(P_0)$ has $d_G(v) \geq (1 - \lambda)n$. Then H contains a 2-cycle C on at least $(1 - \gamma)n$ vertices such that P_0 is a segment of C and $G[V \setminus V(C)]$ contains a perfect matching.*

By combining these two lemmas we obtain the following lemma, which shows that if H is a non-extremal 4-graph whose minimum codegree is close to the Dirac threshold, then H contains a Hamilton 2-cycle, thereby establishing the backwards implication of Theorem 1.11 for non-extremal 4-graphs.

Lemma 2.8. *Suppose that $1/n \ll \varepsilon \ll c$ and that n is even, and let H be a 4-graph of order n with $\delta(H) \geq n/2 - \varepsilon n$. If H is neither c -odd-extremal nor c -even-extremal, then H contains a Hamilton 2-cycle.*

Proof. We introduce constants $\gamma, \lambda, \mu > 0$ such that

$$\frac{1}{n} \ll \varepsilon \ll \gamma \ll \lambda \ll \mu \ll c.$$

By Lemma 2.6 there is a 2-path P_0 in H and a graph G on $V(H)$ such that the properties (i)-(iii) of Lemma 2.6 hold. Then by Lemma 2.7 there is a 2-cycle C on at least $(1 - \gamma)n$ vertices which contains P_0 as a segment and such that $G[V \setminus V(C)]$ contains a

perfect matching e_1, \dots, e_s . Denote the 2-path $C[V(C) \setminus V(P_0)]$ by P' . Since $s \leq \gamma n$, there is a 2-path P^* with the same ends as P_0 such that $V(P^*) = V(P_0) \cup \bigcup_{j=1}^s e_j$, and P^*P' is then a Hamilton 2-cycle in H . \square

2.3.2 Extremal 4-graphs

It remains to prove the backwards implication of Theorem 1.11 in the case where H is either even-extremal or odd-extremal. In each case we have significant structural information about H . Indeed, if $V(H)$ admits an even-extremal bipartition $\{A, B\}$ (respectively odd-extremal), then we know that almost all edges of H are even (respectively odd). However, the assumption that the bipartition of H is even-good (respectively odd-good) yields certain small structures in H_{odd} (respectively H_{even}). By a careful analysis we are able to use this information to find a Hamilton 2-cycle in H in each case. We consider the even-extremal case in Section 2.5 and the odd-extremal case in Section 2.6, culminating in the following two lemmas, which establish the backwards implication of Theorem 1.11 in each extremal case.

Lemma 2.9. *Suppose that $1/n \ll \varepsilon, c \ll 1$ and that n is even, and let H be a 4-graph of order n with $\delta(H) \geq n/2 - \varepsilon n$. If H is c -even-extremal and every bipartition $\{A, B\}$ of $V(H)$ is even-good, then H contains a Hamilton 2-cycle.*

Lemma 2.10. *Suppose that $1/n \ll \varepsilon, c \ll 1$ and that n is even, and let H be a 4-graph of order n with $\delta(H) \geq n/2 - \varepsilon n$. If H is c -odd-extremal and every bipartition $\{A, B\}$ of $V(H)$ is odd-good, then H contains a Hamilton 2-cycle.*

Combining Lemmas 2.8, 2.9 and 2.10 completes the proof of Theorem 1.11 by establishing the backwards implication.

Proof of Theorem 1.11. Fix a constant c small enough for Lemmas 2.9 and 2.10. Having done so, choose ε sufficiently small for us to apply Lemma 2.8 with this choice of c , and

n_0 sufficiently large that we may apply Lemmas 2.8, 2.9 and 2.10 with these choices of c and ε and any even $n \geq n_0$. Let $n \geq n_0$ be even, and let H be a 4-graph on n vertices with $\delta(H) \geq (\frac{1}{2} - \varepsilon)n$, and suppose that every bipartition $\{A, B\}$ of $V(H)$ is both even-good and odd-good. If H is either c -even-extremal or c -odd-extremal then H contains a Hamilton 2-cycle by Lemma 2.9 or 2.10 respectively. On the other hand, if H is neither c -odd-extremal nor c -even-extremal then H contains a Hamilton 2-cycle by Lemma 2.8. This completes the proof of the backwards implication of Theorem 1.11; the proof of the forwards implication was Proposition 2.4. \square

2.4 Hamilton 2-cycles in 4-graphs: non-extremal case

In this section we give the proofs of Lemma 2.6 (the absorbing lemma) and Lemma 2.7 (the long cycle lemma). Throughout this section we only consider 2-paths and 2-cycles, so we suppress the 2 and speak simply of paths, cycles and Hamilton cycles.

2.4.1 A connecting lemma

We begin with a ‘connecting lemma’. This states that if H is a 4-graph whose minimum codegree is close to the Dirac threshold, then either H is even-extremal, or H is well-connected in the sense that any two disjoint pairs of vertices of H are connected by many short paths. This property is encapsulated in the following definition.

Definition 2.11. *For $\kappa > 0$, we say that a 4-graph H of order n is κ -connecting if for every two disjoint pairs $p_1, p_2 \in \binom{V(H)}{2}$ there are either at least κn^2 paths of length 2 or at least κn^4 paths of length 3 whose ends are p_1 and p_2 .*

In the proof of our connecting lemma we will make use of the following result of Goodman [47], that any two-colouring of a complete graph has many monochromatic triangles.

Theorem 2.12 ([47]). *Suppose that $1/n \ll \varepsilon$. If G is a graph on n vertices, then*

$$|\{S \in \binom{V(G)}{3} : G[S] \cong K_3 \text{ or } G[S] \cong \overline{K_3}\}| \geq \frac{1-\varepsilon}{24} n^3.$$

Lemma 2.13 (Connecting lemma). *Suppose that $1/n \ll \varepsilon \ll \kappa \ll c$ and that $H = (V, E)$ is a 4-graph on n vertices with $\delta(H) \geq n/2 - \varepsilon n$. If H is not κ -connecting, then H is c -even-extremal. Moreover, there exists an algorithm *Procedure EvenPartition*(H) which returns a c -even-extremal bipartition $\{A, B\}$ of V in time $O(n^8)$.*

Proof. Introduce further constants $\beta, \eta, \delta > 0$ such that

$$\frac{1}{n} \ll \varepsilon \ll \kappa \ll \beta \ll \eta \ll \delta \ll c.$$

Suppose that H is not κ -connecting. Then we may fix two disjoint pairs $\{a_1, a_2\}$ and $\{b_1, b_2\}$ of vertices of H such that there are fewer than κn^2 paths of length 2 whose ends are $\{a_1, a_2\}$ and $\{b_1, b_2\}$, and fewer than κn^4 paths of length 3 whose ends are $\{a_1, a_2\}$ and $\{b_1, b_2\}$. Observe that there are then at most βn vertices $v \in V$ with $|N(a_1, a_2, v) \cap N(b_1, b_2, v)| \geq \beta n$, as otherwise there are at least $\frac{1}{2}\beta^2 n^2 > \kappa n^2$ paths of length 2 connecting $\{a_1, a_2\}$ and $\{b_1, b_2\}$. Now we colour the edges of the complete graph G on V as follows. For distinct vertices $v, w \in V$ we say that

$$\{v, w\} \text{ is } \begin{cases} \text{red, if } \{a_1, a_2, v, w\} \in E \text{ and } \{b_1, b_2, v, w\} \notin E, \\ \text{blue, if } \{a_1, a_2, v, w\} \notin E \text{ and } \{b_1, b_2, v, w\} \in E, \\ \text{uncoloured, otherwise.} \end{cases}$$

Because of our first observation we have for at least $(1 - \beta)n - 4 \geq (1 - 2\beta)n$ vertices $v \in V$ that

$$d_{\text{red}}(v), d_{\text{blue}}(v) \geq \delta(H) - \beta n \geq \left(\frac{1}{2} - 2\beta\right)n. \quad (2.4.1)$$

Furthermore, for at least $\frac{1}{2}(n-2)\delta(H) \geq (\frac{1}{2}-2\varepsilon)\binom{n}{2}$ pairs $\{v, w\} \subseteq V$ we have $\{a_1, a_2, v, w\} \in E$, and likewise for at least $(\frac{1}{2}-2\varepsilon)\binom{n}{2}$ pairs $\{v, w\} \subseteq V$ we have $\{b_1, b_2, v, w\} \in E$. Since by assumption at most κn^2 pairs $\{x, y\} \subseteq V$ fulfil both $\{a_1, a_2, x, y\} \in E$ and $\{b_1, b_2, x, y\} \in E$, it follows that there are at most

$$2\kappa n^2 + 4\varepsilon \binom{n}{2} < 5\kappa \binom{n}{2} \quad (2.4.2)$$

uncoloured edges. Therefore the number of red edges and the number of blue edges are each at least

$$(\frac{1}{2}-2\varepsilon)\binom{n}{2} - 5\kappa \binom{n}{2} \geq (\frac{1}{2}-6\kappa)\binom{n}{2}. \quad (2.4.3)$$

Now we show that G either contains at least $n^3/50$ red triangles or at least $n^3/50$ blue triangles. To do this, suppose that there are fewer than $n^3/50$ red triangles. Then by Theorem 2.12 there are at least $(1-\varepsilon)\frac{n^3}{24} - \frac{n^3}{50} \geq \frac{n^3}{50} + 3\kappa n^3$ triangles which consist only of blue and uncoloured edges. Since by (2.4.2) there are at most $5\kappa \binom{n}{2} \cdot n \leq 3\kappa n^3$ triangles in G which contain an uncoloured edge, there are then at least $n^3/50$ blue triangles in G . So we may assume without loss of generality that G contains at least $n^3/50$ red triangles. At least $n^3/50 - 2\beta n^3 \geq n^3/100$ of these red triangles contain only vertices which fulfil (2.4.1); we denote the set of such triangles by \mathcal{T} . Furthermore for a triangle $T \in \mathcal{T}$ with vertex set $\{v_1, v_2, v_3\}$ we denote by $N_{\text{red}}(T)$ (respectively $N_{\text{blue}}(T)$) the set of vertices $x \in V$ such that each of the edges $\{v_1, x\}$, $\{v_2, x\}$ and $\{v_3, x\}$ is red (respectively blue). We call an edge $\{x, y, z, w\}$ of H *colourful* if it contains disjoint red and blue edges of G , say $\{x, y\}$ and $\{z, w\}$ respectively. Observe that each colourful edge of H disjoint from $\{a_1, a_2, b_1, b_2\}$ creates at least one path $(a_1, a_2, x, y, z, w, b_1, b_2)$ of length 3 in H with ends $\{a_1, a_2\}$ and $\{b_1, b_2\}$. Since there are at most κn^4 such paths, but no two distinct

colourful edges of H can create the same path, there must be at most

$$\kappa n^4 + 4n^3 < 2\kappa n^4 \quad (2.4.4)$$

colourful edges of H . This observation will prove the following claim.

Claim 2.14. *There exist vertices $v_1, v_2, v_3 \in V$ such that $T^* = \{v_1, v_2, v_3\}$ is a triangle with $T^* \in \mathcal{T}$ and such that the following properties hold.*

(i) $|N_{\text{red}}(T^*)| \geq (\frac{1}{2} - \eta)n,$

(ii) *there are at most ηn vertices $x \in N_{\text{red}}(T^*)$ with $|N_H(v_1, v_2, x) \cap N_{\text{blue}}(T^*)| \geq \eta n,$*

(iii) *there are at most ηn vertices $x \in N_{\text{blue}}(T^*)$ with $|N_H(v_1, v_2, x) \cap N_{\text{red}}(T^*)| \geq \eta n,$*

(iv) *there are at most ηn vertices $x \in V$ with $|N_H(v_1, v_2, x) \cap N_{\text{blue}}(x)| \geq \eta n.$*

Proof of Claim 2.14. Arbitrarily fix, for each $T \in \mathcal{T}$, a labelling of the vertices of T as v_1, v_2 and v_3 . For this labelling we refer to the pair $\{v_1, v_2\}$ as the *specified pair* of T . Now let \mathcal{T}_1 consist of those triangles $T \in \mathcal{T}$ which do not satisfy (i). Likewise, let $\mathcal{T}_2, \mathcal{T}_3$ and \mathcal{T}_4 consist of those triangles $T \in \mathcal{T}$ which do not satisfy (ii), (iii) and (iv) respectively (for our choice of specified pair). Let $T \in \mathcal{T}_1$, so by definition we have $|N_{\text{red}}(T)| < (\frac{1}{2} - \eta)n$. Since $T \in \mathcal{T}$, each vertex of T satisfies (2.4.1) and so is incident to at most $4\beta n$ uncoloured edges of G , so there are at least $|N_H(V(T))| - |N_{\text{red}}(T)| - 3 \cdot 4\beta n \geq (\eta - \varepsilon - 12\beta)n > \frac{1}{2}\eta n$ colourful edges of H containing $V(T)$. So in total H has at least $\frac{1}{4} \cdot |\mathcal{T}_1| \cdot \frac{1}{2}\eta n$ colourful edges; by (2.4.4) we find that $|\mathcal{T}_1| < \frac{16\kappa}{\eta}n^3 < \beta n^3$. Next observe that at least $\frac{1}{n}|\mathcal{T}_2|$ pairs $\{v_1, v_2\} \in \binom{V}{2}$ are the specified pair for some triangle $T \in \mathcal{T}_2$. For each such pair there are at least $\frac{1}{2}\eta^2 n^2$ pairs $\{x, y\} \in \binom{V}{2}$ with $x \in N_{\text{red}}(T)$ and $y \in N_H(v_1, v_2, x) \cap N_{\text{blue}}(T)$, and each 4-tuple $\{v_1, v_2, x, y\}$ formed in this way is a colourful edge of H . Overall this gives at least $\frac{1}{6} \cdot \frac{1}{n}|\mathcal{T}_2| \cdot \frac{1}{2}\eta^2 n^2$ colourful edges in total, so $|\mathcal{T}_2| < \beta n^3$. Essentially the same argument

shows that $|\mathcal{T}_3| < \beta n^3$. Finally, at least $\frac{1}{n}|\mathcal{T}_4|$ pairs $\{v_1, v_2\} \in \binom{V}{2}$ are the specified pair for some triangle $T \in \mathcal{T}_4$. For each such pair there are at least $\frac{1}{2}\eta^2 n^2$ pairs $\{x, y\} \in \binom{V}{2}$ with $y \in N_H(v_1, v_2, x) \cap N_{\text{blue}}(x)$, and each 4-tuple $\{v_1, v_2, x, y\}$ formed in this way is a colourful edge of H . In total this gives at least $\frac{1}{6} \cdot \frac{1}{n}|\mathcal{T}_4| \cdot \frac{1}{2}\eta^2 n^2$ colourful edges, so $|\mathcal{T}_4| < \beta n^3$. So at least $|\mathcal{T}| - \sum_{i=1}^4 |\mathcal{T}_i| > |\mathcal{T}| - 4\beta n^3 > 0$ triangles $T \in \mathcal{T}$ satisfy (i)-(iv). \diamond

Fix a triangle $T^* \in \mathcal{T}$ with vertex set $\{v_1, v_2, v_3\}$ as in Claim 2.14, and define $d_{\min} := (\frac{1}{2} - 2\beta)n$, $N_r^* := N_{\text{red}}(T^*)$ and $N_b^* := N_{\text{blue}}(T^*)$. Since $T^* \in \mathcal{T}$ each v_i fulfills (2.4.1) and so is incident to at least d_{\min} blue edges, none of which has an endvertex in N_r^* , so we have

$$|N_b^*| \geq |V \setminus N_r^*| - 3(|V \setminus N_r^*| - d_{\min}) = 3d_{\min} - 2|V \setminus N_r^*| > (\frac{1}{2} - 3\eta)n, \quad (2.4.5)$$

where the final inequality is by Claim 2.14(i). Also by (2.4.1) we derive the upper bounds

$$|N_r^*| \leq (\frac{1}{2} + 2\beta)n \quad \text{and} \quad |N_b^*| \leq (\frac{1}{2} + 2\beta)n. \quad (2.4.6)$$

Together, (2.4.5) and Claim 2.14(i) tell us that at most $4\eta n$ vertices of V are neither in N_r^* or N_b^* , so by Claim 2.14(ii), all but at most ηn vertices $x \in N_r^*$ satisfy

$$|N_H(v_1, v_2, x) \cap N_r^*| \geq \delta(H) - |N_H(v_1, v_2, x) \cap N_b^*| - 4\eta n \geq (\frac{1}{2} - 6\eta)n. \quad (2.4.7)$$

By the same argument using Claim 2.14(iii), all but at most ηn vertices $x \in N_b^*$ have

$$|N_H(v_1, v_2, x) \cap N_b^*| \geq (\frac{1}{2} - 6\eta)n. \quad (2.4.8)$$

Define $A := N_r^*$ and $B := V \setminus A$; we conclude the proof by showing that $\{A, B\}$ is a c -even-extremal bipartition of V . For this, observe that by (2.4.6), (2.4.7) and Claim 2.14(iv) at

most $2\eta n$ vertices $x \in N_r^*$ have $|N_{\text{blue}}(x) \cap N_r^*| > 8\eta n$. Similarly, by (2.4.6), (2.4.8) and Claim 2.14(iv) at most $2\eta n$ vertices $x \in N_b^*$ have $|N_{\text{blue}}(x) \cap N_b^*| > 8\eta n$. Since at most $5\kappa \binom{n}{2}$ edges of G are uncoloured (see (2.4.2)), it follows that at least

$$\frac{1}{2} \cdot (|N_r^*| - 2\eta n) \cdot (|N_r^*| - 8\eta n) + \frac{1}{2} \cdot (|N_b^*| - 2\eta n) \cdot (|N_b^*| - 8\eta n) - 5\kappa \binom{n}{2} \geq (\frac{1}{2} - \delta) \binom{n}{2}$$

edges of $G[A] \cup G[B]$ are coloured red. Since by Claim 2.14(i) and (2.4.6) we have

$$\binom{|A|}{2} + \binom{|B|}{2} \leq \binom{(\frac{1}{2} + \eta)n}{2} + \binom{(\frac{1}{2} - \eta)n}{2} \leq (\frac{1}{2} + \delta) \binom{n}{2},$$

we conclude that at most $2\delta \binom{n}{2}$ edges of $G[A] \cup G[B]$ are not red. Together with (2.4.3) this shows that at least $(\frac{1}{2} - 6\kappa - 2\delta) \binom{n}{2} \geq (\frac{1}{2} - 3\delta) \binom{n}{2}$ of the at most $\frac{n^2}{4} \leq (\frac{1}{2} + \delta) \binom{n}{2}$ edges of G with one vertex in A and the other in B are blue, and so at most $4\delta \binom{n}{2}$ such edges are not blue. Combining this and (2.4.4), we find that the number of edges of H with an odd number of vertices in A is at most

$$2\delta \binom{n}{2} \cdot (\frac{1}{2} + \delta) \binom{n}{2} + 4\delta \binom{n}{2} \cdot (\frac{1}{2} + \delta) \binom{n}{2} + 2\kappa n^4 < c \binom{n}{4}.$$

Since $(\frac{1}{2} - c)n < |A| < (\frac{1}{2} + c)n$ by Claim 2.14(i) and (2.4.6), it follows that the bipartition $\{A, B\}$ is c -even-extremal. To complete the proof we note that each step of the proof directly translates to an algorithm which returns the desired bipartition in the claimed running-time. \square

2.4.2 A reservoir lemma

Another lemma which we use to prove our ‘long cycle lemma’ is a ‘reservoir lemma’. This states that given a 4-graph H and a graph G on the same vertex set V , we can choose a ‘reservoir set’, that is, a small subset $R \subseteq V$ such that $H[R]$ and $G[R]$ are

representative of H and G . A standard and straightforward probabilistic argument shows that such a subset must exist, but derandomising this argument to give an algorithm which finds such a subset is somewhat more technical. For this we adapt the approach of Karpiński, Ruciński and Szymańska [62] to our setting. We first describe an algorithm, Procedure SelectSet, which is similar to Procedure SelectSubset from [62], but chooses a subset which is representative of two graphs (instead of just one graph) simultaneously. We will use this procedure in this subsection to find a ‘reservoir set’, and also in the next subsection to find an absorbing path. The conditions on the two graphs from which we want to choose the subset are described by the following setup.

Setup 2.15. Fix constants $\beta, \lambda, \tau > 0$ and integers m, M, N and r with $1 \leq r \leq N$. Let U and W be disjoint sets of sizes $|U| = M$ and $|W| = N$. Let G_1 be a graph with vertex set $U \cup W$ such that $G_1[U]$ is empty, $G_1[W]$ has precisely m edges, and $|N_{G_1}(u)| \geq \beta N$ for every $u \in U$. Also let G_2 be a graph with vertex set W such that $|N_{G_2}(w)| \geq (1 - \lambda)N$ for every $w \in W$. Finally, define $\nu := 2mr/N^2$.

Note that specifying the sextuple $(G_1, G_2, r, \beta, \lambda, \tau)$ determines all of the information given in Setup 2.15. Whilst Procedure SelectSet is entirely deterministic, to analyse it we consider a set of r vertices chosen uniformly at random and apply the following Chernoff-type bounds for binomial and hypergeometric distributions.

Theorem 2.16 ([58], Corollary 2.3 and Theorem 2.10). Suppose X has binomial or hypergeometric distribution and $0 < a < 3/2$. Then $\mathbb{P}(|X - \mathbb{E}(X)| \geq a\mathbb{E}(X)) \leq 2 \exp(-\frac{a^2}{3}\mathbb{E}(X))$.

Proposition 2.17. Adopt Setup 2.15. Assume additionally that $\beta > \tau$, that r and N are sufficiently large, and that $M \leq \frac{1}{8} \cdot \exp(\frac{\tau^2 r}{3\beta})$ as well as $N \leq \frac{1}{8} \cdot \exp(\frac{\lambda r}{3})$. Then in time $O(N^4 + MN^3)$ Procedure SelectSet($G_1, G_2, r, \beta, \lambda, \tau$) returns a set $R \subseteq W$ such that

$$(a) \quad (1 - \nu)r \leq |R| \leq r,$$

Procedure SelectSet($G_1, G_2, r, \beta, \lambda, \tau$)

Data: A sextuple $(G_1, G_2, r, \beta, \lambda, \tau)$ as in Setup 2.15.

Result: A ‘reservoir set’ $R \subseteq W$.

Set $U := \{u_1, \dots, u_M\}$, $W := \{w_1, \dots, w_N\}$ and $R' := \emptyset$.

for $k = 1$ **to** r **do**

for $j = 1$ **to** N **do**

 Set $R'_j := R' \cup \{w_j\}$.

 Set $e_j := e(G_1[R'_j])$, $e'_j := e(G_1[R'_j, W \setminus R'_j])$ and $e''_j := e(G_1[W \setminus R'_j])$.

for $i = 1$ **to** M **do**

 Set $d_{i,j} := |N_{G_1}(u_i) \cap R'_j|$ and $d'_{i,j} := |N_{G_1}(u_i) \setminus R'_j|$.

for $i = 1$ **to** N **do**

 Set $f_{i,j} := |N_{G_2}(w_i) \cap R'_j|$ and $f'_{i,j} := |N_{G_2}(w_i) \setminus R'_j|$.

 Find $j_k \in [N] \setminus \{j_1, \dots, j_{k-1}\}$ such that $A + B + C < 1$, where

$$A := \sum_{i=1}^M \sum_{s \leq (\beta - \tau)r - d_{i,j_k}} \frac{\binom{d'_{i,j_k}}{s} \binom{N-k-d'_{i,j_k}}{r-k-s}}{\binom{N-k}{r-k}},$$

$$B := \frac{1}{\nu r} \left(e_{j_k} + e'_{j_k} \frac{r-k}{N-k} + e''_{j_k} \frac{(r-k)(r-k-1)}{(N-k)(N-k-1)} \right),$$

$$C := \sum_{i=1}^N \sum_{s \leq (1-2\lambda)r - f_{i,j_k}} \frac{\binom{f'_{i,j_k}}{s} \binom{N-k-f'_{i,j_k}}{r-k-s}}{\binom{N-k}{r-k}}.$$

 Add w_{j_k} to R' .

Remove one vertex from each edge of $G_1[R']$ and call the resulting set R .

return R .

(b) R is an independent set in G_1 ,

(c) $|N_{G_1}(u) \cap R| \geq (\beta - \tau - \nu)r$ for all $u \in U$ and

(d) $|N_{G_2}(w) \cap R| \geq (1 - 2\lambda - \nu)r$ for all $w \in W$.

Proof. For a set $S \in \binom{W}{r}$, let $X(S)$ denote the number of vertices $u \in U$ with $|N_{G_1}(u) \cap S| \leq (\beta - \tau)r$, let $Y(S)$ denote the number of edges in $G_1[S]$ and let $Z(S)$ denote the number of vertices $w \in W$ with $|N_{G_2}(w) \cap S| \leq (1 - 2\lambda)r$. Now choose uniformly at random a subset $S \in \binom{W}{r}$, and let X, Y and Z denote the random variables $X(S), Y(S)$ and $Z(S)$ respectively. For each $u \in U$ the random variable $|N_{G_1}(u) \cap S|$ has a hypergeometric distribution with expectation at least βr ; by Theorem 2.16 and our assumption that $M \leq \frac{1}{8} \exp(\frac{r^2}{3\beta})$ it follows that the probability of the event $|N_{G_1}(u) \cap S| \leq (\beta - \tau)r$ is at most $2 \exp(-\frac{(\tau/\beta)^2}{3}\beta r) \leq \frac{1}{4M}$. So by linearity of expectation we have $\mathbb{E}(X) < \frac{1}{4}$. Likewise, for each $w \in W$ the random variable $|S \setminus N_{G_2}(w)|$ has a hypergeometric distribution with expectation at most λr , so a similar calculation using our assumption that $N \leq \frac{1}{8} \exp(\frac{\lambda r}{3})$ shows that $\mathbb{E}(Z) < \frac{1}{4}$. By linearity of expectation we also have $\mathbb{E}(Y) = m \cdot \frac{r(r-1)}{N(N-1)} \leq m(\frac{r}{N})^2$, so in particular we have $\frac{1}{\nu r} \mathbb{E}(Y) \leq \frac{1}{2}$. Thus we have in total that

$$\mathbb{E}(X) + \frac{1}{\nu r} \mathbb{E}(Y) + \mathbb{E}(Z) < 1. \quad (2.4.9)$$

Now suppose that for some $1 \leq k \leq r$ we have already chosen vertices $w_{j_1}, \dots, w_{j_{k-1}} \in W$ such that

$$\mathbb{E}(X|j_1, \dots, j_{k-1}) + \frac{1}{\nu r} \mathbb{E}(Y|j_1, \dots, j_{k-1}) + \mathbb{E}(Z|j_1, \dots, j_{k-1}) < 1,$$

where we identify j_i with the event that $w_{j_i} \in S$. Note that the base case $k = 1$ is guaranteed by (2.4.9). Then, by the law of total probability, it is possible to choose

$w_{j_k} \in W \setminus \{w_{j_1}, \dots, w_{j_{k-1}}\}$ such that

$$\mathbb{E}(X|j_1, \dots, j_k) + \frac{1}{\nu r} \mathbb{E}(Y|j_1, \dots, j_k) + \mathbb{E}(Z|j_1, \dots, j_k) < 1.$$

Having chosen w_{j_1}, \dots, w_{j_r} in this way, define $R' := \{w_{j_1}, \dots, w_{j_r}\}$. Then

$$X(R') + \frac{1}{\nu r} Y(R') + Z(R') = \mathbb{E}(X|j_1, \dots, j_r) + \frac{1}{\nu r} \mathbb{E}(Y|j_1, \dots, j_r) + \mathbb{E}(Z|j_1, \dots, j_r) < 1.$$

So $X(R') = Z(R') = 0$, as $X(R')$ and $Z(R')$ must have non-negative integer values. Also $Y(R') < \nu r$, meaning that $G_1[R']$ has at most νr edges. So if we form R from R' by removing one vertex from each of these edges, then R has the properties (a)-(d).

It therefore suffices to show that the choices of j_1, \dots, j_k in Procedure SelectSet are identical to the choices of j_1, \dots, j_k in the above argument, so the resulting sets R' are identical. That is, we must show that for each $1 \leq k \leq r$ we have $A = \mathbb{E}(X|j_1, \dots, j_k)$, $B = \frac{1}{\nu r} \mathbb{E}(Y|j_1, \dots, j_k)$ and $C = \mathbb{E}(Z|j_1, \dots, j_k)$, where A , B and C are the quantities given in Procedure SelectSet. The first two of these equalities were established in [62] (and are straightforward to verify). For the third define $f_{i,j}$ and $f'_{i,j}$ as in Procedure SelectSet. Then we have

$$\begin{aligned} \mathbb{E}(Z|j_1, \dots, j_k) &= \sum_{w \in W} \mathbb{P}(|N_{G_2}(w) \cap S| \leq (1-2\lambda)r \mid j_1, \dots, j_k) \\ &= \sum_{i=1}^N \sum_{s \leq (1-2\lambda)r - f_{i,j_k}} \mathbb{P}(|N_{G_2}(w_i) \cap (S \setminus \{j_1, \dots, j_k\})| = s \mid j_1, \dots, j_k) \\ &= \sum_{i=1}^N \sum_{s \leq (1-2\lambda)r - f_{i,j_k}} \frac{\binom{f'_{i,j_k}}{s} \binom{N-k-f'_{i,j_k}}{r-k-s}}{\binom{N-k}{r-k}} = C, \end{aligned}$$

as required. □

A simple application of Procedure SelectSet gives an algorithm to find a reservoir set.

Lemma 2.18 (Reservoir lemma). *Suppose that $1/n \ll \rho \ll \lambda, \kappa$, that $H = (V, E)$ is a 4-graph of order n which is κ -connecting, and that G is a 2-graph on the same vertex set V with $\delta(G) \geq n - \lambda n$. Then there exists a subset $R \subseteq V$ such that*

$$(a) \quad (1 - 4\rho)\rho n \leq |R| \leq \rho n,$$

$$(b) \quad \text{for every } x \in V \text{ we have } |N_G(x) \cap R| \geq (1 - 35\lambda)|R| \text{ and}$$

$$(c) \quad \text{for every disjoint } p_1, p_2 \in \binom{V}{2} \text{ there are at least } \frac{\kappa}{5}|R| \text{ internally disjoint paths of length at most three in } H[R \cup p_1 \cup p_2] \text{ with ends } p_1 \text{ and } p_2.$$

Moreover, there exists an algorithm, *Procedure SelectReservoir* (H, G, ρ) , which returns such a subset $R \subseteq V$ in time $O(n^{16})$.

Proof. We first define the graphs on which we will use *Procedure SelectSet*. For this set $U := \{\{p, p'\} : p, p' \in \binom{V}{2} \text{ and } p \cap p' = \emptyset\}$ and $W := \binom{V}{4}$. We also set

$$E_1 := \{\{\{p, p'\}, S\} : \{p, p'\} \in U, S \in W \text{ and } H[S \cup p \cup p'] \text{ contains a path with ends } p, p'\},$$

$$E'_1 := \{\{S, S'\} : S, S' \in W \text{ and } S \cap S' \neq \emptyset\}, \text{ and}$$

$$E_2 := \{\{S, S'\} : S, S' \in W \text{ and } \{u, v\} \in E(G) \text{ for all } u \in S, v \in S'\},$$

and define graphs $G_1 := (U \cup W, E_1 \cup E'_1)$ and $G_2 := (W, E_2)$. We now define $M := |U| \leq 3\binom{n}{4}$ and $N := |W| = \binom{n}{4}$, $r := \frac{\rho}{4}n$, $\beta := \kappa$, $\lambda' := 17\lambda$ and $\tau := \rho$. We also define $m := |E'_1|$ and note that we then have $m \leq \frac{1}{2} \cdot \binom{n}{4} \cdot 4 \cdot \binom{n-1}{3}$, from which it follows that

$$\nu := \frac{2mr}{N^2} = \frac{2 \cdot 2\binom{n}{4}\binom{n-1}{3} \cdot \frac{\rho}{4}n}{\binom{n}{4}^2} = 4\rho.$$

Observe that $G_1[U]$ is empty and $G_1[W]$ has precisely m edges. Furthermore, since H is κ -connecting we have $d_{G_1}(\{p, p'\}) \geq \min\{\kappa n^4, \frac{1}{6} \cdot \kappa n^2 \cdot \binom{n-2}{2}\} \geq \kappa \binom{n}{4}$ for every $\{p, p'\} \in U$, or in other words, $|N_{G_1}(u)| \geq \beta N$ for all $u \in U$. Also, since $\delta(G) \geq (1 - \lambda)n$

we have $d_{G_2}(S) \geq \binom{n}{4} - 4\binom{n-1}{3} - 4 \cdot \lambda n \cdot \binom{n-5}{3} \geq (1 - 17\lambda)\binom{n}{4}$ for every $S \in W$, or in other words, $|N_{G_2}(w)| \geq (1 - \lambda')N$ for all $w \in W$. So our chosen graphs and constants satisfy Setup 2.15 with λ' in place of λ , and the conditions of Proposition 2.17 are satisfied (in particular our assumption that $\frac{1}{n} \ll \lambda, \kappa, \rho$ gives $M \leq \frac{1}{8} \cdot \exp(\frac{\tau^2 r}{3\beta})$ and $N \leq \frac{1}{8} \cdot \exp(\frac{\lambda r}{3})$). We may therefore apply Procedure SelectSet to obtain $R' := \text{SelectSet}(G_1, G_2, r, \beta, \lambda', \tau) \subseteq W$ such that

- (a) $(1 - 4\rho)\frac{\rho}{4}n \leq |R'| \leq \frac{\rho}{4}n$,
- (b) R' is an independent set in G_1 ,
- (c) $|N_{G_1}(u) \cap R'| \geq (\kappa - \rho - 4\rho)\frac{\rho}{4}n > \frac{\kappa\rho}{5}n$ for all $u \in U$ and
- (d) $|N_{G_2}(w) \cap R'| \geq (1 - 34\lambda - 4\rho)\frac{\rho}{4}n > (1 - 35\lambda)\frac{\rho}{4}n$ for all $w \in W$.

