Embedding spanning structures in graphs and hypergraphs

by

Fiachra Knox

A thesis submitted to
The University of Birmingham
for the degree of
Doctor of Philosophy

School of Mathematics
The University of Birmingham
September 2012

Abstract

In this thesis we prove three main results on embeddings of spanning subgraphs into graphs and hypergraphs. The first is that for $\log^{50} n/n \leqslant p \leqslant 1 - n^{-1/4} \log^9 n$, a binomial random graph $G \sim G_{n,p}$ contains with high probability a collection of $\lfloor \delta(G)/2 \rfloor$ edge-disjoint Hamilton cycles (plus an additional edge-disjoint matching if $\delta(G)$ is odd), which confirms for this range of p a conjecture of Frieze and Krivelevich. Secondly, we show that any 'robustly expanding' graph with linear minimum degree on sufficiently many vertices contains every bipartite graph on the same number of vertices with bounded maximum degree and sublinear bandwidth. As corollaries we obtain the same result for any graph which satisfies the Ore-type condition $d(x) + d(y) \geqslant (1 + \eta)n$ for non-adjacent vertices x and y, or which satisfies a certain degree sequence condition. Thirdly, for $\gamma > 0$ we give a polynomial-time algorithm for determining whether or not a k-graph with minimum codegree at least $(1/k + \gamma)n$ contains a perfect matching. This essentially answers a question of Rödl, Ruciński and Szemerédi. Our algorithm relies on a strengthening of a structural result of Keevash and Mycroft. Finally and additionally, we include a short note on Maker-Breaker games.

Acknowledgements

I would like to thank my supervisors, Deryk Osthus and Daniela Kühn, for their patient support and reliable advice throughout the course of my studies at Birmingham. I would also like to thank them for their collaboration on the results proved in Chapter 2, which correspond to those in [72] and [73]. I would like to thank Andrew Treglown for his collaboration on the results proved in Chapter 3, which correspond to those in [74], and in proposing Conjecture 1.4.6. I would also like to thank Peter Keevash and Richard Mycroft for their collaboration on the results proved in Chapter 4. Thanks must also go to the referees who provided many helpful comments on [72] and [74]. I was supported in my studies by the EPSRC.

There are many more people at Birmingham with whom I have enjoyed interacting during my time there; among them are Oliver Cooley, Andrew Bailey, Diana Piguet, Nikolaos Fountoulakis, Dan Hefetz, Allan Lo, John Lapinsakas and Katherine Staden. Finally, I would like to thank my parents, who have provided constant encouragement, and my wonderful fiancée Ana Maria, whose love has motivated me in both work and life in general.

Contents

List of Figures

Chapter 1

Introduction

For combinatorial structures G and H (for example, two graphs), one very basic and natural question to ask is whether G contains a substructure which is isomorphic to H. This type of problem is known as an embedding problem. If G indeed contains such a substructure, we say that H *embeds into* G, written $H \subseteq G$. Embedding problems are a very wide class of problems and their character and the techniques used to solve them vary considerably, depending on the type of combinatorial structure involved and the way in which G and H are chosen. For example, G and H could be specific graphs or could be taken from some specified class of graphs. Alternatively, either or both structures could be taken from a specified probability distribution, in which case we are usually interested in the *probability* that H embeds into G.

In this thesis we will consider embedding problems involving graphs and hypergraphs. Such problems have a long history in extremal combinatorics; they go back at least as far as Petersen [100], who proved in 1891 that every regular graph G contains a regular subgraph of degree 2 which is *spanning* (that is, contains all of the vertices of G). Another very old embedding result is Mantel's Theorem [95], proved in 1907, which states that any graph with at least $(n^2 + 1)/4$ edges contains a triangle. This is a special case of Turán's Theorem [118], which gives for each $r \geqslant 3$ the maximum number of edges in a graph which does not contain a copy of K_r.

In both Turán's Theorem and Mantel's Theorem, G may have many more vertices than H. By contrast, in this thesis we shall focus on the case in which the ground sets of G and H have the same size, as in Petersen's Theorem. In this case the subgraph we find in G, isomorphic to H, is a spanning subgraph.

One of the most famous classical results in this area is Dirac's theorem [33], which states that every graph G of minimum degree at least $n/2$ contains a Hamilton cycle (that is, a cycle containing all of the vertices of G). Another famous and beautiful result is the Hajnal–Szemerédi theorem [52], which states that for any $r \geqslant 3$ and any n which is divisible by r, any graph G on n vertices of minimum degree at least $(1 - 1/r)n$ contains a *perfect* K_r*-packing*; that is, the vertices of G may be covered by vertex-disjoint copies of K_r, each of which is a subgraph of G. (Corrádi and Hajnal [30] had earlier proved this result in the case when $r = 3$.) In recent years, the area has seen a great deal of progress in the form of many striking results, as will be discussed in detail in the later sections of this introduction.

1.1 Overview

For now we give a general description of the results in this thesis, which will be stated precisely once some futher definitions have been established. Our first main results, presented in Chapter 2, concern random graphs. We study the binomial random graph model $G_{n,p}$, first introduced by Gilbert [48] and studied extensively by Erdős and Renyi [37, 38]. In this model, a graph G on n vertices is chosen by including each possible edge independently at random with probability p. A property is said to hold *with high probability* (whp) in $G_{n,p}$ (or more generally, in any graph model) if the probability that $G \sim G_{n,p}$ possesses the property tends to 1 as $n \to \infty$. We show that when p is not too large or too small, with high probability G contains a large collection of edge-disjoint Hamilton cycles. In fact, we will show that with high probability the limit on the size of such a collection is determined by the minimum degree of G. This proves a conjecture of Frieze and Krivelevich [43] except when p is very small or very large. Before proving this result, we also establish a weaker 'approximate' result which applies for a somewhat wider range of p. This proof is simpler, but uses many of the same themes as the 'exact' result.

In Chapter 3, we present a very general embedding result for graphs G which have linear minimum degree and which satisfy a certain condition, known as 'robust expansion', which we will describe in detail later on. Roughly speaking, a graph G on n vertices is a robust expander if, for every 'reasonably sized' set $S \subseteq V(G)$, G contains at least $|S| + \Omega(n)$ vertices that are adjacent to 'many' vertices in S. A graph H is said to have 'small bandwidth' if there exists a linear ordering of its vertices such that the endvertices of each edge of H are not too far apart. Conceptually, graphs of small bandwidth can be thought of as being 'path-like'; examples of such graphs are Hamilton cycles and K_r-factors (for small r). We will give a precise definition of both concepts in Section 1.4. We show that if G is a robust expander its minimum degree is not too small, every bipartite graph H of bounded maximum degree and small bandwidth embeds into G. Further, as corollaries we establish similar results when either a condition on the degree seqence of G or an 'Ore-type' condition is given. Our proof uses the Regularity Lemma (described below), and so requires the vertex sets of G and H to be sufficiently large; in fact, this restriction also applies to the results in Chapter 4.

A *hypergraph* $H = (V, E)$ consists of a set V of vertices and a set $E \subseteq P(V)$ of edges. In other contexts hypergraphs are also known as *set systems*. A hypergraph H is *k-uniform* if every edge of H has size k. In this thesis a k-uniform hypergraph will also be referred to as a k-graph. While both Dirac's Theorem and the Hajnal–Szemerédi Theorem have short proofs using elementary methods (see [70] for the latter), generalisations of these results to k-graphs have proven notoriously difficult.

In Chapter 4 we will treat the simplest case of such problems, that of finding a perfect matching. It turns out that even this seemingly simple problem in fact requires a large amount of technical work and theory. Indeed, Karp [61] showed that even the decision problem of determining whether a k-graph contains a perfect matching is NP-complete for $k \geqslant 3$. We will show that for any $\gamma > 0$, every k-graph H on sufficiently many vertices with 'miniumum codegree' at least $(1/k + \gamma)n$ either contains a perfect matching, or has a very specific structure. This structure is restrictive enough that, when H satisfies this codegree condition, we can exploit our result to obtain a polynomial-time algorithm for determining whether or not H contains a perfect matching.

1.2 Techniques in graph and hypergraph theory

In proving these results we use a number of established techniques, in particular probabilistic methods, the concept of 'pseudorandomness', and the Regularity Lemma. Probabilistic methods are an invaluable tool in almost all areas of combinatorics. Most obviously, dealing with random graphs will always require probabilistic tools to some extent. However, the use of probabilistic methods is very far from being restricted to random graphs. Indeed, whenever it is required to prove that a combinatorial object possessing a certain property exists, it suffices to prove that a random object, drawn from a specified distribution, has the property with positive probability. In Chapters 3 and 4 we will use probabilistic methods for precisely this purpose, despite the fact that the questions we address are wholly deterministic. There are many properties for which this method works even though the objects that possess them are very difficult to construct explicitly.

One of the most famous examples of the use of probabilistic methods in combinatorics is the exponential lower bound on Ramsey numbers given by Erdős [36]. To this date no explicit construction of graphs which realise this lower bound, or even which provide any exponential lower bound at all, is known. Due to their huge influence on combinatorics, it would be impossible to give a complete account of the application of probabilistic methods here. Extensive literature has been written on this topic; see for example [5]. Some basic probabilistic tools which we use in this thesis are given in Section 1.8, at the end of this introduction.

On the other hand, it is often useful to have examples of specific graphs which exhibit at least some of the properties typical of random graphs. Such graphs are often referred to as *pseudorandom* graphs. While in general it is not possible to capture all of the typical properties of a random graph in this way, certain important properties can and have been transferred to a deterministic setting. One important example, which we will make use of, is the *edge-distribution* of the graph. This was first treated by Thomason [112], who introduced the notion of *jumbled* graphs (see Definition 2.3.1 for a precise definition).

Later, Chung, Graham and Wilson [27] exhibited a number of seemingly different properties relating to pseudorandomness and proved them to be equivalent. They thus introduced the notion of *quasirandom* graphs. More recently, much attention has been focused on (n, d, λ)-graphs, that is, graphs on n vertices which are regular of degree d and have *second eigenvalue* λ. The significance of the second eigenvalue (which technically refers to the absolute values of the eigenvalues rather than the eigenvalues themselves) is that its value was one of the properties studied in [27], and so a bound on the second eigenvalue implies a number of other properties. While we do not refer directly to (n, d, λ)-graphs in the main body of this thesis, such graphs are known (see, e.g., Chapter 9 of [5]) to have strong edge distribution properties and also strong expansion properties. Thus they are related to the graphs dealt with in Chapters 2 and 3. In particular, Kühn and Osthus [84] observed that for any δ, there exists ε such that for sufficiently large n, any (n, d, λ)-graph with $d \geqslant \delta n$ and $\lambda \leqslant \varepsilon n$ is a 'robust expander'. Results for robust expanders therefore also apply to (n, d, λ)-graphs. A survey by Krivelevich and Sudakov [92] gives a great deal of additional background on pseudorandomness.

Szemerédi's Regularity Lemma, first introduced in [110], is relatively simple to prove but has proven an extremely powerful tool with a wide range of applications in graph theory. Roughly speaking, the lemma states that any sufficiently large graph may be partitioned into a number of 'clusters', such that the edge distribution of the bipartite subgraph between almost every pair of clusters is pseudorandom (or ε-*regular*). The lemma is stated precisely in Chapter 3 as Lemma 3.3.3. Early applications of the Regularity Lemma include the proof of the bounded degree case of the Burr-Erdős conjecture by Chvátal, Rödl, Szemerédi and Trotter [29] and Komlós, Sárközy and Szemerédi's proof of the Pósa-Seymour conjecture in the case of large graphs [77]. Further examples are given in Section 1.4.

The Regularity Lemma is particularly powerful when used in conjunction with the Blow-up Lemma of Komlós, Sárkőzy and Szemerédi [76], which allows the embedding of any bipartite graph of bounded degree into regular pairs. We will (implicitly) use the Blow-up Lemma in Chapter 3 (Lemma 3.8.2 relies upon it). One notable disadvantage of using the Reguarity Lemma is that the minimum size of the graphs on which it holds is very large, as shown by Gowers [50].

More recently, extensions of this method for use on hypergraphs have also proved very useful. The Weak Hypergraph Regularity Lemma, introduced by Chung [26], has a similar, simple proof to that of the Graph Regularity Lemma, but the embedding properties it ensures are much less powerful than in the graph case. Nevertheless, it has seen many applications and we use this version directly when dealing with hypergraphs.

By contrast, the (strong) Hypergraph Regularity Lemma is much more technical even when restricted to 3-graphs, both in its statement and proof, but is also more powerful. In fact, at least three different versions of the Hypergraph Regularity Lemma have been proposed. The first was introduced for 3-graphs by Frankl and Rödl [41], and later extended to arbitrary k-graphs by Rödl and Skokan [107]. Another version of the Hypergraph Regularity Lemma was introduced by Gowers [51], and an algorihmic version was proposed by Haxell, Nagle and Rödl [57]. Keevash [63] proved a hypergraph version of the Blow-up Lemma which, like its graph counterpart, is particularly useful when combined with the Hypergraph Regularity Lemma. While we do not use the Hypergraph Regularity method directly in this thesis, it should be noted that we do use a powerful result (Lemma 4.4.7) which relies on it.

Two other more recent concepts we use are 'robust expansion' and geometric methods. Expansion properties of graphs have been widely considered and studied. 'Robust expansion' is a stronger property which has proved particularly useful in embedding problems. Recall that a graph G on n vertices is a robust expander if, for every 'reasonably sized' set $S \subseteq V(G)$, G contains at least $|S| + \Omega(n)$ vertices that are adjacent to 'many' vertices in S. This definition was introduced by Kühn, Osthus and Treglown [87] in the context of directed graphs, although it was already implicit in papers of Kelly, Kühn and Osthus [66] and Keevash, Kühn and Osthus [64]. Robust expansion is a key definition in Chapter 3; indeed, as previously mentioned Theorem 1.4.8 applies specifically to robustly expanding graphs. It is worth noting that in the remainder of the thesis, while we do not refer to robust expansion itself, we do use properties of a similar flavour (see, e.g., Lemma 2.2.7). Recently, Kühn and Osthus [83] proved a very powerful result concerning Hamilton cycles in robust expanders, which we will mention in Section 1.3.3.

The correspondance between perfect matchings in hypergraphs and geometric structures was first drawn by Rödl, Ruciński and Szemerédi [105] and was developed further by Keevash and Mycroft [65]. Given a hypergraph H and a partition of its vertices, we can assign to each edge of the hypergraph (or more generally, to any set of vertices) an *index vector* according to which parts of the vertex set its vertices fall into. Then given any perfect matching M, the index vectors of the edges in M must sum to the index vector of $V(H)$. We consider the lattice L generated by the index vectors of edges of H, and observe that H can only have a perfect matching if the index vector of $V(H)$ lies in L. By reframing the problem along these lines, geometric techniques (see, e.g., Lemma 4.4.9) can be brought to bear and, as we shall see, prove useful in solving the problem of whether or not a perfect matching exists.

1.3 Edge-disjoint Hamilton cycles in random graphs

The theory of random graphs has been well developed since its beginnings in the late 1950s. While many models of random graphs have been studied, the model $G_{n,p}$ which we use is particularly nice to work with, due to the independence of different edges being present. Erdős and Renyi [37] determined the connectivity threshold for $G_{n,p}$ is $\log n/n$; that is, they proved that when $pn - \log n \to \infty$ as $n \to \infty$ then with high probability $G_{n,p}$ is connected, and if $pn - \log n \to -\infty$ then with high probability $G_{n,p}$ is not connected. (The latter case is simply due to the fact that with high probability one of the vertices of the graph is isolated.) They later determined the threshold values for the appearance of some small subgraphs [38] and various other properties.

Since then a large body of work has been published on the subject; see, e.g., [15, 60] for further details and many remarkable results. One property we will make particular use of is the following: Suppose that $np/\log n \to \infty$ as $n \to \infty$. Then with high probability, the minimum degree of G is $(1 - o(1))np$.

Hamilton cycles are an extensively studied topic in graph theory. As they are very natural structures to examine there are many natural problems which may be posed concerning Hamilton cycles, but these problems are often rather difficult to approach. For example, for an arbitrary graph G the decision problem of determining whether or not G contains a Hamilton cycle is known to be NP-complete [61], making it unlikely that simple necessary and sufficient conditions for the existence of a Hamilton cycle can be found. Simple sufficient conditions *do* exist; recall for example Dirac's Theorem, which shows that $\delta(G) \geqslant n/2$ suffices, where $\delta(G)$ is the minimum degree of G. On the other hand, trivially the condition $\delta(G) \geqslant 2$ is a necessary one.

For a random graph $G \sim G_{n,p}$, however, it turns out that the condition $\delta(G) \geqslant 2$ is also sufficient to guarantee, with high probability, the existence of a Hamilton cycle. Indeed, Bollobás [14] and Ajtai, Komlós and Szemerédi [2] proved that in almost every random graph process where we add edges at random, one by one, to an empty graph on n vertices, the very edge which increases the minimum degree of the graph to 2 also produces a Hamilton cycle.

A natural extension of the problem of finding a Hamilton cycle is the following question: How many Hamilton cycles can be 'packed' into a graph G, so that each cycle is a subgraph of G and all of the cycles are edge-disjoint? As in the case of a single Hamilton cycle, there is a trivial necessary condition on $\delta(G)$: If G contains k edge-disjoint Hamilton cycles, then it must have minimum degree at least $2k$. This leads to the following definition:

Definition 1.3.1 *For a graph G on n vertices, call a matching M in G* optimal *if $|M| = \lfloor n/2 \rfloor$. A graph G has* property $\mathcal{H}$ *if G contains $\lfloor \delta(G)/2 \rfloor$ edge-disjoint Hamilton cycles, together with an additional edge-disjoint optimal matching if $\delta(G)$ is odd.*

In the case that G is regular of even degree and possesses property $\mathcal{H}$, G has a *Hamilton decomposition*; that is, all of the edges of G may be covered by a collection of edge-disjoint Hamilton cycles.

Frieze and Krivelevich [42] conjectured that almost every graph with n vertices and m edges has property $\mathcal{H}$, *regardless* of how we choose $m = m(n)$. They later [43] conjectured that with high probability $G_{n,p}$ has property $\mathcal{H}$.

Conjecture 1.3.2 (Frieze and Krivelevich [43]) *For any $p = p(n)$, with high probability $G_{n,p}$ has property $\mathcal{H}$.*

This conjecture has now essentially been proved. This was done in two parts: a 'dense case', which is mostly covered by results in this thesis, and a 'sparse' case covered by the work of other authors. For completeness we give background on both cases. We also include a brief account of exciting new results in the area which have recently been proved.

1.3.1 Dense random graphs

Frieze and Krivelevich gave a partial result towards Conjecture 1.3.2 in the case where p is constant, namely that G contains an 'approximate' Hamilton decomposition.

Theorem 1.3.3 (Frieze and Krivelevich [42]) *Let $0 < p < 1$ be constant. Then whp $G_{n,p}$ contains $(1 - o(1))np/2$ edge-disjoint Hamilton cycles.*

As mentioned previously, $\delta(G_{n,p}) = (1 - o(1))np$ for this range of p and so the size of the collections is $(1 - o(1))\delta(G_{n,p})/2$. Hence the Hamilton cycles cover all but a small proportion of the edges of $G_{n,p}$. As remarked in [44], the proof of [42] also works as long as p is a little larger than $n^{-1/8}$.

The first result we present in Chapter 2 extends Theorem 1.3.3 to essentially the entire range of p. Our proof builds on ideas from [42] and [91].

Theorem 1.3.4 *For any $\eta > 0$, there exists a constant C such that if $p \geqslant \frac{C \log n}{n}$, then whp $G_{n,p}$ contains $(1 - \eta)np/2$ edge-disjoint Hamilton cycles.*

This result was proven independently by Krivelevich (personal communication). Later in Chapter 2 we build on the ideas used in the proof of Theorem 1.3.4 to prove the following result:

Theorem 1.3.5 *Let $\log^{50} n/n \leqslant p \leqslant 1 - n^{-1/4} \log^9 n$. Then with high probability, $G_{n,p}$ has property $\mathcal{H}$.*

Very recently, Kühn and Osthus [84] covered the 'very dense' case $p \geqslant 2/3$. By this time the sparse case $p \leqslant n^{-1+\varepsilon}$ had also been covered, as detailed in Section 1.3.2.

In fact we prove Theorem 1.3.5 via a deterministic result (Theorem 2.3.62) which we believe to be of independent interest. Hefetz, Kühn, Lapinskas and Osthus [58] proved a 'covering' result for a similar range of p, which states that with high probability all of the edges of $G \sim G_{n,p}$ can be covered with $\lceil \Delta(G)/2 \rceil$ (not necessarily edge-disjoint) Hamilton cycles, each of which is contained in G. Their proof uses Lemma 2.3.60, a key lemma in the proof of Theorem 2.3.62, as a component. This answered a question and improved a previous result of Glebov, Krivelevich and Szabó [49].

Hypergraph versions of Conjecture 1.3.2 were also considered by Frieze and Krivelevich [44], Frieze, Krivelevich and Loh [45] and Bal and Frieze [10]. However, as yet no 'exact' results (in which the number of edge-disjoint Hamilton cycles depends only on the minimum degree, as in Theorem 1.3.5) are known.

1.3.2 Sparse random graphs

After the result of Bollobás [14] and Ajtai, Komlós and Szemerédi [2] for single Hanilton cycles, Bollobás and Frieze [16] showed that that almost every graph with n vertices and $n(\log n + (2k-1)\log\log n)/2$ edges contains a collection of k edge-disjoint Hamilton cycles; with high probability such graphs have minimum degree equal to $2k$. This result immediately deals with the case $np \leqslant \log n + O(\log\log n)$ of Conjecture 1.3.2, and in [43] Frieze and Krivelevich extended this to all p with $np = (1+o(1))\log n$. Ben-Shimon, Krivelevich and Sudakov [13] went further and showed that Conjecture 1.3.2 holds as long as $np \leq 1.02\log n$. Finally, Krivelevich and Samotij [90] covered the range $\log n/n \leqslant p \leqslant n^{-1+\varepsilon}$, which overlaps with the range covered by Theorem 1.3.5.

Hence taken together, Theorem 1.3.5 and the results of [13, 90, 84] essentially prove Conjecture 1.3.2. (Unfortunately, the results of [90, 84] do not include the additional edge-disjoint optimal matching.)

Theorem 1.3.6 *For any $p = p(n)$, with high probability $G \sim G_{n,p}$ contains $\lfloor \delta(G)/2 \rfloor$*

edge-disjoint Hamilton cycles.

1.3.3 Deterministic results for graphs and digraphs

The question of whether a graph possesses an approximate or exact Hamilton decomposition is, of course, far from being restricted to random graphs. In this section we briefly detail some results in this area, including exciting recent results which have resolved a number of long-standing conjectures and made progress on others.

An old construction by Walecki (see e.g. [7, 94]) shows that the complete graph K_n has a Hamilton decomposition if n is odd. It is well known that more generally K_n has property $\mathcal{H}$ (see e.g. [120]). Nash-Williams [97] conjectured that any d-regular graph on n vertices where $d \geqslant n/2$ has $\lfloor d/2 \rfloor$ edge-disjoint Hamilton cycles. Christofides, Kühn and Osthus [25] showed that any regular graph G on $n \geqslant n_0 = n_0(\varepsilon)$ vertices with degree at least $(1+\varepsilon)n/2$ has an approximate Hamilton decomposition. Their proof uses the Regularity Lemma, and so as remarked before n_0 is quite large. Harte and Seacrest [56] later gave an improved version of this result which avoids using the Regularity Lemma, and so holds for smaller graphs.

A *directed graph* or *digraph* $G = (V, E)$ is an analogue of a graph for which V is a set of vertices and E is a set of *directed edges*, i.e., of ordered pairs of vertices in V. A *directed Hamilton cycle* is a cycle which contains all of the vertices and in which every vertex has in- and out-degree exactly 1. (Whenever we refer to a Hamilton cycle in a digraph setting we implictly mean a directed Hamilton cycle.) Hamilton decompositions and approximate Hamilton decompositions in a digraph setting are defined analogously to those in a graph setting.

Tillson [114] proved that a complete digraph on n vertices has a Hamilton decomposition for any $n \neq 4, 6$. A *tournament* is a digraph formed by replacing every edge xy of a complete (undirected) graph by exactly one of the directed edges $\vec{xy}$ and $\vec{yx}$. A famous conjecture in this area is Kelly's conjecture, which states that any regular tournament has a Hamilton decomposition. Kühn, Osthus and Treglown [85] proved an approximate version of Kelly's conjecture.

Recently, Kühn and Osthus [83] proved a very general result with wide implications, namely that every regular digraph which is a robust outexpander (the natural analogue of a robust expander in the digraph case) has a Hamilton decomposition. Their proof relies on another recent result of Osthus and Staden [99] which states that such a digraph at least has an approximate Hamilton decomposition. As a result they were able to prove not only Kelly's conjecture, but also (in [84]) a conjecture of Erdös which states that a random tournament possesses with high probability a digraph analogue of property $\mathcal{H}$, a further partial result towards the Nash-Williams conjecture, and the previously mentioned result for the 'very dense' case of Conjecture 1.3.2. They also proved a strong result concerning collections of edge-disjoint Hamilton cycles in (not necessarily regular) graphs of high minimum degree; this was improved on by Kühn, Lapinskas and Osthus [79], again using the main result of [83].

1.4 Embedding bipartite graphs of small bandwidth

Recall that the Hajnal–Szemerédi theorem [52] gives the minimum degree condition on a graph G which guarantees the existence of a perfect K_r-packing in G. This seminal result has since inspired a number of sucessively wider generalisations. Given a graph F, a perfect F-packing in a graph G is a collection of vertex-disjoint copies of F which covers all the vertices in G. (Perfect F-packings are also referred to as *F-factors* or *perfect F-tilings*.) Kühn and Osthus [80, 82] characterised, up to an additive constant, the minimum degree which ensures a graph G contains a perfect F-packing for an arbitrary graph F. (This improved previous bounds of Alon and Yuster [6] and Komlós, Sárközy and Szemerédi [78].)

In a similar vein, the Pósa–Seymour conjecture (see [39] and [108]) states that any graph G on n vertices with $\delta(G) \geqslant rn/(r+1)$ contains the rth power of a Hamilton cycle. (The rth power of a Hamilton cycle C is obtained from C by adding an edge between every pair of vertices of distance at most r on C.) As mentioned previously, Komlós, Sárközy and Szemerédi [77] proved this conjecture for sufficiently large graphs. Notice that in the case when $r+1$ divides $|G|$, a necessary condition for a graph G to contain the rth power of a Hamilton cycle is that G contains a perfect K_{r+1}-packing. Thus, the Pósa–Seymour conjecture implies the Hajnal–Szemerédi theorem.

Given a graph H on n vertices, the *bandwidth* of H is defined as follows: Consider all vertex orderings $\sigma = (v_1, \dots, v_n)$ of H, and let $b(H, \sigma)$ be the maximum distance between two adjacent vertices according to this ordering. That is, $b(\sigma)$ is the maximum of $|i-j|$ over all pairs $\{i, j\}$ such that $v_i v_j$ is an edge of H. The bandwidth $bw(H)$ is the minimum of $b(H, \sigma)$ over all orderings σ of $V(H)$. Both perfect F-packings and rth powers of Hamilton cycles are examples of graphs with low bandwidth, provided that $|F|$ and r respectively are small compared to n (this case is almost universally the one considered). Indeed, every graph H has bandwidth at most $|H|-1$. Thus, a perfect F-packing has bandwidth at most $bw(F) \leqslant |F|-1$. A cycle has bandwidth 2, and in general the rth power of a cycle has bandwidth at most $2r$. Further, Böttcher, Pruessmann, Taraz and Würfl [20] proved that every planar graph H on n vertices with bounded maximum degree has bandwidth at most $O(n/\log n)$.

The following result of Böttcher, Schacht and Taraz [22] gives a condition on the minimum degree of a graph G on n vertices that ensures G contains *every* r-chromatic graph on n vertices of bounded degree and of bandwidth $o(n)$, thereby proving a conjecture of Bollobás and Komlós [75].

Theorem 1.4.1 (Böttcher, Schacht and Taraz [22]) *Given any $r, \Delta \in \mathbb{N}$ and any $\gamma > 0$, there exist constants $\beta > 0$ and $n_0 \in \mathbb{N}$ such that the following holds. Suppose that*

H is an r-chromatic graph on $n \geqslant n_0$ vertices with $\Delta(H) \leqslant \Delta$ and bandwidth at most βn. If G is a graph on n vertices with

$$\delta(G) \geqslant \left(\frac{r-1}{r} + \gamma\right) n,$$

then G contains a copy of H. □

Prior to the proof of Theorem 1.4.1, Csaba [31] and Hàn [54] proved the case when H is bipartite and Böttcher, Schacht and Taraz [21] proved the case when $\chi(H) = 3$. Our aim in Chapter 3 will be to strengthen Theorem 1.4.1 in the case when H is bipartite.

1.4.1 Degree sequence conditions

Dirac's theorem and the Hajnal–Szemerédi theorem are best possible in the sense that the minimum degree conditions in both these results cannot be lowered. However, this does not mean that one cannot strengthen these results. Indeed, Chvátal [28] gave a condition on the degree sequence of a graph which ensures Hamiltonicity: Suppose that the degrees of the graph G are $d_1 \leqslant \ldots \leqslant d_n$. If $n \geqslant 3$ and $d_i \geqslant i+1$ or $d_{n-i} \geqslant n-i$ for all $i < n/2$ then G is Hamiltonian. Previously, Pósa [102] showed that the stronger condition $d_i \geqslant i+1$ for all $i < n/2$ suffices. Notice that both Pósa's Theorem and Chvátal's Theorem are much stronger than Dirac's theorem since they allow for almost half of the vertices of G to have degree less than $n/2$.

Balogh, Kostochka and Treglown [12] proposed the following two conjectures concerning the degree sequence of a graph which forces a perfect F-packing, which can be seen as analogues of Pósa's Theorem.

Conjecture 1.4.2 (Balogh, Kostochka and Treglown [12]) *Let $n, r \in \mathbb{N}$ such that r divides n. Suppose that G is a graph on n vertices with degree sequence $d_1 \leqslant \ldots \leqslant d_n$ such that:*

- $d_i \geqslant (r-2)n/r + i$ *for all* $i < n/r$*;*

- $d_{n/r+1} \geqslant (r-1)n/r$.

Then G *contains a perfect* K_r*-packing.*

Note that Conjecture 1.4.2, if true, is much stronger than the Hajnal–Szemerédi theorem since the degree condition allows for n/r vertices to have degree less than $(r-1)n/r$.

Conjecture 1.4.3 (Balogh, Kostochka and Treglown [12]) *Suppose* $\gamma > 0$ *and* F *is a graph with* $\chi(F) = r$*. Then there exists an integer* $n_0 = n_0(\gamma, F)$ *such that the following holds. If* G *is a graph whose order* $n \geqslant n_0$ *is divisible by* $|F|$*, and whose degree sequence* $d_1 \leqslant \ldots \leqslant d_n$ *satisfies*

- $d_i \geqslant (r-2)n/r + i + \gamma n$ *for all* $i < n/r$*,*

then G *contains a perfect* F*-packing.*

This conjecture is much more general than Conjecture 1.4.2, but is not exact (due to the error term γ) and hence is weaker in the case when $F = K_r$.

In Chapter 3 we prove the following result which gives a condition on the degree sequence of a graph G on n vertices that ensures G contains every bipartite graph on n vertices of bounded degree and of bandwidth $o(n)$.

Theorem 1.4.4 *Given any* $\Delta \in \mathbb{N}$ *and any* $0 < \gamma < 1/2$*, there exists constants* $\beta > 0$ *and* $n_0 \in \mathbb{N}$ *such that the following holds. Suppose that* H *is a bipartite graph on* $n \geqslant n_0$ *vertices with* $\Delta(H) \leqslant \Delta$ *and bandwidth at most* βn*. Let* G *be a graph on* n *vertices with degree sequence* $d_1 \leqslant \ldots \leqslant d_n$*. If*

- $d_i \geqslant i + \gamma n$ *or* $d_{n-i-\gamma n} \geqslant n - i$ *for all* $i < n/2$

then G *contains a copy of* H*.*

The degree sequence condition in Theorem 1.4.4 is similar to that in Chvátal's theorem, except that now we have two error terms in the condition. Notice that Theorem 1.4.4 is much stronger than the bipartite case of Theorem 1.4.1. Furthermore, in the case when $r = 2$, Conjecture 1.4.3 is implied by Theorem 1.4.4.

Theorem 1.4.4 is, up to the error terms, best-possible for many graphs H. Indeed, suppose that H is a bipartite graph on an even number n of vertices that contains a perfect matching. Suppose that $m \in \mathbb{N}$ is such that $m < n/2$. Let G be a graph on n vertices with vertex classes V_1, V_2, V_3 of sizes m, $m-1$ and $n-2m+1$ respectively and whose edge set contains all possible edges except for those in V_1 and between V_1 and V_3. Let $d_1 \leqslant \ldots \leqslant d_n$ denote the degree sequence of G. Then

- $d_i \geqslant i-1$ and $d_{n-i+2} \geqslant n-i$ for all $2 \leqslant i < n/2$,

but since $|V_1| > |V_2|$, G does not contain a perfect matching or, therefore, H.

1.4.2 Ore-type degree conditions

Ore-type degree conditions consider the sum of the degrees of non-adjacent vertices of a graph. The name comes from Ore's theorem [98], which states that a graph G of order $n \geq 3$ contains a Hamilton cycle if $d(x) + d(y) \geqslant n$ for all non-adjacent $x \neq y \in V(G)$. Recently, Châu [23] proved an Ore-type analogue of the Pósa–Seymour conjecture in the case of the square of a Hamilton cycle (i.e. when $r = 2$).

The following Ore-type result of Kierstead and Kostochka [69] implies the Hajnal–Szemerédi theorem: Let $n, r \in \mathbb{N}$ such that r divides n. Suppose that G is a graph on n vertices such that for all non-adjacent $x \neq y \in V(G)$, $d(x) + d(y) \geqslant 2(r-1)n/r - 1$. Then G contains a perfect K_r-packing. Kühn, Osthus and Treglown [86] characterised, asymptotically, the Ore-type degree condition which ensures a graph G contains a perfect F-packing for an arbitrary graph F.

It is natural to seek an Ore-type analogue of Theorem 1.4.1. The following result provides such an analogue in the case when H is bipartite.

Theorem 1.4.5 *Given any $\Delta \in \mathbb{N}$ and any $\gamma > 0$, there exists constants $\beta > 0$ and $n_0 \in \mathbb{N}$ such that the following holds. Suppose that H is a bipartite graph on $n \geqslant n_0$ vertices with $\Delta(H) \leqslant \Delta$ and bandwidth at most βn. Let G be a graph on n vertices such that, for all non-adjacent $x \neq y \in V(G)$,*

$$d(x) + d(y) \geqslant (1 + \gamma)n.$$

Then G contains a copy of H.

We prove Theorem 1.4.5 by showing that it is a direct consequence of Theorem 1.4.4. Note that Theorem 1.4.5 is best-possible up to the error term for bipartite graphs H on n vertices which do not contain an isolated vertex. Indeed, let G consist of a copy of K_{n-1} and an isolated vertex. Then G does not contain H but $d(x) + d(y) = n - 2$ for all non-adjacent $x \neq y \in V(G)$.

In light of Theorem 1.4.5, we propose the following Ore-type analogue of Theorem 1.4.1.

Conjecture 1.4.6 *Given any $r, \Delta \in \mathbb{N}$ and any $\gamma > 0$, there exists constants $\beta > 0$ and $n_0 \in \mathbb{N}$ such that the following holds. Suppose that H is an r-chromatic graph on $n \geqslant n_0$ vertices with $\Delta(H) \leqslant \Delta$ and bandwidth at most βn. Let G be a graph on n vertices such that, for all non-adjacent $x \neq y \in V(G)$,*

$$d(x) + d(y) \geqslant 2\left(\frac{r-1}{r} + \gamma\right)n.$$

Then G contains a copy of H.

If true, Conjecture 1.4.6 is stronger than Theorem 1.4.1. Böttcher and Müller [18, 19] have proved the conjecture in the case when $r = 3$.

1.4.3 Robustly expanding graphs

We now give a precise definition of the concept of *robust expansion* which was mentioned earlier in this introduction. Let $0 < \nu \leqslant \tau < 1$. Suppose that G is a graph on n vertices and $S \subseteq V(G)$. Then the *ν-robust neighbourhood $RN_{\nu,G}(S)$ of S* is the set of vertices $v \in V(G)$ such that $|N(v) \cap S| \geqslant \nu n$. We say that G is a *robust (ν, τ)-expander* if every $S \subseteq V(G)$ with $\tau n \leqslant |S| \leqslant (1-\tau)n$ satisfies $|RN_{\nu,G}(S)| \geqslant |S| + \nu n$.

The following result is an immediate consequence of Theorem 16 from [87].

Theorem 1.4.7 (Kühn, Osthus and Treglown [87]) *Given positive constants $1/n_0 \ll \nu \leq \tau \ll \eta < 1$, let G be a graph on $n \geq n_0$ vertices with $\delta(G) \geq \eta n$ which is a robust (ν, τ)-expander. Then G contains a Hamilton cycle.* □

(Throughout this thesis, we write $0 < \alpha \ll \beta \ll \gamma$ to mean that we can choose the constants α, β, γ from right to left. More precisely, there are increasing functions f and g such that, given γ, whenever we choose some $\beta \leqslant f(\gamma)$ and $\alpha \leqslant g(\beta)$, all calculations needed in our proof are valid. Hierarchies of other lengths are defined in the obvious way.) Christofides, Keevash, Kühn and Osthus [24] proved that under the conditions of Theorem 1.4.7, a Hamilton cycle in G can be found in polynomial time.

We will use Theorem 1.4.7 to prove the following result concerning embedding bipartite graphs of small bandwidth.

Theorem 1.4.8 *Given $\Delta \in \mathbb{N}$ and positive constants $\nu \leq \tau \ll \eta < 1$ there exist constants $\beta > 0$ and $n_0 \in \mathbb{N}$ such that the following holds. Suppose that H is a bipartite graph on $n \geqslant n_0$ vertices with $\Delta(H) \leqslant \Delta$ and bandwidth at most βn. Let G be a graph on n vertices with $\delta(G) \geq \eta n$ which is a robust (ν, τ)-expander. Then G contains a copy of H.*

Note that Theorem 1.4.8 is very general in the sense that it allows for the graph G to have small minimum degree (although $\delta(G)$ must be linear). Furthermore, there are examples of graphs G that satisfy the hypothesis of Theorem 1.4.8 and whose *maximum degree* is also comparitively small. Indeed, let $0 < \nu \ll \tau \ll \eta < 1$ such that $1/\eta$ is an odd integer. Further choose $n \in \mathbb{N}$ such that $\eta n \in \mathbb{N}$. Define G to be the blow-up of a cycle on $1/\eta$ vertices, such that each vertex class of G contains ηn vertices. Thus, $|G| = n$ and $\delta(G) = \Delta(G) = 2\eta n$. It is easy to check that G is a robust (ν, τ)-expander. Given constants $0 < \nu \ll \tau \ll p \leqslant 1$, it is also easy to see that with high probability $G(n, p)$ is a robust (ν, τ)-expander with minimum degree at least $pn/2$ and maximum degree at most $2pn$.

Theorem 1.4.8 therefore implies that, with high probability for constant p, $G(n, p)$ contains all bipartite graphs H on n vertices of bounded degree and bandwidth $o(n)$. A result of Huang, Lee and Sudakov [59] actually implies that, with high probability for constant $p > 0$, any spanning subgraph G' of $G(n, p)$ with minimum degree $\delta(G') \geqslant (1/2 + o(1))np$ contains all such H.

1.5 Perfect matchings in hypergraphs

Recall that a k-graph is a hypergraph in which every edge has size k. The question of whether a k-graph H contains a perfect matching, while simple to state, is one of the key questions of combinatorics. In the graph case $k = 2$, Tutte's Theorem [116] gives necessary and sufficient conditions for H to contain a perfect matching, and Edmonds' Algorithm [34] gives a solution in polynomial time. However, for $k \geqslant 3$ this problem was one of Karp's celebrated 21 NP-complete problems [61]. Results for perfect matchings have many potential practical applications; one example which has garnered interest in recent years is the 'Santa Claus' allocation problem (see [9]). Since the general problem is intractable provided P $\neq$ NP, it is natural to seek conditions on H which render the problem tractable or even which guarantee that a perfect matching exists. In recent years a substantial amount of progress has been made in this direction.