We now define $R := \bigcup_{S \in R'} S$; it remains to show that R has the desired properties. Indeed, (b) implies that the members of R' are pairwise-disjoint, so $|R| = 4|R'|$, and so (a) immediately implies (i). Now consider some $x \in V$, and arbitrarily choose a set $S \in W$ with $x \in S$. Then by (d) there are at least $(1 - 35\lambda)\frac{\rho}{4}n$ sets $S' \in R'$ such that every $y \in S'$ is a neighbour of x ; since the members of R' are pairwise-disjoint it follows that $|N_G(x) \cap R| \geq (1 - 35\lambda)\rho n \geq (1 - 35\lambda)|R|$, establishing (ii). Finally, (c) implies that for every disjoint $p_1, p_2 \in \binom{V}{2}$ there are at least $\frac{\kappa\rho}{5}n \geq \frac{\kappa}{5}|R|$ sets $S \in R'$ such that $H[S \cup p_1 \cup p_2]$ contains a path of length at most three with ends p_1 and p_2 . Together with the fact that the members of R' are pairwise-disjoint this gives (iii). Furthermore, the dominant term of the running time is that of Procedure SelectSet, which runs in time $O(N^4 + MN^3) = O(n^{16})$. \square

2.4.3 Proof of the absorbing lemma (Lemma 2.6)

We now turn to the proof of the absorbing lemma (Lemma 2.6), for which we make the following general definition of an absorbing structure.

Definition 2.19. Let $\alpha, \beta > 0$ and let $H = (V, E)$ be a 4-graph of order n .

- (a) We say that an ordered octuple $O = (a_1, a_2, c_1, c_2, c_3, c_4, b_1, b_2)$ of distinct vertices of H is an absorbing structure for a pair $p \in \binom{V}{2}$ if there are paths P and P' in H , both with ends $\{a_1, a_2\}$ and $\{b_1, b_2\}$, such that $V(P) = O$ and $V(P') = O \cup p$.
- (b) We say that a pair $p \in \binom{V}{2}$ is β -absorbable if there are at least βn^8 absorbing structures for p in H .
- (c) We say that H is (α, β) -absorbing if at most αn^2 pairs $p \in \binom{V}{2}$ are not β -absorbable.

More specifically, we will work with the two types of absorbing structures given by the next definition.

Definition 2.20. Let $H = (V, E)$ be a 4-graph, and let x and y be distinct vertices of H . We say that an ordered octuple $O = (a_1, a_2, c_1, c_2, c_3, c_4, b_1, b_2)$ of vertices of H , is an absorbing structure of type 1 for the pair $\{x, y\}$ if

$$\{a_1, a_2, c_1, c_2\}, \{c_1, c_2, c_3, c_4\}, \{c_3, c_4, b_1, b_2\}, \{c_1, c_2, x, y\}, \{x, y, c_3, c_4\} \in E .$$

Similarly, we say that O is an absorbing structure of type 2 for $\{x, y\}$ if

$$\{a_1, a_2, c_1, c_2\}, \{c_1, c_2, c_3, c_4\}, \{c_3, c_4, b_1, b_2\}, \{a_1, a_2, x, y\}, \{x, y, c_1, c_4\}, \{c_2, c_3, b_1, b_2\} \in E .$$

The absorbing structures are depicted in Figure 2.1. Observe that if O is an absorbing structure for $\{x, y\}$ of type 1 or type 2, then the sequence $(a_1, a_2, c_1, c_2, c_3, c_4, b_1, b_2)$ forms a path in H with vertex set O and ends $\{a_1, a_2\}$ and $\{b_1, b_2\}$. Moreover, if O is an absorbing structure of type 1 then $(a_1, a_2, c_1, c_2, x, y, c_3, c_4, b_1, b_2)$ is a path in H with vertex set $O \cup \{x, y\}$ and the same ends, whilst if O is an absorbing structure of type 2 then $(a_1, a_2, x, y, c_1, c_4, c_2, c_3, b_1, b_2)$ is a path in H with vertex set $O \cup \{x, y\}$ and the same

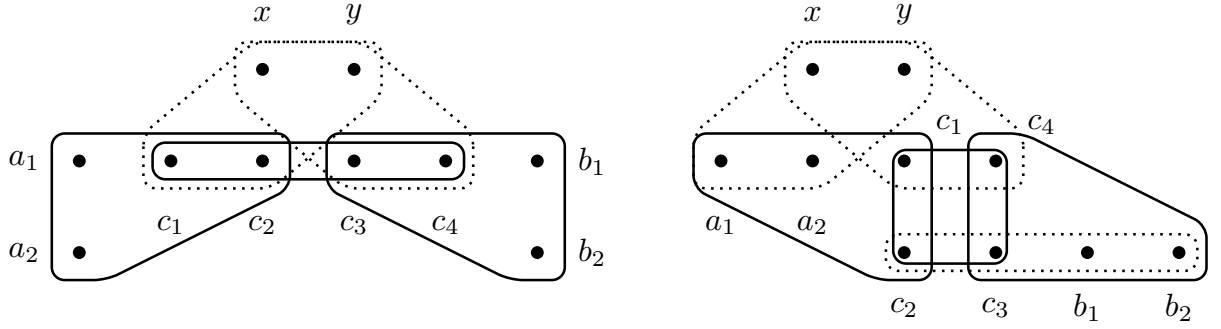


Figure 2.1: Absorbing structures of type 1 (left) and type 2 (right) for the pair $p = \{x, y\}$. In each case the edges of the path not containing p are marked by solid lines, whilst the paths including p use the edges marked by dashed lines together with some of the edges marked by solid lines.

ends. Therefore absorbing structures of type 1 and 2 are indeed absorbing structures according to Definition 2.19. We now show that most pairs in non-odd-extremal 4-graphs are contained in many absorbing structures.

Lemma 2.21. *Suppose that $1/n \ll \varepsilon \ll \beta \ll \alpha \ll c$ and that $H = (V, E)$ is a 4-graph of order n with $\delta(H) \geq n/2 - \varepsilon n$. If H is not (α, β) -absorbing, then H is c -odd-extremal. Moreover, there exists an algorithm $\text{Procedure OddPartition}(H)$ which returns a c -odd-extremal bipartition $\{A, B\}$ of V in time $O(n^6)$.*

Proof. We introduce constants $\omega, \omega', \varphi, \varepsilon_2, \psi, \varepsilon_1 > 0$ such that

$$\frac{1}{n} \ll \varepsilon \ll \beta \ll \omega \ll \omega' \ll \varphi \ll \varepsilon_2 \ll \psi \ll \varepsilon_1 \ll \alpha \ll c.$$

Furthermore, for a pair $\{x, y\} \in \binom{V}{2}$ we define

$$T_{x,y} = \{\{e_1, e_2, e_3\} \subseteq E : e_1 \cap e_2 = \{x, y\} \text{ and } (e_1 \cup e_2) \setminus \{x, y\} = e_3\}.$$

We then can make the following observation.

Claim 2.22. *If $|T_{x,y}| \geq \beta n^4$, then $\{x, y\}$ is β -absorbable. Moreover, there exists an edge $\{x, y, x', y'\}$ of H such that $|T_{x,y}| < \beta n^4$ and $|N_H(x, y) \cap N_H(x', y')| < \frac{\omega}{2} \binom{n}{2}$.*

Proof of Claim 2.22. Suppose that $|T_{x,y}| \geq \beta n^4$ for some $\{x, y\} \in \binom{V}{2}$. Then there are at least $8\beta n^4$ ordered sextuples $S = (c_1, c_2, x, y, c_3, c_4)$ such that each of $\{c_1, c_2, x, y\}$, $\{x, y, c_3, c_4\}$ and $\{c_1, c_2, c_3, c_4\}$ is an edge of H . By the fact that $\delta(H) \geq (\frac{1}{2} - \varepsilon)n$, for each such sextuple S there are at least $(n - 6)((\frac{1}{2} - \varepsilon)n - 4) \geq \frac{3n^2}{8}$ ways to choose an ordered pair (a_1, a_2) of vertices of $V \setminus S$ such that $\{a_1, a_2, c_1, c_2\} \in E(H)$, and then at least $(n - 8)((\frac{1}{2} - \varepsilon)n - 6) \geq \frac{3n^2}{8}$ ways to choose an ordered pair (b_1, b_2) of vertices of $V \setminus (S \cup \{a_1, a_2\})$ such that $\{b_1, b_2, c_3, c_4\} \in E(H)$. Overall this gives at least $8\beta n^4 \cdot (\frac{3n^2}{8})^2 > \beta n^8$ -many 10-tuples $(a_1, a_2, c_1, c_2, x, y, c_3, c_4, b_1, b_2)$, each of which is an absorbing structure of type 1 for $\{x, y\}$. Hence $\{x, y\}$ is β -absorbable, proving the first statement of the claim.

Since by assumption H is not (α, β) -absorbing, we may choose a pair $\{x, y\} \in \binom{V}{2}$ which is not β -absorbable, so $|T_{x,y}| < \beta n^4$. Since $|N_H(x, y)| \geq (\frac{1}{2} - 2\varepsilon) \binom{n}{2}$, there must then exist a pair $\{x', y'\} \in N_H(x, y)$ such that $|N_H(x, y) \cap N_H(x', y')| < \frac{\omega}{2} \binom{n}{2}$, as otherwise we would have $|T_{x,y}| \geq \frac{1}{2} \cdot (\frac{1}{2} - 2\varepsilon) \binom{n}{2} \cdot \frac{\omega}{2} \binom{n}{2} > \beta n^4$. This gives the desired edge $\{x, y, x', y'\}$ of H . \diamond

Fix an edge $e := \{x, y, x', y'\}$ of H as in Claim 2.22. We now colour the edges of the complete 2-graph K on V in the following way: for $\{a, b\} \in \binom{V}{2}$ we say that

$$\{a, b\} \text{ is } \begin{cases} \text{red} & \text{if } \{x, y, a, b\} \in E(H), \{x', y', a, b\} \notin E(H) \text{ and } a, b \notin e, \\ \text{blue} & \text{if } \{x, y, a, b\} \notin E(H), \{x', y', a, b\} \in E(H) \text{ and } a, b \notin e, \\ \text{uncoloured} & \text{otherwise.} \end{cases}$$

Note that if $\{a, b\}$ and $\{a', b'\}$ are disjoint red edges of K such that $\{a, b, a', b'\} \in E$, then $\{\{x, y, a, b\}, \{x, y, a', b'\}, \{a, b, a', b'\}\}$ is an element of $T_{x,y}$. In such a case we call the edge $\{a, b, a', b'\}$ of H a *red hyperedge*. Also, we say that a vertex of V is *normal* if we have

both $d_{\text{red}}(v) \geq (\frac{1}{2} - \omega')n$ and $d_{\text{blue}}(v) \geq (\frac{1}{2} - \omega')n$ in K . We observe that our colouring of K has the following properties.

- (A) There are at most $|T_{x,y}| < \beta n^4$ red hyperedges of H ,
- (B) at least $(\frac{1}{2} - \omega) \binom{n}{2}$ edges of K are coloured red,
- (C) at least $(\frac{1}{2} - \omega) \binom{n}{2}$ edges of K are coloured blue,
- (D) at most $2\omega \binom{n}{2}$ edges of K are uncoloured, and
- (E) all but at most $\omega'n$ vertices are normal.

Indeed, (A) follows immediately from our definition of a red hyperedge and our choice of e . For (B) note that there are at least $\frac{1}{2}(n-2)\delta(H)$ pairs $\{a, b\}$ for which $\{x, y, a, b\}$ is an edge of H . By choice of e at most $\frac{\omega}{2} \binom{n}{2}$ of these pairs have $\{x', y', a, b\} \in E(H)$, whilst at most $4n$ such pairs have $a \in e$ or $b \in e$, so at least $\frac{1}{2}(n-2)\delta(H) - \frac{\omega}{2} \binom{n}{2} - 4n \geq (\frac{1}{2} - \omega) \binom{n}{2}$ edges of K are coloured red. Essentially the same argument proves (C), and (D) follows immediately from (B) and (C). Finally, observe that for any $v \in V \setminus e$ we have $d_H(\{x, y, v\}), d_H(\{x', y', v\}) \geq (\frac{1}{2} - \varepsilon)n$, so for any $v \in V$ we have $d_{\text{red}}(v), d_{\text{blue}}(v) \leq (\frac{1}{2} + \varepsilon)n$. So if there are more than $\frac{\omega'}{2}n$ vertices with $d_{\text{red}}(v) < (\frac{1}{2} - \omega')n$, then the total number of red edges is at most $\frac{1}{2}(\frac{\omega'}{2}n \cdot (\frac{1}{2} - \omega')n + (1 - \frac{\omega'}{2})n \cdot (\frac{1}{2} + \varepsilon)n) < (\frac{1}{2} - \omega) \binom{n}{2}$, contradicting (B). A similar argument shows that there cannot be more than $\frac{\omega'}{2}n$ vertices with $d_{\text{blue}}(v) < (\frac{1}{2} - \omega')n$, establishing (E).

For any triangle T in K with vertex set $\{u, v, w\}$, we write $N_{\text{red}}(T)$ (respectively $N_{\text{blue}}(T)$) for the set of vertices $x \in V$ for which each of ux, vx and wx is a red (respectively blue) edge of K . We then have the following claim.

Claim 2.23. *If the number of red triangles in K is at least φn^3 , then there exists a red triangle T in K with vertex set $\{u, v, w\}$ such that*

(a) $|N_{\text{blue}}(T)| > (\frac{1}{2} - \varphi)n$ and $|N_{\text{red}}(T)| > (\frac{1}{2} - \varphi)n$,

(b) at most φn vertices $x \in N_{\text{red}}(T)$ have $|N_H(v, w, x) \cap N_{\text{red}}(T)| > \varphi n$, and

(c) there are at most $9\varphi \binom{n}{2}$ red edges of K between $N_{\text{red}}(T)$ and $V \setminus N_{\text{red}}(T)$.

Proof of Claim 2.23. By (E) there are at most $\omega'n \cdot n^2 < \frac{\varrho}{3}n^3$ triangles in K which contain a vertex which is not normal, and by (A) there are at most $\frac{\varrho}{3}n^3$ triangles in K which are contained in more than $\omega'n$ red hyperedges, as otherwise there would be at least $\frac{1}{4} \cdot \frac{\varrho}{3}n^3 \cdot \omega'n > \beta n^4$ red hyperedges in total. Similarly, by (A) all but at most $\omega'n^2$ edges of K are contained in at most $\omega'n^2$ red hyperedges, as otherwise there would be at least $\frac{1}{6} \cdot \omega'n^2 \cdot \omega'n^2 > \beta n^4$ red hyperedges in total. We call these edges of K *normal*; observe that at most $\omega'n^2 \cdot n < \frac{\varrho}{3}n^3$ triangles in K contain an edge which is not normal. So we can choose a red triangle T in K with vertex set $\{u, v, w\}$ which contains only normal vertices and normal edges and which is contained in at most $\omega'n$ red hyperedges. Then $|N_H(T) \cap N_{\text{red}}(x)| \leq \omega'n$ for each $x \in \{u, v, w\}$, as otherwise $\{u, v, w\}$ would be contained in more than $\omega'n$ red hyperedges. Also, since u, v and w are normal, for each $x \in \{u, v, w\}$ at most $2\omega'n$ vertices are in neither $N_{\text{red}}(x)$ nor $N_{\text{blue}}(x)$. It follows that for each $x \in \{u, v, w\}$ at most $3\omega'n$ of the vertices inside $N_H(T)$ are not in $N_{\text{blue}}(x)$, so $|N_{\text{blue}}(T)| \geq |N_H(T)| - 3 \cdot 3\omega'n \geq \delta(H) - 9\omega'n > (\frac{1}{2} - \varphi)n$. Similarly, for each $x \in \{u, v, w\}$ we have $|(V \setminus N_H(T)) \cap N_{\text{red}}(x)| \geq d_{\text{red}}(x) - \omega'n \geq (\frac{1}{2} - 2\omega')n$; since $|V \setminus N_H(T)| \leq n - \delta(H) \leq (\frac{1}{2} + \varepsilon)n$ it follows that $|N_{\text{red}}(T)| \geq (\frac{1}{2} + \varepsilon)n - 3 \cdot (2\omega' + \varepsilon)n \geq (\frac{1}{2} - \varphi)n$, proving (a).

Now observe that if (b) does not hold, then there are at least $\frac{1}{2}(\varphi n)^2 > \omega'n^2$ pairs $\{x, y\}$ such that $x \in N_{\text{red}}(T)$ and $y \in N_H(v, w, x) \cap N_{\text{red}}(T)$. Each such pair yields a red hyperedge $\{v, w, x, y\}$ containing $\{v, w\}$, contradicting the fact that $\{v, w\}$ is normal. So (b) holds, and together with (a) and inclusion-exclusion we find that all but at most φn vertices $x \in N_{\text{red}}(T)$ have $|N_{\text{red}}(T) \cup N_H(v, w, x)| \geq (\frac{1}{2} - \varphi)n + \delta(H) - \varphi n \geq n - 3\varphi n$. So

there are at most $4\varphi n^2$ pairs $\{x, y\}$ with $x \in N_{\text{red}}(T)$ and $y \in V \setminus (N_{\text{red}}(T) \cup N_H(v, w, x))$. On the other hand, any red edge $\{x, y\}$ of K with $x \in N_{\text{red}}(T)$ and $y \in N_H(v, w, x)$ yields a red hyperedge $\{v, w, x, y\}$ containing $\{v, w\}$, so there are at most $\omega' n^2$ such edges. It follows that K has at most $4\varphi n^2 + \omega' n^2 \leq 9\varphi \binom{n}{2}$ red edges between $N_{\text{red}}(T)$ and $V \setminus N_{\text{red}}(T)$, proving (c). \diamond

Suppose first that there exists a red triangle $\{u, v, w\}$ in K with the properties given in Claim 2.23, and define $A := N_{\text{red}}(T)$ and $B := V \setminus A$. Then by Claim 2.23(a) we have $(\frac{1}{2} - \varphi)n \leq |A|, |B| \leq (\frac{1}{2} + \varphi)n$, so certainly we have $(\frac{1}{2} - c)n \leq |A|, |B| \leq (\frac{1}{2} + c)n$. Furthermore, by (B) and Claim 2.23(c) there are at least $(\frac{1}{2} - \omega) \binom{n}{2} - 9\varphi \binom{n}{2} > (\frac{1}{2} - 10\varphi) \binom{n}{2}$ red edges in $K[A] \cup K[B]$. Together with the fact that there are $\binom{|A|}{2} + \binom{|B|}{2} \leq \binom{(1/2+\varphi)n}{2} + \binom{(1/2-\varphi)n}{2} \leq (\frac{1}{2} + \varphi) \binom{n}{2}$ edges in $K[A] \cup K[B]$, this implies that there are at most $11\varphi \binom{n}{2}$ edges in $K[A] \cup K[B]$ which are not red. So if there are more than $c \binom{n}{4}$ edges $e \in E$ such that $|e \cap A|$ is even, then there are at least $c \binom{n}{4} - 11\varphi \binom{n}{2} \cdot (\frac{1}{2} + \varphi) \binom{n}{2} > \beta n^4$ red hyperedges, contradicting (A). We may therefore conclude that there are at most $c \binom{n}{4}$ edges $e \in E$ such that $|e \cap A|$ is even, whereupon $\{A, B\}$ is the desired c -odd-extremal bipartition of V .

Now assume that such a red triangle does not exist. We may then by Claim 2.23 assume that there are at most φn^3 red triangles in K . So we may choose a normal vertex $v \in V$ which is contained in at most $4\varphi n^2$ red triangles, as otherwise K would have at least $\frac{1}{3}(1 - \omega')n \cdot 4\varphi n^2 > \varphi n^3$ red triangles in total. There are then at most $4\varphi n^2$ red edges inside $N_{\text{red}}(v)$, so by (E) and the fact that v is normal there are at least

$$\sum_{u \in N_{\text{red}}(v)} d_{\text{red}}(u) - 2 \cdot 4\varphi n^2 \geq (\frac{1}{2} - 2\omega')n \cdot (\frac{1}{2} - \omega')n - 8\varphi n^2 > (\frac{1}{2} - 17\varphi) \binom{n}{2}$$

red edges between $N_{\text{red}}(v)$ and $V \setminus N_{\text{red}}(v)$, so at most $\frac{n^2}{4} - (\frac{1}{2} - 17\varphi) \binom{n}{2} \leq 18\varphi \binom{n}{2}$ edges between $N_{\text{red}}(v)$ and $V \setminus N_{\text{red}}(v)$ are not red. It follows that at most $\varepsilon_2 \binom{n}{4}$ edges $f \in E(H)$

have $|N_{\text{red}}(v) \cap f| = 2$, as otherwise there would be at least $\varepsilon_2 \binom{n}{4} - 18\varphi \binom{n}{2} \cdot \frac{n^2}{4} > \beta n^4$ red hyperedges in total, contradicting (A). So if $H[N_{\text{red}}(v)]$ and $H[V \setminus N_{\text{red}}(v)]$ each contain at most $\varepsilon_1 \binom{n}{4}$ edges of H , then the total number of edges $f \in E(H)$ for which $|f \cap N_{\text{red}}(v)|$ is even is at most $(2\varepsilon_1 + \varepsilon_2) \binom{n}{4} \leq c \binom{n}{4}$, and then taking $A := N_{\text{red}}(v)$ and $B := V \setminus A$ gives the desired c -odd-extremal bipartition $\{A, B\}$ of V , as $(\frac{1}{2} - c)n \leq |N_{\text{red}}(v)| \leq (\frac{1}{2} + c)n$ since v is normal.

This leaves only the cases in which either $H[N_{\text{red}}(v)]$ or $H[V \setminus N_{\text{red}}(v)]$ contains more than $\varepsilon_1 \binom{n}{4}$ edges of H . If the former holds then we set $A := N_{\text{red}}(v)$ and $B := V \setminus N_{\text{red}}(v)$, and otherwise we set $A := V \setminus N_{\text{red}}(v)$ and $B := N_{\text{red}}(v)$; either way this results in a bipartition $\{A, B\}$ of V such that

$$(F) \quad (\frac{1}{2} - \omega')n \leq |A|, |B| \leq (\frac{1}{2} + \omega')n \text{ (since } v \text{ is normal),}$$

$$(G) \quad \text{at least } \varepsilon_1 \binom{n}{4} \text{ edges } f \in E(H) \text{ have } |f \cap A| = 4, \text{ and}$$

$$(H) \quad \text{at most } \varepsilon_2 \binom{n}{4} \text{ edges } f \in E(H) \text{ have } |f \cap A| = 2.$$

In the remaining part of the proof we show that these conditions imply that at most $\alpha \binom{n}{2}$ pairs of vertices are not β -absorbable. This contradicts our assumption that H is not (α, β) -absorbing, and so completes the proof. Recall that a pair is *split* if it has one vertex in A and one in B , and *connate* otherwise. Additionally, we say that a pair $p \in \binom{V}{2}$ is *good* if there are at least $(\frac{1}{2} - \psi) \binom{n}{2}$ pairs $p' \in \binom{V}{2}$ such that $p \cup p'$ is an odd edge of H (so a split pair is good if it forms an edge with most connate pairs, and a connate pair is good if it forms an edge with most split pairs).

Claim 2.24. *At most $\psi \binom{n}{2}$ pairs in $\binom{V}{2}$ are not good.*

Proof of Claim 2.24. First consider a split pair $p \in \binom{V}{2}$ which is not good. Then there are at least $\frac{1}{2}(n-2)\delta(H) \geq (\frac{1}{2} - 2\varepsilon) \binom{n}{2}$ pairs p' for which $p \cup p'$ is an edge of H . Since p is not good it follows that p is contained in at least $(\psi - 2\varepsilon) \binom{n}{2} \geq \frac{\psi}{2} \binom{n}{2}$ even edges of

H , each of which must have precisely two vertices in A (because p is a split pair). So at most $\frac{\psi}{3}\binom{n}{2}$ split pairs in $\binom{V}{2}$ are not good, as otherwise in total there would be at least $\frac{1}{6} \cdot \frac{\psi}{3}\binom{n}{2} \cdot \frac{\psi}{2}\binom{n}{2} > \varepsilon_2\binom{n}{4}$ edges of H with precisely two vertices in A , contradicting (H).

Next observe that for all but at most $\frac{\psi}{3}\binom{n}{2}$ pairs $p \in \binom{B}{2}$ there are at most $\frac{\psi}{3}n$ vertices $w \in A$ such that $|N_H(p \cup \{w\}) \cap A| > \frac{\psi}{3}n$, as otherwise there would be at least $\frac{1}{12} \cdot \frac{\psi}{3}\binom{n}{2} \cdot \frac{\psi}{3}n \cdot \frac{\psi}{3}n > \varepsilon_2\binom{n}{4}$ edges of H with precisely two vertices in A , contradicting (H). For each such p there are at least $(|A| - \frac{\psi}{3}n)(\delta(H) - \frac{\psi}{3}n) \geq (\frac{1}{2} - \omega' - \frac{\psi}{3})(\frac{1}{2} - \varepsilon - \frac{\psi}{3})n^2 > (\frac{1}{2} - \psi)\binom{n}{2}$ split pairs p' such that $p \cup p'$ is an edge of H ; in other words, each such p is good. The same argument with the roles of A and B reversed shows that all but at most $\frac{\psi}{3}\binom{n}{2}$ pairs $p \in \binom{A}{2}$ are good. \diamond

We now show that any good pair is β -absorbable. First consider any good pair $\{x, y\}$ with $x \in A$ and $y \in B$. Since $\{x, y\}$ is good, at most $\binom{|A|}{2} + \binom{|B|}{2} - (\frac{1}{2} - \psi)\binom{n}{2} \leq 2\psi\binom{n}{2}$ pairs $p' \in \binom{A}{2}$ are not in $N_H(\{x, y\})$, and so at most $2\psi\binom{n}{2} \cdot \binom{n}{2}$ sets $S \in \binom{A}{4}$ contain such a pair p' . By (G) it follows that at least $\varepsilon_1\binom{n}{4} - \psi\binom{n}{2}^2 \geq \beta n^4$ edges $\{a, b, c, d\} \in H[A]$ are such that $\{x, y, a, b\}$ and $\{x, y, c, d\}$ are both edges of H . Each such edge yields an element of $T_{x,y}$, so $|T_{x,y}| \geq \beta n^4$, and so by Claim 2.22 $\{x, y\}$ is β -absorbable.

Now consider any good pair $\{x, y\}$ with $x, y \in A$ or $x, y \in B$. Choose (not necessarily distinct) vertices $a_1, c_1, c_2, c_3, b_1, b_2 \in A$ and $a_2, c_4 \in B$ uniformly at random, and observe that with probability at least $1 - \frac{100}{n}$ these eight vertices are distinct from each other and from x and y . Furthermore, since $\{x, y\}$ is good and $|A||B| \leq \frac{n^2}{4}$ we have that

$$\mathbb{P}(\{x, y, a_1, a_2\} \in E(H)) \geq \frac{1}{|A||B|} \cdot \left(\frac{1}{2} - \psi\right) \binom{n}{2} \geq 1 - 3\psi,$$

and likewise the probability that $\{x, y, c_1, c_4\}$ is an edge of H is at least $1 - 3\psi$. Also,

$$\begin{aligned} & \mathbb{P}(\{a_1, a_2, c_1, c_2\} \in E(H)) \\ & \geq \mathbb{P}(\{c_1, c_2\} \text{ is good}) \cdot \mathbb{P}(\{a_1, a_2, c_1, c_2\} \in E(H) \mid \{c_1, c_2\} \text{ is good}) \\ & \geq \frac{2}{|A|^2} \left(\binom{|A|}{2} - \psi \binom{n}{2} \right) \cdot \frac{1}{|A||B|} \left(\frac{1}{2} - \psi \right) \binom{n}{2} \geq 1 - 8\psi, \end{aligned}$$

where we use (F) for the final inequality. Exactly the same calculation shows that the probabilities that $\{c_1, c_2, c_3, c_4\}$ and $\{c_3, c_4, b_1, b_2\}$ are edges of H are each at least $1 - 8\psi$. Finally, by (G) the probability that $\{c_2, c_3, b_1, b_2\}$ is an edge of H is at least $\frac{4!}{|A|^4} \cdot e(H[A]) \geq \varepsilon_1$. Taking a union bound we find that with probability at least $\varepsilon_1 - 30\psi - \frac{100}{n} \geq \frac{\varepsilon_1}{2}$ all of these events occur, in which case $(a_1, a_2, c_1, c_2, c_3, c_4, b_1, b_2)$ is an absorbing structure of type 2 for $\{x, y\}$. So in total there are at least $\frac{\varepsilon_1}{2}|A|^6|B|^2 \geq \beta n^8$ such absorbing structures for $\{x, y\}$, so $\{x, y\}$ is β -absorbable.

We conclude by Claim 2.24 that at most $\psi \binom{n}{2} < \alpha n^2$ pairs in $\binom{V}{2}$ are not β -absorbable. This contradicts our assumption that H is not (α, β) -absorbing and so completes the proof.

Finally, note that each step of the proof directly translates to an algorithm which returns the desired bipartition in the claimed running-time. \square

We are now ready to prove our absorbing lemma (Lemma 2.6) which guarantees the existence of an absorbing path in the non-extremal case. In fact, we actually prove the following stronger statement, in which the assumption that H is non-extremal is replaced by the assumption that H is absorbing and connecting, and which concludes that we can find an absorbing path P in polynomial time (furthermore, the modified condition (iii) allows the absorption of pairs into P to be done greedily). Since Lemmas 2.13 and 2.21 imply that a non-extremal graph H must be absorbing and connecting, this is indeed a stronger statement than Lemma 2.6. We could instead prove Lemma 2.6 directly by a standard random selection argument; we avoid this approach since we will also use the

polynomial-time algorithm given by Lemma 2.25 in the proof of Theorem 1.12.

Lemma 2.25. *Suppose that $1/n \ll \varepsilon \ll \gamma \ll \beta \ll \alpha \ll \lambda \ll \kappa, \mu$. Let H be a 4-graph on n vertices with $\delta(H) \geq n/2 - \varepsilon n$ which is κ -connecting and (α, β) -absorbing. Then there is a path P in H and a graph G on $V(H)$ with the following properties.*

(i) P has at most μn vertices.

(ii) Every vertex of $V(H) \setminus V(P)$ is contained in at least $n - \lambda n$ edges of G .

(iii) For any edge e of G which does not intersect $V(P)$ there are at least $2\gamma n$ vertex-disjoint segments of P which are absorbing structures for e .

Furthermore, there exists an algorithm, *Procedure AbsorbingPath(H)*, which returns such a path P and graph G in time $O(n^{32})$.

Proof. Let $W := V(H)^8$, let U be the set of all β -absorbable pairs of vertices of H , and define the graph $G := (V(H), U)$. Furthermore set $V_1 := U \cup W$ and let E_1 be the set of all pairs $\{p, T\}$ with $p \in U, T \in W$ for which T is an absorbing structure for p . We then set $E'_1 := \{\{T, T'\} : T, T' \in W \text{ and } T \cap T' \neq \emptyset\}$ and define the graph $G_1 := (V_1, E_1 \cup E'_1)$. Set $M := |U|$, so $M \geq \binom{n}{2} - \alpha n^2 > (1 - 3\alpha)\binom{n}{2}$ since H is (α, β) -absorbing, and set $N := |W| = n^8$, $m := |E'_1| < 64n^{15}$, $r = \beta^2 n$ and $\nu = \frac{2mr}{N^2} < 128\beta^2$. Then, taking G_2 to be the empty graph on vertex set W , the conditions of Setup 2.15 and Proposition 2.17 are satisfied (with β playing the same role here as there, and with 1 and β^2 in place of λ and τ respectively), since each pair $p \in U$ is β -absorbable and so has $d_{G_1}(p) \geq \beta n^8$. The call of *Procedure SelectSet($G_1, \emptyset, \beta^2 n, \beta, 1, \beta^2$)* then returns a family $\mathcal{T}' \subseteq W$ of ordered octuples of vertices of H which is an independent set in G_1 such that $(1 - 128\beta^2)\beta^2 n \leq |\mathcal{T}'| \leq \beta^2 n$ and $|\mathcal{T}' \cap N_{G_1}(p)| \geq (\beta - 129\beta^2)|\mathcal{T}'| > \frac{1}{2}\beta^3 n$ for each $p \in U$. If we now delete from \mathcal{T}' every $T \in \mathcal{T}'$ which is not an absorbing structure for some β -absorbable pair $\{x, y\} \in \binom{V(H)}{2}$, then we obtain a pairwise-disjoint family \mathcal{T} satisfying the following properties:

(A) $|\mathcal{T}| \leq \beta^2 n$,

(B) for any β -absorbable pair $\{x, y\} \in \binom{V(H)}{2}$ the family \mathcal{T} contains at least $\frac{1}{2}\beta^3 n$ absorbing structures for $\{x, y\}$, and

(C) every $T \in \mathcal{T}$ is an absorbing structure for some β -absorbable pair $\{x, y\} \in \binom{V(H)}{2}$.