One well-studied class of such conditions are minimum degree conditions. In the graph case, a simple greedy argument shows that a minimum degree of $n/2$ guarantees a perfect matching provided that n is even (in what follows we assume that k divides n, since this is a necessary condition for H to contain a perfect matching). Indeed, Dirac's theorem [33] states that this condition even guarantees that H contains a Hamilton cycle. For $k \geqslant 3$, there are a number of possible definitions for the minimum degree of H. Indeed, for any set $A \subseteq V(H)$, the *degree* $d(A)$ of A the number of edges of H containing A. Then for any $1 \leqslant \ell \leqslant k-1$, the *minimum ℓ-degree* $\delta_\ell(H)$ of H is the minimum of $d(A)$ over all subsets $A \subseteq V(H)$ of size ℓ.

Two cases have received particular attention: the minimum 1-degree $\delta_1(H)$ is also known as the *minimum vertex degree*, and the minimum $(k-1)$-degree $\delta_{k-1}(H)$ as the *minimum codegree* of H. General bounds for the minimum vertex degree which guarantees a perfect matching were obtained by Daykin and Häggvist [32] and improved by Markström and Ruciński [96]. For $k = 3$, the minimum vertex degree was determined independently by Kühn, Osthus and Treglown [88] and Khan [67]. Earlier, Hán, Person and Schacht [55] had given an asymptotic version of these results. Khan [68] also determined the minimum vertex degree which guarantees a perfect matching in a 4-graph, while Alon, Frankl, Huang, Rödl, Ruciński and Szemerédi [3] gave an asymptotic result for 5-graphs. For $k > 5$ this question is one of the main outstanding open problems in extremal hypergraph theory.

1.5.1 Space and divisibility barriers

For sufficiently large n, Rödl, Ruciński and Szemerédi [106] determined the minimum codegree which guarantees a perfect matching in H to be exactly $n/2 - k + c$, where $c \in \{1.5, 2, 2.5, 3\}$ is an explicitly given function of n and k. They also showed that a minimum codegree of just n/k is sufficient to guarantee a matching covering all but k

vertices of H (i.e. one edge away from a perfect matching). This is in sharp contrast with the graph case, where a minimum degree of $\delta(G) \geqslant n/2 - \varepsilon n$ only guarantees the existence of a matching covering at least $n - 2\varepsilon n$ vertices. The extremal example from [106] is as follows: Let A and B be disjoint sets with $|A|$ odd and $|A \cup B| = n$. Let H be the k-graph on $A \cup B$ whose edges are all k-tuples in which every edge intersects A in an even number of vertices. Then any matching in H covers an even number of vertices in A, so is not perfect, but $\delta_{k-1}(H) \geqslant n/2 - k$. Indeed, for suitable choices of $|A|$ and $|B|$, the k-graph formed in this manner has the highest codegree of any k-graph on n vertices with no perfect matching. This construction is the simplest example of a 'divisibility barrier' to a perfect matching.

For general ℓ, it was conjectured by Hán, Person and Schacht [55] that the extremal constructions for H which maximise $\delta_\ell(H)$ and such that H contains no perfect matching have one of two forms. The first, the 'divisibility barrier', is constructed in a similar way to the extremal example from [106] above. However, the construction may additionally be varied by including the edges which intersect A in an odd number of vertices instead of those which intersect A in an even number of vertices and varying the parity of $|A|$ appropriately. (The precise way this is done differs depending on the parity of n/k.) The second construction, the 'space barrier', is constructed as follows: Again we partition a vertex set V into vertex classes A and B, and this time H contains every edge which intersects A in at least one vertex. Then any perfect matching in H must contain at least n/k vertices in A, and so by taking $|A| = n/k - 1$ we ensure that H does not contain a perfect matching.

Treglown and Zhao [115] showed that when k is divisible by 4 and $k/2 \leqslant \ell \leqslant k-1$, there is indeed a 'divisibility barrier' construction which is extremal. However, as remarked in [115], finding the precise value of $|A|$ which actually gives the minimal ℓ-degree is not trivial and in fact is probably a difficult problem in itself. Their result improved upon an asymptotic result of Pikhurko [101]. Much of their proof in fact generalises to the case when k is even or when no divisibility condition on k is assumed; further, the result of [101] is not bound by the restriction that k is divisible by 4. However, in the case $1 < \ell < k/2$ much less is known; in particular, the asymptotics of the problem for general k have yet to be determined. For further results on this topic see the survey [104], as well as recent papers [1] and [93].

1.5.2 Polynomial-time algorithms

Given the result of [106], a natural question to ask is the following: Let $\mathbf{PM}(k, \delta)$ be the decision problem of determining whether a k-graph H with $\delta_{k-1}(H) \geqslant \delta n$ contains a perfect matching. For which values of δ is $\mathbf{PM}(k, \delta)$ in P? The main result of [106] implies that $\mathbf{PM}(k, 1/2)$ is in P. On the other hand, $\mathbf{PM}(k, 0)$ includes no degree restriction on H at all; as stated above, this was shown to be NP-complete by Karp [61]. Szymańska [111] proved that for $\delta < 1/k$ the problem $\mathbf{PM}(k, 0)$ admits a polynomial-time reduction to $\mathbf{PM}(k, \delta)$ and hence $\mathbf{PM}(k, \delta)$ is also NP-complete, while Karpiński, Ruciński and Szymańska [62] showed that there exists $\varepsilon > 0$ such that $\mathbf{PM}(k, 1/2 - \varepsilon)$ is in P. This left a hardness gap for $\mathbf{PM}(k, \delta)$ when $\delta \in [1/k, 1/2 - \varepsilon)$.

Later in this section we give an algorithm, Procedure DeterminePM, which eliminates this hardness gap almost entirely. (The analysis of this algorithm constitutes the majority of Chapter 4.)

Theorem 1.5.1 *Fix $k \geqslant 3$ and $\delta > 1/k$. Then for any k-graph H on n vertices with $\delta_{k-1}(H) \geqslant \delta n$ Procedure DeterminePM determines correctly whether or not H contains a perfect matching. Furthermore, it will do so in time at most $O(n^{f(k)})$, where $f(k) \leqslant k^{k^k}$.*

A consequence of Theorem 1.5.1 is that $\mathbf{PM}(k, \delta)$ is in P for any $\delta > 1/k$. Together with the main result of [111] this determines the complexity status of $\mathbf{PM}(k, \delta)$ for every $\delta \neq 1/k$. The case $\delta = 1/k$ remains open.

1.5.3 Geometric theory

Our algorithm relies heavily on a result of Keevash and Mycroft [65] giving sufficient conditions which ensure a perfect matching in a large family of k-graphs. In this context, the result from [65] essentially states that if H is a k-graph on n vertices, where n is large enough, and $\delta_{k-1}(H) \geqslant n/k + o(n)$, then either H contains a perfect matching, or H is close to a divisibility barrier. The simplest examples of these have already been described in Section 1.5.1. In fact, they proved that even if the codegree condition is weakened to $\delta_{k-1}(H) \geqslant (1/k - \varepsilon)n$ then the only additional possibility is that H close to a space barrier, which again generalises the second class of conjectured extremal examples in Section 1.5.1.

More generally, for any k-graph H and any partition $\mathcal{P}$ of $V(H)$ into d parts, we define the *index vector* $\mathbf{i}_{\mathcal{P}}(S) \in \mathbb{Z}^d$ of a subset $S \subseteq V(H)$ with respect to $\mathcal{P}$ to be the vector whose coordinates are the sizes of the intersections of S with each part of $\mathcal{P}$. Note that for any edge $e \in H$ the index vector $\mathbf{i}_{\mathcal{P}}(e)$ has non-negative co-ordinates which sum to k; we refer to a lattice generated by such vectors as an *edge-lattice*. In particular, we let $L_{\mathcal{P}}(H) \subseteq \mathbb{Z}^d$ denote the edge-lattice generated by the vectors $\mathbf{i}_{\mathcal{P}}(e)$ for $e \in H$. We say that $L_{\mathcal{P}}(H)$ is *complete* if it contains every index vector whose coordinates sum to a multiple of k and *incomplete* otherwise. With this notation, a divisibility barrier is a k-graph H and a partition $\mathcal{P}$ of $V(H)$ such that $L_{\mathcal{P}}(H)$ is incomplete and $\mathbf{i}_{\mathcal{P}}(V(H)) \notin L_{\mathcal{P}}(H)$. Since $\mathbf{i}_{\mathcal{P}}(e) \in L_{\mathcal{P}}(H)$ for any edge $e \in H$, we have $\mathbf{i}_{\mathcal{P}}(V(M)) \in L_{\mathcal{P}}(H)$ for any matching in H, and so H cannot contain a perfect matching. In the case of the above example, the partition $\mathcal{P}$ has parts A and B, and $L_{\mathcal{P}}(H)$ is the lattice of vectors (x, y) for which x is even.

The result from [65] which we use implies that any k-graph H on n vertices with minimum codegree at least $n/k + o(n)$, which does *not* contain a perfect matching must be close to a divisibility barrier; that is, H can be transformed into a divisibility barrier by the removal of $o(n^k)$ edges. However, the converse of this statement does not hold. The majority of Chapter 4 is devoted to proving the next theorem, which gives an 'if and only if' statement for the existence of a perfect matching in such a k-graph, using extremal hypergraph theory. For this we need one more definition: we say that a lattice L is $(1,-1)$-free if it does not contain any difference $\mathbf{u}_i - \mathbf{u}_j$ of two unit vectors with $i \neq j$. Here $\mathbf{u}_i$ is the vector which takes the value 1 in co-ordinate i and the value 0 in all other co-ordinates.

Theorem 1.5.2 *For any integers $k \geqslant 3$ and C_0 and any $\gamma > 0$, there exists $n_0 = n_0(k, \gamma, C_0)$ such that for any C with $k^{k^k} \leqslant C \leqslant C_0$, any $n \geqslant n_0$ divisible by k, and any k-graph H on n vertices with $\delta_{k-1}(H) \geqslant (1/k + \gamma)n$, the following statements are equivalent:*

(i) H has a perfect matching.

(ii) If

- *$\mathcal{P}$ is a partition of $V(H)$ into d parts of size greater than n/k, and*
- *$L \subseteq \mathbb{Z}^d$ is a $(1,-1)$-free edge-lattice such that any matching M in H formed of edges $e \in H$ with $\mathbf{i}_{\mathcal{P}}(e) \notin L$ has size less than C,*

then there exists a matching M' of size at most $k-2$ such that $\mathbf{i}_{\mathcal{P}}(V(H) \setminus V(M')) \in L$.

For $k = 3$ there is only one type of divisibility barrier to consider, so we can restate this result in a simpler form, showing that H has a perfect matching unless every single edge of H satisfies the divisibility condition of the divisibility barrier in the example given earlier.

Theorem 1.5.3 *For any $\gamma > 0$ there exists $n_0 = n_0(\gamma)$ such that the following statement holds. Let H be a 3-graph on $n \geqslant n_0$ vertices, such that 3 divides n and $\delta_2(H) \geqslant (1/3+\gamma)n$,* *and suppose that H does not contain a perfect matching. Then 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.*

Note that the existence of a subset $A \subseteq V(H)$ as in Theorem 1.5.3 in turn implies that H does not contain a perfect matching, so this is indeed an 'if and only if' statement as in Theorem 1.5.2.

Procedure DeterminePM of Theorem 1.5.1 proceeds by checking whether condition (ii) of Theorem 1.5.2 holds.

Procedure DeterminePM

Input: A constant $\gamma > 0$ and a k-graph H whose vertex set V has size n with $\delta_{k-1}(H) \geqslant n/k + \gamma n$.
Output: Determines whether or not H has a perfect matching.
if $|V| < n_0(k, \gamma, k^{k^k+1})$ **then**
  Test every possible perfect matching in H, and halt with appropriate output.
foreach *matching M in H of size at most k^{k^k+1}* **do**
  foreach *integer $1 \leqslant d < k$ and edge-lattice $L \subseteq \mathbb{Z}^d$* **do**
    foreach *partition $\mathcal{P}$ of V into d parts so that any edge $e \in H$ which does not intersect $V(M)$ has $\mathbf{i}_{\mathcal{P}}(e) \in L$* **do**
      if *there is no matching $M' \subseteq H$ of size at most $k-2$ such that $\mathbf{i}_{\mathcal{P}}(V \setminus V(M')) \in L$* **then**
        Output "H does not contain a perfect matching"
Output "H contains a perfect matching"

In Section 4.2 of Chapter 4 we deduce Theorem 1.5.1 from Theorem 1.5.2. To do this, we first show that Procedure DeterminePM does indeed have polynomial running time. For this it is not hard to see that the ranges of M, d, L and M' can each be computed in polynomial time, but not immediately clear that this is true of the range of $\mathcal{P}$. However, Lemma 4.2.1 shows that this is indeed the case. To complete the proof of Theorem 1.5.1 we then deduce from Theorem 1.5.2 that Procedure DeterminePM does indeed determine correctly whether H has a perfect matching.

While the main object of proving Theorem 1.5.2 is to facilitate the construction of the algorithm, we nevertheless believe it to be of independent interest. Our proof relies on the main result of [65] as well as a substantial amount of geometric, probabilistic and combinatorial theory.

1.6 Maker-Breaker games

Given a (finite) hypergraph G, the *Maker-Breaker* game on G is a 2-player game which is defined as follows. Two players, Maker and Breaker, each take turns to claim a (previously unclaimed) vertex of G. Maker wins if he claims all of the vertices of some edge of G, and Breaker wins if he prevents Maker from doing so before all of the vertices have been claimed.

For a given n, let $M(n)$ be the largest integer for which there exists an n-uniform hypergraph (an *n-graph*) G of maximum degree at most $M(n)$ such that Maker wins the Maker-Breaker game on G. The following well-known pairing argument, due to Hales and Jewett [53], shows that $M(n) > n/2$ for each n: Form an auxiliary bipartite graph B on vertex classes $V_1 = V(G)$ and V_2, where V_2 is composed of *two* copies of $E(G)$. Since the degree of every vertex in V_2 is n and the degree of every vertex in V_1 is at most n, Hall's theorem implies that B contains a matching which covers every vertex of V_2. For each $e \in E(G)$, let x_e and y_e be the vertices which were matched to copies of e in V_2. Now whenever Maker plays at one of x_e and y_e, Breaker immediately plays at the other. This prevents Maker from claiming any edge of G and so Breaker wins.

Beck [11, Open Problem 9.1] gives a number of successively weaker conjectures on the values of $M(n)$, which are collectively known as the Neighbourhood Conjecture. Gebauer [46] proved that $M(n) < 2^{n-1}/n$ for sufficiently large n, thus disproving the strongest version of the Neighbourhood Conjecture. A more recent result of Gebauer, Szabó and Tardos [47] implies that $M(n) < (1 + o(1))2^n/en$. Since the general problem seems very difficult to solve, it is hoped that the cases for small n. Trivially $f(2) = 2$. The graph G_3 (see Figure 1.1) demonstrates that $f(3) = 2$ also. Maker's winning strategy

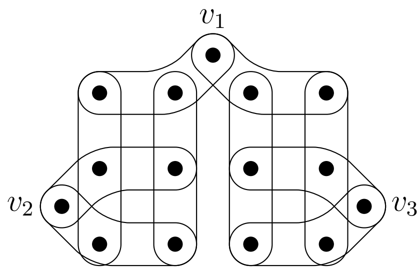

Figure 1.1: The graph G_3

is to play at v_1, and then if Breaker plays to the right of v_1 he plays at v_2; otherwise, he plays at v_3. In either case he wins quickly regardless of Breaker's next move.

We can derive from G_3 a 4-graph which has maximum degree 4, and on which Maker wins, as follows: For each edge $e \in G_3$, add two vertices x_e and y_e and replace e by the two edges $e \cup \{x_e\}$ and $e \cup \{y_e\}$. To win, Maker first claims an edge of G_3 and claims one of x_e and y_e. So $f(4) \leqslant 4$. A question posed by Imre Leader at the Workshop on Probabilistic techniques in Graph Theory, University of Birmingham (March 25, 2012) is whether there exists a 4-graph with maximum degree 3 on which Maker wins the Maker-Breaker game. Since $f(4) > 4/2 = 2$, this would show that $f(4) = 3$.

In Chapter 5 we exhibit a 4-graph G_4 which answers this question in the affirmative. Thus indeed $f(4) = 3$. An illustration of how little is known for this problem is that the exact value for $M(n)$ is unknown for any $n \geqslant 5$. The best known general bounds are the upper bound implied by [47] and the previously stated lower bound $M(n) > n/2$.

1.7 Notation

The following notation is used throughout this thesis. Additional notation required for individual chapters is included at the start of the relevant chapter.

We omit floors and ceilings whenever this does not affect the argument. We write $|G|$ for the order of a graph G, $\delta(G)$ and $\Delta(G)$ for its minimum and maximum degrees respectively and $\chi(G)$ for its chromatic number. $e(G)$ denotes the number of edges of G. The degree of a vertex $x \in V(G)$ is denoted by $d(x)$ and its neighbourhood by $N(x)$. For a spanning subgraph H of G, let $G\backslash H$ denote the graph obtained by removing the edges of H from G. On the other hand, for a set A of vertices $G[A]$ denotes the induced subgraph of G on A, and $G - A$ denotes $G[V(G)\backslash A]$.

Given disjoint $A, B \subseteq V(G)$ the number of edges with one endpoint in A and one endpoint in B is denoted by $e_G(A, B)$. $e_G(A)$ denotes $e_G(A, A)$. We write $(A, B)_G$ for the bipartite subgraph of G with vertex classes A and B whose edges are precisely those edges in G with one endpoint in A and the other in B. Often we will write (A, B), for example, if this is unambiguous.

Recall from Section 1.4.3 that $\alpha \ll \beta \ll \gamma$ means that we can choose the constants γ, β, α from right to left. log denotes the natural logarithm, and we write $\log^a n$ for $(\log n)^a$. We write $[r]$ to denote the set of integers from 1 to r.

1.8 Probabilistic Tools

We will need the following inequality known as the Chernoff bound, as applied to binomial and hypergeometric random variables. We briefly give the standard definitions. Recall that the binomial random variable $Bin(n, p)$ with parameters (n, p) is the sum of n independent copies of the $\{0, 1\}$-valued variable A with $\mathbb{P}(A = 1) = p$. The hypergeometric random variable X with parameters (N, m, n) is defined as $X = |T \cap S|$, where $S \subseteq [N]$ is a fixed set of size m, and $T \subseteq [N]$ is a uniformly random set of size n. If $m = pN$ then both variables have mean pn.

Lemma 1.8.1 *[60, Corollary 2.3, Theorem 2.8 and Theorem 2.10] Suppose X is a sum of independent Bernoulli random variables (e.g., a binomial random variable). Then*

(i) *If $0 < a < 3/2$, then $\mathbb{P}(|X - \mathbb{E}X| \geqslant a\mathbb{E}X) \leq 2\exp\left(-\frac{a^2}{3}\mathbb{E}X\right)$.*

If X is a binomial random variable $Bin(n, p)$, then in addition

(ii) *If* $t \geqslant 7np$, *then* $\mathbb{P}(X \geqslant t) \leqslant e^{-t}$.

We will also need to deal with hypergeometric random variables and sums of the same, for which the following result will be useful.

Corollary 1.8.2 *Suppose* X *is a sum of independent hypergeometric random variables and* $0 < a < 3/2$. *Then* $\mathbb{P}(|X - \mathbb{E}X| \geqslant a\mathbb{E}X) \leq 2\exp\left(-\frac{a^2}{3}\mathbb{E}X\right)$.

Proof. We follow Remark 2.11 of [60]. By Lemma 1 of [119], each of the hypergeometric random variables may be expressed as a sum of independent (but not identically distributed) Bernoulli random variables. Hence Lemma 1.8.1(i) implies the result. □

We will also need two forms of the well-known Azuma-Hoeffding inequality. A sequence $(X_0, \ldots, X_n)$ of random variables is a *martingale* if $\mathbb{E}(|X_j|)$ is finite and $\mathbb{E}(X_j | X_0, \ldots, X_{j-1}) = X_{j-1}$ for any $1 \leqslant j \leqslant n$.

Theorem 1.8.3 *[60, Theorem 2.25] Let* $A_0, A_1, \ldots, A_n$ *be a martingale such that* $|A_i - A_{i-1}| \leqslant c$ *for every* $1 \leqslant i \leqslant n$. *Then for any* $t > 0$ *and any* $0 \leqslant i \leqslant n$,

$$\mathbb{P}(|A_i - A_0| > tc) \leqslant 2\exp\left(-\frac{t^2}{2i}\right).$$

The following alternative form will be used in Chapter 3.

Lemma 1.8.4 *[109, Proposition 1.1] Let* $X_1, \ldots, X_n$ *be random variables taking values in* $[0, 1]$, *such that for each* $1 \leqslant k \leqslant n$,

$$\mathbb{E}[X_k \mid X_{k-1}, \ldots, X_1] \leqslant a_k.$$

Let $\mu := \sum_{k=1}^{n} a_k$. *Then for any* $0 < \delta \leqslant 1$,

$$\mathbb{P}\left[\sum_{k=1}^{n} X_k > (1+\delta)\mu\right] \leqslant e^{-\frac{\delta^2 \mu}{3}}.$$

□

The following lemma will be very useful in conjunction with the second form of Azuma's inequality.

Lemma 1.8.5 *[89, Lemma 2.1] Suppose that $1/k \ll p, 1-p, \varepsilon$, that $n \geqslant k^3/6$, and that $X \sim Bin(n, p)$. Then for any $0 \leqslant r \leqslant k-1$,*

$$\frac{1-\varepsilon}{k} \leqslant \mathbb{P}(X \equiv r \mod k) \leqslant \frac{1+\varepsilon}{k}.$$

□

Finally, we require the following well-known expectation bound, whose proof we include here for completeness.

Lemma 1.8.6 *Suppose that X and Y are integer-valued random variables on a probability space Ω and that B is an event on Ω which has positive probability, such that for each $x, y \in \mathbb{Z}$, $\mathbb{P}[X = x \mid B \cap \{Y = y\}] = \mathbb{P}[X = x \mid Y = y]$. Then*

$$\mathbb{E}[X \mid B] \leqslant \max_{y \in \mathbb{Z}} \mathbb{E}[X \mid Y = y].$$

Proof. Note that for each $x \in \mathbb{Z}$,

$$\mathbb{P}[X = x \mid B] = \sum_{y \in \mathbb{Z}} \mathbb{P}[\{X = x\} \cap \{Y = y\} \mid B].$$

Further,

$$\begin{aligned}
\mathbb{P}[\{X = x\} \cap \{Y = y\} \mid B] &= \frac{\mathbb{P}[\{X = x\} \cap \{Y = y\} \cap B]}{\mathbb{P}[B]} \\
&= \frac{\mathbb{P}[\{X = x\} \cap \{Y = y\} \cap B]}{\mathbb{P}[\{Y = y\} \cap B]} \cdot \frac{\mathbb{P}[\{Y = y\} \cap B]}{\mathbb{P}[B]} \\
&= \mathbb{P}[X = x \mid Y = y] \cdot \mathbb{P}[Y = y \mid B].
\end{aligned}$$

Hence,

$$\begin{aligned}\mathbb{E}[X \mid B] &= \sum_{x\in\mathbb{Z}} x\mathbb{P}[X = x \mid B] = \sum_{x\in\mathbb{Z}} x \sum_{y\in\mathbb{Z}} \mathbb{P}[X = x \mid Y = y] \cdot \mathbb{P}[Y = y \mid B] \\ &= \sum_{y\in\mathbb{Z}} \mathbb{P}[Y = y \mid B] \sum_{x\in\mathbb{Z}} x\mathbb{P}[X = x \mid Y = y] = \sum_{y\in\mathbb{Z}} \mathbb{P}[Y = y \mid B] \cdot \mathbb{E}[X \mid Y = y] \\ &\leqslant \max_{y\in\mathbb{Z}} \mathbb{E}[X \mid Y = y].\end{aligned}$$

□

Chapter 2

Edge-disjoint Hamilton cycles in random graphs

2.1 Notation

In this chapter $N_G(A)$ is always taken to be the external neighbourhood of A, i.e., $N_G(A) = (\bigcup_{x \in A} N(x)) \backslash A$.

Since we are aiming to prove a result with high probability, we may and do assume throughout the chapter that n is always sufficiently large for our estimates to hold. Further we assume large quantities to be integers, whenever this does not have a significant effect on the argument.

2.2 Approximate result

2.2.1 Outline

The goal of this section is to prove Theorem 1.3.4. Roughly speaking, the proof proceeds as follows:

(1) First we choose $0 < \varepsilon \ll \eta$ and remove a random subgraph G_{thin} (say) of density εp from $G_{n,p}$, and call the remaining graph G_{dense}. G_{thin} will be a pseudorandom

graph, which is important in Step 4 below.

(2) Next we apply Tutte's theorem to find an r-regular subgraph G_{reg} of G_{dense}, with $r = (1-\varepsilon)np$.

(3) Using a counting argument, one can show that most 2-factors of G_{reg} have few cycles. We remove such 2-factors one by one to obtain a set of $(1-\varepsilon)r/2$ edge-disjoint 2-factors F_i which cover most edges of G_{reg} (and thus of $G_{n,p}$).

(4) We now use the edges of G_{thin} to transform each F_i into a Hamilton cycle. More precisely, for each i in turn, we swap some edges between G_{thin} and F_i in such a way that the modified 2-factor is in fact a Hamilton cycle. For the next 2-factor, we use the 'current' version of G_{thin}. The fact that F_i has few cycles means that we do not need to swap out many edges of G_{thin} for this, and so G_{thin} retains the pseudorandomness properties which allow us to perform this step.

The outline of the proof is as follows: We begin in Section 2.2.2 by stating and applying some large deviation bounds on the number of edges in certain subgraphs of $G_{n,p}$, which we will use later on. We will split $G_{n,p}$ into G_1 (which plays the role of G_{dense} above) and G_2 (which plays the role of G_{thin}). Then in Section 2.2.3, we will use Tutte's r-factor theorem to show that one may find a regular subgraph of G_1 whose degree is close to np_1 (the average degree of G_1). This regular subgraph plays the role of G_{reg}. In Section 2.2.4 we show that this subgraph can almost be decomposed into 2-factors in such a way that each 2-factor has relatively few cycles. Finally in Section 2.2.5 we convert each of these 2-factors into Hamilton cycles, using the edges of G_2 (along with any edges of G_1 which were not included in our collection of 2-factors). This is achieved using an appropriate variant of the well-known rotation-extension technique, first introduced by Pósa [103]. The fact that each of the 2-factors originally had few cycles will allow us to place an upper bound on the number of edges needed to perform the conversions, and thus to show that the process can be completed before all of the edges of G_2 have been used up.

2.2.2 Large deviation bounds

We can use Lemma 1.8.1 to deduce some facts about the number of edges between subsets of vertices of a random graph, as follows:

Lemma 2.2.1 *Let $G \sim G(n,p)$. Then with probability at least $1 - 1/n^2$, for any disjoint $A, B \subseteq [n]$, the following properties hold: Let $a = |A|$ and $b = |B|$. Then*

(i) *If $\left(\frac{1}{a} + \frac{1}{b}\right)\frac{\log n}{p} \geqslant \frac{7}{2}$, then $e_G(A,B) \leqslant 2(a+b)\log n$, and*

(ii) *If $\left(\frac{1}{a} + \frac{1}{b}\right)\frac{\log n}{p} \leqslant \frac{7}{2}$, then $e_G(A,B) \leqslant 7abp$.*

Proof. (i) Let $X = e_G(A,B)$ and let $t = 2(a+b)\log n$. Since $X \sim Bin(ab,p)$, we have that $t \geqslant 7abp = 7\mathbb{E}X$. If $a+b < 3$ then the result is trivial; otherwise, by Lemma 1.8.1(ii) we have that $\mathbb{P}(X \geqslant t) \leqslant e^{-t} = \left(\frac{1}{n^a n^b}\right)^2 \leqslant \frac{1}{n^3}\left(\frac{1}{n^a n^b}\right)$, and a union bound immediately gives the result.

(ii) Similarly, we have $\mathbb{P}(X \geqslant 7abp) \leqslant e^{-7abp} \leqslant e^{-t}$ and the result follows. $\square$

In an exactly similar way, we can show that

Lemma 2.2.2 *Let $G \sim G_{n,p}$. Then with probability at least $1 - 1/n^2$, for every $A \subseteq [n]$ the following properties hold: Let $a = |A|$. Then*

(i) *If $\frac{\log n}{ap} \geqslant \frac{7}{4}$, then $e_G(A) \leqslant 2a\log n$, and*

(ii) *If $\frac{\log n}{ap} \leqslant \frac{7}{4}$, then $e_G(A) \leqslant \frac{7a^2p}{2}$.*

For larger sets, Lemma 2.2.3 gives a more precise result. Note that we allow $\alpha, \beta \to 0$ in the statement.

Lemma 2.2.3 *Let $G \sim G_{n,p}$. Then whp, for all pairs $A, B \subseteq [n]$ of disjoint sets the following property holds: Let $\alpha = |A|/n$ and $\beta = |B|/n$, and suppose that $\alpha\beta np \geqslant 700$. Then*

$$\frac{13}{14}\alpha\beta n^2 p \leqslant e_G(A, B) \leqslant \frac{15}{14}\alpha\beta n^2 p.$$

Proof. $e_G(A, B) \sim Bin(\alpha\beta n^2, p)$, so by Lemma 1.8.1(i),

$$\mathbb{P}\left(e_G(A, B) < \frac{13}{14}\alpha\beta n^2 p\right) \leqslant 2e^{-\frac{\alpha\beta n^2 p}{3\cdot 14^2}} \leqslant 2e^{-\frac{700n}{588}} \leqslant 2e^{-n\log(3.1)} = \frac{2}{(3.1)^n}.$$

A union bound over all 3^n possibilities (each vertex can lie either in A, in B or in neither, but not in both) now gives the result. The right-hand inequality follows in an exactly similar manner using the opposite tail estimate. $\square$

2.2.3 Regular subgraphs of a random graph

Let $p_0 = p_0(n)$ be such that $p_0 \geqslant C \log n/n$ and let $G_0 \sim G_{n,p_0}$. The graph G_1 is generated by including each edge of G_0 independently at random with probability $(1-\frac{\eta}{4})$. We define $G_2 := G_0 \backslash G_1$. Note that $G_1 \sim G_{n,p_1}$ and $G_2 \sim G_{n,p_2}$, where $p_1 = (1 - \frac{\eta}{4})p_0$ and $p_2 = \frac{\eta p_0}{4}$. (Of course, the distributions of these random graphs are not independent of each other.)

We first show that G_1 contains a regular subgraph of degree at least

$$r_1 = np_0\left(1 - \frac{3\eta}{4}\right),$$

where r_1 is taken to be even.

We do this by using a theorem of Tutte: Let G be an arbitrary graph, r a positive integer and suppose that S, T, U is a partition of $V(G)$. Then define

$$R_r(S, T) = \sum_{v\in T} d(v) - e_G(S, T) + r(|S| - |T|)$$

and let $Q_{r,G}(S,T)$ be the number of *odd* components of $G[U]$, where a component C is odd if and only if $r|C| + e_G(C,T)$ is odd. (In our case it will often suffice to bound $Q_{r,G}(S,T)$ simply by the total number of components of $G[U]$.) Where it is clear from the context which graph G is being referred to, we will usually simply write $Q_r(S,T)$.

Theorem 2.2.4 (Tutte [117]) *Let r be a positive integer. A graph G contains an r-factor if and only if $R_r(S,T) \geqslant Q_r(S,T)$ for every partition S, T, U of $V(G)$.*

In order to apply Theorem 2.2.4 we will need an upper bound on $Q_r(S,T)$. We do this by observing that if $G[U]$ has many components, then it must contain a large *isolated set* of vertices; that is, a set $A \subseteq U$ such that $e_G(A, U \backslash A) = 0$. This becomes useful when looking at a random graph, since (as we will prove in Lemma 2.2.7) it follows that whp A has many neighbours in $S \cup T$. This gives a lower bound on $|S \cup T|$ in terms of $Q_r(S,T)$, and thus gives an upper bound on $Q_r(S,T)$ in terms of $|S \cup T| = |S| + |T|$.

Lemma 2.2.5 *Let G be a graph with v components. Then for any $v' \leqslant \frac{v}{2}$, there exists a set $W \subseteq V(G)$ which is isolated in G, such that $v' \leqslant |W| \leqslant \max\{2v', \frac{2|G|}{v}\}$.*

Proof. Call a component C of G *small* if its order is at most v', and *large* otherwise. Suppose first that the union of all small components of G also has order at most v'. Then the number of small components is at most v', and hence there are at least $v - v' \geqslant \frac{v}{2}$ large components. So one of these components must have order at most $\frac{2|G|}{v}$, and we can set W to be this component.

On the other hand, if the sum of the orders of small components is greater than v' then we can form W by starting with $\emptyset$ and adding small components one by one until $|W| \geqslant v'$. Now since the last component added has size at most v', we have that $|W| \leqslant 2v'$. $\square$

Given a graph G, define the *boundary* $B_G(A)$ of a set $A \subseteq V(G)$ to be the set of vertices which are adjacent (in G) to some vertex of A, but are not themselves elements of A. We will use the following two lemmas to give a lower bound on $|B_{G_1}(A)|$. Set

$$w_0 = \frac{np_0}{\log n}.$$

So using the notation and assumptions of Theorem 1.3.4, $w_0 \geqslant C$. Our results will hold provided that w_0 is sufficiently large depending on η, which we will assume throughout.

Lemma 2.2.6 *Whp,*

(i) $\delta(G_1) \geqslant (1 - \frac{\eta}{2})np_0$,

(ii) $\Delta(G_1) \leqslant np_0$, *and*

(iii) $\delta(G_2) \geqslant \frac{\eta np_0}{5}$.

Proof. Note that for a vertex x of G_1, $d(x) \sim Bin(n-1, p_1)$ and $\mathbb{E}(d(x)) = (1 - \frac{\eta}{4})(n-1)p_0$. By Lemma 1.8.1(i), we have

$$\mathbb{P}\left(d(x) \leqslant \left(1 - \frac{\eta}{2}\right) np_0\right) \leqslant 2e^{-(\frac{\eta}{5})^2 \frac{np_0}{3}} = 2n^{-\frac{\eta^2 w_0}{75}} \leqslant \frac{1}{n^2},$$

and a union bound gives the result. The bound on the maximum degree follows similarly, as does that on the minimum degree of G_2. □

Lemma 2.2.7 *The following holds whp: Let H be a spanning subgraph of G_0, and let $A \subseteq [n]$ be nonempty. Let $\delta_A = \min_{x\in A} d_H(x)$. Then setting $a = |A|$ and $b = |B_H(A)|$, the following properties hold:*

(i) *If $\frac{\log n}{ap_0} \geqslant \frac{7}{2}$, then $b \geqslant \frac{a(\delta_A - 6\log n)}{2\log n}$. In particular, if $H = G_1$, then $b \geqslant a$.*

(ii) *If $\frac{\log n}{ap_0} \leqslant \frac{7}{2}$, then $3a + b \geqslant \frac{\delta_A}{7p_0}$. In particular, if $H = G_1$, then $3a + b \geqslant \frac{n}{14}$.*

Proof. Let $B = B_H(A)$.

(i) Then

$$a\delta_A \leqslant \sum_{x \in A} d_H(x) = e_H(A, B) + 2e_H(A) \leqslant e_{G_0}(A, B) + 2e_{G_0}(A) \tag{2.2.8}$$

which by Lemmas 2.2.1(i) and 2.2.2(i) is at most $2(a+b)\log n + 4a \log n$. The final part follows from Lemma 2.2.6(i).

(ii) We claim that $e_{G_0}(A, B) \leqslant 7ap_0(a+b)$. Indeed, if $\left(\frac{1}{a} + \frac{1}{b}\right)\frac{\log n}{p_0} \leqslant \frac{7}{2}$, then by Lemma 2.2.1(ii), $e_{G_0}(A, B) \leqslant 7abp_0 \leqslant 7ap_0(a+b)$. On the other hand, if $\left(\frac{1}{a} + \frac{1}{b}\right)\frac{\log n}{p_0} \geqslant \frac{7}{2}$, then $e_{G_0}(A, B) \leqslant 2(a+b)\log n \leqslant 7ap_0(a+b)$. Similarly, Lemma 2.2.2 implies that $e_{G_0}(A) \leqslant 7a^2 p_0$. Now

$$a\delta_A \overset{(2.2.8)}{\leqslant} e_{G_0}(A, B) + 2e_{G_0}(A) \leqslant 7ap_0(a+b) + 14a^2 p_0$$

and the result follows immediately. Again the final part follows by Lemma 2.2.6(i). □

We will later use the above lemmas to show that taking successive neighbourhoods of a set will give us a set of size linear in n in a reasonably short time. For now, they allow us to give a bound on the size of $Q_r(S, T)$ in terms of $|S|$ and $|T|$.

Lemma 2.2.9 *In the graph G_1, whp, for any partition S, T, U of $[n]$, $Q_{r_1}(S, T) \leqslant 150(|S| + |T|)$.*

Proof. Let v be the number of components of $G_1[U]$. Consider first the case when $150 \leqslant v \leqslant \frac{n}{75}$. Then by applying Lemma 2.2.5 to the graph $G_1[U]$, we have that there exists a set $W \subseteq [n]$, isolated in $G_1[U]$, such that $\frac{v}{2} \leqslant |W| \leqslant \frac{n}{75}$. Now by Lemma 2.2.7 with $A = W$,

$$|B_{G_1}(W)| \geqslant \min\{|W|, \frac{n}{14} - 3|W|\} \geqslant |W| \geqslant \frac{v}{2}.$$

But if W is isolated in $G_1[U]$, then its boundary in G_1 lies entirely in $S \cup T$. So $\frac{v}{2} \leqslant |B_{G_1}(W)| \leqslant |S| + |T|$, and hence $Q_{r_1}(S,T) \leqslant v \leqslant 2(|S| + |T|)$.

Now consider the case $v \geqslant \frac{n}{75}$. Setting $v' = \frac{n}{150}$ in Lemma 2.2.5, we have a set W, isolated in $G_1[U]$, such that $\frac{n}{150} \leqslant |W| \leqslant \frac{n}{75}$. Again the boundary of W in G_1 has size at least $|W| \geqslant \frac{n}{150}$, and hence $\frac{n}{150} \leqslant |S| + |T|$. So $150(|S| + |T|) \geqslant n$ and the result holds trivially, since $Q_{r_1}(S,T)$ cannot be greater than n.

Finally if $v \leqslant 150$, then $Q_{r_1}(S,T) \leqslant 150 \leqslant 150(|S| + |T|)$ unless we are in the trivial case $|S| = |T| = 0$. But if S, T are both empty then $U = [n]$ and $G_1[U] = G_1$, which whp has only one component (this follows from the fact that a random graph $G_{n,p}$ with $pn \geq 2 \log n$ is whp connected, as proved in [37]). Since r_1 is even this component cannot be odd, whence $Q_{r_1}(S,T) = 0$. $\square$

Lemma 2.2.10 *In the graph G_1, whp, we have that $R_{r_1}(S,T) \geqslant Q_{r_1}(S,T)$ for any partition S, T, U of $[n]$.*

Proof. Let d_S, d_T be the average degrees of the vertices in S, T respectively. Let $\rho = \frac{|T|}{|S|}$, and $s = |S|$. We consider the following cases:

Case 1: $\rho \leqslant \frac{1}{2}$. Then since $e_{G_1}(S,T) \leqslant d_T|T|$ and $|S| \geqslant 2|T|$, we have

$$R_{r_1}(S,T) \geqslant r_1(|S| - |T|) \geqslant \frac{r_1}{3}(|S| + |T|)$$

and for sufficiently large n, $\frac{r_1}{3}(|S| + |T|) \geqslant 150(|S| + |T|) \geqslant Q_{r_1}(S,T)$.

Case 2: $\rho \geqslant 4$. Observe that by the definition of r_1 and by Lemma 2.2.6, we have

$$d_T - r_1 \geqslant \left(1 - \frac{\eta}{2}\right) np_0 - \left(1 - \frac{3\eta}{4}\right) np_0 = \frac{\eta np_0}{4} \tag{2.2.11}$$

and

$$d_S - r_1 \leqslant np_0 - \left(1 - \frac{3\eta}{4}\right) np_0 = \frac{3\eta np_0}{4}.$$

Now since $e_{G_1}(S,T) \leqslant d_S|S|$ and $|T| \geqslant 4|S|$,

$$\begin{aligned} R_{r_1}(S,T) &\geqslant d_T|T| - d_S|S| + r_1(|S| - |T|) = (d_T - r_1)|T| - (d_S - r_1)|S| \\ &\geqslant \frac{\eta n p_0}{4}(|T| - 3|S|) \geqslant \frac{\eta n p_0}{20}(|S| + |T|) \end{aligned}$$

which again is at least $Q_{r_1}(S,T)$ for sufficiently large n.