Enumerate the members of \mathcal{T} as T_1, \dots, T_q , so $q \leq \beta^2 n$ by (A), and for each $1 \leq i \leq q$ let a_1^i, a_2^i, b_1^i and b_2^i be the first, second, seventh and eighth elements of T_i respectively. Then by (C) and the definition of an absorbing structure we may choose, for each i , a path P_i in H with vertex set T_i and with ends $\{a_1, a_2\}$ and $\{b_1, b_2\}$. Let $Q = \bigcup_{i=1}^q T_i$, so $|Q| = 8q \leq 8\beta^2 n$, and let $X \subseteq V \setminus Q$ be the set of vertices not in Q which lie in fewer than $(1 - \lambda)n$ -many β -absorbable pairs. We must have $|X| \leq \lambda n$, as otherwise there would be at least $\frac{1}{2} \cdot \lambda n \cdot (\lambda n - 1) > \alpha \binom{n}{2}$ pairs in H which are not β -absorbable, contradicting the fact that H is (α, β) -absorbing.

We now greedily construct a path P_0 containing every vertex of X . For this write $X = \{x_1, \dots, x_t\}$, so $t = |X| \leq \lambda n$, and greedily choose distinct vertices $y_1, \dots, y_t \in V \setminus (Q \cup X)$ such that $\{x_{i-1}, y_{i-1}, x_i, y_i\}$ is an edge of H for each $2 \leq i \leq t$. This is possible since y_1 can be any vertex of $V \setminus (Q \cup X)$, and when choosing y_i for $2 \leq i \leq t$ there are at least $d_H(x_{i-1}, y_{i-1}, x_i) - |Q| - |X| - (i - 1) \geq \delta(H) - 8\beta^2 n - \lambda n - t > 0$ suitable choices available. Having done so, we let P_0 be the path $(x_1, y_1, \dots, x_t, y_t)$ in H , and set $b_0^1 := x_t$ and $b_0^2 := y_t$. Note that P_0 has $2t \leq 2\lambda n$ vertices and that the paths P_0, P_1, \dots, P_q are pairwise vertex-disjoint.

To complete the proof we use the fact that H is κ -connecting to greedily choose paths Q_i of length at most three which join the paths P_i together into a single path. Suppose for this that we have already chosen paths Q_1, \dots, Q_{i-1} for some $1 \leq i \leq q$, and set $A_i := (Q \cup V(P_0) \cup \bigcup_{j=1}^{i-1} Q_j) \setminus \{b_1^{i-1}, b_2^{i-1}, a_1^i, a_2^i\}$. Since H is κ -connecting, there are either at least κn^2 paths of length two or at least κn^4 paths of length three in H with ends

$\{b_1^{i-1}, b_2^{i-1}\}$ and $\{a_1^i, a_2^i\}$. In either case, since $|A_i| \leq 8\beta^2 n + 2\lambda n + 4q < \kappa n$, at least one of these paths does not intersect A_i . Arbitrarily choose such a path and call it Q_i . Having proceeded in this manner to find paths Q_1, \dots, Q_q we define $P := P_0 Q_1 P_1 \dots Q_q P_q$ and observe that P is a path in H . It remains only to show that P has the desired properties. Indeed, as just shown P has at most $\kappa n \leq \mu n$ vertices, so (i) holds. For (ii), recall that any edge e of G is a β -absorbable pair in H and that by construction of P_0 the paths P included every vertex which was in fewer than $(1 - \lambda)n$ -many β -absorbable pairs. Finally, by (B) there are at least $\frac{1}{2}\beta^3 n \geq 2\gamma n$ paths P_i for which $T_i = V(P_i)$ is an absorbing structure for e , and these paths P_i are vertex-disjoint segments of P . For the running time note that the dominant term is the call of the Procedure SelectSet with $O(N^4 + MN^3) = O(n^{32})$. \square

2.4.4 Proof of the long cycle lemma (Lemma 2.7)

Now we can turn to the proof of Lemma 2.7 for which we need the following result of Erdős [35]. We say that a k -graph H is k -partite if its vertex set may be partitioned into vertex classes V_1, \dots, V_k such that $|e \cap V_i| = 1$ for every $e \in E(H)$ and every $i \in [k]$. We say that H is complete k -partite if additionally every set $e \subseteq V(H)$ such that $|e \cap V_i| = 1$ for every $i \in [k]$ is an edge of H .

Theorem 2.26 ([35]). *Suppose that $1/n \ll d, 1/f, 1/k$. Let F be a k -partite k -graph on f vertices. If H is a k -graph on n vertices with $e(H) \geq d \binom{n}{k}$, then H contains a copy of F . Moreover, such a copy can be found in time $O(n^k)$.*

Actually the original version of this theorem did not consider the running time, but this can be derived by a straightforward argument. First we restrict to a constant size subgraph H' of H whose density is similar to that of H , and then we find a copy of F in H' by exhaustive search. The existence of such a subgraph can be established by a simple probabilistic argument, and this argument can be derandomised to give an algorithm which finds a subgraph H' with density at least as large as that of H .

We will also make use of the following observation of Rödl, Ruciński and Szemerédi [83] (they did not mention the running time, but this follows immediately from their proof).

Theorem 2.27 ([83]). *Given $a > 0$ and $k \geq 2$, every k -graph F on m vertices and with at least $a \binom{m}{k}$ edges contains a tight path on at least am/k vertices. Moreover, such a path can be found in time $O(n^k)$.*

Note that deleting every other edge of a tight path on s vertices in a 4-graph yields a 2-path on at least $s - 1$ vertices, so for $k = 4$ we may replace ‘tight path’ by ‘path’ (*i.e.* 2-path) and am/k by $am/k - 1$ in the statement of Theorem 2.27.

We are now ready to prove our long cycle lemma, Lemma 2.7. Again we actually prove a stronger statement, Lemma 2.28, which assumes instead that H is connecting and states that we can find the cycle C in polynomial time. Since by Lemma 2.13 any 4-graph which is not even-extremal is connecting, this is indeed a stronger statement. Our proof of Lemma 2.28 is based on the proof given by Karpiński, Ruciński and Szymańska [62, Fact 4] for tight cycles, which in turn was based on a similar lemma for tight cycles in 3-graphs given by Rödl, Ruciński and Szemerédi [83, Lemma 2.2]. The principal differences are that our minimum codegree assumption is weaker, and that our absorbing lemma also requires us to consider the auxiliary graph G , in which we must find a perfect matching among the vertices not used in C .

Lemma 2.28. *Suppose that $1/n \ll \varepsilon \ll \gamma \ll \lambda \leq \mu \ll \kappa$ and that n is even. Let $H = (V, E)$ be a 4-graph of order n with $\delta(H) \geq n/2 - \varepsilon n$ which is κ -connecting. Also let P_0 be a 2-path in H on at most μn vertices, and let G be a 2-graph on V such that each vertex $v \in V \setminus V(P_0)$ has $d_G(v) \geq (1 - \lambda)n$. Then there exists a 2-cycle C in H on at least $(1 - \gamma)n$ vertices such that P_0 is a segment of C and $G[V \setminus V(C)]$ contains a perfect matching. Moreover, there exists an algorithm, Procedure $\text{LongCycle}(H, G, P_0)$, which returns such a cycle C in time $O(n^{16})$.*

Proof. First, we introduce a constant $D > 0$ such that

$$\frac{1}{n} \ll \frac{1}{D} \ll \varepsilon \ll \gamma \ll \lambda \leq \mu \ll \kappa.$$

Set $V' = V \setminus V(P_0)$, $H' = H[V']$, $G' = G[V']$ and $n' = |V'|$, so $n' \geq (1 - \mu)n$, $\delta(G') \geq (1 - \lambda - \mu)n \geq (1 - 2\mu)n'$ and $\delta(H') \geq (\frac{1}{2} - \varepsilon - \mu)n \geq (\frac{1}{2} - 2\mu)n'$. Also it follows from the definition of κ -connecting that H' is $\frac{\kappa}{2}$ -connecting. So by Lemma 2.18 (with $2\gamma/3, n', \kappa/2$ and 2μ in place of ρ, n, κ and λ respectively) we can choose a set $R \subseteq V'$ with $\frac{3\gamma}{5}n \leq (1 - \frac{8}{3}\gamma)\frac{2\gamma}{3}n' \leq |R| \leq \frac{2\gamma}{3}n' \leq \frac{2\gamma}{3}n$ such that for any $x \in V'$ we have $|N_{G'}(x) \cap R| \geq (1 - 70\mu)|R| \geq \frac{4\gamma}{7}n$ and for every disjoint $p_1, p_2 \in \binom{V'}{2}$ there are at least $\frac{\kappa}{10}|R| \geq \frac{\kappa\gamma}{20}n$ internally disjoint paths in $H'[R \cup p_1 \cup p_2]$ of length at most three whose ends are p_1 and p_2 .

We first extend P_0 to a path P'_0 in H by adding a single edge at each end. The purpose of this is that the ends of P'_0 will then be pairs in H' to which we can apply the fact that H' is κ -connecting. So let $\{u_1, u_2\}$ and $\{u_3, u_4\}$ be the ends of P_0 . Then since $|V' \setminus R| \geq (1 - 2\mu)n$ and $\delta(H) \geq (\frac{1}{2} - \varepsilon)n$ we may choose distinct vertices $u'_1, u'_2, u'_3, u'_4 \in V' \setminus R$ such that $u'_2 \in N(u_1, u_2, u'_1)$ and $u'_4 \in N(u_3, u_4, u'_3)$. This gives the desired path $P'_0 = (u'_1, u'_2)P_0(u'_3, u'_4)$. Write $q := \{u'_1, u'_2\}$, so q is an end of P'_0 .

We next proceed by an iterative process to extend P'_0 to a path on at least $(1 - \gamma)n$ vertices in H . At any point in this process we write P for the path we have built so far (so initially we take $P = P'_0$), and write $L := V \setminus (V(P) \cup R)$ and $R' := R \setminus V(P)$ (so at any point the sets $V(P)$, R' and L partition V). Moreover, throughout the process P'_0 will be a segment of P which shares an end in common with P , namely q . If at any point in the process we have $|L| \leq \frac{\gamma}{3}n$ then we terminate; observe that we then have $|V(P)| \geq n - |L| - |R| \geq (1 - \gamma)n$, so P is the desired path. We may therefore assume throughout the process that $|L| > \frac{\gamma}{3}n$.

The first stage of the process is while P contains at most $(\frac{1}{2} - \mu)n$ vertices, in which case we have $|V(P) \cup R| \leq (\frac{1}{2} - \mu)n + \frac{2\gamma}{3}n \leq \delta(H) - 2$. So we may use the minimum degree condition as before to extend P by one edge. That is, let p be the end of P other than q , choose any vertex $u \in L$ and any vertex $v \in N(p \cup \{u\}) \cap L$, and extend P by the edge $p \cup \{u, v\}$. We continue to extend P in this way until P has more than $(\frac{1}{2} - \mu)n$ vertices. Note that no vertices of R are added to P during this stage of the process.

Once P contains more than $(\frac{1}{2} - \mu)n$ vertices, we enter the second stage of the process. In each step of the process during this stage we will add at most 8 additional vertices from the reservoir set R to the path P , and this stage of the process will continue for at most $\frac{3n}{D}$ steps. In consequence we will always have $|R'| \geq |R| - 8 \cdot \frac{3n}{D}$. Since $8 \cdot \frac{3n}{D} \leq \frac{\kappa\gamma}{20}n$ it follows from our choice of R that at any point in the process (or immediately after the process terminates), given disjoint pairs p and p' in V' , we can find a path of length at most three in $H[p \cup p' \cup R']$ with ends p and p' .

Suppose first that, at some step of the second stage of the process, we have $e(H[L]) > \mu \binom{|L|}{4}$. Then we can use Theorem 2.27 to find a path P' in $H[L]$ on at least $\frac{\mu|L|}{4} - 1$ vertices. We let $p \in \binom{V}{2}$ be the end of P other than q and let p' be an end of P' , and choose a path Q of length at most three in $H[R' \cup p \cup p']$ with ends p and p' . We then replace P by PQP' and proceed to the next iteration. Note that in this step we added at most 4 vertices from R' to P , and that the total number of vertices added to P is at least $\frac{\mu|L|}{4} - 1 \geq \frac{\mu\gamma}{12}n - 1 \geq \frac{D}{3}$.

Now suppose instead that, at some step of the second stage of the process, we have $e(H[L]) \leq \mu \binom{|L|}{4}$. Then we have the following claim.

Claim 2.29. *There exist sets $J \subseteq I \subseteq V(P) \setminus V(P'_0)$ and a 3-graph H_0 with vertex set L such that $|I| = D$, $P[I]$ is a segment of P , $|J| = \frac{D}{3}$, $e(H_0) \geq 2^{-D\frac{1}{7}} \binom{|L|}{3}$ and for every $e \in E(H_0)$ and every $v \in J$ we have $e \cup \{v\} \in E(H)$.*

Proof of Claim 2.29. Let E_0 be the set of edges of $H[L]$, E_1 be the set of edges of H with

three vertices in L and one vertex in $V(P) \setminus V(P'_0)$ and E'_1 be the set of edges of H with three vertices in L and one vertex in $R' \cup V(P'_0)$. We then have that

$$\left(\frac{1}{2} - \varepsilon\right)n \binom{|L|}{3} \leq \sum_{S \in \binom{L}{3}} d_H(S) = 4|E_0| + |E_1| + |E'_1|.$$

Since $4|E_0| \leq 4\mu \binom{|L|}{4} < \mu|L| \binom{|L|}{3} \leq \mu n \binom{|L|}{3}$ and $|E'_1| \leq \left(\frac{2\gamma}{3}n + \mu n + 4\right) \binom{|L|}{3}$, this yields

$$|E_1| \geq \left(\frac{1}{2} - 3\mu\right)n \binom{|L|}{3}.$$

Let \mathcal{P} be the family of segments of P with precisely D vertices which do not intersect P'_0 , and for each $Q \in \mathcal{P}$ let N_Q be the number of edges of E_1 which intersect $V(Q)$. Since all but at most $2D$ vertices of $V(P) \setminus V(P'_0)$ appear in precisely $\frac{D}{2}$ of the sets $V(Q)$, we have

$$\sum_{Q \in \mathcal{P}} N_Q \geq \left(|E_1| - 2D \binom{|L|}{3}\right) \cdot \frac{D}{2} \geq \left(\frac{1}{2} - 4\mu\right) \binom{|L|}{3} D \cdot \frac{n}{2}$$

Since P has at most n vertices we have $|\mathcal{P}| \leq \frac{n}{2}$, so we may fix a segment $Q \in \mathcal{P}$ with $N_Q \geq \left(\frac{1}{2} - 4\mu\right) \binom{|L|}{3} D$. Write $I := V(Q)$ and let H_1 be the 3-graph on L whose edges are all sets $S \in \binom{L}{3}$ with $|N_H(S) \cap I| \geq \frac{D}{3}$. Then we have $N_Q \leq e(H_1)D + \binom{|L|}{3} \frac{D}{3}$, and it follows that $e(H_1) \geq \left(\frac{1}{2} - 4\mu - \frac{1}{3}\right) \binom{|L|}{3} \geq \frac{1}{7} \binom{|L|}{3}$. Also, since I has at most 2^D subsets, by averaging we may fix a set $J' \subseteq I$ with $|J'| \geq \frac{D}{3}$ such that at least $2^{-D}e(H_1)$ edges $S \in E(H_1)$ have $N_{H_1}(S) = J'$. Let H_0 be the 3-graph on L with all such sets S as edges, and choose any $J \subseteq J'$ with $|J| = \frac{D}{3}$. \diamond

Fix such an I , J and H_0 . Then H_0 contains a complete 3-partite 3-graph K with all vertex classes of size $\frac{D}{3}$ by Theorem 2.26. So let K' be the complete 4-partite subgraph of H whose vertex classes are J and the three vertex classes of K , and let Q be a Hamilton path in K' (so in particular Q is a path in H on $\frac{4D}{3}$ vertices). Since $P[I]$ is a segment

of P , removing the vertices in I from P leaves two vertex-disjoint subpaths of P ; of these let P_1 be the path which has q as an end (so in particular P'_0 is a segment of P_1) and let P_2 be the other path. Let p_1 be the end of P_1 other than q , let p_2 be an end of P_2 , and let q_1 and q_2 be the ends of Q , and choose vertex-disjoint paths Q_1 and Q_2 in $H[p_1 \cup q_1 \cup R']$ and $H[p_2 \cup q_2 \cup R']$ respectively, each of length at most three, so that Q_1 has ends p_1 and q_1 and Q_2 has ends p_2 and q_2 . We now replace P with the path $P_1Q_1Q_2P_2$ and proceed to the next iteration. Note that in this step we added at most eight vertices from R' to P (at most four in each of Q_1 and Q_2), and that the total number of vertices in P increased by at least $|V(Q)| - |I| = \frac{D}{3}$.

Since each step in the second stage of the process increases the number of vertices of P by at least $\frac{D}{3}$, this stage of the process continues for at most $\frac{3n}{D}$ steps, as claimed. When the process terminates the final path P has at least $(1 - \gamma)n$ vertices and has q as an end, and P'_0 is a segment of P . Let p be the end of P other than q ; then we may choose a path Q in $H[R' \cup p \cup q]$ of length at most three and with ends p and q . This gives a cycle $C = PQ$ in H on at least $(1 - \gamma)n$ vertices such that P_0 is a segment of C . It remains only to find a perfect matching in $G^* := G[V \setminus V(C)]$. For this note that $|V(G^*)| \leq \gamma n$ and $|R \setminus V(C)| \geq |R| - \frac{24n}{D} - 4$. Therefore our choice of R implies that $\delta(G^*) \geq \frac{4\gamma}{7}n - \frac{24n}{D} - 4 \geq \frac{\gamma}{2}n \geq \frac{|V(G^*)|}{2}$. Since C is a 2-cycle (so has an even number of vertices) and n is even, we have that $|V \setminus V(C)|$ is even, so G^* contains a perfect matching.

Following this proof gives a polynomial-time algorithm to find a long cycle as in the statement in a κ -connecting k -graph of high minimum codegree. Indeed, Lemma 2.18 gives a reservoir set R as required in time $O(n^{16})$, and by Theorem 2.26 we may find the complete 3-partite 3-graph K in time $O(n^3)$, whilst Theorem 2.27 allows the choice of the path P' in time $O(n^4)$, and it is clear that the remaining steps of the proof can be carried out in polynomial time (e.g. by exhaustive search to find a path of length at most three). \square

2.5 Hamilton 2-cycles in 4-graphs: even-extremal case

In this section we give a detailed proof of Lemma 2.9 which states that Theorem 1.11 holds for even-extremal 4-graphs. As in the previous section, all paths and cycles we consider in this section are 2-paths and 2-cycles, therefore we will again omit the 2 and speak simply of paths and cycles. Also, for most of the section we work within the following setup.

Setup 2.30. *Fix constants satisfying $1/n \ll \varepsilon, c \ll \gamma \ll \beta \ll \beta_2 \ll \beta_1 \ll \rho \ll \mu \ll 1$. Let H be a 4-graph of order n , and let $V = V(H)$.*

Recall that if H is an even-extremal 4-graph on n vertices, with a corresponding even-extremal bipartition $\{A, B\}$ of $V(H)$, then H has very few odd edges. If $\delta(H)$ is close to $n/2$ then it follows from this that $H[A]$ and $H[B]$ are both very dense, and also that H has very high density of edges with precisely two vertices in A . Furthermore recall that we call a pair $p \in \binom{V}{2}$ a split pair if $|p \cap A| = 1$ and a connate pair otherwise.

One strategy for finding a Hamilton cycle in such an H is as follows. We first find short paths P and Q in H each joining a connate pair in A to a connate pair in B . That is, the ends p and p' of P and the ends q and q' of Q have $p, q \in \binom{A}{2}$ and $p', q' \in \binom{B}{2}$. Moreover P and Q are chosen so that $A' := A \setminus (V(P) \cup V(Q))$ and $B' := B \setminus (V(P) \cup V(Q))$ each have even size. We then use the high density of $H[A]$ and $H[B]$ to find a Hamilton path P_A in $H[A' \cup p \cup q]$ with ends p and q and a Hamilton path P_B in $H[B' \cup p' \cup q']$ with ends p' and q' . Together P, P_A, Q and P_B then form a Hamilton cycle in H . Another strategy for finding a Hamilton cycle in such an H is to first find a short path P in H whose ends p and q are both split pairs such that $V' := V(H) \setminus V(P)$ satisfies $|V' \cap A| = |V' \cap B|$. We then use the high density of edges of H with precisely two vertices in A to find a Hamilton path P' in $H[V' \cup p \cup q]$ with ends p and q which consists of a sequence of split pairs. Together P' and P then form a Hamilton cycle. We give the necessary preliminaries for implementing this strategy in Subsection 2.5.1, culminating in Lemma 2.34 which gives

sufficient conditions for us to find P' as desired. We will also use Lemma 2.34 to find the paths P_A and P_B in the very dense case.

Finally, in Subsection 2.5.2 we complete the proof of Lemma 2.9 by distinguishing various cases; in each case we apply one of the two strategies described above to find a Hamilton cycle in H .

2.5.1 Hamilton paths of split pairs

In this subsection we consider 4-graphs H admitting a bipartition $\{A, B\}$ of $V(H)$ such that H has a very high density of edges with two vertices in A and two vertices in B , motivating the following definitions.

Definition 2.31. *Under Setup 2.30, for a fixed bipartition $\{A, B\}$ of V , we say that*

- (i) *a triple $S \in \binom{V}{3}$ is γ -good if it is contained in at least $(\frac{1}{2} - \gamma)n$ even edges,*
- (ii) *a pair $p \in \binom{V}{2}$ is γ -good if it is contained in at least $(\frac{1}{2} - \gamma^3)\binom{n}{2}$ even edges,*
- (iii) *a vertex $v \in V$ is γ -good if it is contained in at least $(\frac{1}{2} - \gamma^5)\binom{n}{3}$ even edges,*
- (iv) *a pair $p \in \binom{V}{2}$ is β_2 -medium if it is contained in at least $\beta_2\binom{n}{2}$ even edges,*
- (v) *a pair $p \in \binom{V}{2}$ is β_2 -bad if it is not β_2 -medium,*
- (vi) *a vertex $v \in V$ is (β_1, β_2) -medium if it is contained in at least $\beta_1 n$ -many β_2 -medium pairs, and*
- (vii) *a vertex $v \in V$ is (β_1, β_2) -bad if it is not (β_1, β_2) -medium.*

The following elementary proposition shows that good vertices and pairs lie in many good pairs and triples.

Proposition 2.32. *Assume Setup 2.30, and fix a bipartition $\{A, B\}$ of V such that $n/2 - cn \leq |A| \leq n/2 + cn$. Then the following statements hold.*

(i) If $v \in V$ is γ -good, then v is contained in at least $(1 - \gamma)n$ -many γ -good pairs.

(ii) If $p \in \binom{V}{2}$ is a γ -good pair, then p is contained in at least $(1 - \gamma)n$ -many γ -good triples.

Proof. First, we make the following observation. Each set $S \subseteq V$ of size $i \in \{1, 2, 3\}$ is contained in at least $(\frac{1}{2} - 2c)\binom{n}{4-i}$ and at most $(\frac{1}{2} + 2c)\binom{n}{4-i}$ even 4-sets. To prove (i) note that since v is γ -good, there are at most $(\frac{1}{2} + 2c)\binom{n}{3} - (\frac{1}{2} - \gamma^5)\binom{n}{3} \leq (\gamma^5 + 2c)\binom{n}{3}$ even 4-sets which contain v and which do not form an edge in H . Now assume that there are more than γn vertices w such that $\{v, w\}$ is not γ -good. Then there are at least $\frac{1}{3} \cdot \gamma n \cdot ((\frac{1}{2} - 2c)\binom{n}{2} - (\frac{1}{2} - \gamma^3)\binom{n}{2}) > (\gamma^5 + 2c)\binom{n}{3}$ even 4-sets which contain v and do not form an edge in H , a contradiction. A similar reasoning proves (ii). \square

Proposition 2.33. *Assume Setup 2.30, and fix a bipartition $\{A, B\}$ of V such that $n/2 - cn \leq |A| \leq n/2 + cn$. Also let $R \subseteq V$ be such that $|A \setminus R| \geq \mu n$ and $|B \setminus R| \geq \mu n$. Then*

(i) *for any two disjoint γ -good split pairs s_1 and s_2 there exists a split pair $s_3 \in \binom{V \setminus R}{2}$ such that $s_1 \cup s_3 \in E(H)$ and $s_3 \cup s_2 \in E(H)$, and*

(ii) *for any two disjoint γ -good connate pairs p_1 and p_2 there exists a connate pair $p_3 \in \binom{A \setminus R}{2}$ such that $p_1 \cup p_3 \in E(H)$ and $p_3 \cup p_2 \in E(H)$.*

Now suppose additionally that at most ρn vertices of V are not γ -good. Then

(iii) *for any γ -good split pair s_1 there exists a γ -good split pair $s_2 \in \binom{V \setminus R}{2}$ such that $s_1 \cup s_2 \in E(H)$, and*

(iv) *for any γ -good connate pair p_1 there exists a γ -good connate pair $p_2 \in \binom{A \setminus R}{2}$ such that $p_1 \cup p_2 \in E(H)$.*

Note that by symmetry (ii) and (iv) remain valid with B in place of A .

Proof. For (i), since s_1 and s_2 are γ -good, there are at most $2\left(\frac{n^2}{4} - \left(\frac{1}{2} - \gamma^3\right)\binom{n}{2}\right) < 3\gamma^3\binom{n}{2}$ split pairs which do not form an edge with both s_1 and s_2 . Since there are at least $\mu^2 n^2 > 3\gamma^3\binom{n}{2}$ split pairs which do not contain a vertex of R , we may choose s_3 as required. Similarly, for (iii), there are at most $\frac{n^2}{4} - \left(\frac{1}{2} - \gamma^3\right)\binom{n}{2} < \frac{1}{2}\mu^2 n^2$ split pairs which do not form an edge with s_1 , and since at most ρn vertices are not γ -good, by Proposition 2.32(i) the total number of pairs which are not γ -good is at most $\rho n \cdot n + n \cdot \gamma n < \frac{1}{2}\mu^2 n^2$, and so we may choose a split pair $s_2 \in \binom{V \setminus R}{2}$ as required. The arguments for (ii) and (iv) are very similar, so we omit them. \square

Note that we can list all γ -good pairs in H in time $O(n^4)$, and having done so we can find pairs as in Proposition 2.33(i)-(iv) in time $O(n^2)$ by exhaustive search. We now state and prove our Hamilton path connecting lemma for this setting.

Lemma 2.34. *Assume Setup 2.30, and fix a bipartition $\{A, B\}$ of V with $|A| = |B|$. Suppose also that every vertex of V is γ -good. If s_1 and s_2 are disjoint γ -good split pairs, then there exists a Hamilton path in H with ends s_1 and s_2 . Moreover, such a path can be found in time $O(n^4)$.*

Proof. Set $m := \lceil \frac{(1-\rho)n}{8} \rceil$. Throughout this argument we only use edges $e \in E$ such that $|e \cap A| = |e \cap B| = 2$. Consequently, every path we construct contains the same number of vertices from A and B . The proof consists of the following three steps.

- (I) We define a notion of a palatable vertex, and build a grid $L \subseteq V \setminus (s_1 \cup s_2)$ which consists of $3m + 2$ vertices from A and $3m + 2$ vertices from B , contains two γ -good split pairs $q_1, q_2 \in \binom{L}{2}$ and which can swallow any set $S \subseteq V \setminus L$ of $2m$ palatable vertices with $|S \cap A| = |S \cap B| = m$, meaning that for any such S there is a path in H with vertex set $L \cup S$ and with ends q_1 and q_2 .
- (II) Next, we construct disjoint paths Q_1 and Q_2 in $H[(V \setminus L) \cup q_1 \cup q_2]$ such that Q_1

has ends s_1 and q_1 , Q_2 has ends q_2 and s_2 , the set $R := V \setminus (L \cup Q_1 \cup Q_2)$ satisfies $|R \cap A| = |R \cap B| = m$ and every vertex in R is palatable.

(III) Finally, since L can swallow R , there is a path P in H with vertex set $R \cup L$ and with ends q_1 and q_2 , and $Q_1 P Q_2$ is then a Hamilton path in H with ends s_1 and s_2 .

To construct the grid we first greedily choose a set $L'_1 = \{x_1, y_1, x_2, y_2, \dots, y_m, x_{m+1}\}$ of distinct vertices of $A \setminus (s_1 \cup s_2)$ such that for each $i \in [m]$ both $\{x_i, y_i\}$ and $\{y_i, x_{i+1}\}$ are γ -good. We then greedily select distinct vertices $z_1, \dots, z_{m+1} \in B \setminus (s_1 \cup s_2)$ such that $\{x_1, z_1\}$ and $\{x_{m+1}, z_{m+1}\}$ are γ -good pairs and such that for any $i \in [m]$ both $\{z_i, x_i, y_i\}$ and $\{y_i, x_{i+1}, z_{i+1}\}$ are γ -good triples. Having done this, we set $L_1 = L'_1 \cup \{z_1, \dots, z_{m+1}\}$ and continue to form a ‘mirror image’ L_2 as follows. We greedily choose a set $L'_2 = \{x'_1, y'_1, x'_2, y'_2, \dots, y'_m, x'_{m+1}\}$ of distinct vertices of $B \setminus (L_1 \cup s_1 \cup s_2)$ such that $\{x_{m+1}, z_{m+1}, x'_1\}$ is γ -good and for each $i \in [m]$ both $\{x'_i, y'_i\}$ and $\{y'_i, x'_{i+1}\}$ are γ -good. Finally, we greedily select distinct vertices $z'_1, \dots, z'_{m+1} \in A \setminus (L_1 \cup s_1 \cup s_2)$ such that $\{x_{m+1}, z_{m+1}, x'_1, z'_1\} \in E(H)$, the pair $\{x'_{m+1}, z'_{m+1}\}$ is γ -good and for any $i \in [m]$ both $\{z'_i, x'_i, y'_i\}$ and $\{y'_i, x'_{i+1}, z'_{i+1}\}$ are γ -good triples. Let $L_2 := L'_2 \cup \{z'_1, \dots, z'_{m+1}\}$; our grid is then $L := L_1 \cup L_2$, and we take $q_1 := \{x_1, z_1\}$ and $q_2 := \{x'_{m+1}, z'_{m+1}\}$. To confirm that it is possible to make these greedy selections, observe that by the definition of a γ -good triple and Proposition 2.32(i) and (ii) at least $|B| - 2\gamma n$ vertices are suitable for each choice of a vertex from B and at least $|A| - 2\gamma n$ vertices are suitable for each choice of a vertex from A . Since in total we choose $3m + 2$ vertices from each of A and B , and $|s_1 \cup s_2| = 4$, there are always at least $n/2 - (3m + 2) - 4 - 2\gamma n \geq 1$ suitable vertices which have not previously been chosen.