Case 3: $\frac{1}{2} \leqslant \rho \leqslant 4$ and $\left(\frac{1}{s} + \frac{1}{\rho s}\right) \frac{\log n}{p_1} \geqslant \frac{7}{2}$. In this case by Lemma 2.2.1 we have that $e_{G_1}(S,T) \leqslant 2(\rho+1)s\log n$, and so it suffices to prove that

$$\rho s(d_T - 2\log n - r_1 - 150) + s(r_1 - 2\log n - 150) \geqslant 0. \tag{2.2.12}$$

(2.2.12) holds if $d_T - 2\log n - r_1 - 150 \geqslant 0$ and $r_1 - 2\log n - 150 \geqslant 0$. But the latter inequality holds since $r_1 = \left(1 - \frac{3\eta}{4}\right) np_0 = \left(1 - \frac{3\eta}{4}\right) w_0 \log n$, and the former since

$$d_T - r_1 \overset{(2.2.11)}{\geq} \frac{\eta n p_0}{4} = \frac{\eta w_0 \log n}{4} \geqslant 3\log n,$$

as $w_0 \geqslant \frac{12}{\eta}$.

Case 4: $\frac{1}{2} \leqslant \rho \leqslant 4$ and $(\frac{1}{s} + \frac{1}{\rho s})\frac{\log n}{p_1} \leqslant \frac{7}{2}$ and $\rho s \leqslant \frac{n}{30}$. In this case by Lemma 2.2.1 we have that $e_{G_1}(S,T) \leqslant 7\rho s^2 \left(1 - \frac{\eta}{4}\right) p_0$, and so it suffices to prove that

$$\rho s(d_T - r_1 - 150) + s(r_1 - 7\rho s \left(1 - \frac{\eta}{4}\right) p_0 - 150) \geqslant 0,$$

and hence it suffices that $d_T - r_1 - 150 \geqslant 0$ and $r_1 - 7\rho s(1 - \frac{\eta}{4})p_0 - 150 \geqslant 0$. But the former inequality holds as before, and the latter since

$$r_1 - 150 \geqslant \frac{14}{15} r_1 = \frac{14}{15}\left(1 - \frac{3\eta}{4}\right) np_0 \geqslant 28\rho s \left(1 - \frac{3\eta}{4}\right) p_0 \geqslant 7\rho s \left(1 - \frac{\eta}{4}\right) p_0.$$

Case 5: $\frac{1}{2} \leqslant \rho \leqslant 4$ and $\rho s \geqslant \frac{n}{30}$. Note that we still have $\frac{s}{n} \leqslant \frac{1}{\rho+1}$, since S,T are disjoint. Now by Lemma 2.2.3,

$$e_{G_1}(S,T) \leqslant \frac{15}{14}\rho s^2\left(1-\frac{\eta}{4}\right)p_0 \leqslant \frac{15}{14}\frac{\rho}{\rho+1}sn\left(1-\frac{\eta}{4}\right)p_0 \leqslant \frac{6}{7}sn\left(1-\frac{\eta}{4}\right)p_0.$$

So it suffices to prove that

$$\rho sn\left(1-\frac{\eta}{2}\right)p_0 - \frac{6}{7}sn\left(1-\frac{\eta}{4}\right)p_0 + (1-\rho)sn\left(1-\frac{3\eta}{4}\right)p_0 - 150(\rho+1)s \geqslant 0,$$

i.e., that $\frac{\eta\rho}{4}+(1-\frac{3\eta}{4})-\frac{6}{7}(1-\frac{\eta}{4})-\frac{150(\rho+1)}{np_0} \geqslant 0$, which is true if η is not too large (which we can assume without loss of generality). □

Corollary 2.2.13 *Whp, G_1 contains an even-regular subgraph of degree $r_1 = (1-\frac{3\eta}{4})np_0$.*

Proof. This follows immediately from Theorem 2.2.4 and Lemma 2.2.10. □

2.2.4 2-factors of regular subgraphs of a random graph

In this section we will show that any even-regular subgraph of G_0 of sufficiently large degree contains a 2-factor with fewer than $\frac{\kappa n}{\log n}$ cycles, where

$$\kappa = 2\log\left(\frac{16}{\eta}\right). \tag{2.2.14}$$

It will follow immediately that we can decompose almost all of our regular subgraph into 2-factors with at most this many cycles. Roughly, our strategy will be to show that the number of 2-factors with many cycles in the original graph is rather small; smaller, in fact, than the minimum number of 2-factors which an even-regular graph of degree r_1 must contain. Let $k_0 = \frac{\kappa n}{\log n}$.

Lemma 2.2.15 *Whp, for any r-regular subgraph $H \subseteq G_0$ with $r \geqslant 2np_0e^{-\frac{\kappa}{2}}$, H contains a 2-factor with at most k_0 cycles.*

To prove Lemma 2.2.15 we will need a number of further lemmas. We use Lemmas 2.2.16 and 2.2.17 to bound the number of 2-factors in $G_{n,p}$ with many cycles, while Lemma 2.2.20 gives a bound on the total number of 2-factors in H.

Lemma 2.2.16 *For any k and for $n \geqslant 3k$, we have*

$$\sum \prod_{i=1}^{k} \frac{1}{a_i} \leqslant \frac{k}{n}(\log n)^{k-1},$$

where the sum is taken over all ordered k-tuples $(a_1, a_2, \ldots, a_k)$ such that $a_1 + \ldots + a_k = n$ and $a_i \geqslant 3$ for each $i \in [n]$.

Proof. We proceed by induction on k. The case $k = 1$ is trivial since both sides equal $\frac{1}{n}$. Supposing that the result holds for $k-1$, we have

$$\sum \prod_{i=1}^{k} \frac{1}{a_i} = \sum_{a_k=3}^{n-3(k-1)} \frac{1}{a_k} \sum \prod_{i=1}^{k-1} \frac{1}{a_i},$$

where again the second sum on the right-hand side is taken over all ordered $(k-1)$-tuples $(a_1, \ldots, a_{k-1})$ such that $a_1 + \ldots + a_{k-1} = n - a_k$ and $a_i \geqslant 3$ for all $i \in [k-1]$. By induction, this is bounded above by

$$\begin{aligned}
&\sum_{a_k=3}^{n-3(k-1)} \frac{1}{a_k} \frac{k-1}{n-a_k}(\log(n-a_k))^{k-2} \\
&= \frac{k-1}{n} \sum_{a_k=3}^{n-3(k-1)} \left(\frac{1}{a_k} + \frac{1}{n-a_k}\right)(\log(n-a_k))^{k-2} \\
&\leqslant \frac{k-1}{n}\left((\log n)^{k-2}\left(\sum_{a_k=3}^{n-3} \frac{1}{a_k}\right) + \sum_{a_k=3}^{n-3} \frac{1}{n-a_k}(\log(n-a_k))^{k-2}\right) \\
&\leqslant \frac{k-1}{n}\left((\log n)^{k-1} + \frac{1}{k-1}(\log n)^{k-1}\right) = \frac{k}{n}(\log n)^{k-1},
\end{aligned}$$

where the last inequality follows from the fact that $\log n = \int_1^n \frac{1}{x}\,dx$ and $\frac{1}{k-1}(\log n)^{k-1} = \int_1^n \frac{1}{x}(\log x)^{k-2}\,dx$. $\square$

Lemma 2.2.17 *Let $G \sim G_{n,p}$. Then whp, for any $k \geqslant \log n$ the number A_k of 2-factors in G with at least k cycles satisfies*

$$A_{k+1} < \frac{n!(\log n)^{2k}p^n}{k!2^k}.$$

Proof. Note that it suffices to show that if A'_k is the number of 2-factors in G with *exactly* k cycles, then

$$\mathbb{E}(A'_k) \leqslant \frac{(n-1)!(\log n)^{k-1}p^n}{(k-1)!2^k}. \tag{2.2.18}$$

Indeed, we then have

$$\begin{aligned}\mathbb{E}(A_{k+1}) &= \sum_{i=k+1}^{\frac{n}{3}} \mathbb{E}(A'_i) \leqslant \sum_{i=k+1}^{n} \frac{(n-1)!(\log n)^{i-1}p^n}{(i-1)!2^i} \\ &\leqslant \sum_{i=k+1}^{n} \frac{(n-1)!(\log n)^k p^n}{k!2^i} \leqslant \frac{(n-1)!(\log n)^k p^n}{k!2^k}\end{aligned}$$

and Markov's inequality implies that

$$\mathbb{P}\left(A_{k+1} \geqslant \frac{n!(\log n)^{2k}p^n}{k!2^k}\right) \leqslant \frac{1}{n(\log n)^k} \leqslant \frac{1}{n^2}.$$

A union bound now gives that whp the result holds for all $\log n \leqslant k \leqslant \frac{n}{3}$.

To prove (2.2.18), it suffices to show that K_n contains at most $\frac{(n-1)!(\log n)^{k-1}}{(k-1)!2^k}$ 2-factors with exactly k cycles. We can count these as follows: Define an *ordered 2-factor* to be a 2-factor together with an ordering of its cycles. We can count the number of ordered 2-factors by first choosing some k-tuple $(a_1, a_2, \ldots, a_k)$ and counting those ordered 2-

factors whose cycles have lengths $a_1, a_2, \ldots, a_k$ in that order. This can be done by simply ordering $V(G)$ and placing vertices 1 to a_1 in the first cycle, vertices a_1+1 to a_1+a_2 in the second, etc. This procedure will count each ordered 2-factor of the appropriate type $(2a_1)(2a_2)\cdots(2a_k)$ times, and hence the number of these ordered 2-factors is $\frac{n!}{2^k a_1 a_2 \cdots a_k}$. Summing over all valid k-tuples, we have that the total number of ordered 2-factors with k cycles is

$$\sum \frac{n!}{2^k} \prod_{i=1}^{k} \frac{1}{a_i} \leqslant \frac{n!}{2^k} \frac{k}{n} (\log n)^{k-1}$$

by Lemma 2.2.16. But the number of ordered 2-factors (with k cycles) is simply $k!$ times the total number of such 2-factors, and the result follows immediately. □

We now need a lower bound on the total number of 2-factors. To do this we use the Egorychev-Falikman-Waerden theorem along with a well known consequence (see e.g. the proof of Lemma 2 in [42]).

Theorem 2.2.19 (Egorychev [35], Falikman [40]) *Let $r \leqslant n$ be positive integers and let B be an r-regular bipartite graph with vertex classes of size n. Then the number of perfect matchings of B is at least $(\frac{r}{n})^n n!$.*

Lemma 2.2.20 *Let r be even, and let H be an r-regular graph on n vertices. Then H contains at least $(\frac{r}{2n})^n n!$ 2-factors.*

Proof. It is easy to see that the number of perfect matchings of a d-regular bipartite graph B with vertex classes of size n equals the permanent of the incidence matrix of B. So we take an orientation of H in which every vertex has in- and out-degree $\frac{r}{2}$. Form a bipartite graph B whose vertex classes X, Y are each copies of $V(H)$, with an edge xy for each $x \in X$, $y \in Y$ such that $\vec{xy}$ is an edge of the orientation of H. Now B is $\frac{r}{2}$-regular and hence by Theorem 2.2.19 has at least $\left(\frac{r}{2n}\right)^n n!$ perfect matchings. But any perfect matching in B yields a 2-factor in H, and distinct matchings yield distinct 2-factors. □

Proof of Lemma 2.2.15. It suffices to show that whp $A_{k_0+1} < (\frac{r}{2n})^n n!$, and hence by Lemma 2.2.17 it suffices that

$$\left(\frac{r}{2n}\right)^n n! \geqslant \frac{n!(\log n)^{2k_0} p_0^n}{k_0! 2^{k_0}},$$

which holds as long as

$$\left(\frac{2np_0}{r}\right)^n \leqslant \frac{2^{k_0} k_0!}{(\log n)^{2k_0}}.$$

Noting that $k_0! \geqslant (\frac{k_0}{e})^{k_0}$, it suffices that

$$\begin{aligned} n \log \frac{2np_0}{r} &\leqslant k_0(\log k_0 + \log 2 - 2 \log\log n - 1) \\ &= \frac{\kappa n}{\log n}(\log n - 3 \log\log n + \log \kappa + \log 2 - 1). \end{aligned}$$

Since $\log \kappa + \log 2 - 1 > 0$, this follows immediately from

$$\log \frac{2np_0}{r} \leqslant \frac{\kappa}{2} \leqslant \kappa \left(1 - \frac{3 \log\log n}{\log n}\right).$$

□

Corollary 2.2.21 *Let $t = \frac{1}{2}(1-\eta)np_0$. Then G_1 contains a collection of at least t edge-disjoint 2-factors, each with at most k_0 cycles.*

Proof. By Corollary 2.2.13, G_1 contains a regular subgraph H of degree $r_1 = \left(1 - \frac{3\eta}{4}\right) np_0$. By Lemma 2.2.15 (noting that H is also a regular subgraph of G_0), we can remove 2-factors with at most k_0 cycles from H one by one as long as the degree of the resulting graph remains above $2np_0 e^{-\frac{\kappa}{2}}$. Recalling by (2.2.14) that $e^{-\frac{\kappa}{2}} = \frac{\eta}{16}$, this gives us a collection of $\frac{1}{2}\left(r_1 - \frac{\eta np_0}{8}\right) \geq \frac{1}{2}(1-\eta)np_0 = t$ 2-factors, each with at most k_0 cycles. □

2.2.5 Converting 2-factors into Hamilton cycles

Applying Corollary 2.2.21 yields a collection $F_1, F_2, \ldots, F_m$ of edge-disjoint 2-factors in G_1, each with at most k_0 cycles, where

$$m = \frac{1}{2}(1-\eta)np_0 \quad \text{and} \quad k_0 = \frac{\kappa n}{\log n}. \tag{2.2.22}$$

Now we wish to convert these 2-factors into Hamilton cycles. Our proof develops ideas from Krivelevich and Sudakov [91]. Our strategy will be to show that for each F_i, we can connect the cycles of F_i into a Hamilton cycle using edges of $G_0 \backslash (F_1 \cup F_2 \cup \ldots \cup F_m)$. We do this by incorporating the cycles of F_i one by one into a long path, and then finally closing this path to a Hamilton cycle.

Definition 2.2.23 *Let $P = v_1 v_2 \ldots v_\ell$ be a path. A* rotation *of P with* pivot *v_i is the operation of deleting the edge $v_{i-1}v_i$ from P and adding the edge $v_i v_1$ to form a new path $v_{i-1}v_{i-2}\ldots v_1 v_i v_{i+1} \ldots v_\ell$ with endpoints v_{i-1} and v_ℓ. Call $v_{i-1}v_i$ the* broken edge *of the rotation, and $v_i v_1$ the* new edge.

Throughout this section all rotations will have new edges in a spanning subgraph Γ of G_0, which is edge-disjoint from P. Call a spanning subgraph F of G_0 a *broken 2-factor* if F consists of a collection of vertex-disjoint cycles together with a vertex-disjoint path, which we call the *long path* of F. The key to our proof of Theorem 1.3.4 is the following lemma. Let

$$E_0 = \frac{\log n}{\log(\frac{\eta w_0}{20})} \quad \text{and} \quad E_1 = \frac{\log n}{\log(\frac{\eta^2 w_0}{10^5})}.$$

Lemma 2.2.24 *Let P be a path in G_0. Let Γ be a spanning subgraph of G_0 whose edges are disjoint from those of P, such that*

$$e(G_2 \backslash \Gamma) \leqslant \frac{\eta^6 n^2 p_0}{10^{17}} \tag{2.2.25}$$

and

$$\delta(\Gamma) \geqslant \frac{n\eta p_0}{4}. \tag{2.2.26}$$

Then there exists a sequence of at most $2E_0 + 2E_1$ rotations which can be performed on P, using edges of Γ, to produce a new path P', such that at least one of the following holds:

(i) $|P| \geqslant \frac{\eta n}{200}$ *and the endpoints x, y of P' are joined by an edge of Γ.*

(ii) *One of the endpoints x, y of P' is joined to a vertex outside P' by an edge of Γ.*

Before we prove Lemma 2.2.24, we will first show how it is used to prove Theorem 1.3.4. Our aim will be to convert each 2-factor F_i into a Hamilton cycle H_i in turn.

For this, we use the following algorithm: Let F^* be a broken 2-factor formed by removing an edge of F_i arbitrarily. Let j be the number of *steps* performed so far during the conversion process (both on F^* and on the 2-factors which have already been converted into Hamilton cycles). Here a step is taken to mean a single application of Lemma 2.2.24 to either obtain a broken 2-factor with fewer cycles or to close a Hamilton path to a cycle. Let

$$\Gamma_j = G_0 \backslash (H_1 \cup \ldots \cup H_{i-1} \cup F^* \cup F_{i+1} \cup \ldots \cup F_m).$$

Then since $\delta(G_0) \geqslant \delta(G_1) \geqslant (1 - \frac{\eta}{2})np_0$ by Lemma 2.2.6, and

$$\Delta(H_1 \cup \ldots \cup H_{i-1} \cup F^* \cup F_{i+1} \cup \ldots \cup F_m) \leqslant 2m = (1-\eta)np_0,$$

we have that $\delta(\Gamma_j) \geqslant \frac{\eta n p_0}{4}$. Assume also that $e(G_2 \backslash \Gamma_j) \leqslant 4jE_1$, and that

$$j \leqslant k_0 m \overset{(2.2.22)}{\leqslant} \frac{\kappa n^2 p_0}{2 \log n}. \tag{2.2.27}$$

Then

$$e(G_2 \backslash \Gamma_j) \overset{(2.2.27)}{\leqslant} \frac{2\kappa n^2 p_0}{\log(\frac{\eta^2 w_0}{10^5})} \overset{(2.2.14)}{\leqslant} \frac{\eta^6 n^2 p_0}{10^{17}}. \tag{2.2.28}$$

Let P^* be the long path of F^*. If P^* is a Hamilton path, then Lemma 2.2.24 applied with $\Gamma = \Gamma_j$ and $P = P^*$ shows that after at most $2E_0 + 2E_1$ rotations we can close P^* to a Hamilton cycle H_i. We then move on to the next 2-factor F_{i+1}. If there are no 2-factors remaining (i.e., if $i = m$), then we have constructed the required set of t edge-disjoint Hamilton cycles.

Otherwise by Lemma 2.2.24, after at most $2E_0 + 2E_1$ rotations we can either join an endpoint of P^* to a vertex x outside P^*, or we can close P^* to form a cycle C^*. In the first case x will be a vertex of some cycle C_x of F^*, and we can delete one of the edges of C_x incident to x to form a new path P^{**} which incorporates C_x. We then redefine F^* to be the union of P^{**} with the remaining cycles of F_i; this is a broken 2-factor with long path P^{**}, which has one cycle fewer than before. The algorithm then proceeds to the next step.

In the second case we have that $|C^*| = |P^*| \geqslant \frac{\eta n}{200}$. Now if $|C^*| \leqslant n - \frac{\eta n}{200}$, then by Lemma 2.2.3 we have $e_{G_2}(V(C^*), [n] \backslash V(C^*)) \geqslant \frac{\eta^2 n^2 p_0}{50000}$. Since

$$e(G_2 \backslash \Gamma_j) \overset{(2.2.28)}{\leqslant} \frac{\eta^6 n^2 p_0}{10^{17}} < \frac{\eta^2 n^2 p_0}{50000},$$

there must exist an edge in Γ_j from some vertex y of C^* to a vertex outside C^*. On the other hand, if $|C^*| \geqslant n - \frac{\eta n}{200}$ then applying Lemma 2.2.7 with $H = \Gamma_j$ and $A = [n] \backslash V(C^*)$ implies the same. We then delete one of the edges of C^* incident to y and extend the resulting path as in the first case.