We now set $A' = A \setminus L$ and $B' = B \setminus L$, and define two bipartite auxiliary graphs G_A and G_B . Indeed, we take G_A to be the bipartite graph with vertex classes A' and $Y' := \{y'_1, \dots, y'_m\}$ whose edges are those pairs $\{y'_i, w\}$ with $i \in [m]$ and $w \in A'$ such that $\{x'_i, z'_i, y'_i, w\}, \{y'_i, w, x'_{i+1}, z'_{i+1}\} \in E(H)$. By construction both $\{x'_i, z'_i, y'_i\}$ and

$\{y'_i, x'_{i+1}, z'_{i+1}\}$ are γ -good for each $i \in [m]$, and it follows that $d_{G_A}(y'_i) \geq |A'| - 2\gamma n > (1 - 16\gamma)|A'|$. In particular the graph G_A has at least $m(1 - 16\gamma)|A'|$ edges. Likewise we take G_B to be the bipartite graph with vertex classes B' and $Y := \{y_1, \dots, y_m\}$ whose edges are those pairs $\{y_i, w\}$ with $i \in [m]$ and $w \in B'$ such that $\{x_i, z_i, y_i, w\}, \{y_i, w, x_{i+1}, z_{i+1}\} \in E(H)$; by the same argument we have $d_{G_B}(y'_i) > (1 - 16\gamma)|B'|$ for each $i \in [m]$, and so in particular the graph G_B has more than $m(1 - 16\gamma)|B'|$ edges. We call a vertex $a \in A'$ palatable if $d_{G_A}(a) \geq 0.9m$ and a vertex $b \in B'$ palatable if $d_{G_B}(b) \geq 0.9m$. Let $M_A \subseteq A'$ be the set of non-palatable vertices in A' and let $M_B \subseteq B'$ be the set of non-palatable vertices in B' . Then $|M_A| < \frac{\rho}{100}n$, as otherwise the number of edges in G_A would be at most

$$\left(|A'| - \frac{\rho}{100}n\right)m + \frac{\rho}{100}n \cdot 0.9m < m(1 - 16\gamma)|A'|,$$

contradicting our previous lower bound. For the same reason we have $|M_B| < \frac{\rho}{100}n$. Observe that given a set $S \subseteq A' \cup B'$ of $2m$ palatable vertices with $|S \cap A'| = |S \cap B'| = m$ the subgraphs $G'_A := G_A[(A' \cap S) \cup Y]$ and $G'_B := G_B[(B' \cap S) \cup Y]$ contain perfect matchings $\{\{y'_i, a_i\} \mid i \in [m]\}$ and $\{\{y_i, b_i\} \mid i \in [m]\}$ respectively, as $\delta(G'_A), \delta(G'_B) \geq 0.9m$. It follows that

$$(x_1, z_1, y_1, b_1, \dots, y_m, b_m, x_{m+1}, z_{m+1}, x'_1, z'_1, y'_1, a_1, \dots, y'_m, a_m, x'_{m+1}, z'_{m+1})$$

is a path in H with ends q_1 and q_2 . This demonstrates that L can swallow any set S of $2m$ palatable vertices with $|S \cap A| = |S \cap B| = m$, and so completes Step (I) of the proof.

We now construct Q_1 and Q_2 . We first use Proposition 2.33(i) to find $s'_1 \in \binom{V \setminus (L \cup s_2)}{2}$ such that $Q_1 := s_1 s'_1 q_1$ is a path. It then remains to construct a path Q_2 with ends q_2 and s_2 , of length $\ell := \frac{n}{2} - 4m - 4$, with $V(Q_2) \subseteq V \setminus (L \cup Q_1)$, so that Q_2 contains all non-palatable vertices not in Q_1 and such that $|V(Q_2) \cap A| = |V(Q_2) \cap B|$. We do this in the following way. Let M be the set of all non-palatable vertices not in Q_1 or s_2 , so

$|M| \leq |M_A| + |M_B| \leq \frac{\rho}{50}n$, and write $M = \{g_1, \dots, g_k\}$ and $V' := V \setminus (L \cup V(Q_1) \cup s_2)$. Now greedily choose distinct vertices $h_1, \dots, h_k \in V' \setminus M$ such that $\{g_i, h_i\}$ is a γ -good split pair for each $1 \leq i \leq k$. This is possible as each g_i is γ -good and $k \leq \frac{\rho}{50}n$. Write $p_0 := s_2$, and for each $i \in [k]$ let $p_{2i} := \{g_i, h_i\}$. By repeated application of Proposition 2.33(i) we may then choose split pairs $p_i \in \binom{V'}{2}$ for each odd $i \in [2k]$ such that the pairs p_i are all disjoint and $p_{i-1}p_i \in E(H)$ for each $i \in [k]$. We then use Proposition 2.33(iii) repeatedly to obtain γ -good split pairs $p_{2k+1}, p_{2k+2}, \dots, p_{\ell-2} \in \binom{V'}{2}$ which are pairwise-disjoint and disjoint from p_0, \dots, p_{2k} and such that $p_{i-1}p_i \in E(H)$ for each $k+1 \leq i \leq \ell-2$. Finally, we set $p_\ell := q_2$ and apply Proposition 2.33(i) to choose a split pair $p_{\ell-1} \in \binom{V'}{2}$ disjoint from $\bigcup_{i=0}^{\ell-2} p_i$ so that $p_{\ell-2}p_{\ell-1} \in E(H)$ and $p_{\ell-1}p_\ell \in E(H)$, and $Q_2 := p_0p_1 \dots p_\ell$ is then a path with the desired properties, completing Step (II). In particular, the set $L \cup Q_1 \cup Q_2$ contains precisely $\frac{1}{2}(|L| + |V(Q_1) \setminus L| + |V(Q_2) \setminus L|) = (3m+2) + 2 + \ell = \frac{n}{2} - m$ vertices from each of A and B , so $R := V \setminus (L \cup Q_1 \cup Q_2)$ satisfies $|R \cap A| = |R \cap B| = m$, and therefore can be swallowed by L for Step (III). This calculation also justifies that it was possible to choose vertices h_i and to use Proposition 2.33 as claimed above.

Finally, for the algorithmic statement observe that for Step (I) we can list the γ -good vertices, pairs and triples, greedily construct the grid L , form the auxiliary bipartite graphs G_A and G_B and list all palatable vertices in time $O(n^4)$. Then, for Step (II), we can form the paths Q_1 and Q_2 in time $O(n^3)$ by repeated exhaustive search of which pair to add next. Finally, for Step (III) we need to find a perfect matching in G'_A and G'_B , and we can do so in time $O(n^4)$ by using Edmonds's blossom algorithm [34]. \square

2.5.2 Proof of Lemma 2.9

We now turn to the proof of Lemma 2.9, for which we first give two preliminary results. The first of these states that if $\{A, B\}$ is an even-extremal bipartition of V , then almost all vertices and pairs are good, and that pairs which are not bad must form an even edge

with a good pair, even if we forbid a small number of vertices from being used for this.

Proposition 2.35. *Assume Setup 2.30, and suppose that $\delta(H) \geq n/2 - \varepsilon n$ and that $\{A, B\}$ is a c -even-extremal bipartition of V . Then*

- (a) *there are at most $\frac{c}{\gamma^3 - 2\varepsilon} \binom{n}{2}$ pairs which are not γ -good,*
- (b) *there are at most $\frac{c}{\gamma^5 - 2\varepsilon} n$ vertices which are not γ -good,*
- (c) *there are at most $\frac{c}{(1 - \beta_1)(\frac{1}{2} - \beta_2 - 2\varepsilon)} n$ vertices which are (β_1, β_2) -bad, and*
- (d) *if $R \subseteq V$ satisfies $|R| \leq \frac{1}{3}\beta n$, then for every β -medium pair p_1 there exists a γ -good pair $p_2 \in \binom{V \setminus R}{2}$ such that $p_1 \cup p_2$ is an even edge.*

Proof. For (a) note that by our minimum degree condition every pair forms an edge with at least $(\frac{1}{2} - 2\varepsilon) \binom{n}{2}$ other pairs, so if the assertion is not true, then there are more than $\frac{1}{6} \cdot \frac{c}{\gamma^3 - 2\varepsilon} \binom{n}{2} \cdot (\gamma^3 - 2\varepsilon) \binom{n}{2} \geq c \binom{n}{4}$ odd edges. Similarly, for (b) note that every vertex forms an edge with at least $(\frac{1}{2} - 2\varepsilon) \binom{n}{3}$ triples, so if the assertion is not true, then there are more than $\frac{1}{4} \cdot \frac{c}{\gamma^5 - 2\varepsilon} n \cdot (\gamma^5 - 2\varepsilon) \binom{n}{3} \geq c \binom{n}{4}$ odd edges. For (c), if we assume otherwise, then there are more than $\frac{1}{12} \cdot \frac{c}{(1 - \beta_1)(\frac{1}{2} - \beta_2 - 2\varepsilon)} n \cdot (1 - \beta_1)n \cdot (\frac{1}{2} - \beta_2 - 2\varepsilon) \binom{n}{2} \geq c \binom{n}{4}$ odd edges in H . In each case we have a contradiction to the fact that the bipartition $\{A, B\}$ is c -even-extremal. Finally, for (d) note that by (a) there are at most $\frac{c}{\gamma^3 - 2\varepsilon} \binom{n}{2} < \frac{\beta}{4} \binom{n}{2}$ pairs which are not γ -good, and at most $\frac{\beta}{3} n^2 < \frac{3\beta}{4} \binom{n}{2}$ pairs contain a vertex of R , but since p_1 is β -medium there are at least $\beta \binom{n}{2}$ pairs which form an even edge with p_1 . \square

Recall the proof strategies outlined at the start of this section, both of which begin by choosing short paths whose ends we can connect by the Hamilton path connecting lemma. We will use the following lemma to obtain suitable short paths.

Lemma 2.36. *Assume Setup 2.30, and suppose that $\delta(H) \geq n/2 - \varepsilon n$ and that $\{A, B\}$ is a c -even-extremal bipartition $\{A, B\}$ of V . Suppose also that every vertex of H is (β_1, β_2) -*

medium. If H contains an odd edge, then H contains a path of length 3 whose ends are a γ -good split pair and a γ -good connate pair.

Proof. It suffices to show that H contains an odd edge e which is the union of two β_2 -medium pairs. Indeed, we can then write $e = p \cup s$ where p is a β_2 -medium connate pair and s is a β_2 -medium split pair, following which two applications of Proposition 2.35(d) yield a γ -good connate pair p' and a γ -good split pair s' such that $p'ps'$ is a path.

By assumption we may choose an odd edge $e \in H$. If e is the union of two β_2 -medium pairs, then we are done by our first observation. So we may assume that there are vertices $x, y \in e$ such that $\{x, y\}$ is β_2 -bad. This means that $\{x, y\}$ is contained in at most $\beta_2 \binom{n}{2}$ even edges, and so (using the minimum degree condition) there are at least $(\frac{1}{2} - 2\beta_2) \binom{n}{2}$ pairs $\{z, w\}$ such that $\{x, y, z, w\}$ is an odd edge of H . By Proposition 2.35(b) at most $\frac{c}{\gamma^5 - 2\epsilon} n \cdot n < \gamma \binom{n}{2}$ pairs $\{z, w\}$ contain a vertex which is not γ -good. Furthermore, since y is (β_1, β_2) -medium, at least $\beta_1 n$ pairs containing y are β_2 -medium. Therefore there are at least $\beta_1 n \cdot ((\frac{1}{2} - c)n - \beta_1 n - 1) > \frac{\beta_1}{2} \binom{n}{2}$ pairs $\{z, w\}$ for which $\{y, z\}$ and $\{y, w\}$ are not both β_2 -bad. Since there are at most $(\frac{1}{2} + c) \binom{n}{2}$ pairs $\{z, w\}$ such that $\{x, y, z, w\}$ is an odd 4-tuple, and $(\frac{1}{2} + c) - (\frac{1}{2} - 2\beta_2) + \gamma < \frac{\beta_1}{2}$, it follows that there exist γ -good vertices z and w such that $e := \{x, y, z, w\}$ is an odd edge of H and $\{y, z\}$ is a β_2 -medium pair. If $\{x, w\}$ is β_2 -medium, then e is the union of two β_2 -medium pairs and we are done by our first observation, whilst if $\{x, w\}$ is β_2 -bad, then we have a β_2 -bad pair containing the γ -good vertex w . We then proceed as follows.

The minimum degree condition of H , combined with the fact that $\{w, x\}$ is β_2 -bad, implies that there are at least $(\frac{1}{2} - 2\beta_2) \binom{n}{2}$ pairs $\{z', w'\}$ such that $\{w, x, z', w'\}$ is an odd edge of H . Since w is a γ -good vertex, by Proposition 2.32(i) at most γn pairs containing w are not γ -good. So certainly there are at most $\gamma n^2 \leq 3\gamma \binom{n}{2}$ pairs $\{z', w'\}$ for which $\{w, z'\}$ is β_2 -bad or $\{w, w'\}$ is β_2 -bad. Moreover, since x is (β_1, β_2) -medium (since we assumed all vertices are), at least $\beta_1 n$ pairs containing x are β_2 -medium. Therefore there are at least

$\beta_1 n \cdot ((\frac{1}{2} - c)n - \beta_1 n - 1) > \frac{\beta_1}{2} \binom{n}{2}$ pairs $\{z', w'\}$ such that $\{w, x, z', w'\}$ is an odd 4-tuple and such that $\{x, z'\}$ and $\{x, w'\}$ are not both β_2 -bad. Since there are at most $(\frac{1}{2} + c) \binom{n}{2}$ pairs $\{z', w'\}$ such that $\{w, x, z', w'\}$ is an odd 4-tuple, and $(\frac{1}{2} + c) - (\frac{1}{2} - 2\beta_2) + 3\gamma < \frac{\beta_1}{2}$, it follows that there exist vertices z' and w' such that $\{w, x, z', w'\}$ is an odd edge of H containing β_2 -medium pairs $\{w, z'\}$ and $\{x, w'\}$, and we saw already that this is sufficient. \square

Now we have all the tools we need to prove Lemma 2.9. In fact, we actually prove the following stronger statement, which additionally asserts that given an even-extremal bipartition as input, we can find a Hamilton 2-cycle in polynomial time.

Lemma 2.37. *Suppose that $1/n \ll \varepsilon, c' \ll 1$ and that n is even, and let H be a 4-graph of order n with $\delta(H) \geq n/2 - \varepsilon n$. If H is c' -even-extremal and every bipartition $\{A, B\}$ of $V(H)$ is even-good, then H contains a Hamilton 2-cycle. Moreover, there exists an algorithm *Procedure HamCycleEven*($H, \{A, B\}$) which, given as input a c' -even-extremal bipartition $\{A, B\}$ of $V(H)$, returns a Hamilton 2-cycle in H in time $O(n^{10})$.*

Proof. For this proof we introduce further constants such that

$$\varepsilon, c' \ll c \ll \gamma \ll \beta \ll \beta_2 \ll \beta'_2 \ll \beta_1 \ll \beta'_1 \ll 1.$$

Since H is c' -even-extremal we may fix a bipartition $\{A', B'\}$ of $V := V(H)$ with $(\frac{1}{2} - c')n \leq |A'| \leq (\frac{1}{2} + c')n$ which induces at most $c' \binom{n}{4}$ odd edges. We begin by moving all vertices which are (β'_1, β'_2) -bad to the other side. More precisely, we define $A^{\text{bad}} := \{a \in A' : a \text{ is } (\beta'_1, \beta'_2)\text{-bad}\}$ and $B^{\text{bad}} := \{b \in B' : b \text{ is } (\beta'_1, \beta'_2)\text{-bad}\}$, and set $A := (A' \setminus A^{\text{bad}}) \cup B^{\text{bad}}$ and $B := (B' \setminus B^{\text{bad}}) \cup A^{\text{bad}}$; we say that the vertices of $A^{\text{bad}} \cup B^{\text{bad}}$ are *moved*. By Proposition 2.35(c) at most $3c'n$ vertices are moved in total.

Claim 2.38. *$\{A, B\}$ is a c -even-extremal bipartition of V with respect to which every vertex of H is (β_1, β_2) -medium. Moreover, every vertex of H is contained in at least βn connate pairs which are β -medium with respect to $\{A, B\}$.*

Proof of Claim 2.38. Since at most $3c'n$ vertices were moved, we have $(\frac{1}{2} - c)n \leq (\frac{1}{2} - c' - 3c')n \leq |A|(\frac{1}{2} + c' + 3c')n \leq (\frac{1}{2} + c)n$, and the number of edges of H which are odd with respect to $\{A, B\}$ is at most $c' \binom{n}{4} + 3c'n^4 \leq c \binom{n}{4}$, so $\{A, B\}$ is c -even-extremal. Now, since $\beta_2 \ll \beta'_2 \ll \beta_1 \ll \beta'_1$, every unmoved vertex which was (β'_1, β'_2) -medium under $\{A', B'\}$ is (β_1, β_2) -medium under $\{A, B\}$, whilst every moved vertex was in at least $(1 - \beta'_1)n$ pairs which were not β'_2 -medium under $\{A', B'\}$ and therefore is in at least $(1 - \beta'_1 - 3c')n$ -many β_2 -medium pairs under $\{A, B\}$. Thus every vertex is (β_1, β_2) -medium with respect to $\{A, B\}$.

Now suppose that some vertex $x \in V$ is contained in fewer than βn connate pairs which are β -medium. Since x is (β_1, β_2) -medium it follows that x is contained in at least $(\beta_1 - \beta)n \geq \frac{\beta_1}{2}n$ split pairs which are β_2 -medium, so the number N of even edges $e \in E(H)$ with $|e \cap A| = |e \cap B| = 2$ which contain x satisfies $N \geq \frac{1}{3} \cdot \frac{\beta_1}{2}n \cdot \beta_2 \binom{n}{2} \geq 2\beta n^3$ even edges e with $|e \cap A| = |e \cap B| = 2$. But then there must be at least βn connate pairs containing x which are contained in at least $\beta \binom{n}{2}$ even edges of H , as otherwise we would have $N < \beta n \cdot \binom{n}{2} + n \cdot \beta \binom{n}{2} < 2\beta n^3$, and each of these pairs is β -medium, giving a contradiction. \diamond

We henceforth exclusively use the terms odd, even, γ -good, (β_1, β_2) -medium and so forth with respect to the bipartition $\{A, B\}$ of V . Observe that H, V, n and the constants $\varepsilon, c, \gamma, \beta, \beta_1$ and β_2 are as in Setup 2.30.

We distinguish five cases which are related to the cases in the definition of an even-good bipartition. Case A assumes only that H contains two disjoint odd edges. All other cases assume that there are no two disjoint odd edges in H , and in addition Case B assumes that $|A|$ and $|B|$ are even, Case C assumes that $|A|$ and $|B|$ are odd and that there are two odd edges whose intersection is a split pair, Case D assumes that $|A| = |B|$ is odd and that there are no two odd edges whose intersection is a split pair and Case E assumes that $|A| = |B| + 2$ and that there are two odd edges whose intersection is a connate

pair in A . By symmetry Case E also covers the case where $|B| = |A| + 2$ and there are two odd edges whose intersection is a connate pair in B . Since $\{A, B\}$ is even-good by assumption, at least one of these cases must hold, so to prove the lemma it suffices to construct a Hamilton cycle in H in each case.

We begin with Cases A–C, for which we construct a Hamilton cycle as follows.

- (I) We build a ‘bridge’ Q , which is a path on at most γn vertices whose ends are a γ -good pair $q_A \in \binom{A}{2}$ and a γ -good pair $q_B \in \binom{B}{2}$ such that $|A \setminus V(Q)|$ and $|B \setminus V(Q)|$ are both even and so that Q contains all vertices of H which are not γ -good.
- (II) Next, we choose γ -good pairs $p_A \in \binom{A \setminus V(Q)}{2}$ and $p_B \in \binom{B \setminus V(Q)}{2}$ such that $p_A \cup p_B \in E(H)$, and take P to be the path consisting of this single edge.
- (III) Finally, we apply Lemma 2.34 twice to form a Hamilton path P_A in $H[(A \setminus V(Q)) \cup q_A]$ with ends p_A and q_A and a Hamilton path P_B in $H[(B \setminus V(Q)) \cup q_B]$ with ends p_B and q_B , and then $QP_B P P_A$ is a Hamilton cycle in H .

It suffices to prove the existence of the bridge Q in each case. Indeed, having constructed the bridge Q we may choose p_A and p_B as in Step (II) by choosing any γ -good pair $p_A \in \binom{A \setminus V(Q)}{2}$ and then using Proposition 2.33(iv) to obtain a γ -good pair $p_B \in \binom{B \setminus V(Q)}{2}$ with $p_A \cup p_B \in E(H)$. We then just need to explain how to apply Lemma 2.34 in Step (III).

For this define $A^* := (A \setminus V(Q)) \cup q_A$, $H_A := H[A^*]$ and $n_A := |A^*|$. Then by choice of Q we have that n_A is even and $(\frac{1}{2} - 2\gamma)n \leq n_A \leq (\frac{1}{2} + c)n$, so we can choose an arbitrary bipartition $A^* = S_1 \cup S_2$ of H_A such that $|S_1| = |S_2|$ and such that p_A and q_A are split pairs in relation to (S_1, S_2) . Since each vertex $v \in A^*$ is γ -good in H , it follows that there are at most $(\frac{1}{2} + 2c)\binom{n}{3} - (\frac{1}{2} - \gamma^5)\binom{n}{3}$ even 4-tuples $S \in \binom{V}{4}$ which contain v and do not form an edge of H . Then v is contained in at least $n_A/2 \cdot (n_A/2 - 1) - (2c + \gamma^5)\binom{n}{3} > (1/2 - \beta^3)\binom{n_A}{3}$ edges in H_A , and so v is β -good in H_A . Similar calculations show that, since each of p_A and q_A is a γ -good pair in H , each is a β -good pair in H_A . Thus we can apply Lemma 2.34

to obtain a Hamilton path in H_A with ends p_A and q_A as required. The same argument shows that we can find a Hamilton path in H_B with ends p_B and q_B also.

We now show how to construct the bridge Q in each of Cases A–C.

Case A: Assume there are two disjoint odd edges $e_1, e_2 \in E$. To construct Q we first use Lemma 2.36 to find a path Q'_1 of length three in $H[V \setminus e_2]$ whose ends are a γ -good connate pair p'_1 and a γ -good split pair s_1 . Next, we use Lemma 2.36 again to find a path Q_2 of length three in $H[V \setminus V(Q'_1)]$ whose ends are a γ -good split pair s_2 and a γ -good connate pair p_2 . By Proposition 2.33(iv) we may then choose a γ -good connate pair $p_1 \in \binom{A}{2}$ disjoint from $V(Q'_1) \cup V(Q'_2)$ such that $p_1 \cup p'_1 \in E(H)$, to give a path $Q_1 := p_1 p'_1 Q'_1 s_1$ of length four which is vertex-disjoint from Q_2 . Write $V' := V \setminus (V(Q_1) \cup V(Q_2))$, and let $X = \{x_1, \dots, x_k\}$ be the set of all vertices not in Q_1 or Q_2 which are not γ -good, so $k := |X| \leq \frac{c}{\gamma^5 - 2\varepsilon} n$ by Proposition 2.35(b). Since every vertex is in at least βn -many β -medium connate pairs by Claim 2.38, and $\beta n > \frac{2c}{\gamma^5 - 2\varepsilon} n + 18 \geq 2k + |V(Q_1)| + |V(Q_2)|$, we may then greedily choose distinct vertices $\{y_1, \dots, y_k\} \in V' \setminus X$ such that the pair $r_i := \{x_i, y_i\}$ is a β -medium connate pair for each $i \in [k]$.

We next choose γ -good connate pairs $g_i, g'_i \in \binom{V'}{2}$ with $g_i \cup r_i \in E(H)$ and $g'_i \cup r_i \in E(H)$ for each $i \in [k]$, and then choose connate pairs $h_i \in \binom{V'}{2}$ with $h_i \cup g_i \in E(H)$ and $h_i \cup g'_{i+1} \in E(H)$ for each $i \in [k-1]$. We additionally require that the pairs g_i, g'_i and h_i are disjoint from each other and from the pairs r_i . By Proposition 2.35(d) it is possible to choose the pairs g_i and g'_i greedily with these properties, whilst Proposition 2.33(ii) ensures that we may choose the pairs h_i greedily also. Similarly, we may also apply Proposition 2.33(iv) to choose a γ -good connate pair $h_k \in \binom{V' \cap B}{2}$ such that $h_k \cup g_k \in E(H)$ which is disjoint from all previously-chosen pairs. Observe that, having made these choices,

$$Q_3 := g'_1 r_1 g_1 h_1 g'_2 r_2 g_2 h_2 g'_3 r_3 g_3 h_3 g'_4 \dots h_{k-1} g'_k r_k g_k h_k$$

is a path in H . By Proposition 2.33(ii) we may then choose a connate pair p^* such that $p^* \cup p_2 \in E(H)$ and $p^* \cup g'_1 \in E(H)$, so ‘connecting’ the paths Q_2 and Q_3 .

Write $V'' := V \setminus (V(Q_1) \cup V(Q_2) \cup V(Q_3) \cup p^*)$. If $|V'' \cap A|$ is odd then we use Proposition 2.33(i) to choose a split pair $s^* \in \binom{V''}{2}$ such that $s_1 \cup s^* \in E(H)$ and $s_2 \cup s^* \in E(H)$, and define the path $Q := p_1 Q_1 s_1 s^* s_2 Q_2 p_2 p^* g'_1 Q_3 h_k$. On the other hand, if $|V'' \cap A|$ is even then we first use Proposition 2.33(iii) to choose a γ -good split pair $s' \in \binom{V''}{2}$ such that $s_1 \cup s' \in E(H)$ and then use Proposition 2.33(i) to choose a split pair $s^* \in \binom{V''}{2}$ such that $s' \cup s^* \in E(H)$ and $s_2 \cup s^* \in E(H)$; we then define the path $Q := p_1 Q_1 s_1 s' s^* s_2 Q_2 p_2 p^* g'_1 Q_3 h_k$. Either way Q is a path in H whose ends $q_A := p_1 \in \binom{A}{2}$ and $q_B := h_k \in \binom{B}{2}$ are γ -good pairs such that $|A \setminus V(Q)|$ is even; since n and $V(Q)$ are both even it follows that $|B \setminus V(Q)|$ is even also. Moreover Q contains all non- γ -good vertices by choice of Q_3 , and the total number of vertices in Q is at most $|V(Q_1)| + |V(Q_2)| + |V(Q_3)| + 6 \leq 22 + 8k \leq \gamma n$, so Q has the required properties.

Case B: Assume that there are no two disjoint odd edges, but $|A|$ and $|B|$ are even. Then there must be at most one non- γ -good vertex, as otherwise we would have two disjoint odd edges. Let x be such a vertex, if it exists; otherwise choose $x \in V$ arbitrarily. Then by Claim 2.38 we may choose a β -medium connate pair p which contains x . We now apply Proposition 2.35(d) twice to find disjoint γ -good connate pairs p_1 and p_2 such that $p_1 \cup p \in E(H)$ and $p_2 \cup p \in E(H)$. By symmetry we may assume that $p_2 \in \binom{B}{2}$, and by Proposition 2.33(iv) we may choose a γ -good connate pair $p_0 \in \binom{A}{2}$ disjoint from p_1, p and p_2 such that $p_0 \cup p_1 \in E(H)$. We may then take $Q := p_0 p_1 p p_2$ (in particular Q has an even number of vertices in each of A and B , so $|A \setminus V(Q)|$ and $|B \setminus V(Q)|$ are both even, as required).

Case C: Assume that $|A|$ and $|B|$ are odd and there are no two disjoint odd edges, but there are two odd edges e_1 and e_2 whose intersection is a split pair. That is, we may write $e_1 = p_1 \cup s$ and $e_2 = s \cup p_2$ where s is a split pair and p_1 and p_2 are connate pairs. Then p_1

and p_2 must be γ -good pairs, as otherwise we would have two disjoint odd edges. For the same reason all vertices in $V \setminus (s \cup p_1 \cup p_2)$ must be γ -good. By Proposition 2.33(iv) we may choose disjoint γ -good connate pairs $q_1 \in \binom{A}{2}$ and $q_2 \in \binom{B}{2}$ which do not intersect s, p_1 or p_2 so that $p_1 \cup q_1 \in E(H)$ and $p_2 \cup q_2 \in E(H)$, and we may then take $Q := q_1 p_1 s p_2 q_2$ (in particular Q has an odd number of vertices in each of A and B , so $|A \setminus Q|$ and $|B \setminus Q|$ are even, as required).

We now turn to cases D and E , for which we use the following, similar strategy to construct a Hamilton cycle.

- (I) We construct a path P_0 on at most six vertices whose ends s_1 and s_2 are both γ -good split pairs such that P_0 contains all non- γ -good vertices of H and $|A \setminus V(P_0)| = |B \setminus V(P_0)|$.
- (II) Write $V' := (V \setminus V(P_0)) \cup s_1 \cup s_2$. Then by our choice of P_0 we have $|A \cap V'| = |B \cap V'|$, and $H[V']$ contains only γ -good vertices. So by Lemma 2.34 there is a Hamilton path P_1 in $H[V']$ with ends s_1 and s_2 , and then $P_0 P_1$ is a Hamilton cycle in H .

Hence it suffices to construct P_0 in each case.

Case D: Assume that $|A| = |B|$ is odd, there are no two disjoint odd edges and there are no two distinct odd edges whose intersection is a split pair. Then there is at most one non- γ -good vertex, as otherwise we would have two disjoint odd edges. If there is such a vertex, we denote it by x and assume without loss of generality that $x \in A$; if not then we choose a vertex $x \in A$ arbitrarily. For every $b \in B$ the pair $\{x, b\}$ must be γ -good, as otherwise there would be two distinct odd edges whose intersection is the split pair $\{x, b\}$. So we may choose $b \in B$ such that $s_1 := \{x, b\}$ is γ -good and then use Proposition 2.33(iii) to obtain a γ -good split pair s_2 such that $s_1 \cup s_2 \in E(H)$. We may then take P_0 to be the single edge $s_1 \cup s_2$.

Case E: Assume that $|A| = |B| + 2$ and there are no two disjoint odd edges, but there are odd edges e_1 and e_2 whose intersection is a connate pair in A . That is, we may write $e_1 = s_1 \cup p$ and $e_2 = p \cup s_2$ where s_1 and s_2 are split pairs and p is a connate pair in A . Then s_1 and s_2 must both be γ -good pairs, as otherwise we would have two disjoint odd edges. For the same reason all vertices in $V \setminus (s_1 \cup p \cup s_2)$ must be γ -good, so we may take $P_0 := s_1 p s_2$ (in particular, P_0 contains two more vertices from A than B , so $|A \setminus V(P_0)| = |B \setminus V(P_0)|$, as required).

Finally, for the ‘moreover’ part of the lemma statement, note that if we are given a 4-graph H as in the lemma, it is not clear that we can find an even-extremal partition $\{A', B'\}$ of H in polynomial time. However, if such a partition is also given, then the remaining steps of the proof can be carried out efficiently. Indeed, we can identify the sets A^{bad} and B^{bad} and form the partition $\{A, B\}$ in time $O(n^4)$, and then we can identify in time $O(n^8)$ which of Cases A–E holds for $\{A, B\}$ (since $\{A, B\}$ is even-good at least one of the cases must hold). In Cases B–E the path Q or P_0 (according to the case) has at most 10 vertices, so can be found by exhaustive search in time $O(n^{10})$, whilst the greedy argument given in case A constructs Q in time $O(n^4)$. This completes Step (I) in each case. In Cases A–C we can then find an edge f as in Step (II) in time $O(n^4)$ by exhaustive search, and Lemma 2.34 states that we can then complete Step (III) in time $O(n^4)$ also. Similarly, in Cases D and E we can complete Step (II) in time $O(n^4)$. \square

2.6 Hamilton 2-cycles in 4-graphs: odd-extremal case

In this section we give a detailed proof of Lemma 2.10 which states that Theorem 1.11 holds for odd-extremal 4-graphs. As in the previous two sections, all paths and cycles we consider in this section are 2-paths and 2-cycles, therefore we will again omit the 2 and speak simply of paths and cycles. Throughout this section we will work within the following setup.

Setup 2.39. Fix constants satisfying $1/n \ll \varepsilon, c \ll \gamma \ll \psi \ll \beta_2 \ll \beta_1$. Let H be a 4-graph on n vertices with $\delta(H) \geq n/2 - \varepsilon n$, write $V := V(H)$ and let $\{A, B\}$ be a c -odd-extremal bipartition of V .

Our strategy is broadly similar to the one outlined in the previous section. Indeed, the minimum degree condition on H , combined with the fact that H has very few even edges (since the bipartition $\{A, B\}$ is odd-extremal), implies that almost all possible odd edges are present in H . To find a Hamilton cycle in H , we will find a short path P in H whose ends s_1 and s_2 are split pairs so that, using the high density of odd edges of H , we can then find a Hamilton path Q in $H[(V \setminus V(P)) \cup s_1 \cup s_2]$ with ends s_1 and s_2 . Together P and Q then form a Hamilton cycle in H . To implement this strategy, we begin by establishing some necessary preliminaries, then prove the Hamilton path connecting lemma (Lemma 2.42) which we use to find Q , before proceeding to give the proof of Lemma 2.10.