We run this algorithm until the last 2-factor F_m has been converted into a Hamilton cycle. Now since each step either reduces the number of cycles in a broken 2-factor or closes a Hamilton path to a Hamilton cycle, the algorithm will terminate after at most $k_0 m$ steps. It remains to justify our assumption that $e(G_2 \backslash \Gamma_j) \leqslant 4jE_1$, for each j (i.e., at each step). We can prove this by induction: $G_2 \subseteq \Gamma_0$, and since at most $2E_0 + 2E_1$

rotations are performed at each step, it follows that $e(\mathit{\Gamma}_j \backslash \mathit{\Gamma}_{j+1}) \leqslant 2E_0 + 2E_1 + 2 \leqslant 4E_1$. So $e(G_2 \backslash \mathit{\Gamma}_{j+1}) \leqslant 4jE_1 + 4E_1 = 4(j+1)E_1$, as required. □

It remains to prove Lemma 2.2.24. Our strategy will be as follows: We can assume that whenever we have an endpoint x of a path P' obtainable by fewer than $2E_0 + 2E_1$ rotations of P, then all of its neighbours lie on P' (otherwise (ii) holds). So assuming this, we try to form some large sets A, B, such that for any $a \in A, b \in B$, we can obtain a path P' with endpoints a, b. Then Lemma 2.2.3 together with (2.2.25) will allow us to close P' to a cycle.

We will (eventually) obtain the sets A, B by dividing the path P into two segments, and showing that we can perform a large number of rotations using only those pivots which lie all in one half or all in the other. This will ensure that the rotations involving the first endpoint do not interfere with those involving the second endpoint, and vice versa. In order to do this we need to show two things: Firstly, there exists a subset C_1 of the first segment of the path and a subset C_2 of the second segment of the path, such that for $i = 1, 2$, each vertex in C_i has many neighbours which also lie in C_i. In fact since we are concerned with the successors or predecessors of the neighbours rather than the neighbours themselves, we will require the neighbours to lie in the interior (taken along P) of C_i. Secondly we will show that we can force the endpoints of the path to actually lie in these subsets.

We can accomplish the latter property by showing that the subsets are sufficiently large and by performing rotations until each endpoint lies in its corresponding subset. The obvious problem with this is that as we perform these rotations, C_1 and C_2 will cease to lie in their respective segments. So instead of defining C_1, C_2 immediately, we construct a subset C of $V(P)$ with certain properties; then after rotating so that a, b lie in C, we will define C_1, C_2 to be subsets of C, and the properties of C will ensure that the vertices of each C_i have many neighbours in $int(C_i)$. Here the *interior* $int(C_i)$ of C_i is the set of elements x of C_i such that both of the vertices adjacent to x along P also lie in C_i.

We start with the following lemma, where $k = \log n$.

Lemma 2.2.29 *Let $\varepsilon = \frac{\eta}{600}$, and $P \subseteq G_0$ be a path, $n' := |P| \geqslant \frac{\eta n}{200}$. Let Γ be a spanning subgraph of G_0, edge-disjoint from P, which satisfies (2.2.25). Let $W_1, W_2, \ldots, W_k$ be a partition of P into segments whose lengths are as equal as possible. Then there exists $S \subseteq [k]$ with $|S| \geqslant (1-\varepsilon)k$, and subsets $W_i' \subseteq W_i$ for each $i \in S$ with $|int(W_i')| \geqslant (1-\varepsilon)\frac{n'}{k}$, such that for any $x \in W_i'$, and for at least $|S| - \varepsilon k$ of the sets W_j', $|N_\Gamma(x) \cap int(W_j')| \geqslant \frac{\eta p_0 n'}{20k}$.*

Proof. We start with $S = [k]$ and $W_i' = W_i$, and as long as there exists $i \in S$ and a vertex $x \in W_i'$, such that $|N_\Gamma(x) \cap int(W_j')| \leqslant \frac{\eta p_0 n'}{20k}$ for at least εk values of $j \in S$, we remove x. (In this case, call x *weakly connected* to W_j'.) Further, if at any stage there exists $i \in S$ such that $|int(W_i')| \leqslant (1-\varepsilon)\frac{n'}{k}$, then we remove i from S.

We claim that this process must terminate before $\frac{\varepsilon^2 n'}{4}$ vertices are removed. Indeed, suppose we have removed $\frac{\varepsilon^2 n'}{4}$ vertices and let R be the set of removed vertices. Now $|R| = \frac{\varepsilon^2 n'}{4}$, and so $\sum_{i=1}^k |int(W_i')| \geqslant (1 - \frac{3\varepsilon^2}{4})n'$. Hence we have $|int(W_i')| \geqslant (1-\varepsilon)\frac{n'}{k}$ for at least $\left(1 - \frac{3\varepsilon}{4}\right)k$ values of i, i.e., at most $\frac{3\varepsilon k}{4}$ indices have been removed from our original set S. So each $x \in R$ is still weakly connected to at least $\frac{\varepsilon k}{4}$ sets W_i' with $i \in S$. For each $i \in S$, let $WC(i)$ be the set of vertices $x \in R$ which are weakly connected to W_i'.

Now consider the set $S_0 = \{i \in S \mid |WC(i)| \geqslant \frac{\varepsilon^3 n'}{32}\}$. Note that if $i \in S_0$, then

$$\frac{|int(W_i')|}{n}\frac{|WC(i)|}{n} np_2 \geqslant \frac{(1-\varepsilon)\varepsilon^3 (n')^2 \eta w_0 \log n}{128kn^2} \geqslant \frac{\eta^6 w_0}{10^{16}} \geqslant 700.$$

So Lemma 2.2.3 implies that the number of edges of G_2 between $int(W_i')$ and $WC(i)$ is at least

$$\frac{13}{14}\frac{\eta p_0 |WC(i)|(1-\varepsilon)n'}{4k} \geqslant \frac{\eta p_0 |WC(i)| n'}{10k}.$$

But by the definition of $WC(i)$, Γ contains at most $\frac{\eta p_0 |WC(i)| n'}{20k}$ edges between $int(W_i')$ and $WC(i)$, and hence $G_2 \backslash \Gamma$ contains at least this many edges between $int(W_i')$ and $WC(i)$.

Observe that $\sum_{i\in S}|WC(i)| \geqslant |R|\frac{\varepsilon k}{4} = \frac{\varepsilon^3 n' k}{16}$, since each $x \in R$ is weakly connected to W_i' for at least $\frac{\varepsilon k}{4}$ values of $i \in S$. But since $\sum_{i\in S\setminus S_0}|WC(i)| \leqslant \frac{\varepsilon^3 n' k}{32}$, we have that $\sum_{i\in S_0}|WC(i)| \geqslant \frac{\varepsilon^3 n' k}{32}$. Hence $G_2\backslash\Gamma$ contains at least

$$\frac{\eta n' p_0}{20k}\sum_{i\in S_0}|WC(i)| \geqslant \frac{\eta\varepsilon^3(n')^2 p_0}{640} \geqslant \frac{\eta^3\varepsilon^3 n^2 p_0}{64\cdot 4\cdot 10^5} \geqslant \frac{\eta^6 n^2 p_0}{10^{16}}$$

edges, which would contradict (2.2.25). This proves the claim, and now we consider the sets W_i' as they are at the point at which the process terminates. It is immediate that for each $i \in S$, W_i' satisfies the requirements of the lemma. But since we have removed at most $\frac{3\varepsilon k}{4}$ indices from our original set S, we also have $|S| \geqslant (1-\varepsilon)k$. $\square$

Let $C = \bigcup_{i\in S} W_i'$ and note that $|C| \geqslant (1-\varepsilon)^2 n'$. We now need to show that the set of vertices which we can make into endpoints of P with relatively few rotations is of size at least $2\varepsilon n$. Doing this gives immediately that one of these endpoints must be an element of C. For this we use the following definition and lemma.

Definition 2.2.30 *Let P be a path and H a graph with $V(P) \subseteq V(H)$, such that P and H are edge-disjoint. Let $Q \subseteq V(H)$ and let $\tau \geqslant 1$ be an integer. Then a vertex v of P is* (H, Q, τ)-reachable *if there exists a sequence of at most τ rotations which make v into the first endpoint of P, with all of the new edges being edges of H and all of the pivots being elements of Q.*

The main reason we include the set Q in this definition is that there will be certain edges of P that we do not want to break. We achieve this by ensuring that none of the endpoints of these edges lie in Q.

Lemma 2.2.31 *Let H be a graph on n vertices, and let $P = x\dots y$ be a path on a subset of $V(H)$ which is edge-disjoint from H. Let $Q \subseteq V(P)$, and let U_τ be the set of (H, Q, τ)-reachable vertices of P. Then $|U_{\tau+1}| \geqslant \frac{1}{2}|N_H(U_\tau) \cap Q| - |U_\tau|$.*

Proof. For a vertex $v \in P$, let v^- and v^+ be the predecessor and successor, respectively, of v along P. Let $T = \{v \in N_H(U_\tau) \cap Q \mid v^-, v^+ \notin U_\tau\}$. If $v \in T$, then since neither v nor either of its neighbours on P are in U_τ, the neighbours of v are preserved by every sequence of at most τ rotations of P with pivots in Q; i.e., v^+ and v^- are adjacent to v along any path obtained from P by at most τ rotations with pivots in Q. It follows that one of v^- and v^+ must be in $U_{\tau+1}$. Indeed, starting from P, we can obtain by performing at most τ rotations a path with endpoints z and y, such that $z \in U_\tau$ and zv is an edge of H. Now by one further rotation with pivot v and broken edge either vv^+ or vv^-, we obtain a path whose endpoints are either v^+, y or v^-, y.

Now let $T^+ = \{v^+ \mid v \in T, v^+ \in U_{\tau+1}\}$ and $T^- = \{v^- \mid v \in T, v^- \in U_{\tau+1}\}$. It follows from the above that either $|T^+| \geqslant |T|/2$ or $|T^-| \geqslant |T|/2$, and both T^+ and T^- are subsets of $U_{\tau+1}$. Hence $|U_{\tau+1}| \geqslant |T|/2 \geqslant (|N_H(U_\tau) \cap Q| - 2|U_\tau|)/2$. $\square$

Note that in this section P will always be a path in G_0 and we will apply Lemma 2.2.31 with $H = \Gamma$.

Corollary 2.2.32 *Either $|U_{E_0}| \geqslant \frac{\eta n}{200}$, or some element of U_{E_0} has a neighbour in Γ lying outside P (or both).*

Proof. It suffices to show that as long as $|U_t| \leqslant \frac{\eta n}{200}$, and assuming no element of U_t has a neighbour outside P, we have that $|U_{t+1}| \geqslant \min\{\frac{\eta w_0}{20}|U_t|, \frac{\eta n}{200}\}$. We apply Lemma 2.2.7, setting $H = \Gamma$ and $A = U_t$. Now in the notation of Lemma 2.2.7, $\delta_A \geqslant \delta(H[V(P)]) \geqslant \frac{\eta n p_0}{4}$ by (2.2.26), and so we have that either (i) $|B_\Gamma(U_t)| \geqslant (\frac{\eta w_0}{8} - 3)|U_t|$, or (ii) $3|U_t| + |B_\Gamma(U_t)| \geqslant \frac{\eta n}{28}$. If (i) holds then

$$|U_{t+1}| \geqslant \frac{1}{2}|B_\Gamma(U_t)| - |U_t| \geqslant (\frac{\eta w_0}{20} + 1)|U_t| - |U_t| = \frac{\eta w_0}{20}|U_t|.$$

On the other hand if (ii) holds then

$$|U_{t+1}| \geqslant \frac{1}{2}|B_\Gamma(U_t)| - |U_t| \geqslant \frac{\eta n}{56} - \frac{5}{2}|U_t| \geqslant \frac{\eta n}{200}.$$

$\square$

Corollary 2.2.32 implies that if our path P in Lemma 2.2.24 satisfies $|P| < \frac{\eta n}{200}$, then alternative (ii) of Lemma 2.2.24 holds. So suppose that $|P| \geq \frac{\eta n}{200}$. Then we can apply Lemma 2.2.29 to obtain a set $C = \bigcup_{i\in S} W_i'$. Now since $\frac{\eta n}{200} + |C| > n' = |P|$, we have that either alternative (ii) of Lemma 2.2.24 holds, or we can obtain in at most E_0 rotations a path with endpoints a', b such that $a' \in C$. Suppose we are in the latter case. Repeating the argument for b gives us a path P''' with endpoints $a', b' \in C$ which is obtained from P by at most $2E_0$ rotations.

Call a segment W_i of P *unbroken* if none of the rotations by which P''' is obtained had their pivot in W_i. Note that each unbroken segment is still a segment of P''' in the sense that the vertices are consecutive and their adjacencies along the path are preserved. Since we have arrived at the path P''' by at most $2E_0$ rotations, there are at least $k - 2E_0$ unbroken segments W_i, and for at least $k - 2E_0 - \varepsilon k$ of these we have that $i \in S$. Noting that $E_0 \leqslant \frac{k}{10}$, we are still left with at least $\frac{3k}{5}$ unbroken segments W_i for which $i \in S$. Let us relabel these segments W_i according to their order along P''', and take $C_1 = \bigcup_{i\leqslant \frac{3k}{10}} W_i'$ and $C_2 = \bigcup_{i>\frac{3k}{10}} W_i'$. Note that for any $x \in C$ (and in particular for $x \in C_1$ and for a'),

$$|N_\Gamma(x) \cap int(C_1)| \geqslant \frac{\eta p_0 n'}{20k}\left(\frac{3k}{10} - \varepsilon k\right) \geqslant \frac{\eta p_0 n'}{70} \geqslant \frac{\eta^2 n p_0}{14000}. \qquad (2.2.33)$$

Now let x_0 be a vertex separating C_1, C_2 along P''', and let x_0 divide P''' into paths $P_{a'}, P_{b'}$. Let U_t be the set of endpoints of paths obtainable by at most t rotations about a' with pivots lying only in $int(C_1)$. So these rotations affect only $P_{a'}$, and $P_{b'}$ is left intact in each of the resulting paths.

Lemma 2.2.34 *Suppose that $|U_t| \leqslant \frac{\eta^2 n}{10^6}$. Then*

$$|B_\Gamma(U_t) \cap int(C_1)| \geqslant \min\left\{\frac{\eta^2 n}{150000}, \frac{\eta^2 w_0}{40000}|U_t|\right\}.$$

Proof. Let $u = |U_t|$ and $u' = |B_\Gamma(U_t) \cap int(C_1)|$. Consider the case $\frac{\log n}{up_0} \geqslant \frac{7}{2}$. Then similarly to the proof of Lemma 2.2.7,

$$\frac{u\eta^2 np_0}{14000} \overset{(2.2.33)}{\leqslant} \sum_{x\in U_t} |N_\Gamma(x) \cap int(C_1)| \leqslant e_\Gamma(U_t, B_\Gamma(U_t) \cap int(C_1)) + 2e_\Gamma(U_t)$$
$$\leqslant 2(u+u')\log n + 4u\log n,$$

whence

$$u' \geqslant \frac{(\eta^2 np_0 - 84000\log n)u}{28000\log n} = \frac{(\eta^2 w_0 - 84000)u}{28000} \geqslant \frac{\eta^2 w_0}{40000}|U_t|.$$

On the other hand, if $\frac{\log n}{up_0} \leqslant \frac{7}{2}$ then

$$\frac{u\eta^2 np_0}{14000} \leqslant \sum_{x\in U_t} |N_\Gamma(x) \cap int(C_1)| \leqslant e_\Gamma(U_t, B_\Gamma(U_t) \cap int(C_1)) + 2e_\Gamma(U_t)$$
$$\leqslant 7up_0(u+u') + 14u^2 p_0$$

and so $3u + u' \geqslant \frac{\eta^2 n}{98000}$. Hence $u' \geqslant \frac{\eta^2 n}{150000}$. □

Corollary 2.2.35 $|U_{E_1}| \geqslant \frac{\eta^2 n}{10^6}$.

Proof. It suffices to prove that for each t such that $|U_t| \leqslant \frac{\eta^2 n}{10^6}$, either $|U_{t+1}| \geqslant \frac{\eta^2 n}{10^6}$ or $|U_{t+1}| \geqslant \frac{\eta^2 w_0}{10^5}|U_t|$ (or both). Similarly to Lemma 2.2.31, we have that $|U_{t+1}| \geqslant \frac{1}{2}|B_\Gamma(U_t) \cap int(C_1)| - |U_t|$, and now Lemma 2.2.34 immediately gives the result. □

Proof of Lemma 2.2.24. Suppose first that $|P| \leqslant \frac{\eta n}{200}$. Then Corollary 2.2.32 immediately implies that we can obtain a path, one of whose endpoints has a neighbour in Γ lying outside P, in at most E_0 rotations. So we may assume that $|P| \geqslant \frac{\eta n}{200}$, and hence the conditions of Lemma 2.2.29 are satisfied. Now we proceed as above to obtain a path $P''' = P_{a'} \cup P_{b'}$, with endpoints a', b' and with sets C_1, C_2 satisfying (2.2.33), such that $a' \in C_1 \subseteq P_{a'}$ and $b' \in C_2 \subseteq P_{b'}$.

Let $U = U_{E_1}$. Now similarly, we can rotate about b' using only pivots in C_2, to obtain another set U' of endpoints in another E_1 rotations, such that $|U'| \geqslant \frac{\eta^2 n}{10^6}$. Now by Lemma 2.2.3, there are at least $\frac{13}{14}\frac{\eta^4 n^2 p_0}{10^{12}}$ edges of G_2 between U and U', and since by (2.2.25) $G_2 \backslash \Gamma$ contains fewer edges than this, it follows that there exists an edge xy of Γ between U and U'. Now by the definition of U and U', we can obtain a path P'' with endpoints x, b' from P''' by a sequence of at most E_1 rotations, none of which affect the second half $P_{b'}$ of the path P'''. From P'' we can obtain a path P' with endpoints x, y by at most E_1 rotations. $\square$

2.3 Exact result

2.3.1 Outline

The goal of this section is to prove Theorem 1.3.5. Our approach uses the ideas from Section 2.2 (see Section 2.2.1), but some obvious obstacles arise. For simplicity, assume that n and $\delta(G_{n,p})$ are both even in what follows. We can use Tutte's theorem to show that $G_{n,p}$ has an r-regular subgraph G_{reg} with $r = \delta(G_{n,p})$ (this is harder than Step 2 but not too difficult for our range of p). So we can decompose G_{reg} into $r/2$ 2-factors F_i and would like to transform each F_i into a Hamilton cycle. At first sight, this looks infeasible for 2 reasons:

(a) The counting argument in Step 3 only works for reasonably dense subgraphs of $G_{n,p}$. So the 2-factors produced by the later iterations in Step 3 will have too many cycles.

(b) We no longer have a graph G_{thin} to use in Step 4.

For (a), it turns out that one can extend the counting argument so that it also works for sparser subgraphs of $G_{n,p}$, provided these subgraphs are pseudorandom (see Lemmas 2.3.21 and 2.3.25).

A very useful observation which goes some way in solving problem (b) is the following: Let x_0 be the vertex of minimum degree in $G_{n,p}$. Then whp there is a small but significant gap between $d(x_0)$ and $d(x)$ for any $x \neq x_0$ in $G_{n,p}$ (the gap has size close to $\sqrt{np}$). Let G_{left} be the subgraph of $G_{n,p}$ consisting of all the edges not in G_{reg}. Then G_{left} has density about $\sqrt{p/n}$, and we could try to use G_{left} to merge 2-factors instead of the graph G_{thin} in Step 4. One problem here is that G_{left} is not actually pseudorandom as it is just the 'leftover' from the Tutte application.

We use the following idea to overcome this problem. Right at the start, we remove the edges of a random subgraph G_5 of density p_5 from $G_{n,p}$, where $np_5 \ll \sqrt{np}$. Let x_0 be the vertex of minimum degree of the remaining subgraph G_1 of $G_{n,p}$. The choice of p_5 implies that x_0 will also be the vertex of minimum degree in $G_{n,p}$. Now add all edges of G_5 which are incident to x_0 to G_1. Then $\delta(G_1) = \delta(G_{n,p})$ and it turns out that we can still apply Tutte's theorem to obtain an r-regular subgraph G_{reg} of G_1 (see Lemma 2.2.4). Again we obtain a decomposition of G_{reg} into 2-factors F_i.

The key advantage of this method is that $G_5 - x_0$ will be pseudorandom, and so one can use G_5 in the same way as G_{thin} to merge the F_i into Hamilton cycles. However, it turns out that G_5 is far too sparse to complete the process: if e.g. p is a constant, then our bound on the total number of cycles in the F_i is significantly larger than $n^{3/2}$. On the other hand, G_5 has significantly fewer than $n^{3/2}$ edges. To reduce the number of cycles in F_i by 1, we need about $\log n$ edges from G_5. So even if we return the same number of edges from F_i to G_5 each time, we cannot assume G_5 to be pseudorandom after dealing with even a small proportion of the F_i. We remark that while performing additional iterations would allow us to improve the lower bound on p in Theorem 1.3.5 by a polylogarithmic factor, we cannot eliminate the polylogarithmic term entirely merely by increasing the number of iterations. Conversely, with fewer iterations one would still obtain a weaker version of Theorem 1.3.5.

Our main idea is to use an iterative approach to overcome this. Choose p_2, p_3 and p_4 so that $p_5 \ll p_4 \ll p_3 \ll p_2 \ll p$. Then choose randomly edge-disjoint subgraphs G_i of $G_{n,p}$, of density p_i, for $i = 2, 3, 4, 5$. (For simplicity, assume that G_2, G_3 and G_4 are regular of even degree in what follows.) Let G_1 be the remaining subgraph of $G_{n,p}$ (with all of the edges of G_5 which are incident to x_0 added to G_1) and let F_i be a set of $\delta(G_1)/2$ edge-disjoint 2-factors of G_1. Then as our first iteration we transform all the F_i into Hamilton cycles, where G_2 plays the role of G_{thin}. In the second iteration, we do the same for (the leftover of) G_2, with G_3 playing the role of G_{thin}, and so on until we reach G_5. (Note that there is no need to do anything with the leftover of G_5, as G_5 contains no edges incident to x_0.) In total, this gives us a set of $\lfloor(\delta(G_1) + \delta(G_2) + \delta(G_3) + \delta(G_4))/2\rfloor = \lfloor d(x_0)/2\rfloor$ edge-disjoint Hamilton cycles which together contain all of the edges incident to x_0.

The new problem that we now face is that e.g. in the second iteration the graph G_2 is a 'leftover' from the first iteration. So its pseudorandom properties are not as strong as we need them to be (e.g. in order to apply the strengthened version of the counting argument in Step 3) due to the existence of some 'bad edges' which were moved into G_2 from the F_i during the first iteration.

In order to deal with this, we perform some intermediate steps (about $\tau = \log n/\log\log n$ per iteration) using yet more random graphs $G_{(2,j)}$ (which we must also remove from $G_{n,p}$ initially). So first $G_{(2,1)}$ plays the role of G_{thin} when transforming the 2-factors of G_1 into Hamilton cycles. We then transform (the leftover of) $G_{(2,1)}$ into Hamilton cycles using $G_{(2,2)}$ and $G_{(2,3)}$, and then the leftover of $G_{(2,2)}$ and $G_{(2,3)}$ using $G_{(2,4)}$ and $G_{(2,5)}$, etc. Roughly speaking, after τ iterations we will have replaced $G_{(2,1)}$ by a graph $G_{(2,2\tau+1)}$ of almost the same density as $G_{(2,1)}$, but containing no bad edges (see Lemma 2.3.60). So it is now possible to carry out the second iteration (as described in the previous paragraphs) with $G_{(2,2\tau+1)}$ in place of G_2 and $G_{(3,1)}$ in place of G_3, and similarly the third and further iterations.

This section is organized as follows: In Section 2.3.2, we define the pseudorandomness properties we need in the proof and show that with very high probability they are satisfied by $G_{n,p}$. In Section 2.3.3 we use Tutte's theorem find an analogue of G_{reg}. In Section 2.3.4 we use our extended counting argument to split this analogue of G_{reg} into 2-factors. In Section 2.3.5 we show how to merge the cycles in each of these 2-factors into a Hamilton cycle. Finally in Section 2.3.6 we combine our results so far to prove Theorem 1.3.5, using the iterative approach discussed above.

2.3.2 Pseudorandom graphs

Our aim in this section is to establish several properties of random graphs which we will use later on, mainly regarding the degree sequence and expansion of small sets. For many of these, the fact that they hold whp in $G_{n,p}$ is well known; however, we need them to hold in $\log n/\log\log n$ random spanning subgraphs of $G_{n,p}$ simultaneously. To accomplish this we first show that they hold in $G_{n,p}$ with probability $1 - O(1/\log n)$, and then take a union bound.

The following useful definition is due to Thomason [112]. It involves tighter bounds than the similar and more common notion of ε-regularity. We will rely on these in the proof of Theorem 1.3.5.

Definition 2.3.1 *Let $p, \beta \geqslant 0$ with $p \leqslant 1$. A graph G is (p,β)-*jumbled *if $|e_G(S) - p\binom{s}{2}| \leqslant \beta s$ for every $S \subseteq V(G)$ with $|S| = s$.*

We will often use the following immediate consequence of Definition 2.3.1: Let G be a (p, β)-jumbled graph and let $S, T \subseteq V(G)$ be disjoint. Then

$$|e_G(S,T) - p|S||T|| \leqslant 2\beta(|S| + |T|). \tag{2.3.2}$$

To see this, note that $e_G(S,T) = e_G(S \cup T) - e_G(S) - e_G(T)$; now applying Definition 2.3.1 and using the triangle inequality implies (2.3.2).

The following two definitions formalise the notion of 'degree gap', which we need in our proof.

Definition 2.3.3 *Let G be a graph on n vertices with a vertex x_0 of minimum degree, and let $u \leqslant n$. Then G is* u-jumping *if every vertex of G apart from x_0 has degree at least $\delta(G) + u$.*

Definition 2.3.4 *Let G be a graph on n vertices. For a set $T \subseteq V(G)$, let $\overline{d}_G(T)$ be the average degree of the vertices of T in G. Then G is* strongly 2-jumping *if $\overline{d}_G(T) \geqslant \delta(G) + \min\{|T| - 1, \log^2 n\}$ for every $T \subseteq V(G)$.*

Note that if G is strongly 2-jumping then it is also 2-jumping. In addition to these three properties we will use several other bounds concerning the degree sequence and edge distribution of a random graph. The following definition collects these properties together.

Definition 2.3.5 *Call a graph G on n vertices* p-pseudorandom *if all of the following hold:*

(a) *G is $(p, 2\sqrt{np(1-p)})$-jumbled.*

(b) *For any disjoint sets $S, T \subseteq V(G)$ with $|S| = s$ and $|T| = t$,*

   (i) *if $\left(\frac{1}{s} + \frac{1}{t}\right)\frac{\log n}{p} \geqslant \frac{7}{2}$, then $e_G(S,T) \leqslant 2(s+t)\log n$,*

   (ii) *if $\left(\frac{1}{s} + \frac{1}{t}\right)\frac{\log n}{p} \leqslant \frac{7}{2}$, then $e_G(S,T) \leqslant 7stp$,*

(iii) *if* $\frac{\log n}{sp} \geqslant \frac{7}{4}$, *then* $e_G(S) \leqslant 2s \log n$, *and*

(iv) *if* $\frac{\log n}{sp} \leqslant \frac{7}{4}$, *then* $e_G(S) \leqslant 7s^2p/2$.

(c) $np - 2\sqrt{np \log n} \leqslant \delta(G) \leqslant np - 200\sqrt{np(1-p)}$ *and* $\Delta(G) \leqslant np + 2\sqrt{np \log n}$.

(d) G *is strongly* 2*-jumping.*

Definition 2.3.5(a) gives good bounds on the densities of large subgraphs but not on those of very small subgraphs, which is why we need (b).

The remainder of this section will be mainly devoted to showing that the random graphs we consider in the rest of the proof are in fact pseudorandom (see Lemma 2.3.12). We will need the following large deviation bounds on the binomial distribution, proved in [60] (as Theorem 2.1, Corollary 2.3 and Corollary 2.4 respectively):

Lemma 2.3.6 *Let* $X \sim Bin(n,p)$. *Then the following properties hold:*

(i) *If* $h > 0$, *then* $\mathbb{P}[X \leqslant np - h] < e^{-h^2/2np}$.

(ii) *If* $c > 1$ *and* $c' = \log c - 1 + 1/c$, *then for any* $a \geqslant cnp$, $\mathbb{P}[X \geqslant a] < e^{-c'a}$.

The following lemma states that the property in Definition 2.3.5(a) holds with very high probability in $G_{n,p}$ for the desired range of p.

Lemma 2.3.7 *Suppose that* $np/\log^2 n \to \infty$ *and* $n(1-p)/\log^2 n \to \infty$, *and let* $G \sim G_{n,p}$. *Then the probability that* G *is not* $(p, 2\sqrt{np(1-p)})$*-jumbled is at most* $2/n^2$.

Proof. Without loss of generality we may assume that $0 < p \leqslant 1/2$ (by considering the complement of G if $p > 1/2$). For each set $S \subseteq V(G)$ with $|S| = s$, we have that $e(S) \sim Bin(\binom{s}{2}, p)$. Let

$$\varepsilon = \frac{4\sqrt{n(1-p)}}{(s-1)\sqrt{p}} \text{ and } N = \binom{s}{2}$$

and note that $\varepsilon pN = 2\sqrt{np(1-p)}s$. Call S *bad* if $|e_G(S) - pN| \geqslant \varepsilon pN$. We consider the following cases:

Case 1: $s \geqslant \frac{8}{3}\sqrt{\frac{n(1-p)}{p}} + 1$. Then $\varepsilon \leqslant 3/2$, and hence by Lemma 1.8.1(i) we have that the probability of S being bad is at most $4e^{-\varepsilon^2 pN/3}$. But since $p \leqslant 1/2$, we have $\varepsilon^2 pN/3 = \frac{8ns(1-p)}{3(s-1)} \geqslant \frac{4n}{3}$. So the probability that S is bad is at most e^{-n}.

Case 2: $s \leqslant \frac{8}{3}\sqrt{\frac{n(1-p)}{p}}+1$. Since $\varepsilon \geqslant 3/2$ in this case, we have that $\varepsilon pN \geqslant pN$ and hence $\mathbb{P}[e_G(S) < pN - \varepsilon pN] = 0$. Let $c = \varepsilon + 1$, so $c \geqslant 5/2$. Let $c' = \log c - 1 + 1/c$ and note that since c' is an increasing function of c for $c > 1$, we have $c' \geqslant \log(5/2) - 1 + 2/5 > 1/4$. Now by Lemma 2.3.6(ii) applied with $a = (1+\varepsilon)pN$, the probability that S is bad is at most $e^{-c'(1+\varepsilon)pN} \leqslant e^{-2c'\sqrt{np(1-p)}s} \leqslant e^{-4s\log n} = n^{-4s}$.

We now take a union bound on the probability that there exists a bad set S. Firstly for any $s \leqslant \frac{8}{3}\sqrt{\frac{n(1-p)}{p}}+1$, the probability that there is some bad set S with $|S| = s$ is at most $\binom{n}{s}n^{-4s} < n^{-3s} \leqslant n^{-3}$. Summing over all such s, we have an error bound of $n^{1-3} = 1/n^2$. Further the probability that there exists a bad set S with $|S| = s \geqslant \frac{8}{3}\sqrt{\frac{n(1-p)}{p}} + 1$ is at most $2^n e^{-n} \leqslant 1/n^2$. Now adding these two bounds completes the proof. $\square$

To check the properties in Definition 2.3.5(b) we can use Lemmas 2.2.1 and 2.2.2 from the Section 2.2.

It remains to establish the bounds in Definition 2.3.5(c) on the minimum and maximum degree of $G_{n,p}$ (see Lemma 2.3.9), and the fact that $G_{n,p}$ is strongly 2-jumping with probability $1-O(1/\log n)$ (see Lemma 2.3.10). For this we need estimates on the binomial distribution which do not follow from standard Chernoff bounds (see Lemma 2.3.8). These estimates use the following notation, where $X \sim Bin(n-1,p)$:

- $b(r) = \mathbb{P}[X = r] = \binom{n-1}{r}p^r(1-p)^{n-r-1}$,
- $B(m_1, m_2) = \mathbb{P}[m_1 \leqslant X \leqslant m_2]$, and
- $B(m) = \mathbb{P}[X \leqslant m]$.

$b'(r)$, $B'(m_1, m_2)$ and $B'(m)$ are defined similarly for $X \sim Bin(n-2,p)$.

Lemma 2.3.8 *Suppose that $np(1-p) \to \infty$. Let $m_1 = np - 2\sqrt{np\log n}$, $m_2 = np - 200\sqrt{np(1-p)}$ and $\lambda = 1 - 1/(8\log^3 n)$. Then*

(i) $nB(m_2) \geqslant nb(m_2) \geqslant \sqrt{n}/\log n$, *and*

(ii) $nB(m_1) \leqslant 1/\sqrt{n}$.

Suppose in addition that $p \geqslant 48^2 \log^7 n/n$ and $1-p \geqslant 36n^{-1/2}\log^{7/2} n$. Then

(iii) $\frac{b(r-1)}{b(r)} \geqslant \lambda$ *for each $r \geqslant m_1 - 8\log^3 n$,*

(iv) $\frac{B(m_1 - 8\log^3 n, r)}{b(r)} \geqslant 4\log^3 n$ *for each $r \geqslant m_1$, and*

(v) $\frac{b'(r)}{b(r)} \leqslant 1 + 1/\log n$ *for each $r \geqslant m_1$.*

Proof. (i) Let $X \sim Bin(n-1,p)$, $\sigma = \sqrt{(n-1)p(1-p)}$ and $m_2' = np - 201\sqrt{np(1-p)}$. Now the de Moivre-Laplace Theorem (see e.g., Theorem 1.6 of [15]) states that if $\sigma \to \infty$ and $x_2 > x_1$ are constants, then

$$B((n-1)p + x_1\sigma, (n-1)p + x_2\sigma) = (1+o(1))(\phi(x_2) - \phi(x_1)),$$

where $\phi(x)$ is the cumulative density function of the normal distribution. Clearly

$$\begin{aligned}&B(np + x_1\sqrt{np(1-p)}, np + x_2\sqrt{np(1-p)})\\&= (1+o(1))B((n-1)p + x_1\sigma, (n-1)p + x_2\sigma).\end{aligned}$$

To see this, note that the boundaries of the interval change by at most 2 and that $b(r) \to 0$ for any r. Hence we have that

$$B(m_2', m_2) = (1+o(1))c,$$

where $c = \phi(-200) - \phi(-201)$ is constant. Clearly $b(r) \leqslant b(m_2)$ for each $m_2' \leqslant r \leqslant m_2$. So

$$nb(m_2) \geqslant \frac{(1+o(1))cn}{\sqrt{np(1-p)}+1} \geqslant \frac{\sqrt{n}}{\log n}.$$

(ii) Let $X \sim Bin(n-1,p)$. By Lemma 2.3.6(i) we have that

$$\begin{aligned}\mathbb{P}\left[X \leqslant np - 2\sqrt{np\log n}\right] &< \mathbb{P}\left[X \leqslant (n-1)p - 7\sqrt{np\log n}/4\right] \\ &\leqslant e^{-(7\sqrt{np\log n}/4)^2/2np} \leqslant e^{-\frac{3}{2}\log n} = \frac{1}{n^{3/2}}.\end{aligned}$$

(iii) We have

$$\frac{b(r-1)}{b(r)} = \frac{\binom{n-1}{r-1}p^{r-1}(1-p)^{n-r}}{\binom{n-1}{r}p^r(1-p)^{n-r-1}} = \frac{(n-1)!r!(n-r-1)!(1-p)}{(n-1)!(r-1)!(n-r)!p} = \frac{r(1-p)}{(n-r)p},$$

and so

$$\begin{aligned}1 - \frac{b(r-1)}{b(r)} = \frac{np-r}{(n-r)p} &\leqslant \frac{2\sqrt{np\log n} + 8\log^3 n}{(n - np + 2\sqrt{np\log n} + 8\log^3 n)p} \\ &\leqslant \frac{3\sqrt{np\log n}}{np - np^2} = \frac{3\sqrt{\log n}}{(1-p)\sqrt{np}}.\end{aligned}$$

Now if $p \geqslant 1/2$ then $(1-p)\sqrt{np} \geqslant (36n^{-1/2}\log^{7/2} n)(\sqrt{n/2})$ and the result follows. On the other hand if $p \leqslant 1/2$ then $(1-p)\sqrt{np} \geqslant (48\log^{7/2} n)/2$ and again the result follows.

(iv) Note that

$$\begin{aligned}\frac{B(m_1 - 8\log^3 n, r)}{b(r)} &\geqslant 1 + \frac{b(r-1)}{b(r)} + \frac{b(r-2)}{b(r)} + \ldots + \frac{b(r - 8\log^3 n)}{b(r)} \\ &\geqslant 1 + \lambda + \lambda^2 + \ldots + \lambda^{8\log^3 n} = \frac{1-\lambda^{8\log^3 n + 1}}{1-\lambda} \\ &\geqslant \frac{1-e^{-1}}{1/(8\log^3 n)} \geqslant 4\log^3 n,\end{aligned}$$

where in the final line we use that $\lambda \leqslant e^{-1/(8\log^3 n)}$.

(v)

$$\begin{aligned}\frac{b'(r)}{b(r)} &= \frac{\binom{n-2}{r}p^r(1-p)^{n-r-2}}{\binom{n-1}{r}p^r(1-p)^{n-r-1}} = \frac{(n-2)!r!(n-r-1)!}{(n-1)!r!(n-r-2)!(1-p)} \\ &= \frac{n-r-1}{(n-1)(1-p)} = 1 + \frac{np-p-r}{(n-1)(1-p)} \leqslant 1 + \frac{2\sqrt{np\log n}}{(n-1)(1-p)} \\ &\leqslant 1 + \frac{3\sqrt{p\log n}}{\sqrt{n}(1-p)} \leqslant 1 + \frac{1}{\log n},\end{aligned}$$

as desired. $\square$

Lemma 2.3.9 *Let $G \sim G_{n,p}$. Suppose that $np(1-p) \to \infty$. Then*

(i) $\mathbb{P}\left[\delta(G) \leqslant np - 2\sqrt{np\log n}\right] \leqslant 1/\sqrt{n}$, *and*

(ii) $\mathbb{P}\left[\Delta(G) \geqslant np + 2\sqrt{np\log n}\right] \leqslant 1/\sqrt{n}$.

Suppose further that $p \geqslant 48^2 \log^7 n/n$ and $1-p \geqslant 36n^{-1/2}\log^{7/2} n$. Let $m_1 = np - 2\sqrt{np\log n}$, $m_2 = np - 200\sqrt{np(1-p)}$ and let $m_1 \leqslant m \leqslant m_2$ be such that $nB(m) \geqslant \log n$ and $nB(m-1) \leqslant \log n$. Then

(iii) $\mathbb{P}\left[\delta(G) > m\right] \leqslant 4/\log n$, *and*

(iv) $\mathbb{P}\left[\delta(G) > np - 200\sqrt{np(1-p)}\right] \leqslant 4/\log n$.

Note that the results in Chapter 3 of [15] give sharper bounds which hold with high probability. However as mentioned earlier, we need our error probability to be smaller. Since this does seem to affect the precise results we need to prove the bounds explicitly. Also note that by Lemma 2.3.8(i) and (ii), $nB(m_1) \leqslant \log n \leqslant nB(m_2)$ (with room to spare), and so there exists $m_1 \leqslant m \leqslant m_2$ such that $nB(m) \geqslant \log n$, but $nB(m-1) \leqslant \log n$. Thus (iii) is not vacuous.

Proof. (i) By taking a union bound we have that

$$\mathbb{P}\left[\delta(G) \leqslant np - 2\sqrt{np\log n}\right] \leqslant nB(np - 2\sqrt{np\log n}) \leqslant 1/\sqrt{n},$$

where the last inequality follows from Lemma 2.3.8(ii).

(ii) Apply (i) to the complement of G.

(iii) Let Y be the number of vertices $v \in V(G)$ such that $m_1 \leqslant d(v) \leqslant m$. As $nB(m_1) \leqslant 1/\sqrt{n}$ by Lemma 2.3.8(ii), we have

$$\mathbb{E}(Y) = nB(m_1, m) \geqslant nB(m) - nB(m_1) \geqslant \log n - 1/\sqrt{n} \geqslant 2\log n/3$$

and

$$\mathbb{E}_2(Y) = n(n-1)(pB'(m_1-1, m-1)^2 + (1-p)B'(m_1, m)^2) \leqslant n^2 B'(m_1, m)^2.$$

Hence

$$\frac{\sqrt{\mathbb{E}_2(Y)}}{\mathbb{E}(Y)} \leqslant \frac{\sum_{r=m_1}^{m} nb'(r)}{\sum_{r=m_1}^{m} nb(r)} = \frac{\sum_{r=m_1}^{m} nb(r)\frac{b'(r)}{b(r)}}{\sum_{r=m_1}^{m} nb(r)} \leqslant 1 + \frac{1}{\log n},$$

where the last inequality follows from Lemma 2.3.8(v). Hence

$$\begin{aligned} Var(Y) &= \mathbb{E}_2(Y) + \mathbb{E}(Y) - \mathbb{E}(Y)^2 \leqslant (1 + 1/\log n)^2 \mathbb{E}(Y)^2 + \mathbb{E}(Y) - \mathbb{E}(Y)^2 \\ &= (2/\log n + 1/\log^2 n)\mathbb{E}(Y)^2 + \mathbb{E}(Y). \end{aligned}$$

So by Chebyshev's inequality,

$$\mathbb{P}(Y = 0) \leqslant \frac{Var(Y)}{\mathbb{E}(Y)^2} \leqslant \frac{2}{\log n} + \frac{1}{\log^2 n} + \frac{1}{\mathbb{E}(Y)} \leqslant \frac{4}{\log n}.$$

(iv) This follows immediately from (iii) and the remark after the lemma statement.