Definition 2.40. Under Setup 2.39, we say that

- (i) a triple $\{x, y, z\} \in \binom{V}{3}$ is γ -good if it is contained in at least $(\frac{1}{2} - \gamma)n$ odd edges,
- (ii) a pair $\{x, y\} \in \binom{V}{2}$ is γ -good if it is contained in at least $(\frac{1}{2} - \gamma^3)\binom{n}{2}$ odd edges,
- (iii) a vertex $v \in V$ is γ -good if it is contained in at least $(\frac{1}{2} - \gamma^5)\binom{n}{3}$ odd edges,
- (iv) a pair $\{x, y\} \in \binom{V}{2}$ is β_2 -medium if it is contained in at least $\beta_2\binom{n}{2}$ odd edges,
- (v) a pair $\{x, y\} \in \binom{V}{2}$ is β_2 -bad if it is not β_2 -medium,
- (vi) a vertex $v \in V$ is (β_1, β_2) -medium if there are at least $\beta_1 n$ vertices in A which form a β_2 -medium pair with v and also at least $\beta_1 n$ vertices in B which form a β_2 -medium pair with v , and
- (vii) a vertex $v \in V$ is (β_1, β_2) -bad if it is not (β_1, β_2) -medium.

Note that definitions (i)-(v) above are identical to those of the previous section (Definition 2.31) with ‘odd’ in place of ‘even’, but (vi) and (vii) (the definitions of (β_1, β_2) -medium and (β_1, β_2) -bad vertices) differ significantly.

Proposition 2.41. *Adopt Setup 2.39. Then the following statements hold.*

- (a) *If $v \in V$ is γ -good, then v is contained in at least $(1 - \gamma)n$ -many γ -good pairs.*
- (b) *If a pair $p \in \binom{V}{2}$ is γ -good, then p is contained in at least $(1 - \gamma)n$ -many γ -good triples.*
- (c) *At most $\frac{c}{\gamma^3 - 2\varepsilon} \binom{n}{2}$ pairs are not γ -good.*
- (d) *At most $\frac{c}{\gamma^3 - 2\varepsilon} n$ vertices are not γ -good.*
- (e) *At most $5cn$ vertices are (β_1, β_2) -bad.*

Now suppose also that $R \subseteq V$ is such that $|R \cap A|, |R \cap B| \leq n/2 - \psi n$. Then the following statements hold.

- (f) *For every two disjoint γ -good connate pairs p_1 and p_2 there exists a split pair $s \in \binom{V \setminus R}{2}$ such that $p_1 s p_2$ is a path of length 2 in H .*
- (g) *For every two disjoint γ -good split pairs s_1 and s_2 there exists a connate pair $p \in \binom{A \setminus R}{2}$ such that $s_1 p s_2$ is a path of length 2 in H .*
- (h) *For every γ -good connate pair p there exists a γ -good split pair $s \in \binom{V \setminus R}{2}$ such that $p \cup s \in E(H)$.*
- (i) *For every γ -good split pair s there exists a γ -good connate pair $p \in \binom{A \setminus R}{2}$ such that $s \cup p \in E(H)$.*

Proof. The proofs of (a) and (b) are identical to those of Proposition 2.32 with the words ‘odd’ and ‘even’ interchanged, the proofs of (c) and (d) are similarly identical to those

of Proposition 2.35(a) and (b). For (e), observe that any vertex which is (β_1, β_2) -bad is contained in at least $(\frac{1}{2} - c - \beta_1)n$ -many β_2 -bad pairs, each of which is contained in fewer than $\beta_2 \binom{n}{2}$ odd edges of H . Since any pair is contained in at least $(\frac{1}{2} - 2\varepsilon) \binom{n}{2}$ edges of H , and H has at most $c \binom{n}{4}$ even edges, as $\{A, B\}$ is c -odd-extremal, it follows that the number of (β_1, β_2) -bad vertices is at most

$$\frac{4 \cdot 3 \cdot c \binom{n}{4}}{(\frac{1}{2} - c - \beta_1)n \cdot (\frac{1}{2} - 2\varepsilon - \beta_2) \binom{n}{2}} < 5cn .$$

For (f), note that since p_i is γ -good for $i \in \{1, 2\}$, there are at least $(\frac{1}{2} - \gamma) \binom{n}{2}$ split pairs which form an edge with p_i . Since there are at most $\frac{n^2}{4}$ split pairs in total, it follows that there are at most $3\gamma n^2$ split pairs s for which $p_1 \cup s$ and $p_2 \cup s$ are not both edges. Furthermore there are at least $(\psi - c)n$ vertices in each of $A \setminus R$ and $B \setminus R$, so there are at least $(\psi - c)^2 n^2$ split pairs which do not contain a vertex of R . Since $3\gamma n^2 < (\psi - c)^2 n^2$, it follows that there exists a split pair $s \in \binom{V \setminus R}{2}$ such that $p_1 \cup s$ and $p_2 \cup s$ are both edges, as required. Likewise (h) follows, since by (c) at most $\frac{c}{\gamma^3 - 2\varepsilon} \binom{n}{2} < \gamma n^2$ split pairs are not γ -good. The proofs of (g) and (i) are very similar, so we omit them. \square

The following lemma is our Hamilton path connecting lemma which, under the assumption that every vertex of H is good, allows us to choose a Hamilton path in H with specified ends. From a broad perspective the proof is similar to that of Lemma 2.34, but the construction of the ‘grid’ is quite different, reflecting the fact that H is odd-extremal rather than even-extremal. Since many calculations are similar, we will be more concise and primarily emphasise the differences.

Lemma 2.42. *Adopt Setup 2.39, and suppose also that every vertex of H is γ -good, and that*

(i) $n \equiv 6 \pmod{8}$ and $|A| - |B| \equiv 2 \pmod{4}$, or

(ii) $n \equiv 2 \pmod{8}$ and $|A| - |B| \equiv 0 \pmod{4}$.

If s_1 and s_2 are disjoint γ -good split pairs, then there exists a Hamilton path in H whose ends are s_1 and s_2 . Moreover, such a path can be found in time $O(n^4)$.

Proof. Set $m := \lceil \frac{(1-\psi)n}{8} \rceil$. We begin by using Proposition 2.41(a) and our assumption that every vertex is γ -good to greedily choose sets $L'_1 = \{x_1, y_1, \dots, x_{m+1}, y_{m+1}\} \subseteq A \setminus (s_1 \cup s_2)$ and $L'_2 = \{x'_1, y'_1, \dots, x'_{m+1}, y'_{m+1}\} \subseteq B \setminus (s_1 \cup s_2)$ such that both $\{x_i, y_i\}$ and $\{x'_i, y'_i\}$ are γ -good for every $i \in [m+1]$. Next we use Proposition 2.41(b) to greedily choose sets $Z = \{z_1, \dots, z_m\} \subseteq A \setminus (L'_1 \cup s_1 \cup s_2)$ and $Z' = \{z'_1, \dots, z'_m\} \subseteq B \setminus (L'_2 \cup s_1 \cup s_2)$ such that for each $i \in [m]$ the triples $\{x_i, y_i, z_i\}$, $\{z_i, x_{i+1}, y_{i+1}\}$, $\{x'_i, y'_i, z'_i\}$ and $\{z'_i, x'_{i+1}, y'_{i+1}\}$ are all γ -good. Finally by Proposition 2.41(f) we can then choose $v \in A \setminus (L'_1 \cup Z \cup s_1 \cup s_2)$ and $v' \in B \setminus (L'_2 \cup Z' \cup s_1 \cup s_2)$ such that $\{x_{m+1}, y_{m+1}, v, v'\}$ and $\{v, v', x'_1, y'_1\}$ are both edges of H . Our ‘grid’ is then $L := L'_1 \cup Z \cup \{v, v'\} \cup L'_2 \cup Z'$, and we take $p_0 := \{x_1, y_1\}$ and $q_0 := \{x'_{m+1}, y'_{m+1}\}$. Observe in particular that $|L \cap A| = |L \cap B| = 3m + 3$.

Define $A' = A \setminus (L \cup s_1 \cup s_2)$ and $B' = B \setminus (L \cup s_1 \cup s_2)$. Let G_A be the bipartite graph with vertex classes Z' and A' whose edges are all pairs $\{z'_i, w\}$ with $i \in [m]$ and $w \in A'$ for which both $\{x'_i, y'_i, z'_i, w\}$, $\{w, z'_i, x'_{i+1}, y'_{i+1}\} \in E(H)$, and let G_B be the bipartite graph with vertex classes Z and B' and whose edges are all pairs $\{z_i, w\}$ with $i \in [m]$ and $w \in B'$ such that $\{x_i, y_i, z_i, w\}$, $\{w, z_i, x_{i+1}, y_{i+1}\} \in E(H)$. We call a vertex $a \in A'$ *palatable* if $d_{G_A}(a) \geq 0.9m$ and a vertex $b \in B'$ *palatable* if $d_{G_B}(b) \geq 0.9m$. Essentially the same argument as in the proof of Lemma 2.34 then shows that A' and B' each contain at most $\frac{\psi}{100}n$ non-palatable vertices, and furthermore that L can *swallow* any set $S \subseteq V \setminus L$ of $2m$ palatable vertices with $|S \cap A| = |S \cap B| = m$, meaning that for any such S there is a path in H with vertex set $L \cup S$ and ends p_0 and q_0 .

Let $\ell := |A| - \frac{n}{2} = \frac{n}{2} - |B| = \frac{1}{2}(|A| - |B|)$. We then have $|\ell| \leq cn$ since $\{A, B\}$ is a c -odd-extremal bipartition of V , and our divisibility assumptions ensure that both $k_A := \frac{1}{8}(n - 8m - 10 + 4\ell)$ and $k_B := \frac{1}{8}(n - 8m - 10 - 4\ell)$ are integers. Observe also that $k_A, k_B \geq$

$\frac{\psi}{10}n$. Since the number of non-palatable vertices in each of A' and B' is at most $\frac{\psi}{100}n$, we may choose sets $U_A = \{a_1, \dots, a_{k_A-1}\} \subseteq A'$ and $U_B = \{b_1, \dots, b_{k_B-1}\} \subseteq B'$ of distinct vertices so that $U_A \cup U_B$ contains all non-palatable vertices. Now recall our assumption that every vertex of H is γ -good. So by repeated application of Proposition 2.41(a) and (i) we may greedily choose vertex-disjoint γ -good connate pairs $p_1, \dots, p_{k_A} \in \binom{A'}{2}$ and $q_1, \dots, q_{k_B} \in \binom{B'}{2}$ so that $a_i \in p_i$ for each $i \in [k_A - 1]$ and $b_j \in q_j$ for each $j \in [k_B - 1]$, and also so that each of $s_1 \cup p_{k_A}$ and $s_2 \cup q_{k_B}$ is an edge of H . Write $W := \bigcup_{i \in [k_A]} p_i \cup \bigcup_{j \in [k_B]} q_j$. Then by repeated application of Proposition 2.41(f) we may greedily choose vertex-disjoint split pairs $p'_1, \dots, p'_{k_A}, q'_1, \dots, q'_{k_B} \in \binom{(A' \cup B') \setminus W}{2}$ such that $p_{i-1} \cup p'_i$ and $p'_i \cup p_i$ are both edges of H for each $i \in [k_A]$ and $q_{i-1} \cup q'_i$ and $q'_i \cup q_i$ are both edges of H for each $i \in [k_B]$. We then define the paths $Q_1 := p_0 p'_1 p_1 p'_2 \dots p_{k_A-1} p'_{k_A} p_{k_A} s_1$ and $Q_2 := q_0 q'_1 q_1 q'_2 \dots q_{k_B-1} q'_{k_B} q_{k_B} s_2$. Observe that we then have $|(V(Q_1) \setminus L) \cap A| = 3k_A + 1$ and $|(V(Q_1) \setminus L) \cap B| = k_A + 1$, and likewise that $|(V(Q_2) \setminus L) \cap A| = k_B + 1$ and $|(V(Q_2) \setminus L) \cap B| = 3k_B + 1$. Since $|L \cap A| = |L \cap B| = 3m + 3$, we conclude that $R := V \setminus (L \cup Q_1 \cup Q_2)$ satisfies

$$\begin{aligned} |R \cap A| &= |A| - (3m + 3) - (3k_A + 1) - (k_B + 1) \\ &= \frac{n}{2} + \ell - 3m - 5 - \frac{3}{8}(n - 8m - 10 + 4\ell) - \frac{1}{8}(n - 8m - 10 - 4\ell) = m \end{aligned}$$

and $|R \cap B| = m$ by a similar calculation. Also every vertex of R is palatable, since by construction Q_1 and Q_2 cover all non-palatable vertices. It follows that L can swallow R , so there is a path P in H with vertex set $R \cup L$ and with ends p_0 and q_0 . This gives a Hamilton path $Q_1 P Q_2$ in H with ends s_1 and s_2 . The argument for the running time is essentially identical to that in the proof of Lemma 2.34. \square

Recalling the proof strategy at the start of the section, we will use the following simple proposition to find the short path P whose ends are to be connected using the Hamilton path connecting lemma (Lemma 2.42).

Proposition 2.43. *Adopt Setup 2.39, and suppose also that every vertex of V is (β_1, β_2) -medium. Also let $R \subseteq V$ have size $|R| \leq \psi n$. Then the following statements hold.*

- (i) *For every β_2 -medium connate pair p there exists a γ -good split pair $s \in \binom{V \setminus R}{2}$ such that $p \cup s \in E(H)$.*
- (ii) *For every β_2 -medium split pair s there exists a γ -good connate pair $p \in \binom{V \setminus R}{2}$ such that $s \cup p \in E(H)$.*
- (iii) *If there exists a β_2 -bad connate pair $p \in \binom{V \setminus R}{2}$, then there exist disjoint β_2 -medium connate pairs $p_1, p_2 \in \binom{V \setminus R}{2}$ such that $p_1 \cup p_2 \in E(H)$.*
- (iv) *If there exists a β_2 -bad split pair $s \in \binom{V \setminus R}{2}$, then there exist disjoint β_2 -medium split pairs $s_1, s_2 \in \binom{V \setminus R}{2}$ such that $s_1 \cup s_2 \in E(H)$.*

Proof. For (i) observe that, since p is β_2 -medium, there are $\beta_2 \binom{n}{2}$ split pairs $s \in \binom{V}{2}$ such that $p \cup s$ is an edge of H . By Proposition 2.41(c) at most $\frac{c}{\gamma^3 - 2\varepsilon} \binom{n}{2}$ such pairs are not γ -good and at most ψn^2 such pairs contain a vertex from R . Since $\beta_2 \binom{n}{2} > \frac{c}{\gamma^3 - 2\varepsilon} \binom{n}{2} + \psi n^2$, there exists a γ -good split pair $s \in \binom{V \setminus R}{2}$ such that $p \cup s \in E(H)$. A similar argument proves (ii).

For (iii), assume without loss of generality that $p \in \binom{A \setminus R}{2}$, and observe that there are at least $(\frac{1}{2} - 2\varepsilon) \binom{n}{2}$ pairs q for which $p \cup q \in E(H)$, but since p is β_2 -bad at most $\beta_2 \binom{n}{2}$ such pairs q are split pairs. It follows that $p \cup q \in E(H)$ for all but at most $2\beta_2 \binom{n}{2}$ connate pairs $q \in \binom{A}{2}$. Write $p = \{x_1, x_2\}$. Since both x_1 and x_2 are (β_1, β_2) -medium there are at least $\beta_1^2 \binom{n}{2}$ pairs $\{y_1, y_2\} \in \binom{A}{2}$ such that $\{x_1, y_1\}$ and $\{x_2, y_2\}$ are both β_2 -medium. Since at most $2\psi \binom{n}{2}$ pairs contain a vertex from R , and $2\beta_2 + 2\psi < \beta_1^2$, we can choose $\{y_1, y_2\}$ such that $e' = \{x_1, x_2, y_1, y_2\}$ is an even edge in $H[V \setminus R]$ containing two disjoint β_2 -medium connate pairs. A similar argument proves (iv). \square

The following definition is helpful in the proof of Lemma 2.10.

Definition 2.44. We say that an edge $e \in E(H)$ is

- (i) an even connate edge, if $|e \cap A| \in \{0, 4\}$,
- (ii) an even split edge, if $|e \cap A| = 2$.

Observe that for pairs p and q , if $p \cup q$ is an even connate edge then both p and q must be connate pairs, but if $p \cup q$ is an even split edge then either p and q are both connate pairs or p and q are both split pairs.

We are now ready to prove Lemma 2.10; in fact, we actually prove the following stronger algorithmic version of the lemma.

Lemma 2.45. *Suppose that $1/n \ll \varepsilon, c' \ll 1$ and that n is even, and let H be a 4-graph of order n with $\delta(H) \geq n/2 - \varepsilon n$. If H is c' -odd-extremal and every bipartition $\{A, B\}$ of $V(H)$ is odd-good, then H contains a Hamilton 2-cycle. Moreover, there exists an algorithm Procedure HamCycleOdd($H, \{A, B\}$) which, given as input a c' -odd-extremal bipartition $\{A, B\}$ of $V(H)$, returns a Hamilton 2-cycle in H in time $O(n^{12})$.*

Proof. First we introduce further constants such that

$$\varepsilon, c' \ll c \ll \gamma \ll \beta_2 \ll \beta_2' \ll \beta_1 \ll \beta_1' \ll \mu \ll 1.$$

Since H is c' -odd-extremal there exists a bipartition $\{A', B'\}$ with $n/2 - c'n \leq |A'| \leq n/2 + c'n$ for which there are at most $c' \binom{n}{4}$ even edges. We begin by moving all vertices which are (β_1', β_2') -bad to the other side. To be precise define $A^{\text{bad}} = \{a \in A' \mid a \text{ is } (\beta_1', \beta_2')\text{-bad}\}$ and $B^{\text{bad}} = \{b \in B' \mid b \text{ is } (\beta_1', \beta_2')\text{-bad}\}$ and we set $A := (A' \setminus A^{\text{bad}}) \cup B^{\text{bad}}$ and $B := (B' \setminus B^{\text{bad}}) \cup A^{\text{bad}}$; we say that the vertices of $A^{\text{bad}} \cup B^{\text{bad}}$ are *moved*.

Claim 2.46. $\{A, B\}$ is a c -odd-extremal bipartition of V with respect to which every vertex of H is (β_1, β_2) -medium.

Proof of Claim 2.46. By Proposition 2.41(e) at most $5c'n$ vertices are moved in total, so $n/2 - cn \leq n/2 - 6c'n \leq |A|, |B| \leq n/2 + 6c'n \leq n/2 + cn$, and at most $c' \binom{n}{4} + 5c'n \binom{n}{3} \leq c \binom{n}{4}$ edges of H are even with respect to $\{A, B\}$. This proves that $\{A, B\}$ is c -odd-extremal. Also, any vertex which was not moved was (β'_1, β'_2) -medium with respect to $\{A', B'\}$, and so is (β_1, β_2) -medium with respect to $\{A, B\}$. So it remains to show that each moved vertex v is also (β_1, β_2) -medium with respect to $\{A, B\}$. Without loss of generality assume that $v \in A^{\text{bad}}$. First consider the case that there are fewer than $\beta_1 n$ vertices $a \in A$ for which the pair $\{v, a\}$ was β_2 -medium with respect to $\{A', B'\}$. Then v was contained in at least $1/3 \cdot (n/2 - 2\beta_1 n) \cdot (1/2 - 2\beta_2) \binom{n}{2} > (1/4 - \mu) \binom{n}{3}$ edges which were even with respect to $\{A', B'\}$. Since there can be at most $(1/8 + \mu) \binom{n}{3}$ even connate edges in total, this implies that v was contained in at least $(1/8 - 2\mu) \binom{n}{3}$ even split edges (with respect to $\{A', B'\}$). Now consider the other case that there are fewer than $\beta_1 n$ vertices $b \in B$ for which $\{v, b\}$ was β_2 -medium with respect to $\{A', B'\}$. Then v was contained in at least $1/3 \cdot (1/2 - 2\beta_1)n \cdot (1/2 - 2\beta_2) \binom{n}{2} > (1/4 - \mu) \binom{n}{3}$ even split edges (with respect to $\{A', B'\}$). In either case we conclude that v , now after the moving, is contained in at least $(1/8 - 3\mu) \binom{n}{3}$ edges which have precisely three vertices in B ; since $v \in B$ it follows that v is (β_1, β_2) -medium with respect to $\{A, B\}$. \diamond

For the rest of the proof we will not use the constants c', β'_1 and β'_2 , and we use the terms even, odd, split, connate, γ -good, β_2 -medium, (β_1, β_2) -medium and so forth exclusively with respect to the partition $\{A, B\}$. Observe that H, A, B and the remaining constants satisfy the conditions of Setup 2.39. Fix $m \in \{0, 2, 4, 6\}$ and $d \in \{0, 2\}$ with $m \equiv |V| \pmod{8}$ and $d \equiv |A| - |B| \pmod{4}$. We consider separately four cases for the pair (m, d) as in the definition of an odd-good bipartition (Definition 2.1). In each case we proceed by the following steps.

- (I) We use the fact that $\{A, B\}$ is odd-good to construct a path P_0 with at most 30 vertices whose ends s_1 and s_2 are γ -good split pairs such that, writing $V^* =$

$(V \setminus V(P_0)) \cup s_1 \cup s_2$, $A^* = A \cap V^*$ and $B^* = B \cap V^*$, we have either $|V^*| \equiv 6 \pmod{8}$ and $|A^*| - |B^*| \equiv 2 \pmod{4}$ or $|V^*| \equiv 2 \pmod{8}$ and $|A^*| - |B^*| \equiv 0 \pmod{4}$.

(II) We extend the path P_0 to a path P which contains all non- γ -good vertices such that $|V(P)| \equiv |V(P_0)| \pmod{8}$ and $|V(P) \setminus (V(P_0) \cap A)| \equiv |V(P) \setminus (V(P_0) \cap B)| \pmod{4}$ and whose ends are γ -good split pairs s_1 and s_3 .

(III) Finally, we apply Lemma 2.42 to find a Hamilton path Q in $H' = H[(V \setminus V(P)) \cup s_1 \cup s_3]$ with ends s_1 and s_3 . This gives a Hamilton cycle PQ in H .

It suffices to show that we can construct the path P_1 in each case. Indeed, having constructed the path P_0 , let X be the set of all non- γ -good vertices in V^* ; by adding a single further vertex to X if necessary, we may assume that $q := |X|$ is even. Every vertex of X is (β_1, β_2) -medium by Claim 2.46 and by Proposition 2.41(d) we have $q \leq \frac{c}{\gamma^5 - 2\epsilon}n + 1 < \gamma n / 25$. Write $X = \{x_1, \dots, x_q\}$, and greedily choose distinct vertices $y_1, \dots, y_q \in V^* \setminus (s_1 \cup s_2 \cup X)$ so that for each $i \in [q]$ the pair $\{x_i, y_i\}$ is β_2 -medium. We now form a path Q by the following iterative process. Write $g_0 := s_2$. Then, for each $i \in [q]$ in turn we proceed as follows to choose connate pairs $f_{3i-2}, f_{3i-1}, f_{3i}$ and split pairs $g_{3i-2}, g_{3i-1}, g_{3i}$. If $\{x_i, y_i\}$ is a connate pair, set $f_{3i-1} := \{x_i, y_i\}$, and apply Proposition 2.43(i) twice to obtain γ -good split pairs g_{3i-2} and g_{3i-1} such that $g_{3i-2} \cup f_{3i-1}$ and $f_{3i-1} \cup g_{3i-1}$ are both edges of H . Next choose a γ -good split pair g_{3i} , and apply Proposition 2.41(g) twice to choose connate pairs f_{3i-2} and f_{3i} such that each of $g_{3i-3}f_{3i-2}g_{3i-2}$ and $g_{3i-1}f_{3i}g_{3i}$ is a path of length two in H . On the other hand, if $\{x_i, y_i\}$ is a split pair, set $g_{3i-1} := \{x_i, y_i\}$, and apply Proposition 2.43(ii) twice to obtain γ -good connate pairs f_{3i-1} and f_{3i} such that $g_{3i-1} \cup f_{3i-1}$ and $f_{3i} \cup g_{3i-1}$ are both edges of H . Next apply Proposition 2.41(h) twice to obtain γ -good split pairs g_{3i} and g_{3i-2} such that $g_{3i-2} \cup f_{3i-1}$ and $f_{3i} \cup g_{3i}$ are both edges of H , and finally apply Proposition 2.41(g) to choose a connate

pair f_{3i-2} such that $g_{3i-3}f_{3i-2}g_{3i-2}$ is a path of length two in H . If we choose each pair to be disjoint from $V(P'_0)$ and from all previously-chosen pairs, having made these choices for every $i \in [q]$ we obtain the desired path $P_0 = P'_0g_0f_1g_1f_2g_2 \dots f_{3q}g_{3q}$ with ends s_1 and $s_3 := g_{3q}$. Observe that we then have $|V(P_0) \setminus V(P'_0)| = 12q$, and that each f_i is a connate pair and each g_i is a split pair. Since q is even it follows that $|V(P)| \equiv |V(P_0)| \pmod{8}$ and $|V(P) \cap A| - |V(P) \cap B| \equiv |V(P_0) \cap A| - |V(P_0) \cap B| \pmod{4}$, as required. This completes Step (II). Finally, since $|V(P)| \leq |V(P_0)| + 12q \leq \gamma n/2$, we may then apply Lemma 2.42 to find a Hamilton path Q as claimed in Step (III).

We now show how to construct the path P_0 in each case.

Case A: $(m, d) = (0, 0)$ or $(m, d) = (4, 2)$. Using Proposition 2.41(c) we can choose a γ -good split pair s_1 . Then by Proposition 2.41(i) we can find a γ -good connate pair $p \in \binom{A}{2}$ such that $p \cup s_1$ is an edge of H . Finally, using Proposition 2.41(h) there is a γ -good split pair $s_2 \in \binom{V \setminus s_1}{2}$ such that $p \cup s_2$ is an edge of H , and then $P_0 = s_1ps_2$ is the desired path. Observe that we then have either $|V^*| \equiv 0 - 2 \equiv 6 \pmod{8}$ and $|A^*| - |B^*| \equiv 0 - 2 \equiv 2 \pmod{4}$, or $|V^*| \equiv 4 - 2 \equiv 2 \pmod{8}$ and $|A^*| - |B^*| \equiv 2 - 2 \equiv 0 \pmod{4}$.

Case B: $(m, d) = (2, 2)$ or $(m, d) = (6, 0)$. Since the bipartition $\{A, B\}$ of $V(H)$ is odd-good, in this case H must contain an even edge e . If e contains a β_2 -bad connate pair, then we apply Proposition 2.43(iii) with $R = \emptyset$ to obtain two disjoint β_2 -medium connate pairs p_1 and p_2 such that $p_1 \cup p_2$ is an edge of H . On the other hand, if e does not contain a β_2 -bad connate pair, then we may write $e = p_1 \cup p_2$ where p_1 and p_2 are disjoint connate pairs. In either case we obtain β_2 -medium connate pairs p_1 and p_2 such that $p_1 \cup p_2$ is an edge of H . So by Proposition 2.43(i) there are γ -good split pairs s_1, s_2 such that $s_1 \cup p_1$ and $s_2 \cup p_2$ are edges in H , and then $P_0 = s_1p_1p_2s_2$ is the desired path. Note that either $|V^*| \equiv 2 - 4 \equiv 6 \pmod{8}$ and $|A^*| - |B^*| \equiv 2 - 0 \equiv 2 \pmod{4}$ or $|V^*| \equiv 6 - 4 \equiv 2 \pmod{8}$ and $|A^*| - |B^*| \equiv 0 - 0 \equiv 0 \pmod{4}$.

Case C: $(m, d) = (4, 0)$ or $(m, d) = (0, 2)$. Since the bipartition $\{A, B\}$ of $V(H)$ is odd-

good, in this case H_{even} must have total 2-pathlength at least two. That is, H contains even edges e_1 and e_2 such that either e_1 and e_2 are disjoint or $|e_1 \cap e_2| = 2$.

Suppose first that e_1 and e_2 are disjoint. Similarly as in Case B, if e_1 contains a β_2 -bad connate pair, then we use Proposition 2.43(iii) with $R = e_2$ to obtain β_2 -medium connate pairs p_1 and p_2 such that $e'_1 := p_1 \cup p_2$ is an even edge of H . By replacing e_1 with e'_1 if necessary, we may assume that $e_1 = p_1 \cup p_2$ where p_1 and p_2 are β_2 -medium connate pairs, and the same argument applied to e_2 shows that we may assume that $e_2 = p_3 \cup p_4$ where p_3 and p_4 are β_2 -medium connate pairs. Then by Proposition 2.43(i) there are γ -good split pairs s_1, s_2, s_3 and s_4 such that $s_1 p_1 p_2 s_3$ and $s_4 p_3 p_4 s_2$ are vertex-disjoint paths in H . Then, by applying Proposition 2.41(i) twice, followed by Proposition 2.41(f), we obtain γ -good connate pairs q_1, q_2 and a split pair s_5 such that $P_0 = s_1 p_1 p_2 s_3 q_1 s_5 q_2 s_4 p_3 p_4 s_2$ is the desired path. Note that then either $|V^*| \equiv 4 - 18 \equiv 2 \pmod{8}$ and $|A^*| - |B^*| \equiv 0 - 0 \equiv 0 \pmod{4}$, or $|V^*| \equiv 0 - 18 \equiv 6 \pmod{8}$ and $|A^*| - |B^*| \equiv 2 - 0 \equiv 2 \pmod{4}$.

Now suppose that $|e_1 \cap e_2| = 2$, and write $f_1 = e_1 \setminus e_2$, $f_2 = e_1 \cap e_2$ and $f_3 = e_2 \setminus e_1$, so $f_1 f_2 f_3$ is a path in H . Since e_1 and e_2 are both even edges, f_1, f_2 and f_3 are either all connate pairs or all split pairs. If f_1 is not β_2 -medium, then we may apply Proposition 2.43(iii) or (iv) with $R = e_2$ to obtain an even edge e'_1 which is disjoint from e_2 ; we may then proceed as in the previous case with e'_1 and e_2 in place of e_1 and e_2 . So we may assume that f_1 is β_2 -medium, and by the same argument applied to f_3 we may assume that f_3 is β_2 -medium. If each of f_1, f_2 and f_3 is a connate pair, then we may apply Proposition 2.43(i) to obtain γ -good split pairs s_1, s_2 such that $s_1 \cup f_1$ and $s_2 \cup f_3$ are both edges of H . Then we may take $P_0 = s_1 f_1 f_2 f_3 s_2$, since we have either $|V^*| \equiv 4 - 6 \equiv 6 \pmod{8}$ and $|A^*| - |B^*| \equiv 0 - 2 = 2 \pmod{4}$, or $|V^*| \equiv 0 - 6 \equiv 2 \pmod{8}$ and $|A^*| - |B^*| \equiv 2 - 2 = 0 \pmod{4}$. On the other hand, if each of f_1, f_2 and f_3 is a split pair, then we may apply Proposition 2.43(ii) followed by Proposition 2.43(i) to first find γ -good connate pairs p_1 and p_2 , and then γ -good split pairs s_1 and s_2 , such that

$P_0 = s_1 p_1 f_1 f_2 f_3 p_2 s_2$ is a path in H ; we then have either $|V^*| \equiv 4 - 10 \equiv 2 \pmod{8}$ and $|A^*| - |B^*| \equiv 0 - 0 \equiv 0 \pmod{4}$, or $|V^*| \equiv 0 - 10 \equiv 6 \pmod{8}$ and $|A^*| - |B^*| \equiv 2 - 0 \equiv 2 \pmod{4}$.

Case D: $(m, d) = (6, 2)$ or $(m, d) = (2, 0)$. Since the bipartition $\{A, B\}$ of $V(H)$ is odd-good, in this case either H contains an even split edge or H_{even} has total 2-pathlength at least three.

Suppose first that e is an even split edge of H . If e contains a β_2 -bad split pair then we may apply Proposition 2.43(iv) to obtain disjoint β_2 -medium split pairs s'_1 and s'_2 such that $e' := s'_1 \cup s'_2$ is an even edge of H . By replacing e with e' if necessary, we may assume that $e = s'_1 \cup s'_2$ where s'_1 and s'_2 are β_2 -medium split pairs. We next apply Proposition 2.43(ii) twice to obtain γ -good connate pairs p_1 and p_2 such that $p_1 s'_1 s'_2 p_2$ is a path, and then Proposition 2.41(h) twice to obtain γ -good split pairs s_1 and s_2 such that $P_0 = s_1 p_1 s'_1 s'_2 p_2 s_2$ is the desired path in H . Note that then either $|V^*| \equiv 6 - 8 \equiv 6 \pmod{8}$ and $|A^*| - |B^*| \equiv 2 - 0 = 2 \pmod{4}$, or $|V^*| \equiv 2 - 8 \equiv 2 \pmod{8}$ and $|A^*| - |B^*| \equiv 0 - 0 = 0 \pmod{4}$.

From now on, for the rest of Case D, we may assume that H does not contain an even split edge and therefore any even edge is connate. So suppose now that H contains three disjoint even connate edges e_1, e_2 and e_3 . By using Proposition 2.43(iii) as in previous cases to replace e_1, e_2 or e_3 if necessary, we may assume that we can write $e_1 = q_1 \cup q'_1$, $e_2 = q_2 \cup q'_2$ and $e_3 = q_3 \cup q'_3$ where each of $q_1, q_2, q_3, q'_1, q'_2$ and q'_3 is a β_2 -medium connate pair. Then by several applications of Proposition 2.43(i) and Proposition 2.41(g) we may choose connate pairs p_1, p_2 and γ -good split pairs s_1, \dots, s_6 such that $P_0 = s_1 q_1 q'_1 s_3 p_1 s_4 q_2 q'_2 s_5 p_2 s_6 q_3 q'_3 s_2$ is the desired path in H . Note that then either $|V^*| \equiv 6 - 24 \equiv 6 \pmod{8}$ and $|A^*| - |B^*| \equiv 2 - 0 = 2 \pmod{4}$, or $|V^*| \equiv 2 - 24 \equiv 2 \pmod{8}$ and $|A^*| - |B^*| \equiv 0 - 0 = 0 \pmod{4}$.

Next suppose that H contains even connate edges e_1, e_2 and e_3 such that $|e_1 \cap e_2| = 2$ and e_3 is disjoint from $e_1 \cup e_2$. Then exactly as in Case C we may form a path $P'_0 =$

$s_1 f_1 f_2 f_3 s'_1$ where s_1 and s'_1 are γ -good split pairs and f_1, f_2 and f_3 are γ -good connate pairs, and we may do this so that e_3 is disjoint from $V(P'_0)$. Using Proposition 2.43(iii) as in previous cases to replace e_3 if necessary, we may then assume that $e_3 = q_1 \cup q_2$ where q_1 and q_2 are disjoint β_2 -medium connate pairs. By two applications of Proposition 2.43(i) we then choose γ -good split pairs s_2 and s'_2 such that $s'_2 \cup q_1$ and $s_2 \cup q_2$ are both edges of H ; finally, using Proposition 2.41(g) we obtain a connate pair p such that $s'_1 p s'_2$ is a path of length two in H . This gives the desired path $P_0 = s_1 f_1 f_2 f_3 s'_1 p s'_2 q_1 q_2 s_2$. Note that then either $|V^*| \equiv 6 - 16 \equiv 6 \pmod{8}$ and $|A^*| - |B^*| \equiv 2 - 0 = 2 \pmod{4}$, or $|V^*| \equiv 2 - 16 \equiv 2 \pmod{8}$ and $|A^*| - |B^*| \equiv 0 - 0 = 0 \pmod{4}$.

Finally, suppose that H contains a path of three even connate edges e_1, e_2 and e_3 (appearing in that order). Let $f_1 := e_1 \setminus e_2$, $f_2 := e_1 \cap e_2$, $f_3 := e_2 \cap e_3$, and $f_4 := e_3 \setminus e_2$, so each of f_1, f_2, f_3 and f_4 is a connate pair. If the pair f_1 is β -bad, then by applying Proposition 2.43(iii) to f_1 with $R = e_2 \cup e_3$ we obtain an even edge e'_1 of H disjoint from the path $f_2 f_3 f_4$ of two even connate edges. Since we assumed that H contains no even split edge this edge e'_1 is also an even connate edge, and we may then proceed as in the previous case. So we may assume that f_1 is β -medium, and likewise that f_4 is β -medium. Using Proposition 2.43(i) twice we find γ -good split pairs s_1, s_2 such that $P_0 = s_1 f_1 f_2 f_3 f_4 s_2$ is the desired path. Note that then either $|V^*| \equiv 6 - 8 \equiv 6 \pmod{8}$ and $|A^*| - |B^*| \equiv 2 - 0 = 2 \pmod{4}$, or $|V^*| \equiv 2 - 8 \equiv 2 \pmod{8}$ and $|A^*| - |B^*| \equiv 0 - 0 = 0 \pmod{4}$.

For the ‘moreover’ part of the statement, suppose that we are given a 4-graph H as in the lemma, and also an odd-extremal partition $\{A', B'\}$ of $V(H)$. We can then identify the sets A^{bad} and B^{bad} and form the partition $\{A, B\}$ in time $O(n^4)$, and in time $O(n)$ we can identify which of cases A–D holds for $\{A, B\}$. Furthermore we can find the at most three even edges which we use to begin the construction of P_0 in time at most $O(n^{12})$, and for each application of Proposition 2.41 or Proposition 2.43 to choose a pair we can

find such a pair by exhaustive search in time $O(n^2)$. In this way we can form P_0 as in Step (I) in time $O(n^{12})$ and then extend P_0 to P as in Step (II) in time $O(n^3)$. Finally, the application of Lemma 2.42 for Step (III) gives the desired path Q in time $O(n^4)$. \square

The running-time for Lemmas 2.37 and 2.45 could be improved by more careful arguments, but we abstain from this here, since these procedures do not provide the dominant term for the running-time of our main algorithm in Section 2.7.

2.7 An algorithm to find a Hamilton 2-cycle in a dense 4-graph

In this section we complete the proof of Theorem 1.12 by describing an algorithm which finds a Hamilton 2-cycle in a dense 4-graph in polynomial time, or certifies that no such cycle exists. First we show that we can find a Hamilton 2-cycle in a dense, connecting and absorbing 4-graph in polynomial time. For this we use Edmonds's well known algorithm [34] which finds a maximum matching in a 2-graph G of order n in time $O(n^4)$; we refer to this procedure here as `MaximumMatching(G)`.

Proposition 2.47. *Suppose that $1/n \ll \varepsilon \ll \beta \ll \alpha \ll \kappa$ and that n is even. There exists an algorithm `Procedure NonExtremalCase(H)` such that the following holds. If H is a 4-graph of order n with $\delta(H) \geq n/2 - \varepsilon n$ which is κ -connecting and (α, β) -absorbing, then `Procedure NonExtremalCase(H)` returns a Hamilton 2-cycle in H in time $O(n^{32})$.*

Proof. Introduce constants satisfying $1/n \ll \varepsilon \ll \gamma \ll \beta \ll \alpha \ll \lambda \ll \mu \ll \kappa$. By combining the Procedures `AbsorbingPath` and `LongCycle` from Lemmas 2.25 and 2.28 we can find a long cycle C in H and a graph G on $V(H) \setminus V(C)$ which contains a perfect matching M with $|M| \leq \gamma n$, and for each edge $e \in M$ there will be at least $2\gamma n$ vertex-disjoint segments of C which are absorbing structures for e . We can find this perfect matching by using `MaximumMatching(G)`. We can now for each $e \in M$ find

(by exhaustive search) a segment P_e which is an absorbing structure for e and which is disjoint from each segment $P_{e'}$ chosen for each previously-considered $e' \in M$. There will always be a segment available, as each of the fewer than γn previously-chosen segments $P_{e'}$ intersects at most two of the segments which could be chosen for e . Replacing each P_e in C by a path with vertex set $V(P_e) \cup e$ and the same ends as P_e yields a Hamilton 2-cycle in H . For the complexity note that we can find an absorbing structure for each of the at most n edges in M in time $O(n)$ by exhaustive search. Therefore the dominant term for the running time is determined by the Procedure AbsorbingPath. \square

Using the above proposition together with the results of Sections 2.5 and 2.6 we can now describe a polynomial-time algorithm, Procedure HamCycle(H), which, given a 4-graph H with $\delta(H) \geq n/2 - \varepsilon n$ (where $\varepsilon > 0$ is a fixed constant), either finds a Hamilton 2-cycle in H or certifies that there is no such cycle. So the existence of this algorithm proves Theorem 1.12. In the latter case the certificate is a bipartition $\{A, B\}$ of $V(H)$ which is not both even-good and odd-good, as by Theorem 1.11 the existence of such a bipartition demonstrates that H has no Hamilton 2-cycle. Since it is straightforward to verify in polynomial time whether a given partition is even-good or odd-good, our choice for the certificate is justified. We now restate and prove Theorem 1.12.

Theorem 1.12 ([41]). *There exist a constant $\varepsilon > 0$ and an algorithm which, given a 4-graph H on n vertices with $\delta(H) \geq n/2 - \varepsilon n$, runs in time $O(n^{32})$ and returns either a Hamilton 2-cycle in H or a certificate that no such cycle exists (that is, a bipartition of $V(H)$ which is either not even-good or not odd-good).*

Proof of Theorem 1.12. Introduce constants satisfying $1/n \ll \varepsilon \ll \beta \ll \alpha \ll \kappa \ll c \ll 1$, and let H be a 4-graph of order n with $\delta(H) \geq n/2 - \varepsilon n$. By Corollary 2.3 there is an algorithm with running time $O(n^{25})$ which tests whether H contains a Hamilton 2-cycle and which, if this condition fails, returns a bipartition $\{A, B\}$ of $V(H)$ which is not

both even-good and odd-good. We first run this algorithm and, if H does not contain a Hamilton 2-cycle, then we return a certifying bipartition. We may therefore assume that H does contain a Hamilton 2-cycle and that every bipartition of $V(H)$ is both even-good and odd-good, and in particular that H is of even order. Note that we can test in time $O(n^8)$ whether H is κ -connecting by counting, for each of the $3\binom{n}{4}$ possible disjoint pairs $p_1, p_2 \in \binom{V}{2}$, the number of paths of length two or three with ends p_1 and p_2 . Similarly, we can test in time $O(n^{10})$ whether H is (α, β) -absorbing by counting, for each of the $\binom{n}{2}$ possible pairs $p \in \binom{V}{2}$, the number of octuples of vertices from V which form an absorbing structure for p . Suppose that H is not κ -connecting. Then H must be c -even-extremal by Lemma 2.13, and Procedure EvenPartition (defined in Lemma 2.13) returns a c -even-extremal partition of $\{A, B\}$ in time $O(n^8)$. Procedure HamCycleEven (see Lemma 2.37) then returns a Hamilton 2-cycle in time $O(n^{10})$. Likewise, if H is not (α, β) -absorbing then H must be c -odd-extremal by Lemma 2.21, and Procedure OddPartition (defined in Lemma 2.21) returns a c -odd-extremal partition of $\{A, B\}$ in time $O(n^6)$. We can then use Procedure HamCycleOdd (see Lemma 2.45) to return a Hamilton 2-cycle in time $O(n^{12})$. This leaves only the case when H is both κ -connecting and (α, β) -absorbing; in this case we can use Procedure NonExtremalCase (see Proposition 2.47) to return a Hamilton 2-cycle in time $O(n^{32})$. So in each case we can find a Hamilton 2-cycle in H in time at most $O(n^{32})$, as claimed. \square

2.8 Tight Hamilton cycles in k -graphs

In this section we describe the proof of Theorem 1.9, for which we use the following notation. For a function $f(n)$, we write $\text{HC}(k, f)$ (respectively $\text{HP}(k, f)$) to denote the k -graph tight Hamilton cycle (respectively tight Hamilton path) decision problem restricted to k -graphs H with minimum codegree $\delta(H) \geq f(|V(H)|)$. On the other hand, for an integer D , we write $\overline{\text{HC}}(k, D)$ (respectively $\overline{\text{HP}}(k, D)$) to denote the k -graph tight

Hamilton cycle (respectively tight Hamilton path) decision problem restricted to k -graphs H with maximum codegree $\Delta(H) \leq D$. Our starting point is the following theorem of Garey, Johnson and Stockmeyer [44] on subcubic graphs (we say that a graph G is *subcubic* if G has maximum degree $\Delta(G) \leq 3$).

Theorem 2.48 ([44]). *The problem of determining whether a subcubic graph admits a Hamilton cycle (i.e. $\overline{\text{HC}}(2, 3)$) is NP-complete.*

Fix any integers $k \geq 2$ and D . We first show that there are polynomial-time reductions from $\overline{\text{HC}}(k, D)$ to $\overline{\text{HC}}(2k - 1, 2D)$ and from $\overline{\text{HC}}(k, D)$ to $\overline{\text{HC}}(2k, D)$. For this, let H be a k -graph with vertex set V , let A and B be disjoint copies of V , and let $\varphi_A : A \rightarrow V$ and $\varphi_B : B \rightarrow V$ be the corresponding bijections. We define \overline{H}_{2k-1} to be the $(2k - 1)$ -graph with vertex set $A \cup B$ whose edges are all sets $e \in \binom{A \cup B}{2k-1}$ such that either $\varphi_A(e \cap A) \in E(H)$ and $\varphi_B(e \cap B) \subseteq \varphi_A(e \cap A)$, or $\varphi_B(e \cap B) \in E(H)$ and $\varphi_A(e \cap A) \subseteq \varphi_B(e \cap B)$. Likewise we define \overline{H}_{2k} to be the $2k$ -graph with vertex set $A \cup B$ whose edges are all sets $e \in \binom{A \cup B}{2k}$ such that $\varphi_A(e \cap A) \in E(H)$, $\varphi_B(e \cap B) \in E(H)$. Then it is easy to check that

- (a) either H , \overline{H}_{2k-1} and \overline{H}_{2k} all contain tight Hamilton cycles, or none of them does, and
- (b) if $\Delta(H) \leq D$, then $\Delta(\overline{H}_{2k-1}) \leq 2D$ and $\Delta(\overline{H}_{2k}) \leq D$,

so this construction gives the desired reductions.

We next show that there are polynomial-time reductions from $\overline{\text{HP}}(k, D)$ to $\text{HC}(2k - 1, \lfloor \frac{n}{2} \rfloor - k(D + 1))$ and from $\overline{\text{HC}}(k, D)$ to $\text{HC}(2k, \frac{n}{2} - k(D + 1))$. For the first reduction, let H be a k -graph on n vertices; an elementary reduction shows that we may assume without loss of generality that k divides n . Set $\ell := n/k$ and $U := V(H)$, and let X be a set of size $|X| = \frac{(k-1)n}{k} = \ell(k - 1)$ which is disjoint from U . Next, set $A := U \cup X$, and let B be a set of size $|B| = |A| + 1 = \ell(2k - 1) + 1$ which is disjoint from A . Define H_{2k-1} to be the $(2k - 1)$ -graph with vertex set $A \cup B$ whose edges are all sets $e \in \binom{A \cup B}{2k-1}$ with

$|A \cap e| \notin \{k, k+1\}$, or with $|A \cap e| = k$ and $A \cap e \in E(H)$, or with $|A \cap e| = k+1$ and such that $e' \not\subseteq A \cap e$ for every $e' \in E(H)$. For the second reduction let H again be a k -graph on n vertices, let $S := V(H)$ and let T be a set of size n which is disjoint from S , and define H_{2k} to be the $2k$ -graph with vertex set $S \cup T$ whose edges are all sets $e \in \binom{S \cup T}{2k}$ such that $|S \cap e| \notin \{k, k+1\}$, or such that $|S \cap e| = k$ and $S \cap e \in E(H)$, or such that $|S \cap e| = k+1$ and $e' \not\subseteq S \cap e$ for every $e' \in E(H)$. Observe that then

- (c) H_{2k-1} contains a tight Hamilton cycle if and only if H contains a tight Hamilton path,
- (d) H_{2k} contains a tight Hamilton cycle if and only if H contains a tight Hamilton cycle, and
- (e) if $\Delta(H) \leq D$, then $\delta(H_{2k-1}) \geq |A| - k(D+1)$ and $\delta(H_{2k}) \geq n - (D+1)k$.

Since H_{2k-1} has precisely $|A| + |B| = 2|A| + 1$ vertices and H_{2k} has precisely $2n$ vertices, this establishes the desired reductions.

Finally, observe that there are elementary polynomial-time reductions from $\overline{\text{HC}}(k, D)$ to $\overline{\text{HP}}(k, D)$ and from $\overline{\text{HP}}(k, D)$ to $\overline{\text{HC}}(k, D)$; together with the above reductions and Theorem 2.48 this observation completes the proof of Theorem 1.9.

2.9 Algorithmic details for Sections 2.4 and 2.7

In this section we give further details of each of the procedures from Sections 2.4 and 2.7, to enable the reader to more easily verify the correctness and claimed running-time of each such procedure. In each case we will adopt the notation of the proof of the corresponding lemma, and so the algorithm provided here should be read in conjunction with the corresponding proof. Furthermore, we adopt the following constant hierarchy for all procedures in this section.

$$1/n \ll 1/D \ll \varepsilon \ll \gamma \ll \beta \ll \omega \ll \varphi \ll \alpha \ll \lambda \ll \mu \ll \kappa \ll \eta \ll c \ll 1.$$

Lemma 2.13: Let H be a 4-graph on n vertices which satisfies $\delta(H) \geq n/2 - \varepsilon n$. If H is not κ -connecting, then Procedure EvenPartition(H) returns a c -even-extremal bipartition $\{A, B\}$ of $V(H)$ in time $O(n^8)$.

Procedure EvenPartition(H)

Data: A dense 4-graph H which is not κ -connecting.

Result: A c -even-extremal bipartition $\{A, B\}$ of $V(H)$.

By exhaustive search, find disjoint pairs $\{a_1, a_2\}, \{b_1, b_2\} \in \binom{V(H)}{2}$ such that H contains fewer than κn^2 paths of length 2 and fewer than κn^4 paths of length 3 whose ends are $\{a_1, a_2\}$ and $\{b_1, b_2\}$.

Construct the edge-coloured complete 2-graph G on $V(H)$ as in the proof of Lemma 2.13.

Find a monochromatic triangle T^* fulfilling Claim 2.14 (for the given value of η) by exhaustive search.

return $A := N_{\text{red}}(T^*)$ and $B := V(H) \setminus A$.

Lemma 2.18: Suppose that $1/n \ll \rho \ll \lambda, \kappa$. If $H = (V, E)$ is a 4-graph of order n which is κ -connecting, and G is a 2-graph on the same vertex set V with $\delta(G) \geq n - \lambda n$, then Procedure SelectReservoir(H, G, ρ) returns in time $O(n^{16})$ a subset $R \subseteq V$ such that

- (a) $(1 - 4\rho)\rho n \leq |R| \leq \rho n$,
- (b) for every $x \in V$ we have $|N_G(x) \cap R| \geq (1 - 35\lambda)|R|$ and
- (c) for every disjoint $p_1, p_2 \in \binom{V}{2}$ there are at least $\frac{\kappa}{5}|R|$ internally disjoint paths of length at most three in $H[R \cup p_1 \cup p_2]$ with ends p_1 and p_2 .

Procedure SelectReservoir(H, G, ρ)

Data: A 4-graph H with vertex set V , a 2-graph G with vertex set V and a constant $\rho > 0$.

Result: A reservoir set $R \subseteq V$.

$U := \{\{p, p'\} \subseteq \binom{V}{2} \mid p \cap p' = \emptyset\}$ and $W := \binom{V}{4}$.

$E_1 := \{\{\{p, p'\}, S\} \mid \{p, p'\} \in U, S \in W \text{ and } H[S \cup p \cup p'] \text{ contains a path with ends } p, p'\}$.

$E'_1 := \{\{S, S'\} \in \binom{W}{2} \mid S \cap S' \neq \emptyset\}$.

$E_2 := \{\{S, S'\} \in \binom{W}{2} \mid \{u, v\} \in E(G) \text{ for all } u \in S, v \in S'\}$.

Construct graphs $G_1 := (U \cup W, E_1 \cup E'_1)$ and $G_2 := (W, E_2)$.

$R' := \text{SelectSet}(G_1, G_2, \frac{\rho}{4}n, \kappa, 17\lambda, \rho)$.

$R := \bigcup_{S \in R'} S$.

return R .

Lemma 2.21: Let H be a 4-graph on n vertices which satisfies $\delta(H) \geq n/2 - \varepsilon n$. If H is not (α, β) -absorbing, then Procedure $\text{OddPartition}(H)$ returns a c -odd-extremal bipartition $\{A, B\}$ of $V(H)$ in time $O(n^6)$.

Procedure $\text{OddPartition}(H)$

Data: A dense 4-graph H which is not (α, β) -absorbing.

Result: A c -odd-extremal bipartition $\{A, B\}$ of $V(H)$.

By exhaustive search, find an edge $\{x, y, x', y'\}$ of H as in Claim 2.22.

Form the edge-coloured complete graph K on $V(H)$ as in the proof of Lemma 2.21.

if *there is a monochromatic triangle T satisfying the conditions of Claim 2.23* **then**

| **return** $A := N_{\text{red}}(T)$ and $B := V(H) \setminus A$.

else

| By exhaustive search, find a normal vertex $v \in V(H)$ which is contained in at

| most $4\varphi n^2$ red triangles.

| **return** $A := N_{\text{red}}(v)$ and $B := V(H) \setminus A$.

Lemma 2.25: If H is a 4-graph of order n with $\delta(H) \geq n/2 - \varepsilon n$ which is κ -connecting and (α, β) -absorbing, then Procedure $\text{AbsorbingPath}(H)$ returns a path P and a graph G on $V(H)$ with the following properties in $O(n^{32})$.

- (i) P has at most μn vertices.
- (ii) Every vertex of $V(H) \setminus V(P)$ is contained in at least $n - \lambda n$ edges of G .
- (iii) For any edge e of G which does not intersect $V(P)$ there are at least $2\gamma n$ vertex-disjoint segments of P which are absorbing structures for e .

Procedure AbsorbingPath(H)

Data: A dense, κ -connecting and (α, β) -absorbing 4-graph H .

Result: An absorbing path P in H and a graph G on $V(H)$.

Set $W := V(H)^8$, set U to be the set of all β -absorbable pairs of vertices of H , and form the graph $G := (V(H), U)$.

Set $V_1 := U \cup W$,

$E_1 := \{\{p, T\} : p \in U, T \in W \text{ and } T \text{ is an absorbing structure for } p\}$,

$E'_1 := \{\{T, T'\} : T, T' \in W \text{ and } T \cap T' \neq \emptyset\}$ and form the graph

$G_1 := (V_1, E_1 \cup E'_1)$.

Set $\mathcal{T} := \text{SelectSet}(G_1, \emptyset, \beta^2 n, \beta, 1, \beta^2)$.

Delete all elements from \mathcal{T} which are not an absorbing structure for some β -absorbable pair.

Enumerate \mathcal{T} as $\{T_1, \dots, T_q\}$ and choose corresponding paths P_i for each $1 \leq i \leq q$.

Set $Q := \bigcup_{i=1}^q T_i$ and $X := \{v \in V(H) \setminus Q : d_G(v) < (1 - \lambda)n\}$.

Greedy form a path P_0 with $X \subseteq V(P_0) \subseteq V(H) \setminus Q$ as in the proof of Lemma 2.25.

Greedy choose paths Q_1, \dots, Q_q of length at most three connecting the paths P_0, \dots, P_q ; each Q_i may be chosen by exhaustive search.

return $P := P_0 Q_1 P_1 \dots Q_q P_q$ and G .

Theorem 2.26: Here we give a proof of the algorithmic statement. For this we define a *multi- k -graph* H to consist of a vertex set V and a multiplicity function $m_H : \binom{V}{k} \rightarrow \{0\} \cup \mathbb{N}$. We call $m_H(e)$ the *multiplicity* of e , and always count ‘with multiplicity’, so, for example, the number of edges of H is $e(H) := \sum_{e \in \binom{V}{k}} m_H(e)$, and the degree of a vertex $v \in V$ is $d_H(v) := \sum_{e \in \binom{V}{k}: v \in e} m_H(e)$. We begin with the following proposition.

Proposition 2.49. *Suppose that $1/n \ll 1/\ell, 1/r \ll d, 1/k$. Let H be a multi- k -graph on n vertices with $e(H) \geq d \binom{n}{k}$ in which all multiplicities are at most r . Then there exists a set $X \subseteq V(H)$ of size $k\ell$ such that $e(H[X]) \geq d \binom{\ell}{k}$. Moreover, such a set can be found in time $O(n^k)$.*

Proof. We proceed by induction on k ; the base case $k = 1$ is trivial. Suppose therefore that $k \geq 2$ and the proposition holds with $k-1$ in place of k . In time $O(n^k)$ we may choose vertices v_1, \dots, v_ℓ of H each with degree at least $\frac{d}{2} \binom{n}{k-1} + r\ell \binom{n}{k-2}$ in H ; such vertices must exist as otherwise we would have

$$\ell \cdot r \binom{n}{k-1} + n \cdot \left(\frac{d}{2} \binom{n}{k-1} + r\ell \binom{n}{k-2} \right) \geq \sum_{v \in V} d_H(v) = ke(H) \geq dk \binom{n}{k},$$

a contradiction. Observe that each v_i then lies in at least $\frac{d}{2} \binom{n}{k-1}$ edges which do not contain any v_j with $j \neq i$. Now form a multi- $(k-1)$ -graph H' with vertex set $V' := V(H) \setminus \{v_1, \dots, v_\ell\}$ by taking each $(k-1)$ -set $S \in \binom{V'}{k-1}$ to have multiplicity $m_{H'}(S) := \sum_{i=1}^{\ell} m_H(S \cup \{v_i\})$. Then by choice of the vertices v_i we have $e(H') \geq \ell \cdot \frac{d}{2} \binom{n}{k-1}$. So by our induction hypothesis (with $r\ell$ in place of r) we can find in time $O(n^{k-1})$ a set $X' \subseteq V'$ of size $(k-1)\ell$ such that $H'[X']$ has at least $\ell \frac{d}{2} \binom{\ell}{k-1} \geq d \binom{\ell}{k}$ edges; taking $X := X' \cup \{v_1, \dots, v_\ell\}$ gives the desired set. \square

The following corollary is the algorithmic version of Theorem 2.26.

Corollary 2.50 ([35]). *Suppose that $1/n \ll d, 1/f, 1/k$. Let F be a k -partite k -graph on f vertices. If H is a k -graph on n vertices with $e(H) \geq d \binom{n}{k}$, then we can find a copy of F in H in time $O(n^k)$.*

Proof. Introduce a constant ℓ with $1/n \ll 1/\ell \ll d, 1/f, 1/k$. By Proposition 2.49 (with $r = \ell$) we may find in time $O(n^k)$ a set $X \subseteq V(H)$ of size $k\ell$ such that $e(H[X]) \geq d \binom{\ell}{k}$. The non-algorithmic part of Theorem 2.26 then implies that $H[X]$ contains a copy of F , and we can find such a copy in constant time by exhaustive search. \square

Lemma 2.28: Suppose that n is even. Let H be a 4-graph of order n with $\delta(H) \geq n/2 - \varepsilon n$ which is κ -connecting. Also let P_0 be a 2-path in H on at most μn vertices, and let G be a 2-graph on $V(H)$ such that each vertex $v \in V(H) \setminus V(P_0)$ has $d_G(v) \geq (1 - \lambda)n$. Then Procedure LongCycle(H, G, P_0) returns in time $O(n^{16})$ a 2-cycle C on at least $(1 - \gamma)n$ vertices such that P_0 is a segment of C and $G[V(H) \setminus V(C)]$ contains a perfect matching.

Procedure LongCycle(H, G, P_0)

Data: A dense and κ -connecting 4-graph H , a very dense 2-graph G on $V(H)$, and a 2-path P_0 in H .

Result: A long cycle C which contains P_0 as a segment and such that

$G[V(H) \setminus V(C)]$ contains a perfect matching.

$V' := V(H) \setminus V(P_0)$, $H' := H[V']$, $G' := G[V']$, $R := \text{SelectReservoir}(H', G', 2\gamma/3)$.

Extend P_0 by a single edge at each end to a path P'_0 .

Set $L := V(H) \setminus (V(P'_0) \cup R)$ and $R' := R \setminus V(P'_0)$.

Extend P'_0 greedily with vertices only from L to a path P of length $\geq (1/2 - \mu)n$.

Set $L := V(H) \setminus (V(P) \cup R)$ and $R' := R \setminus V(P)$.

while $|L| > \frac{2}{3}n$ **do**

if $e(H[L]) > \mu \binom{|L|}{4}$ **then**

 Use Theorem 2.27 to find a path P' in $H[L]$ on at least $\frac{\mu|L|}{4} - 1$ vertices.

 Only using vertices from R' find a path Q of length at most three (by exhaustive search) such that $P := PQP'$ is a path.

 Set $L := V(H) \setminus (V(P) \cup R)$ and $R' := R \setminus V(P)$.

else

 Find I, J , and H_0 as in Claim 2.29 by exhaustive search.

 Use Theorem 2.26 to find a complete 3-partite 3-graph K with all vertex classes of size $D/3$.

 Derive the complete 4-partite 4-graph K' from K and J .

 Find a Hamilton path Q in K' .

 Delete $P[I]$ from P and call the resulting two segments P_1 and P_2 .

 Find by exhaustive search at most 8 vertices in R' forming two paths Q_1 and Q_2 such that $P := P_1Q_1QQ_2P_2$ is a path.

 Set $L := V(H) \setminus (V(P) \cup R)$ and $R' := R \setminus V(P)$.

By exhaustive search find at most 4 vertices in R' to close P to a cycle C .

return C .

Lemma 2.47: Suppose that n is even. If H is a 4-graph of order n with $\delta(H) \geq n/2 - \varepsilon n$ which is κ -connecting and (α, β) -absorbing, then Procedure NonExtremalCase(H) returns a Hamilton cycle in H in time $O(n^{32})$.

Procedure NonExtremalCase(H)

Data: A dense, κ -connecting and (α, β) -absorbing 4-graph H .

Result: A Hamilton cycle C in H .

Set $(P, G) := \text{AbsorbingPath}(H)$.

Set $C := \text{LongCycle}(H, G, P)$.

Set $M := \text{MaximumMatching}(G[V(H) \setminus V(C)])$.

for $e \in M$ **do**

Find a segment P_e in P which is an absorbing structure for e and which is
disjoint from each segment $P_{e'}$ chosen for each previously-considered $e' \in M$.
Replace P_e in C by a path with vertex set $V(P_e) \cup e$ and the same ends as P_e .

return C .

Theorem 1.12: If H is a 4-graph of order n with $\delta(H) \geq n/2 - \varepsilon n$, then Procedure HamCycle($H, \{A, B\}$) either returns a Hamilton cycle in H or a non-even-good or non-odd-good bipartition of $V(H)$ in time $O(n^{32})$.

Procedure $\text{HamCycle}(H)$

Data: A dense 4-graph H .

Result: A Hamilton 2-cycle in H or a certificate for non-existence.

if H contains a Hamilton cycle (Corollary 2.3) **then**

if H is not κ -connecting **then**

 Set $\{A, B\} := \text{EvenPartition}(H)$.

return $\text{HamCycleEven}(H, \{A, B\})$.

else if H is not (α, β) -absorbing **then**

 Set $\{A, B\} := \text{OddPartition}(H)$.

return $\text{HamCycleOdd}(H, \{A, B\})$.

else

return $\text{NonExtremalCase}(H)$.

else

return A bipartition $\{A, B\}$ of $V(H)$ which is not both even-good and odd-good.

2.10 Full proofs for the reductions claimed in Section 2.8

In this section we give full details of the proofs of the the correctness of the polynomial-time reductions claimed in the proof of Theorem 1.9 in Section 2.8. We do this through the following propositions. Throughout this section we write simply ‘path’ and ‘cycle’ to mean tight path and tight cycle, as these are the only types of paths and cycles considered here.

Proposition 2.51. *For any $k \geq 2$ and any D , there is a polynomial-time reduction from $\overline{\text{HP}}(k, D)$ to $\overline{\text{HC}}(k, D)$ and a polynomial-time reduction from $\overline{\text{HC}}(k, D)$ to $\overline{\text{HP}}(k, D)$.*

Proof. First, we show that there is a polynomial-time reduction from $\overline{\text{HP}}(k, D)$ to $\overline{\text{HC}}(k, D)$. If $D < 2$, then $\overline{\text{HP}}(k, D)$ is trivial, so assume that $D \geq 2$. Let H be a k -graph with $\Delta(H) \leq D$. For each of the at most n^{2k} ordered $2k$ -tuples $(x_1, \dots, x_k, y_1, \dots, y_k)$ such that $x = (x_1, \dots, x_k)$ and $y = (y_1, \dots, y_k)$ are both edges of H we construct an altered graph $H_{(x,y)}$ by deleting all edges $e \in E(H)$ with $x_k \in e$ or $y_1 \in e$ and then adding the edges $\{x_{i+1}, x_{i+2}, \dots, x_k, y_1, \dots, y_i\}$ for $0 \leq i \leq k$, so $(x_1, \dots, x_k, y_1, \dots, y_k)$ is a path in $H_{(x,y)}$. We then test each $H_{(x,y)}$ for a Hamilton cycle. The original graph H contains a Hamilton path starting with y and ending with x if and only if the altered graph $H_{(x,y)}$ contains a Hamilton cycle, and since $D \geq 2$ and $\Delta(H) \leq D$ we have $\Delta(H_{(x,y)}) \leq D$.

Now, we show that there is a polynomial-time reduction for the other direction as well. Given a k -graph H with $\Delta(H) \leq D$, for each of the at most n^{2k} sequences of $2k$ vertices $S = (x_1, \dots, x_k, y_k, \dots, y_1)$ such that S is a path in H we construct a graph H_S by deleting every edge e of H which intersects S except for those edges e such that $e \cap S = \{x_1, \dots, x_\ell\}$ or $e \cap S = \{y_1, \dots, y_\ell\}$ for some $\ell \leq k$. Afterwards we test each H_S for a Hamilton path. Observe that we have $\Delta(H_S) \leq \Delta(H) \leq D$, and that there is

a Hamilton cycle in H if and only if one of the altered graphs H_S contains a Hamilton path. \square

Proposition 2.52. *For any $k \geq 2$ and any D , there is a polynomial-time reduction from $\overline{\text{HC}}(k, D)$ to $\overline{\text{HC}}(2k - 1, 2D)$.*

Proof. Given a k -graph H with vertex set $V = \{v_1, \dots, v_n\}$, we construct a $(2k - 1)$ -graph \overline{H}_{2k-1} as follows. Fix disjoint copies $A = \{v_1^A, \dots, v_n^A\}$ and $B = \{v_1^B, \dots, v_n^B\}$ of V , and let $\varphi_A : A \rightarrow V$ and $\varphi_B : B \rightarrow V$ be the natural bijections (so $\varphi(v_i^A) = v_i = \varphi(v_i^B)$). For convenience we will not always mention the explicit bijections; instead we say that vertices $a \in A$, $b \in B$ and $x \in V$ correspond if $\varphi_A(a) = \varphi_B(b) = x$. We take $A \cup B$ to be the vertex set of \overline{H}_{2k-1} , and the edges of \overline{H}_{2k-1} are the sets $e \in \binom{A \cup B}{2k-1}$ such that either $\varphi_A(e \cap A)$ is an edge of H and $\varphi_B(e \cap B) \subseteq \varphi_A(e \cap A)$, or $\varphi_B(e \cap B)$ is an edge of H and $\varphi_A(e \cap A) \subseteq \varphi_B(e \cap B)$. It is then sufficient to show that

- (i) H contains a Hamilton cycle if and only if \overline{H}_{2k-1} contains a Hamilton cycle, and
- (ii) if $\Delta(H) \leq D$, then $\Delta(\overline{H}_{2k-1}) \leq 2D$.

For (i), first observe that if $C = (v_1, \dots, v_n)$ is a Hamilton cycle in H , then $\overline{C} := (v_1^A, v_1^B, \dots, v_n^A, v_n^B)$ is a Hamilton cycle in \overline{H}_{2k-1} . Indeed, every consecutive subsequence S in \overline{C} of length $2k - 1$ either contains k vertices from A or contains k vertices from B . These k vertices correspond to k consecutive vertices of C and therefore correspond to some edge $e \in E(H)$, and the remaining $k - 1$ vertices of S correspond to a subset of e , so S forms an edge of \overline{H}_{2k-1} . Now suppose that some cyclic ordering of the vertices of \overline{H}_{2k-1} gives a Hamilton cycle C in \overline{H}_{2k-1} . It suffices to show that if we delete every vertex of B from this sequence, then every k consecutive vertices (a_1, \dots, a_k) in the remaining subsequence correspond to an edge of H , since this subsequence would then correspond to a Hamilton cycle in H . For this, let Q be the subsequence of length $2k - 1$ in C beginning with a_1 , and let Q' be the subsequence of length $2k - 1$ in C beginning with the vertex

subsequent to a_1 in C . Suppose first that Q contains all of the vertices a_1, \dots, a_k . In this case $\{a_1, \dots, a_k\}$ must correspond to an edge of H , since $Q \in E(\overline{H}_{2k-1})$. Now suppose instead that Q does not contain a_k . The fact that $Q \in E(\overline{H}_{2k-1})$ then implies that Q contains k vertices of B which correspond to an edge of H , and that a_1, \dots, a_{k-1} each correspond to vertices in $B \cap Q$. Since $Q \setminus Q' = \{a_1\}$, and Q' is also an edge of $E(\overline{H}_{2k-1})$, we must have $Q' \setminus Q = \{a_k\}$. It follows that a_k corresponds to a vertex of $B \cap Q' = B \cap Q$. Thus the k vertices a_1, \dots, a_k correspond to the k vertices of $Q \cap B$, and so correspond to an edge of H .

For (ii), fix a set $S \in \binom{A \cup B}{2k-2}$. It is only possible that S is included in an edge of \overline{H}_{2k-1} if either $|A \cap S| = k-2$, $|B \cap S| = k$ and the vertices of $B \cap S$ each correspond to vertices of $A \cap S$, or the same holds with the roles of A and B reversed, or $|A \cap S| = |B \cap S| = k-1$ and at least $k-2$ vertices of $A \cap S$ correspond to vertices of $B \cap S$. In the first case we have $d_{\overline{H}_{2k-1}}(S) = 2$, since there are precisely two vertices which can be added to S to form an edge of \overline{H}_{2k-1} , namely the vertices of $A \setminus S$ which correspond to vertices of $B \cap S$. Likewise we also have $d_{\overline{H}_{2k-1}}(S) = 2$ in the second case. Finally, suppose that $|A \cap S| = |B \cap S| = k-1$. If exactly $k-2$ vertices of $A \cap S$ correspond to vertices of $B \cap S$, then again there are two edges of \overline{H}_{2k-1} containing S , formed by adding the vertex of $A \setminus S$ which corresponds to a vertex of $B \cap S$ or vice versa. If instead all $k-1$ vertices of $A \cap S$ correspond to vertices of $B \cap S$, then we can form an edge of \overline{H}_{2k-1} containing S by adding any vertex of A or B which corresponds to a neighbour in H of the corresponding $k-1$ vertices of H . Since $\Delta(H) \leq D$ there are at most $2D$ such vertices. So in all cases we have $d_{\overline{H}_{2k-1}}(S) \leq 2D$, as required. \square

Proposition 2.53. *For any $k \geq 2$ and any D , there is a polynomial-time reduction from $\overline{\text{HC}}(k, D)$ to $\overline{\text{HC}}(2k, D)$.*

Proof. Given a k -graph H with vertex set $V = \{v_1, \dots, v_n\}$, we take disjoint copies $A = \{v_1^A, \dots, v_n^A\}$ and $B = \{v_1^B, \dots, v_n^B\}$ of V and define φ_A and φ_B as in the proof of

Proposition 2.52. We then construct a $2k$ -graph \overline{H}_{2k} as follows: the vertex set of \overline{H}_{2k} is $A \cup B$, and the edges of \overline{H}_{2k} are all sets $e \in \binom{A \cup B}{2k}$ such that $\varphi_A(e \cap A)$ and $\varphi_B(e \cap B)$ are both edges of H . It is then sufficient to show that

- (i) H contains a Hamilton cycle if and only if \overline{H}_{2k} contains a Hamilton cycle, and
- (ii) $\Delta(\overline{H}_{2k}) \leq \Delta(H)$.

To show (i), first assume that H contains a Hamilton cycle $C = (v_1, \dots, v_n)$. Then $\overline{C} := (v_1^A, v_1^B, \dots, v_n^A, v_n^B)$ is a Hamilton cycle in \overline{H}_{2k} , since for every consecutive subsequence S of length $2k$ in \overline{C} , each of the sets $S \cap A$ and $S \cap B$ corresponds to a consecutive subsequence in C of length k , that is, an edge of H , so S is an edge of \overline{H}_{2k} . For the other direction, suppose that $C = (v_1, \dots, v_{2n})$ is a Hamilton cycle in \overline{H}_{2k} . Then for any $v_i \in A$ the set $Q = \{v_i, \dots, v_{i+2k-1}\}$ is an edge of \overline{H}_{2k} , so $Q \cap A$ contains exactly k vertices from A , and moreover these k vertices correspond to an edge of H . So if we delete all vertices of B from the sequence (v_1, \dots, v_{2n}) , the resulting subsequence corresponds to a Hamilton cycle in H .

For (ii), let S be a set of $2k - 1$ vertices of \overline{H}_{2k} . It is only possible that S is included in an edge of H if $S \cap A$ is an edge of H and $|S \cap B| = k - 1$, or the same with the roles of A and B reversed; without loss of generality we assume the former. A necessary condition for $\{x\} \cup S$ to be an edge of \overline{H}_{2k} is then that $x \in B$ and that $\varphi_B(\{x\} \cup (S \cap B)) \in E(H)$. The number of such vertices x is at most $d_H(\varphi_B(S \cap B)) \leq \Delta(H)$, so $d_{\overline{H}_{2k}}(S) \leq \Delta(H)$. \square

It is convenient to note that the problem of deciding whether a k -graph with $\Delta(H) \leq D$ contains a Hamilton path reduces to the problem of whether a k -graph H with $\Delta(H) \leq D$ whose order is divisible by k contains a Hamilton path. We refer to the latter problem as $\overline{\text{HP}}(k, D)^\perp$.

Proposition 2.54. *For any $k \geq 2$ and any D , there is a polynomial-time reduction from $\overline{\text{HP}}(k, D)$ to $\overline{\text{HP}}(k, D)^\perp$.*

Proof. Let H be a k -graph on n vertices; we may assume that $\Delta(H) \geq 2$, as otherwise the Hamilton path problem is trivial. Let $0 \leq r \leq k-1$ and $\ell \geq 0$ be such that $n = \ell k + r$. For every ordered $(k-1)$ -set of vertices $T = (v_1, \dots, v_{k-1})$ in H we form a k -graph H_T on $(\ell+1)k$ vertices by adding $k-r$ new vertices x_1, \dots, x_{k-r} and new edges $\{x_i, \dots, x_{k-r}, v_1, \dots, v_{r+i-1}\}$ for $i \in [k-r]$, so $(x_1, \dots, x_{k-r}, v_1, \dots, v_{k-1})$ is a path in H_T . The resulting graph H_T has order divisible by k and maximum codegree $\Delta(H_T) = \Delta(H)$, and H contains a Hamilton path if and only if one of the altered graphs H_T created in this way contains a Hamilton path. \square

Proposition 2.55. *For any $k \geq 2$ and any D there is a polynomial-time reduction from $\overline{\text{HP}}(k, D)$ to $\text{HC}(2k-1, \lfloor \frac{n}{2} \rfloor - (D+1)k)$.*

Proof. By Proposition 2.54 it suffices to give a polynomial-time reduction from $\overline{\text{HP}}(k, D)^\perp$ to $\text{HC}(2k-1, \lfloor \frac{n}{2} \rfloor - (D+1)k)$. So let H be a k -graph on n vertices, where k divides n , and set $\ell := n/k$. We define a $(2k-1)$ -graph H_{2k-1} as follows. Let $U := V(H)$ and let X be a set of size $|X| = \frac{(k-1)n}{k} = \ell(k-1)$ which is disjoint from U . Set $A := U \cup X$, and let B be a set of size $|B| = |A| + 1 = \ell(2k-1) + 1$ which is disjoint from A . We take the vertex set of H_{2k-1} to be $A \cup B$, and the edges of H_{2k-1} to be all sets $e \in \binom{A \cup B}{2k-1}$ with $|A \cap e| \notin \{k, k+1\}$, or with $|A \cap e| = k$ and $A \cap e \in E(H)$, or with $|A \cap e| = k+1$ and such that $e' \not\subseteq A \cap e$ for every $e' \in E(H)$. Since H_{2k-1} has $|A| + |B| = 2|A| + 1$ vertices, it then suffices to show that

- (a) H contains a Hamilton path if and only if H_{2k-1} contains a Hamilton cycle, and
- (b) if $\Delta(H) \leq D$, then $\delta(H_{2k-1}) \geq |A| - k(D+1)$.

For convenience we define the *type* of an edge $e \in E(H_{2k-1})$ to be the pair $(|e \cap A|, |e \cap B|)$.

For (a), first assume that there is a Hamilton path $P = (h_1, \dots, h_n)$ in H , and write

$X = \{x_1, \dots, x_{\ell(k-1)}\}$ and $B = \{b_1, \dots, b_{\ell k}, b'_1, \dots, b'_{\ell(k-1)+1}\}$. Then

$$C = (x_1, \dots, x_{k-1}, b_1, \dots, b_k, \dots, x_{(\ell-1)(k-1)+1}, \dots, x_{\ell(k-1)}, b_{(\ell-1)k+1}, \dots, b_{\ell k}, \\ h_1, \dots, h_k, b'_1, \dots, b'_{k-1}, \dots, h_{n-k+1}, \dots, h_n, b'_{(\ell-1)(k-1)+1}, \dots, b'_{\ell(k-1)}, b'_{\ell(k-1)+1}) .$$

is a Hamilton cycle C in H . That is, C begins with edges of type $(k-1, k)$ which include all the vertices of X , followed by edges of type $(k, k-1)$ which include all of the vertices of U . The final vertex from B then returns C to edges of type $(k-1, k)$ to close the cycle. Note that all vertices of H_{2k-1} are contained in C and that every consecutive subsequence of $2k-1$ vertices of C forms an edge of H_{2k-1} .

Now assume instead that H_{2k-1} contains a Hamilton cycle C . We introduce the following auxiliary bipartite graph G with vertex classes A and B . For every $a \in A$ let S_a be the consecutive subsequence of C of length $2k-1$ starting with a (so S_a is an edge of H_{2k-1}). We define the edge set of G to be $E(G) := \{\{a, b\} \mid a \in A, b \in S_a \cap B\}$, and also define $U' := \{u \in U \mid d_G(u) \geq k\}$ and $B' := \{b \in B \mid d_G(b) \leq k-1\}$.

Claim 2.56. *The following properties hold for G .*

- (i) $d_G(a) \geq k-1$ for all $a \in A$,
- (ii) $d_G(x) \geq k$ for all $x \in X$,
- (iii) $d_G(b) \leq k$ for all $b \in B$,
- (iv) $|B'| \geq n+1$.

Proof of Claim 2.56. The main observation is that if e is an edge of C such that $|e \cap A| \geq k$, then $e \cap A$ is an edge of H , so e is of type $(k, k-1)$ and contains no vertex of X . To see this, assume otherwise that there exists an edge e for which this observation does not hold. Then $|e \cap A| > k$, and by construction of H_{2k-1} the edge e' preceding e in C fulfils $|e' \cap A| > k$ as well. Consequently every edge of C has $|e \cap A| > |e \cap B|$, contradicting the fact that $|B| \geq |A|$. This observation immediately implies (i), (ii) and (iii). To show (iv),

observe that if every edge of C contains a vertex of X , then for the same reason it would follow that $d_G(a) \geq k$ for every $a \in A$ and $d_G(b) \leq k - 1$ for every $b \in B$, and so we would have $|A|k \leq |E(G)| \leq |B|(k - 1) = (|A| + 1)(k - 1)$, contradicting the fact that $|A| \geq k$. So some edge e of C does not contain a vertex of X , and it follows that we may choose a set \mathcal{S} of disjoint consecutive subsequences of C of length $2k - 1$, each beginning with a vertex of X , such that $X \subseteq \bigcup \mathcal{S}$. Then each $S \in \mathcal{S}$ is an edge of C containing a vertex of X , so has $|S \cap A| \leq k - 1$, and so we have $|\mathcal{S}| \geq |X|/(k - 1) = \ell$. Furthermore, we have $\bigcup \mathcal{S} \cap B \subseteq B'$, so if $|\bigcup \mathcal{S} \cap B| \geq n + 1$ then we are done. We may therefore assume that $|\bigcup \mathcal{S} \cap B| = n = k\ell$, so $|\mathcal{S}| = \ell$. Since C has more than $(4k - 3)\ell$ vertices, there must exist $S \in \mathcal{S}$ such that the $2k - 2$ vertices immediately prior to S in C are not in $\bigcup \mathcal{S}$. Let $x \in X$ be the first vertex of S , let S' be the consecutive subsequence of C of length $2k - 1$ ending with x , and let b be the last member of B in S' . Since S' is an edge of C containing a vertex of X , we have $|S' \cap B| \geq k$. It follows that $b \in B'$, and consequently $|B'| \geq n + 1$. \diamond

By making use of Claim 2.56 we now can double-count the edges of G to get

$$|E(G)| \geq k|X| + (k - 1)|U| + |U'| = (k - 1)n + (k - 1)n + |U'|, \text{ and}$$

$$|E(G)| \leq k|B| - |B'| \leq (2k - 1)n + k - (n + 1).$$

Hence we have $|U'| \leq k - 1$. For every $x \in X$ the previous $k - 1$ vertices of U in C are vertices of U' and therefore $|U'| \geq k - 1$. Hence $|U'| = k - 1$ and all inequalities used in the above calculation are in fact equalities. Therefore the vertices of U have to appear in a consecutive order in C only interrupted by vertices of B . Since there are exactly $k - 1$ vertices in U' , the minimal segment of C containing all the vertices of U is a path consisting only of edges of the type $(k, k - 1)$. So deleting all vertices of B from this segment yields a Hamilton path in H .

For (b) let S be a set of $2k-2$ vertices of H_{2k-1} . If $|A \cap S| < k-1$ or $|A \cap S| = k+1$, then we can add any vertex of $A \setminus S$ to S to form an edge of H_{2k-1} , so $d_{H_{2k-1}}(S) \geq |A| - k - 1$. If instead $|A \cap S| > k+1$ or $|A \cap S| = k-1$, then we can add any vertex of $B \setminus S$ to form an edge of H_{2k-1} , so $d_{H_{2k-1}}(S) \geq |B| - k = |A| - k + 1$. The same is true if $|A \cap S| = k$ and $A \cap S$ is an edge of H . Finally, if $|A \cap S| = k$ and $A \cap S$ is not an edge of H , then $S \cup \{a\}$ is an edge of H_{2k-1} for any $a \in A \setminus S$ such that $S' \cup \{a\}$ is not an edge of H for any $S' \subseteq S$ of size $k-1$. Since there are k such sets S' , each of which has at most $\Delta(H) \leq D$ neighbours in H , it follows that $d_{H_{2k-1}}(S) \geq |A| - (D+1)k$. So in all cases we have $d_{H_{2k-1}}(S) \geq |A| - (D+1)k$ as claimed. \square

Proposition 2.57. *For any $k \geq 2$ and any D there is a polynomial-time reduction from $\overline{\text{HC}}(k, D)$ to $\text{HC}(2k, \frac{n}{2} - (D+1)k)$.*

Proof. Let H be a k -graph on n vertices, let $A := V(H)$ and let B be a set of size n which is disjoint from A . We form a $2k$ -graph H_{2k} with vertex set $V := A \cup B$ whose edges are all sets $e \in \binom{V}{2k}$ such that $|A \cap e| \notin \{k, k+1\}$, or such that $|A \cap e| = k$ and $A \cap e \in E(H)$, or such that $|A \cap e| = k+1$ and $e' \not\subseteq A \cap e$ for every $e' \in E(H)$. Since H_{2k} has precisely $2n$ vertices, it then suffices to show that

- (i) H contains a Hamilton cycle if and only if H_{2k} contains a Hamilton cycle, and
- (ii) if $\Delta(H) \leq D$, then $\delta(H_{2k}) \geq n - (D+1)k$.

As in Proposition 2.55, we define the *type* of an edge $e \in E(H_{2k})$ to be the pair $(|e \cap A|, |e \cap B|)$.

For (i) assume first that H_{2k} has a Hamilton cycle C . Observe that our construction of H_{2k} ensures that if e is an edge of H_{2k} of type (k, k) , and e' is an edge of H_{2k} of type $(k+1, k-1)$, then $|e \cap e'| < 2k-1$. It follows that C either has no edges with $|e \cap A| > |e \cap B|$ or has no edges with $|e \cap A| \leq |e \cap B|$. Since $|A| = |B|$ we conclude that every edge in C must have type (k, k) . Let C' be the subsequence of C obtained by

deleting all vertices of B from C , and let S be a subsequence of k consecutive vertices of C' , so in particular $S \subseteq A$. Then S is included in an edge of C , which is an edge of type (k, k) in H_{2k} , and so S is an edge of H by construction of H_{2k} . We conclude that C' is a Hamilton cycle in H . For the other direction assume that H contains a Hamilton cycle (v_1, \dots, v_n) . Then for any enumeration of B as b_1, \dots, b_n , the sequence $(v_1, b_1, \dots, v_n, b_n)$ is a Hamilton cycle in H_{2k} , since every consecutive subsequence consisting of $2k$ vertices contains exactly k vertices from A , which form an edge of H .

For (ii), fix a set S of $2k - 1$ vertices of H_{2k} , and observe that if $|S \cap A| \neq k$, then S is included in at least $n - (k + 1)$ edges of H_{2k} . If instead $|S \cap A| = k$ and $S \cap A$ is an edge of H , then S is included in at least $n - (k - 1)$ edges of H_{2k} , since we may add any vertex of $B \setminus S$ to form an edge of H_{2k} . Finally, if $|S \cap A| = k$ and $S \cap A$ is not an edge of H , then $S \cup \{a\}$ is an edge of H_{2k} for every vertex $a \in A \setminus S$ except for those vertices a such that $S' \cup \{a\}$ is an edge of H for some $S' \subseteq S \cap A$ of size $k - 1$. Since there are k such subsets S' , each of which has $d_H(S') \leq \Delta(H) \leq D$ neighbours in H , it follows that S is included in at least $n - k - kD$ edges of H_{2k} . So in all cases we have $d_{H_{2k}}(S) \geq n - k - kD$ as claimed. \square

CHAPTER 3

CONTAGIOUS SETS IN BOOTSTRAP PERCOLATION

This chapter contains the proofs of Theorem 1.15 and Theorem 1.17.

3.1 Proof of Theorem 1.15

The underlying idea we use to prove Theorem 1.15 is the following. First, we suppose the existence of a minimal counterexample to our assertion and investigate its structure, categorising the vertices according to their degree in relation to ρ . In this way we can distinguish between vertices according to the number of infected neighbours necessary to cause their own infection. Next, using the vertex- and edge-minimality of this counterexample, we derive several degree conditions; in particular we bound the number of edges between low degree and high degree vertices. Finally, we show that we can infect the whole graph by a certain random selection of vertices just among the vertices of high degree. This will contradict the existence of a minimal counterexample. Our random selection follows the approach of Ackerman, Ben-Zwi and Wolfowitz [1] used to establish the bound (1.2.1), whilst the idea of distinguishing between low and high degree vertices was used by Chang [25] to prove the bound (1.2.3). The principal difference in our proof is

that our analysis of the degree conditions is more detailed, as we distinguish more classes of degree, and that we take advantage of the structure of a minimal counterexample. During the proof we will work with the following easy observation.

Proposition 3.1. *Suppose that $\rho \in [0, 1]$ and G is a graph. Let $S \subseteq V(G)$ and set $G' = G - S$. If $A \subseteq V(G')$ is a contagious set for G' with respect to ρ , then $A \cup S$ is a contagious set for G with respect to ρ .*

Proof. Let (X_i) be the infecting sequence in G induced by $A \cup S$ with respect to ρ and let (Y_i) be the infecting sequence in G' induced by A with respect to ρ . We show by induction on i that $Y_i \cup S \subseteq X_i$ and therefore $V(G) = H_{\rho, G'}(A) \cup S \subseteq H_{\rho, G}(A \cup S)$. For the base case we certainly have $Y_0 \cup S = A \cup S = X_0$. So assume the statement is true for $i \in \mathbb{N}$. Let $v \in Y_{i+1} \setminus Y_i$. We then have that $|N_{G'}(v) \cap Y_i| \geq \rho d_{G'}(v)$. We show that $|N_G(v) \cap X_i| \geq \rho d_G(v)$ and therefore $v \in X_{i+1}$. For this define $M(v) := N_G(v) \cap S$ and note that $N_G(v) = N_{G'}(v) \cup M(v)$ by the definition of G' . In particular we have that $d_G(v) = d_{G'}(v) + |M(v)|$. Since $S \cup (N_{G'}(v) \cap Y_i) \subseteq X_i$ by the induction hypothesis, we can deduce that

$$|N_G(v) \cap X_i| \geq |M(v)| + |N_{G'}(v) \cap Y_i| \geq |M(v)| + \rho d_{G'}(v) \geq \rho(|M(v)| + d_{G'}(v)) = \rho d_G(v) .$$

□

We can now turn to the proof of Theorem 1.15 which we first restate.

Theorem 1.15 ([39]). *For every $\rho \in (0, 1]$, every connected graph G of order n has $h_\rho(G) = 1$ or*

$$h_\rho(G) < 2\rho n .$$

Proof of Theorem 1.15. Observe first that if G is a connected graph of order $n \leq 1/\rho + 1$ then $h_\rho(G) = 1$. Indeed, every vertex v in such a graph has $d(v) \leq 1/\rho$ and so would

require only one infected neighbour for its infection; since G is connected it would then follow that any vertex of G forms a contagious set of size one, so $h_\rho(G) = 1$.

It therefore suffices to prove that every connected graph G of order $n > 1/(2\rho)$ fulfils $h_\rho(G) < 2\rho n$, and the rest of the proof is devoted to proving this statement by considering a minimal counterexample (the apparently weaker-than-necessary bound on n in this statement will prove convenient in appealing to minimality). We may assume for this that $\rho \in (0, 1/2)$, as otherwise the statement is trivial. So fix $\rho \in (0, 1/2)$, and let \mathcal{G} be the set of *counterexamples* for this value of ρ (in other words, \mathcal{G} is the set of all connected graphs G with $n(G) > 1/(2\rho)$ such that $h_\rho(G) \geq 2\rho n(G)$). Suppose for a contradiction that \mathcal{G} is non-empty (i.e. that a counterexample exists). Let $n = \min_{G \in \mathcal{G}} n(G)$, and choose $G \in \mathcal{G}$ with $n(G) = n$ such that $e(G) \leq e(H)$ for every $H \in \mathcal{G}$ with $n(H) = n$. Thus G is a minimal counterexample according to its number of vertices and (among those) to its number of edges. Moreover, we must have

$$n > 1/\rho + 1, \tag{3.1.1}$$

as otherwise our initial observation and the fact that $n > 1/(2\rho)$ together imply that $h_\rho(G) = 1 < 2\rho n$, so G would not be a counterexample. The vertex-minimality of G will be used to prove the following claim. For this recall that $H_{\rho,G}(v)$ is the maximal set of vertices of G which becomes eventually effected after the initial infection of v .

Claim 3.2. *For every $v \in V(G)$ we have that $|H_{\rho,G}(v)| < 1/(2\rho)$.*

Proof of Claim 3.2. For the sake of a contradiction assume that there exists $v \in V(G)$ such that $|H_{\rho,G}(v)| \geq 1/(2\rho)$. We set $G' := G - H_{\rho,G}(v)$ and denote by \mathcal{C} the set of connected components of G' . Note first that $\{v\}$ cannot infect all vertices of G , as otherwise we would have $h_\rho(G) = 1 = 2\rho \cdot 1/(2\rho) < 2\rho n$ by (3.1.1), contradicting our assumption that G is a counterexample. So \mathcal{C} is non-empty. Now assume that there is a

component $C \in \mathcal{C}$ which contains at most $1/(2\rho)$ vertices. As G is connected, there exists a vertex $u \in V(C)$ with at least one neighbour in $H_{\rho,G}(v)$. Define $m(u) := |N_G(u) \cap H_{\rho,G}(v)|$, so $m(u) \geq 1$. By the assumption on the size of C we derive the inequality

$$1/(2\rho) + m(u) \geq |N_G(u) \cap C| + m(u) = d_G(u) .$$

The fact that $u \notin H_{\rho,G}(v)$ together with $|N_G(u) \cap H_{\rho,G}(v)| \geq m(u)$ justifies that

$$d_G(u) > 1/\rho \cdot m(u) .$$

As $m(u) \geq 1$, both inequalities combined lead to $\rho > 1/2$, which is a contradiction. Therefore each connected component of G' contains more than $1/(2\rho)$ vertices. Since G was a minimal counterexample, it follows that for each $C \in \mathcal{C}$ there exists a subset $A_C \subseteq V(C)$ with $|A_C| < 2\rho|V(C)|$ which is a contagious set, that is, such that $H_{\rho,C}(A_C) = V(C)$. We then have $H_{\rho,G'}(\bigcup_{C \in \mathcal{C}} A_C) = V(G')$. Since v infects $H_{\rho,G}(v)$ and, by Proposition 3.1 we have that $\bigcup_{C \in \mathcal{C}} A_C \cup H_{\rho,G}(v)$ infects G , it follows that $H_{\rho,G}(\bigcup_{C \in \mathcal{C}} A_C \cup \{v\}) = V(G)$. Hence

$$h_\rho(G) \leq 1 + \sum_{C \in \mathcal{C}} |A_C| < 2\rho \left(|H_{\rho,G}(v)| + \sum_{C \in \mathcal{C}} |V(C)| \right) \leq 2\rho n ,$$

contradicting our assumption that G is a counterexample. \diamond

We now classify the vertices of G according to their degree by defining

$$V_1 := \{v \in V(G) \mid d_G(v) \leq 1/\rho\},$$

$$V_{\geq 2} := V(G) \setminus V_1 , \text{ and}$$

$$V_2 := \{v \in V_{\geq 2} \mid d_G(v) \leq 2/\rho\} .$$

Note that V_1 is the set of vertices which become infected if just one of their neighbours is

infected, $V_{\geq 2}$ is the set of vertices which cannot be infected by just one infected neighbour, and V_2 is the set of vertices which can be infected by two infected neighbours, but not by just one infected neighbour. Claim 3.2 together with the minimality of G will imply the following claim about the size of neighbourhoods of vertices in these sets.

Claim 3.3. *The following holds.*

- (i) $|N_G(v) \cap V_1| < 1/(2\rho)$ for every $v \in V$.
- (ii) The induced subgraph $G[N(v) \cap V_2]$ is a clique for every $v \in V_1$.
- (iii) $|N_G(v) \cap V_2| < 1/(2\rho)$ for every $v \in V_1$.

Proof of Claim 3.3. To prove (i) assume for a contradiction that there exists $v \in V$ such that $|N(v) \cap V_1| \geq 1/(2\rho)$. If v is initially infected, then in the first time step every vertex of $N(v) \cap V_1$ will become infected. So $|H_{\rho,G}(v)| \geq |\{v\} \cup (N(v) \cap V_1)| > 1/(2\rho)$ contradicting Claim 3.2.

For (ii), assume for a contradiction that there are distinct $x, y \in N(v) \cap V_2$ such that $\{x, y\} \notin E(G)$. Let H be the graph formed from G by removing the edges $\{v, x\}$ and $\{v, y\}$ and adding the edge $\{x, y\}$, and let C be the connected component of H containing x and y .

We claim that every contagious set $A \subseteq V(C)$ for C also infects $V(C)$ in G . To see this, let (X_i) be the infecting sequence in G induced by A with respect to ρ and let (Y_i) be the infecting sequence in C induced by A with respect to ρ . We now prove by induction on i that $Y_i \subseteq X_{2i}$ for each $i \in \mathbb{N}$. The base case is given by $Y_0 = A = X_0$, so assume that $Y_i \subseteq X_{2i}$ for some $i \in \mathbb{N}$ and consider the following cases for each $w \in Y_{i+1} \setminus Y_i$.

- (a) If $w \notin \{x, y, v\}$, then by construction of H we have $N_G(w) = N_C(w)$, and by our induction hypothesis we can conclude that at least $\rho d_G(w) = \rho d_C(w)$ neighbours of w are in $Y_i \subseteq X_{2i}$. Hence $w \in X_{2i+1} \subseteq X_{2(i+1)}$ as required.

- (b) If $w = v$, then $v \in V(C)$ and by construction of H we have that $N_C(v) \subseteq N_G(v)$. Again since $v \in Y_{i+1}$, we have that $\rho d_C(v) \leq |N_C(v) \cap Y_i|$ and therefore by also making use of the induction hypothesis there exists $u \in (N_C(v) \cap Y_i) \subseteq (N_G(v) \cap X_{2i})$. Since $d_G(v) \leq 1/\rho$, we have that $|N_G(v) \cap X_{2i}| \geq 1 \geq \rho d_G(v)$ and therefore $v \in X_{2i+1} \subseteq X_{2(i+1)}$.
- (c) If $w \in \{x, y\}$, then we may assume without loss of generality that $w = x$ (since the following arguments hold equally well with y in place of x). Note that by construction of H we have that $N_C(x) = (N_G(x) \setminus \{v\}) \cup \{y\}$ and in particular that $d_G(x) = d_C(x)$. Furthermore, since $x \in Y_{i+1} \setminus Y_i$, we have that $|N_C(x) \cap Y_i| \geq \rho d_C(x)$. Setting $U := N_C(x) \setminus \{y\}$, we first assume that $|U \cap Y_i| \geq \rho d_G(x)$. By the induction hypothesis $U \cap Y_i \subseteq X_{2i}$ and since $U \subseteq N_G(x)$ we have that $x \in X_{2i+1} \subseteq X_{2(i+1)}$. On the other hand, if $|U \cap Y_i| < \rho d_G(x)$, then $y \in Y_i \subseteq X_{2i}$. Since $d_G(v) \leq 1/\rho$ and $y \in N_G(v) \cap X_{2i}$, we have that $v \in X_{2i+1}$. Furthermore $v \notin U$ and $U \subseteq X_{2i} \subseteq X_{2i+1}$. Hence $|N_G(x) \cap X_{2i+1}| \geq |(U \cap Y_i) \cup \{v\}| \geq \rho d_C(x) = \rho d_G(x)$ and therefore $x \in X_{2(i+1)}$.

This completes the induction argument, so we have $Y_i \subseteq X_{2i}$ for each $i \in \mathbb{N}$, and therefore $V(C) = H_{\rho,C}(A) \subseteq H_{\rho,G}(A)$. This proves our claim that every contagious set $A \subseteq V(C)$ for C also infects $V(C)$ in G .

Since G is connected, the construction of H ensures that H has at most two connected components. Suppose first that H is connected and therefore that $H = C$. Since we assumed that G has no contagious set of size less than $2\rho n$, it follows from our above claim that C has no contagious set of size less than $2\rho n$. Thus $C = H$ is also a counterexample, contradicting the edge-minimality of G . We conclude that H must have exactly two connected components C and C' . Observe that since $d_H(x) = d_G(x) > 1/\rho$ we have that $|V(C)| > 1/(2\rho)$. Suppose now that $|V(C')| \leq 1/(2\rho)$. Note that then every vertex $w \in V(C')$ has $d_G(w) \leq 1/\rho$. Indeed, if $w = v$ then this is implied by $v \in V_1$ and if

$w \neq v$ then w cannot have any neighbour in C , as otherwise H would be connected, and therefore $|N_G(w)| \leq |C'| \leq 1/\rho$. Now let $A \subseteq V(C)$ be a contagious set for C . Then by our above claim A infects $V(C)$ in G , so $V(G) \setminus H_{\rho,G}(A) \subseteq V(C')$. It follows that every vertex $w \in V(G)$ not infected by A has $d_G(w) \leq 1/\rho$, and since G is connected it follows that $H_{\rho,G}(A) = V(G)$. In other words any contagious set for C is also a contagious set for G , and therefore $h_\rho(C) \geq 2\rho|V(C)|$. Thus C is also a counterexample, contradicting the vertex-minimality of G .

We therefore conclude that C and C' each have more than $1/(2\rho)$ vertices. By minimality of G it follows there are contagious sets $A \subseteq V(C)$ and $A' \subseteq V(C')$ for C and C' respectively with $|A| < 2\rho|V(C)|$ and $|A'| < 2\rho|V(C')|$. Then A infects $V(C)$ in G by our above claim. Moreover, A' infects $V(C')$ in G since every vertex $w \in V(C')$ with $w \neq v$ has $N_G(w) = N_{C'}(w)$, and $d_G(v) < 1/\rho$. We conclude that $A \cup A'$ is a contagious set in G of size $|A \cup A'| \leq |A| + |A'| < 2\rho|V(C)| + 2\rho|V(C')| = 2\rho n$, contradicting our assumption that G was a counterexample and so completing the proof of (ii).

Finally, to prove (iii), assume for a contradiction that there exists $v \in V_1$ such that $|N(v) \cap V_2| \geq 1/(2\rho)$, and choose any $w \in N(v) \cap V_2$. Since $G[N(v) \cap V_2]$ is a clique by (ii), every other vertex of $N(v) \cap V_2$ is a neighbour of both v and w . Since each vertex of V_2 will become infected once it has two infected neighbours, it follows that if w is initially infected, then v will become infected at the first time step, and then all remaining vertices of $N(v) \cap V_2$ will become infected at the second time step. So $|H_{\rho,G}(w)| \geq |\{v\} \cup (N(v) \cap V_2)| \geq 1/(2\rho)$, contradicting Claim 3.2. \diamond

We now introduce some further notation. Let M be the number of edges with one endvertex in V_1 and one endvertex in V_2 . Also, for each $v \in V(G)$, write $d'(v) := |N(v) \cap V_1|$ and define $x_v := \rho d(v)$ and $y_v := \rho d'(v)$. Note that, by Claim 3.3(i), for every $v \in V_{\geq 2}$ we have

$$0 \leq y_v < 1/2. \tag{3.1.2}$$

Furthermore, double-counting the edges between V_1 and V_2 together with Claim 3.3(iii) gives us

$$\sum_{v \in V_2} y_v = \rho \sum_{v \in V_2} d'(v) = \rho M = \rho \sum_{v \in V_1} |N(v) \cap V_2| < \frac{1}{2} |V_1|. \quad (3.1.3)$$

We now choose an order σ of the vertices of $V_{\geq 2}$ uniformly at random. Let A be the set of vertices $v \in V_{\geq 2}$ for which fewer than $\lceil \rho d(v) \rceil$ neighbours of v precede v in the order σ . We claim that A is a contagious set for G with respect to ρ . Indeed, since G is connected and each vertex of V_1 becomes infected once it has one infected neighbour, it suffices for this to show that A infects $V_{\geq 2}$. Suppose for a contradiction that this is not the case, and let $v \in V_{\geq 2}$ be the first vertex in the order σ which is not infected by A . Then $v \notin A$, so v has at least $\lceil \rho d(v) \rceil$ neighbours which precede v . Our choice of v implies that all of these neighbours are infected by A , but this contradicts the fact that v is not infected by A . So A is indeed a contagious set for G with respect to ρ . Moreover, for any $v \in V_{\geq 2}$ we have $v \in A$ if and only if v is one of the first $\lceil \rho d(v) \rceil$ members of $\{v\} \cup (N(v) \cap V_{\geq 2})$ in the order σ . It follows that $\mathbb{P}[v \in A] = \frac{\lceil \rho d(v) \rceil}{1 + |N(v) \cap V_{\geq 2}|}$, so the expected size of A is

$$\begin{aligned} \mathbb{E}[|A|] &= \sum_{v \in V_{\geq 2}} \frac{\lceil \rho d(v) \rceil}{1 + d(v) - d'(v)} < \sum_{v \in V_{\geq 2}} \frac{1 + \rho d(v)}{d(v) - d'(v)} = \rho \sum_{v \in V_{\geq 2}} \frac{1 + x_v}{x_v - y_v} \\ &= \rho \left(\sum_{v \in V_2} \frac{1 + x_v}{x_v - y_v} + \sum_{v \in V_{\geq 2} \setminus V_2} \frac{1 + x_v}{x_v - y_v} \right) \\ &\stackrel{(*)}{\leq} \rho \left(\sum_{v \in V_2} \frac{1 + 1}{1 - y_v} + \sum_{v \in V_{\geq 2} \setminus V_2} \frac{1 + 2}{2 - 1/2} \right) \\ &= 2\rho \left(\sum_{v \in V_2} \frac{1}{1 - y_v} + |V_{\geq 2} \setminus V_2| \right) =: T, \end{aligned}$$

where the inequality $(*)$ follows from (3.1.2), the fact that $x_v \geq 1$ for $v \in V_2$ and $x_v \geq 2$ for $x \in V_{\geq 2} \setminus V_2$, and the fact that if $0 \leq y < a \leq x$ are real numbers, then $\frac{1+x}{x-y} \leq \frac{1+a}{a-y}$. For the final argument we will use the following claim.

Claim 3.4. Let $n_1, n_2 \in \mathbb{N}$, $K = \{y \in [0, \frac{1}{2}]^{n_2} \mid \sum_{i=1}^{n_2} y_i \leq \frac{n_1}{2}\}$ and $f(y) = \sum_{i=1}^{n_2} \frac{1}{1-y_i}$. We have that

$$\max_{y \in K} f(y) \leq n_1 + n_2 .$$

Proof of Claim 3.4. Suppose first that $n_1 \geq n_2$, and let $u := (\frac{1}{2}, \dots, \frac{1}{2})$. Then $u \in K$, and it is easy to see that $f(u) \geq f(y)$ for every $y \in K$. Furthermore $f(u) = n_2 \cdot \frac{1}{1-1/2} = 2n_2 \leq n_1 + n_2$.

Now suppose instead that $n_1 < n_2$. Since f is a continuous function it attains its maximum on the compact space K . Fix $u \in K$ such that this maximum is attained, and (subject to this) so that as many as possible of the u_i are equal to $1/2$. Clearly we then have $\sum_{i=1}^{n_2} u_i = \frac{n_1}{2}$. Suppose first that there exist distinct $i, j \in [n_2]$ such that $0 < u_i \leq u_j < 1/2$. Then set $\varepsilon := \min\{u_i, 1/2 - u_j\}$ and define $z \in \mathbb{R}^{n_2}$ by

$$z_k := \begin{cases} u_i - \varepsilon, & \text{if } k = i \\ u_j + \varepsilon, & \text{if } k = j \\ u_k, & \text{otherwise} \end{cases} .$$

Note that $z \in K$. We then get by the convexity of the function $g(x) = \frac{1}{1-x}$ on $[0, \frac{1}{2}]$ that

$$f(u) = g(u_i) + g(u_j) + \sum_{k \neq i, j} g(u_k) < g(u_i - \varepsilon) + g(u_j + \varepsilon) + \sum_{k \neq i, j} g(u_k) = f(z) .$$

This contradicts the assumption that $f(u)$ was the maximal value. We therefore conclude that all but at most one of the u_i are equal to 0 or $1/2$; since n_1 is an integer and $\sum_{i=1}^{n_2} u_i = \frac{n_1}{2}$, this implies that in fact every u_i is equal to 0 or $1/2$ (in other words u has n_1 coordinates equal to $1/2$ and the rest equal to 0). Hence $\max_{y \in K} f(y) = f(u) = n_1 \cdot \frac{1}{1-1/2} + (n_2 - n_1) \cdot \frac{1}{1-0} = n_1 + n_2$. \diamond

By setting $n_1 = |V_1|$ and $n_2 = |V_2|$ and by using (3.1.2) and (3.1.3) the assumptions

of Claim 3.4 are fulfilled and we can conclude our calculation with

$$\begin{aligned}\mathbb{E}[|A|] &< T \leq 2\rho(|V_1| + |V_2| + |V_{\geq 2} \setminus V_2|) \\ &= 2\rho n .\end{aligned}$$

We may therefore fix an order σ of $V_{\geq 2}$ such that the contagious set A given by this order has size less than $2\rho n$. This contradicts our assumption that G was a counterexample and so proves the theorem. \square

3.2 Proof of Theorem 1.17.

To prove our second main theorem we use the following modified version of a theorem of Gentner and Rautenbach [45]. For a constant $\rho > 0$ and a graph G we again use the notation $V_1 = \{v \in V(G) \mid d(v) \leq 1/\rho\}$ and $V_{\geq 2} = V \setminus V_1$.

Lemma 3.5 ([45]). *For every $\varepsilon > 0$ there exists $\rho_0 > 0$ such that for any $\rho \in (0, \rho_0)$ the following holds. Let G be a connected graph of order n with girth at least 5 and $\Delta(G) \geq 1/\rho$. If $|N(u) \cap V_1| < \frac{1}{1+\varepsilon}d(u)$ for every $u \in V_{\geq 2}$, then*

$$h_\rho(G) < (1 + \varepsilon)\rho n .$$

The lemma given above follows immediately from the proof of the original version [45, Theorem 1], which omitted the condition on the neighbourhood in V_1 of vertices in $V_{\geq 2}$, and had the weaker conclusion that $h_\rho(G) \leq (2 + \varepsilon)\rho n$. Loosely speaking, in this proof Gentner and Rautenbach used a random selection process to select an initial set X and showed that vertices satisfying certain properties are infected with high probability by X . They then bounded the number of vertices which do not have these properties, and simply added all such vertices to X to obtain a contagious set of the claimed size.

More precisely, they first fixed a sufficiently small constant δ , in particular with $\delta < \varepsilon$, and found a set $X_0 \subseteq V_{\geq 2}$ with two properties:

(a) $|X_0| \leq (1 + \delta)\rho n$, and

(b) $|N(u) \cap (V_1 \setminus H_{\rho,G}(X_0))| \leq \frac{1}{1+\delta}d(u)$ for every $u \in V_{\geq 2} \setminus X_0$.

Then, by a series of random selections which exploit the girth property to ensure independence of the random variables considered, they added at most $(1 + \varepsilon - \delta)\rho n$ vertices to X_0 to gain a set $X_0 \subseteq Y \subseteq V(G)$ such that $|Y| \leq (2 + \varepsilon)n$ and $H_{\rho,G}(Y) = V(G)$.

To prove Lemma 3.5 we follow exactly the same argument with $X_0 := \emptyset$; the additional condition of Lemma 3.5 ensures that X_0 satisfies (a) and (b). We then obtain a set $Y \subseteq V(G)$ such that $|Y| \leq (1 + \varepsilon - \delta)\rho n < (1 + \varepsilon)\rho n$ and $H_{\rho,G}(Y) = V(G)$.

We can now prove Theorem 1.17 which we first restate. The main idea is to show that a minimal counterexample must satisfy the conditions of Lemma 3.5, and therefore is not in fact a counterexample.

Theorem 1.17 ([39]). *For every $\varepsilon > 0$ there exists $\rho_0 > 0$ such that for every $\rho \in (0, \rho_0)$ and every connected graph G of order n and girth at least 5 we have $h_\rho(G) = 1$ or*

$$h_\rho(G) < (1 + \varepsilon)\rho n .$$

Proof of Theorem 1.17. Given $\varepsilon > 0$, let ρ_0 be small enough for Lemma 3.5 and also such that $\rho_0 < \frac{\varepsilon}{1+\varepsilon}$, and fix $\rho \in (0, \rho_0)$. We will show that if G is a connected graph of order $n > \frac{1}{(1+\varepsilon)\rho}$ with girth at least 5, then $h_\rho(G) < (1 + \varepsilon)\rho n$. This suffices to prove Theorem 1.17 as, by a similar argument as in the proof of Theorem 1.15, every connected graph G of order at most $\frac{1}{(1+\varepsilon)\rho}$ has $h_\rho(G) = 1$. So suppose that this assertion is false, let G be a minimal counterexample according to the number of vertices (for these values of ρ and ε), and let n be the order of G . Then G is a connected graph on $n > \frac{1}{(1+\varepsilon)\rho}$ vertices

with girth at least five and $h_\rho(G) \geq (1 + \varepsilon)\rho n$. We may assume that $\Delta(G) > 1/\rho$, as otherwise (since G is connected) we have $h_\rho(G) = 1 < (1 + \varepsilon)\rho n$, a contradiction. We show the following claim.

Claim 3.6. $|N(v) \cap V_1| < \frac{1}{(1+\varepsilon)\rho}$ for every $v \in V_{\geq 2}$.

Proof of Claim 3.6. Suppose for a contradiction that there exists $v \in V_{\geq 2}$ such that $|N(v) \cap V_1| \geq \frac{1}{(1+\varepsilon)\rho}$. Then $|H_{\rho,G}(v)| \geq |\{v\} \cup (N(v) \cap V_1)| > \frac{1}{(1+\varepsilon)\rho}$. We set $G' := G - H_{\rho,G}(v)$ and denote by \mathcal{C} the set of connected components of G' . Note first that $\{v\}$ cannot infect $V(G)$, as otherwise $h_\rho(G) = 1 < (1+\varepsilon)\rho n$, a contradiction. Now assume that there is a component $C \in \mathcal{C}$ which contains at most $\frac{1}{(1+\varepsilon)\rho}$ vertices. As G is connected there exists a vertex $u \in V(C)$ with a neighbour in $H_{\rho,G}(v)$. Let $m(u) := |N(u) \cap H_{\rho,G}(v)| \geq 1$; since $u \notin H_{\rho,G}(v)$, we derive the inequality

$$\frac{1}{(1+\varepsilon)\rho} + m(u) > d(u) > m(u) \cdot \frac{1}{\rho}.$$

This leads to

$$1 - \frac{\varepsilon}{1+\varepsilon} = \frac{1}{1+\varepsilon} > (1-\rho)m(u) \geq 1-\rho \geq 1-\rho_0,$$

which contradicts our choice of ρ_0 . Therefore each connected component of G' contains more than $\frac{1}{(1+\varepsilon)\rho}$ vertices. Since G was a minimal counterexample, it follows that each component $C \in \mathcal{C}$ has $h_\rho(C) < (1+\varepsilon)\rho|V(C)|$. By Proposition 3.1 we then have

$$h_\rho(G) \leq 1 + \sum_{C \in \mathcal{C}} h_\rho(C) < (1+\varepsilon)\rho \left(|H_{\rho,G}(v)| + \sum_{C \in \mathcal{C}} |V(C)| \right) = (1+\varepsilon)\rho n,$$

a contradiction. ◇

So for every vertex $u \in V_{\geq 2}$ we have $|N(u) \cap V_1| < \frac{1}{(1+\varepsilon)\rho} < \frac{1}{1+\varepsilon}d(u)$. By Lemma 3.5, it follows that $h_\rho(G) < (1+\varepsilon)\rho n$. This contradicts our assumption that G was a coun-

terexample, and so completes the proof.

□

LIST OF REFERENCES

- [1] Eyal Ackerman, Oren Ben-Zwi, and Guy Wolfowitz. Combinatorial model and bounds for target set selection. *Theoretical Computer Science*, 411(44):4017–4022, 2010.
- [2] Michael Aizenman and Joel L. Lebowitz. Metastability effects in bootstrap percolation. *Journal of Physics A: Mathematical and General*, 21(19):3801, 1988.
- [3] Hamed Amini and Nikolaos Fountoulakis. What I tell you three times is true: Bootstrap percolation in small worlds. In Paul W. Goldberg, editor, *Internet and Network Economics: 8th International Workshop, WINE 2012, Liverpool, UK, December 10-12, 2012. Proceedings*, pages 462–474. Springer, Berlin, Heidelberg, 2012.
- [4] Hamed Amini and Nikolaos Fountoulakis. Bootstrap percolation in power-law random graphs. *Journal of Statistical Physics*, 155(1):77–92, 2014.
- [5] Omer Angel and Brett Kolesnik. Minimal contagious sets in random graphs. arXiv:1705.06815.
- [6] József Balogh and Béla Bollobás. Bootstrap percolation on the hypercube. *Probability Theory and Related Fields*, 134(4):624–648, 2006.
- [7] József Balogh, Béla Bollobás, Hugo Duminil-Copin, and Robert Morris. The sharp threshold for bootstrap percolation in all dimensions. *Transactions of the American Mathematical Society*, 364:2667–2701, 2012.
- [8] József Balogh, Béla Bollobás, and Robert Morris. Bootstrap percolation in three dimensions. *The Annals of Probability*, 37(4):1329–1380, 2009.
- [9] József Balogh, Béla Bollobás, and Robert Morris. Majority bootstrap percolation on the hypercube. *Combinatorics, Probability and Computing*, 18:17–51, 2009.
- [10] József Balogh, Béla Bollobás, and Robert Morris. Bootstrap percolation in high dimensions. *Combinatorics, Probability and Computing*, 19(5-6):643–692, 2010.
- [11] József Balogh, Béla Bollobás, Robert Morris, and Oliver Riordan. Linear algebra and bootstrap percolation. *Journal of Combinatorial Theory, Series A*, 119(6):1328 – 1335, 2012.

- [12] József Balogh and Gábor Pete. Random disease on the square grid. *Random Structures & Algorithms*, 13(3-4):409–422, 1998.
- [13] József Balogh and Boris G. Pittel. Bootstrap percolation on the random regular graph. *Random Structures & Algorithms*, 30(1-2):257–286, 2007.
- [14] Josefran de Oliveira Bastos, Guilherme Oliveira Mota, Mathias Schacht, Jakob Schnitzer, and Fabian Schulenburg. Loose Hamiltonian cycles forced by large $(k-2)$ -degree—approximate version. *SIAM Journal on Discrete Mathematics*, 31(4):2328–2347, 2017.
- [15] Josefran de Oliveira Bastos, Guilherme Oliveira Mota, Mathias Schacht, Jakob Schnitzer, and Fabian Schulenburg. Loose Hamiltonian cycles forced by large $(k-2)$ -degree—sharp version. *Electronic Notes in Discrete Mathematics*, 61:101–106, 2017.
- [16] Eli Berger. Dynamic monopolies of constant size. *Journal of Combinatorial Theory, Series B*, 83:191–200, 2001.
- [17] Jean-Claude Bermond, Adrien Germa, Marie-Claude Heydemann, and Dominique Sotteau. Hypergraphes hamiltoniens. In *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*, volume 260 of *Colloq. Internat. CNRS*, pages 39–43. CNRS, Paris, 1978.
- [18] Larry Blume, David Easley, Jon Kleinberg, Robert Kleinberg, and Éva Tardos. Which networks are least susceptible to cascading failures? In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pages 393–402, 2011.
- [19] Enno Buß, Hiệp Hàn, and Mathias Schacht. Minimum vertex degree conditions for loose Hamilton cycles in 3-uniform hypergraphs. *Journal of Combinatorial Theory, Series B*, 103(6):658–678, 2013.
- [20] Damon Centola, Víctor M. Egeluz, and Michael W. Macy. Cascade dynamics of complex propagation. *Physica A: Statistical Mechanics and its Applications*, 374(1):449–456, 2007.
- [21] Damon Centola and Michael Macy. Complex contagions and the weakness of long ties. *American Journal of Sociology*, 113(3):702–734, 2007.
- [22] Raphaël Cerf and Emilio N. M. Cirillo. Finite size scaling in three-dimensional bootstrap percolation. *The Annals of Probability*, 27(4):1837–1850, 1999.
- [23] Raphaël Cerf and Francesco Manzo. The threshold regime of finite volume bootstrap percolation. *Stochastic Processes and their Applications*, 101(1):69–82, 2002.
- [24] John Chalupa, Paul L. Leath, and Gary R. Reich. Bootstrap percolation on a Bethe lattice. *Journal of Physics C: Solid State Physics*, 12(1):L31–L35, 1979.

- [25] Ching-Lueh Chang. Triggering cascades on undirected connected graphs. *Information Processing Letters*, 111(19):973–978, 2011.
- [26] Ching-Lueh Chang and Yuh-Dauh Lyuu. Triggering cascades on strongly connected directed graphs. *Theoretical Computer Science*, 593:62–69, 2015.
- [27] Amin Coja-Oghlan, Uriel Feige, Michael Krivelevich, and Daniel Reichman. Contagious sets in expanders. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '15, pages 1953–1987, 2015.
- [28] Gennaro Cordasco, Luisa Gargano, Marco Mecchia, Adele A. Rescigno, and Ugo Vaccaro. Discovering small target sets in social networks: A fast and effective algorithm. arXiv:1610.03721.
- [29] Andrzej Czygrinow and Theodore Molla. Tight codegree condition for the existence of loose Hamilton cycles in 3-graphs. *SIAM Journal on Discrete Mathematics*, 28(1):67–76, 2014.
- [30] Elias Dahlhaus, Peter Hajnal, and Marek Karpiński. On the parallel complexity of Hamiltonian cycle and matching problem on dense graphs. *Journal of Algorithms*, 15:367–384, 1993.
- [31] Michael Dairyko, Michael Ferrara, Bernard Lidický, Ryan R. Martin, Florian Pfender, and Andrew J. Uzzell. Ore and Chvátal-type degree conditions for bootstrap percolation from small sets. arXiv:1610.04499.
- [32] Gabriel A. Dirac. Some theorems on abstract graphs. *Proceedings of the London Mathematical Society*, 2:69–81, 1952.
- [33] Paul A. Dreyer, Jr. and Fred S. Roberts. Irreversible k -threshold processes: Graph-theoretical threshold models of the spread of disease and of opinion. *Discrete Applied Mathematics*, 157(7):1615–1627, 2009.
- [34] Jack Edmonds. Paths, trees, and flowers. *Canadian Journal of Mathematics*, 17:449–467, 1965.
- [35] Pál Erdős. On extremal problems of graphs and generalized graphs. *Israel Journal of Mathematics*, 2:183–190, 1964.
- [36] Pál Erdős, András Gyárfás, and László Pyber. Vertex coverings by monochromatic cycles and trees. *Journal of Combinatorial Theory, Series B*, 51(1):90–95, 1991.
- [37] Uriel Feige, Michael Krivelevich, and Daniel Reichman. Contagious sets in random graphs. *The Annals of Applied Probability*, 27(5):2675–2697, 2017.
- [38] Daniel Freund, Matthias Poloczek, and Daniel Reichman. Contagious sets in dense graphs. *European Journal of Combinatorics*, 68:66–78, 2018.

- [39] Frederik Garbe, Andrew McDowell, and Richard Mycroft. Contagious sets in a degree-proportional bootstrap percolation process. *Random Structures & Algorithms*, to appear. arXiv:1610.06144.
- [40] Frederik Garbe and Richard Mycroft. The complexity of the Hamilton cycle problem in hypergraphs of high minimum codegree. In Nicolas Ollinger and Heribert Vollmer, editors, *33rd Symposium on Theoretical Aspects of Computer Science (STACS 2016)*, volume 47 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 38:1–38:13, Dagstuhl, Germany, 2016. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.
- [41] Frederik Garbe and Richard Mycroft. Hamilton cycles in hypergraphs below the Dirac threshold. *Journal of Combinatorial Theory, Series B*, to appear. arXiv:1609.03101.
- [42] Martin Gardner. Mathematical games – the fantastic combinations of John Conway’s new solitaire game ‘life’. *Scientific American*, 223:120–123.
- [43] Michael R. Garey and David S. Johnson. *Computers and intractability: a guide to the theory of NP-completeness*. W.H. Freeman, 1979.
- [44] Michael R. Garey, David S. Johnson, and Larry Stockmeyer. Some simplified NP-complete graph problems. *Theoretical Computer Science*, 1:237–267, 1976.
- [45] Michael Gentner and Dieter Rautenbach. Dynamic monopolies for degree proportional thresholds in connected graphs of girth at least five and trees. *Theoretical Computer Science*, 667:93–100, 2017.
- [46] Roman Glebov, Yury Person, and Wilma Weps. On extremal hypergraphs for Hamiltonian cycles. *European Journal of Combinatorics*, 33(4):544–555, 2012.
- [47] Adolph W. Goodman. On sets of acquaintances and strangers at any party. *American Mathematical Monthly*, 66(66):778–783, 1959.
- [48] Mark Granovetter. Threshold models of collective behavior. *American Journal of Sociology*, 83(6):1420–1443, 1978.
- [49] Janko Gravner, Christopher Hoffman, James Pfeiffer, and David Sivakoff. Bootstrap percolation on the Hamming torus. *The Annals of Applied Probability*, 25(1):287–323, 2015.
- [50] Janko Gravner, Alexander E. Holroyd, and Robert Morris. A sharper threshold for bootstrap percolation in two dimensions. *Probability Theory and Related Fields*, 153(1):1–23, 2012.
- [51] Hiep Hàn and Mathias Schacht. Dirac-type results for loose Hamilton cycles in uniform hypergraphs. *Journal of Combinatorial Theory, Series B*, 100(3):332–346, 2010.

- [52] Jie Han. Decision problem for perfect matchings in dense k -uniform hypergraphs. *Transactions of the American Mathematical Society*, 369(7):5197–5218, 2017.
- [53] Jie Han and Andrew Treglown. The complexity of perfect matchings and packings in dense hypergraphs. arXiv:1609.06147.
- [54] Jie Han and Yi Zhao. Minimum codegree threshold for Hamilton ℓ -cycles in k -uniform hypergraphs. *Journal of Combinatorial Theory, Series A*, 132(0):194–223, 2015.
- [55] Jie Han and Yi Zhao. Minimum vertex degree threshold for loose Hamilton cycles in 3-uniform hypergraphs. *Journal of Combinatorial Theory, Series B*, 114:70–96, 2015.
- [56] Jie Han and Yi Zhao. Forbidding Hamilton cycles in uniform hypergraphs. *Journal of Combinatorial Theory, Series A*, 143:107–115, 2016.
- [57] Alexander E. Holroyd. Sharp metastability threshold for two-dimensional bootstrap percolation. *Probability Theory and Related Fields*, 125(2):195–224, 2003.
- [58] Svante Janson, Tomasz Łuczak, and Andrzej Ruciński. *Random Graphs*. Wiley Series in Discrete Mathematics and Optimization. Wiley, 2011.
- [59] Svante Janson, Tomasz Łuczak, Tatyana Turova, and Thomas Vallier. Bootstrap percolation on the random graph $G_{n,p}$. *The Annals of Applied Probability*, 22(5):1989–2047, 2012.
- [60] Mihyun Kang and Tamas Makai. A simple proof of almost percolation on $G(n, p)$. arXiv:1608.00800v1.
- [61] Richard M. Karp. Reducibility among combinatorial problems. In Raymond E. Miller, James W. Thatcher, and Jean D. Bohlinger, editors, *Complexity of Computer Computations*, pages 85–103. Springer US, 1972.
- [62] Marek Karpiński, Andrzej Ruciński, and Edyta Szymańska. Computational complexity of the Hamiltonian cycle problem in dense hypergraphs. In *LATIN 2010: Theoretical Informatics, 9th Latin American Symposium, Oaxaca, Mexico, April 19–23, 2010. Proceedings*, pages 662–673, 2010.
- [63] Gyula Y. Katona and Hal A. Kierstead. Hamiltonian chains in hypergraphs. *Journal of Graph Theory*, 30(3):205–212, 1999.
- [64] Peter Keevash, Fiachra Knox, and Richard Mycroft. Polynomial-time perfect matchings in dense hypergraphs. *Advances in Mathematics*, 269:265–334, 2014.
- [65] Peter Keevash, Daniela Kühn, Richard Mycroft, and Deryk Osthus. Loose Hamilton cycles in hypergraphs. *Discrete Mathematics*, 311(7):544–559, 2011.

- [66] Peter Keevash and Richard Mycroft. A geometric theory for hypergraph matching. *Memoirs of the American Mathematical Society*, 233(1098):1–101, 2015.
- [67] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, pages 137–146, 2003.
- [68] Daniela Kühn, Richard Mycroft, and Deryk Osthus. Hamilton ℓ -cycles in uniform hypergraphs. *Journal of Combinatorial Theory, Series A*, 117:910–927, 2010.
- [69] Daniela Kühn and Deryk Osthus. Loose Hamilton cycles in 3-uniform hypergraphs of high minimum degree. *Journal of Combinatorial Theory, Series B*, 96(6):767–821, 2006.
- [70] Daniela Kühn and Deryk Osthus. Matchings in hypergraphs of large minimum degree. *Journal of Graph Theory*, 51(4):269–280, 2006.
- [71] Willem Mantel. Problem 28 (solution by H. Gouwentak, W. Mantel, J. Teixeira de Mattes, F. Schuh and W. A. Wythoff). *Wiskundige Opgaven*, 10:60–61, 1907.
- [72] Natasha Morrison and Jonathan A Noel. Extremal bounds for bootstrap percolation in the hypercube. *Journal of Combinatorial Theory, Series A*, 156:61–84, 2018.
- [73] David Peleg. Size bounds for dynamic monopolies. *Discrete Applied Mathematics*, 86(2):263–273, 1998.
- [74] Gábor Pete. How to make the cube weedy? (in Hungarian). *Polygon*, V11:1:69–80.
- [75] Daniel Reichman. New bounds for contagious sets. *Discrete Mathematics*, 312(10):1812–1814, 2012.
- [76] Christian Reiher, Vojtěch Rödl, Andrzej Ruciński, Mathias Schacht, and Endre Szemerédi. Minimum vertex degree condition for tight Hamiltonian cycles in 3-uniform hypergraphs. arXiv:1611.03118.
- [77] Vojtěch Rödl and Andrzej Ruciński. Dirac-type questions for hypergraphs - a survey (or more problems for Endre to solve). In *An irregular mind*, pages 561–590. Springer, 2010.
- [78] Vojtěch Rödl and Andrzej Ruciński. Families of triples with high minimum degree are Hamiltonian. *Discussiones Mathematicae Graph Theory*, 34:361–381, 2014.
- [79] Vojtěch Rödl, Andrzej Ruciński, Mathias Schacht, and Endre Szemerédi. On the Hamiltonicity of triple systems with high minimum degree. arXiv:1605.00773.

- [80] Vojtěch Rödl, Andrzej Ruciński, and Endre Szemerédi. A Dirac-type theorem for 3-uniform hypergraphs. *Combinatorics, Probability and Computing*, 15(1-2):229–251, 2006.
- [81] Vojtěch Rödl, Andrzej Ruciński, and Endre Szemerédi. An approximate Dirac-type theorem for k -uniform hypergraphs. *Combinatorica*, 28(2):229–260, 2008.
- [82] Vojtěch Rödl, Andrzej Ruciński, and Endre Szemerédi. Perfect matchings in large uniform hypergraphs with large minimum collective degree. *Journal of Combinatorial Theory, Series A*, 116(3):613–636, 2009.
- [83] Vojtěch Rödl, Andrzej Ruciński, and Endre Szemerédi. Dirac-type conditions for Hamiltonian paths and cycles in 3-uniform hypergraphs. *Advances in Mathematics*, 227:1225–1299, 2011.
- [84] Edyta Szymańska. The complexity of almost perfect matchings and other packing problems in uniform hypergraphs with high codegree. *European Journal of Combinatorics*, 34(3):632–646, 2013.
- [85] Pál Turán. On an extremal problem in graph theory. *Matematikai s Fizikai Lapok*, 48:436–452, 1941.
- [86] Stanisław Ulam. Random processes and transformations. *Proceedings of International Congress of Mathematicians*, pages 264–275, 1950.
- [87] John von Neumann. *Theory of self-reproducing automata*. University of Illinois Press, Champaign, 1966.
- [88] Duncan J. Watts. A simple model of global cascades on random networks. *Physical Sciences - Applied Mathematics*, 99(9):5766–5771, 2002.