$\square$

Lemma 2.3.10 *Let $G \sim G_{n,p}$. Suppose that $p \geqslant 48^2 \log^7 n/n$ and $1-p \geqslant 36n^{-1/2}\log^{7/2} n$. Then with probability at least $1 - 6/\log n$, G is strongly 2-jumping.*

Similarly to Lemma 2.3.9, note that Theorem 3.15 in [15] would imply Lemma 2.3.10 if we only required the statement to hold with high probability.

Proof. Let $m_1 = np - 2\sqrt{np\log n}$, $m_2 = np - 200\sqrt{np(1-p)}$ and $\lambda = 1 - 1/(8\log^3 n)$. Let $m_1 \leqslant m \leqslant m_2$ be such that $nB(m) \geqslant \log n$, but $nB(m-1) \leqslant \log n$. By Lemma 2.3.9(iii) there exists a vertex of degree at most m with probability at least $1 - 4/\log n$. So it suffices to show that with probability at least $1 - 2/\log n$ there are no two vertices each of degree at most $m + 2\log^2 n$ whose degrees differ by at most 1.

Let Z be the number of (unordered) pairs v_1, v_2 of vertices such that $m_1 \leqslant \min\{d(v_1), d(v_2)\} \leqslant m + 2\log^2 n$ and $|d(v_1) - d(v_2)| \leqslant 1$. We have

$$\begin{aligned}\mathbb{E}(Z) &\leqslant \binom{n}{2} \sum_{r=m_1}^{m+2\log^2 n} pb'(r-1)^2 + 2pb'(r-1)b'(r) + (1-p)b'(r)^2 \\ &\qquad\qquad + 2(1-p)b'(r)b'(r+1) \\ &\leqslant \frac{n^2}{2} \sum_{r=m_1}^{m+2\log^2 n} 3b'(r)b'(r+1) \leqslant \frac{3n^2}{2}\left(1 + \frac{1}{\log n}\right)^2 \cdot \sum_{r=m_1}^{m+2\log^2 n} b(r)b(r+1) \\ &\leqslant 2n^2 b(m + 2\log^2 n + 1)B(m_1, m + 2\log^2 n),\end{aligned}$$

where the third inequality follows from Lemma 2.3.8(v). Note that

$$\lambda^{2\log^2 n+2} \geqslant 1 - \frac{2\log^2 n + 2}{8\log^3 n} \geqslant \frac{6}{7}. \tag{2.3.11}$$

Now by Lemma 2.3.8(iii),

$$\begin{aligned}
nB(m_1, m + 2\log^2 n + 1) &= \sum_{r=m_1}^{m+2\log^2 n+1} nb(r) \\
&\leqslant \sum_{r=m_1}^{m+2\log^2 n+1} \frac{nb(r - 2\log^2 n - 2)}{\lambda^{2\log^2 n+2}} \\
&= \frac{nB(m_1 - 2\log^2 n - 2, m-1)}{\lambda^{2\log^2 n+2}} \leqslant \frac{nB(m-1)}{\lambda^{2\log^2 n+2}} \\
&\leqslant \frac{\log n}{\lambda^{2\log^2 n+2}} \overset{(2.3.11)}{\leqslant} \frac{7\log n}{6}.
\end{aligned}$$

So Lemma 2.3.8(iv) implies that

$$\begin{aligned}
nb(m + 2\log^2 n + 1) &\leqslant \frac{nB(m_1 - 8\log^3 n, m + 2\log^2 n + 1)}{4\log^3 n} \\
&= \frac{nB(m_1 - 8\log^3 n, m_1 - 1) + nB(m_1, m + 2\log^2 n + 1)}{4\log^3 n} \\
&\leqslant \frac{8\log^3 n \cdot nb(m_1) + \frac{7}{6}\log n}{4\log^3 n}
\end{aligned}$$

and now since $b(m_1) \leqslant B(m_1)$, Lemma 2.3.8(ii) implies that

$$nb(m + 2\log^2 n + 1) \leqslant \frac{2}{\sqrt{n}} + \frac{7}{24\log^2 n} \leqslant \frac{1}{3\log^2 n}.$$

Hence

$$\mathbb{E}(Z) \leqslant 2 \cdot \frac{7\log n}{6} \cdot \frac{1}{3\log^2 n} \leqslant \frac{1}{\log n}.$$

So by Markov's inequality $\mathbb{P}[Z \geqslant 1] \leqslant 1/\log n$. Now by Lemma 2.3.9(i), the probability that there are any vertices at all of degree at most m_1 is at most $1/\sqrt{n}$. So with probability at least $1 - 1/\log n - 1/\sqrt{n} \geqslant 1 - 2/\log n$, there are no pairs of vertices of degree at most $m + 2\log^2 n$, whose degrees differ by at most 1. $\square$

We now combine the above results to prove that our desired pseudorandomness conditions hold whp in $G_{n,p}$.

Lemma 2.3.12 *Let $G \sim G_{n,p}$. Suppose that $p \geqslant 48^2 \log^7 n/n$ and $1-p \geqslant 36n^{-1/2}\log^{7/2} n$. Then the probability that G is* not *p-pseudorandom is at most $11/\log n$.*

Proof. By Lemmas 2.3.7, 2.2.1, 2.2.2, 2.3.9(i), 2.3.9(ii), 2.3.9(iv) and 2.3.10 the probability that G is not p-pseudorandom is at most

$$\frac{2}{n^2} + \frac{1}{n^2} + \frac{1}{n^2} + \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{n}} + \frac{4}{\log n} + \frac{6}{\log n} < \frac{11}{\log n}.$$

□

Lemma 2.3.13 *Let $\mathcal{I}$ denote the interval $[48^2 \log^7 n/n, 1 - 36n^{-1/2}\log^{7/2} n]$. Let $p_0 \in \mathcal{I}$ and let $G_0 \sim G_{n,p_0}$. Let $p_1, \ldots, p_5 \in \mathcal{I}$ be positive reals such that $p_1 + p_2 + p_3 + p_4 + p_5 = p_0$. Let $t_2, t_3, t_4 \leqslant \log n/\log\log n$ be positive integers. For each $i = 2, 3, 4$, let $p_{(i,1)}, \ldots, p_{(i,2t_i+1)} \in \mathcal{I}$ be positive reals such that $p_{(i,1)} + p_{(i,2)} + \ldots + p_{(i,2t_i+1)} = p_i$.*

Suppose that the quantities $p_i + p_{i+1}$ for $i = 2, 3, 4$ all lie in $\mathcal{I}$. Suppose further that the edges of G are distributed among graphs G_1, G_5 and $G_{(i,1)}, \ldots, G_{(i,2m_i+1)}$ for $i = 2, 3, 4$, as follows. For each edge ab of G_0:

- *With probability p_1/p_0 let ab be an edge of G_1.*
- *For $i = 2, 3, 4$ and $1 \leqslant j \leqslant 2m_i + 1$, with probability $p_{(i,j)}/p_0$ let ab be an edge of $G_{(i,j)}$.*
- *Otherwise (i.e., with probability p_5/p_0), let ab be an edge of G_5.*

For $i = 2, 3, 4$, let $G_i = \bigcup_{j=1}^{2t_i+1} G_{(i,j)}$. Then whp, the following properties hold:

(i) *G_i is p_i-pseudorandom for each $0 \leqslant i \leqslant 5$,*

(ii) *$G_{(i,j)}$ is $p_{(i,j)}$-pseudorandom for $i = 2, 3, 4$ and $1 \leqslant j \leqslant 2t_i + 1$, and*

(iii) $G_i \cup G_{i+1}$ *is* $(p_i + p_{i+1})$*-pseudorandom for* $i = 2, 3, 4$.

Proof. First note that $G_{(i,j)} \sim G_{n,p_{(i,j)}}$ for each i, j and that $G_i \sim G_{n,p_i}$ for each i. Further, $G_i \cup G_{i+1} \sim G_{n,(p_i+p_{i+1})}$ for $i = 2, 3, 4$.

Hence by Lemma 2.3.12 for each graph, the probability that it is not p-pseudorandom for the relevant value of p is at most $11/\log n$, and so by taking a union bound we have that the probability that at least one of these graphs is not pseudorandom is at most $67/\log\log n$, which tends to 0 as $n \to \infty$. $\square$

At one point in the proof of Theorem 1.3.5 we will need $G_{n,p}$ to be u-jumping for some $u > 2$. For this we use the following result.

Lemma 2.3.14 *Let* $G \sim G_{n,p}$ *with* $np/\log n \to \infty$ *and* $n(1-p)/\log n \to \infty$*. Then with high probability,* G *is* $8\sqrt{np(1-p)}/\log^{3/4} n$*-jumping.*

We will apply this lemma only once (in Section 2.3.6), so in this case a whp estimate is sufficient.

Proof. The conditions of the lemma immediately imply that $np(1-p)/\log n \to \infty$. Let $t = \min\{(np(1-p)/\log n)^{1/5}, \log^{1/16} n\}$ and let $\alpha = 8\log^{-1/16} n$; note that $t \to \infty$ and $\alpha \to 0$. Applying Theorem 3.15 in [15] to the complement of G, we have that G is u-jumping where

$$u = \frac{\alpha}{m^2}\left(\frac{np(1-p)}{\log n}\right)^{1/2} \geqslant \frac{8\sqrt{np(1-p)}}{\log^{3/4} n}.$$

$\square$

2.3.3 Constructing regular spanning subgraphs

The first step in our general strategy for finding a large collection of Hamilton cycles in a p-pseudorandom graph G is to construct a regular spanning subgraph of G of degree $\delta(G)$ if $\delta(G)$ is even, or of degree $\delta(G) - 1$ if $\delta(G)$ is odd. The aim of this section is to establish that this is always possible for our range of p. As in Section 2.2.3, we will use Tutte's r-factor theorem (see Theorem 2.2.4). For the particular case of a 1-factor (that is, a perfect matching), we use the following simpler result:

Theorem 2.3.15 (Tutte [116]) *Let G be a graph. Then G has a perfect matching if and only if for every $S \subseteq V(G)$, the number of components of $G - S$ which have an odd number of vertices is at most $|S|$.*

Recall from Section 2.2.3 that $Q_{r,G}(S,T)$ is the number of *odd* components of $G[U]$, where a component C is odd if and only if $r|C| + e_G(C,T)$ is odd. In order to make use of Theorem 2.2.4 we first need to bound $Q_{r,G}(S,T)$. For this we use the following lemma:

Lemma 2.3.16 *Let G be an r_G/n-pseudorandom graph on n vertices such that $r_G \geqslant \log n$. Let $r_H \geqslant \frac{549}{550} r_G$, and let H be a spanning subgraph of G such that $\delta(H) = r_H$. Let $W \subseteq V(G)$, and suppose that $\delta(H[W]) \geqslant r_H/3$. Then for any nonempty $B \subseteq W$, the number of components of $H[W \backslash B]$ is at most $|B|$. Moreover, $H[W]$ is connected.*

Proof. Suppose that a set B violates the first assertion.

Claim: *$H[W \backslash B]$ cannot have two disjoint isolated sets (i.e., unions of components of $H[W \backslash B]$) each of size at least $n/32$.*

To prove the claim, it suffices to show that for disjoint sets $S, T \subseteq W \backslash B$ such that $|S|, |T| \geqslant n/32$, we have $e_H(S,T) > 0$. To see that this holds, note that since G is $(r_G/n, 2\sqrt{r_G})$-jumbled by Definition 2.3.5(a), (2.3.2) implies that

$$e_G(S,T) \geqslant \frac{r_G|S||T|}{n} - 4\sqrt{r_G}(|S| + |T|) \geqslant \frac{r_G n}{1024} - 4\sqrt{r_G} n.$$

But since $r_H \geqslant \frac{549}{550} r_G$, it follows that

$$|E(G) \backslash E(H)| = e(G) - e(H) \leqslant \frac{r_G n}{2} + 2\sqrt{r_G} n - \frac{549 r_G n}{1100} = \frac{r_G n}{1100} + 2\sqrt{r_G} n,$$

and hence $e_H(S,T) > 0$, which proves the claim.

Let $\mathcal{C}$ be the set of components of $H[W\backslash B]$ of size less than $n/32$, and let $A = \bigcup \mathcal{C}$. Suppose that $|A| \geqslant n/8$. Then we can form an isolated set S such that $n/32 \leqslant |S| \leqslant n/16$ by taking successive unions of components in $\mathcal{C}$. Similarly we can form another isolated set T with $n/32 \leqslant |T| \leqslant n/16$ using the remaining components in $\mathcal{C}$. But now S and T are disjoint, which contradicts the claim. Hence $|A| < n/8$. Moreover the claim implies that $H[W\backslash B]$ has exactly one component which is not in $\mathcal{C}$.

Now since by assumption, $H[W\backslash B]$ has more than $|B|$ components, we have $|A| \geqslant |\mathcal{C}| \geqslant |B|$. Now since A is isolated in $H[W\backslash B]$, it follows that the $H[W]$-neighbourhood of A lies entirely in B. Hence every edge of $H[W]$ which is incident to some vertex of A lies in $E_H(A) \cup E_H(A,B)$. It follows that $\sum_{v \in A} N_{H[W]}(v) \leqslant 2e_H(A) + e_H(A,B)$, noting that edges in $e_H(A)$ will be counted twice on the left-hand side. So

$$\begin{aligned} r_H|A|/3 &\leqslant |A|\delta(H[W]) \leqslant 2e_H(A) + e_H(A,B) \leqslant 2e_G(A) + e_G(A,B) \\ &\leqslant r_G|A|^2/n + 4\sqrt{r_G}|A| + r_G|A||B|/n + 4\sqrt{r_G}(|A| + |B|) \\ &\leqslant 2r_G|A|^2/n + 12\sqrt{r_G}|A| \leqslant \left(\frac{1}{4} + \frac{12}{\sqrt{r_G}}\right) r_G|A|, \end{aligned}$$

which is a contradiction unless $|A| = 0$. But if $|A| = 0$ then we have $B = \emptyset$, which proves the assertion. The moreover part follows from the special case where $|B| = 1$. $\square$

Corollary 2.3.17 *Let G be an r_G/n-pseudorandom graph with $r_G \geqslant \log^2 n$ and let G' be a graph obtained from G by first deleting an arbitrary matching M and then adding an arbitrary set of additional edges. Let r be an even integer. Then setting $Q_r(S,T) = Q_{r,G'}(S,T)$, any disjoint subsets S, T of $V(G)$ satisfy $Q_r(S,T) \leqslant |S| + |T|$.*

Proof. By Definition 2.3.5(c) we have that $\delta(G\backslash M) \geqslant r_G - 2\sqrt{r_G \log n} - 1 \geqslant \frac{549}{550} r_G$. Note that the number of components of $G' - B$ is always at most that of $(G\backslash M) - B$ for any $B \subseteq V(G)$. Now if $S \cup T \neq \emptyset$ then applying Lemma 2.3.16 with $G = G$, $H = G \backslash M$, $W = V(G)$ and $B = S \cup T$ implies that $H - (S \cup T)$ has at most $|S \cup T|$ components. It follows that $Q_r(S,T) \leqslant |S \cup T| = |S| + |T|$. On the other hand, if $S = T = \emptyset$ then Lemma 2.3.16 implies that H (and thus G') is connected. Let C be the unique component of G' and note that $r|C| + e_{G'}(C,T) = r|C|$, which is even. So in this case $Q_r(S,T) = 0 = |S| + |T|$. □

We are now ready to prove that every pseudorandom graph G has a regular spanning subgraph whose degree is equal to $2\lfloor \delta(G)/2 \rfloor$. In fact we will prove the following slightly stronger statement, which gives the same result even if G is modified slightly from being pseudorandom.

Lemma 2.3.18 *Let G be a p-pseudorandom graph on n vertices such that $np(1-p)/\log^2 n \to \infty$.*

(i) *For any vertex $x \in V(G)$, G contains an optimal matching which covers x.*

(ii) *Let $u \leqslant 4\sqrt{np(1-p)}$, $u \neq 1$ and suppose in addition that G is $2u$-jumping. Let G' be formed from G by adding u edges at the vertex x_0 of minimum degree, and if $\delta(G) + u$ is odd, also removing an arbitrary matching M. Let r be the greatest even integer which is at most $\delta(G) + u$. Then G' has an r-regular spanning subgraph.*

Note that if the matching we remove in (ii) covers x_0, then $r = d_{G'}(x_0)$. The case of Lemma 2.3.18(ii) when $u = 0$ is the only place in the proof of Theorem 1.3.5 where we use the fact that a pseudorandom graph is strongly 2-jumping (i.e., Definition 2.3.5(d)). One can probably replace this by a weaker condition. On the other hand, Defintion 2.3.5(a), (b) and (c) are probably not in themselves sufficient to prove Lemma 2.3.18.

Proof. (i) If n is even then Theorem 2.3.15 and Lemma 2.3.16 together imply that G has a perfect matching, and the result follows immediately. If n is odd remove any vertex

$y \neq x$ from G. Now by Lemma 2.3.16 the number of components of $G - (S \cup \{y\})$ is at most $|S| + 1$ for any $S \subseteq V(G)$. But if the number of components is exactly $|S| + 1$, then at least one component must have an even number of vertices (otherwise we have $n - |S| - 1 \equiv |S| + 1 \mod 2$ which cannot hold). Hence we can apply Theorem 2.3.15 to $G - y$ and the result follows.

(ii) For disjoint sets $S, T \subseteq V(G)$, let $s = |S|$ and $\rho = |T|/|S|$. Throughout the remainder of the proof let $R_r(S,T)$ and $Q_r(S,T)$ be defined with respect to the graph G'. By Corollary 2.3.17, $Q_r(S,T) \leqslant |S| + |T| = s(\rho + 1)$ and so by Theorem 2.2.4 it suffices to show that $R_r(S,T) \geqslant s(\rho + 1)$ for all S and T. If T is nonempty, let $\overline{d}_G(T)$ be the average degree, in G, of the vertices of T, and define $\overline{d}_G(S)$, $\overline{d}_{G'}(T)$ and $\overline{d}_{G'}(S)$ similarly. Note that

$$R_r(S,T) = \overline{d}_{G'}(T)|T| - e_{G'}(S,T) + rs(1 - \rho).$$

Note also that $d_{G'}(x_0) \geqslant r$. We claim that $d_{G'}(x) \geqslant r + 2$ for all $x \neq x_0$. Indeed, if $u = 0$ and $\delta(G)$ is even, we have

$$d_{G'}(x) = d_G(x) \geqslant \delta(G) + 2 = r + 2$$

since G is 2-jumping by Definition 2.3.5(d). If $u = 0$ and $\delta(G)$ is odd, then similarly we have

$$d_{G'}(x) \geqslant d_G(x) - 1 \geqslant \delta(G) + 1 = r + 2.$$

If $u \geqslant 2$ and $\delta(G) + u$ is even, we have

$$d_{G'}(x) = d_G(x) \geqslant \delta(G) + 2u \geqslant r + 2,$$

and finally if $u \geqslant 2$ and $\delta(G) + u$ is odd, we have

$$d_{G'}(x) \geqslant d_G(x) - 1 \geqslant \delta(G) + 2u - 1 \geqslant r + 2.$$

This proves the claim. It follows immediately that unless $T = \{x_0\}$, we have $\overline{d}_{G'}(T) \geqslant r + 1$. Now we consider the following cases:

Case 1: $|S| = 0$. Observe that $R_r(S,T) = \overline{d}_{G'}(T)|T| - r|T|$ and $Q_r(S,T) \leqslant |T|$ by Corollary 2.3.17. If either $|T| = 0$ or $\overline{d}_{G'}(T) \geqslant r+1$ then trivially $R_r(S,T) \geqslant |T| \geqslant Q_r(S,T)$. So we are left with the case in which $T = \{x_0\}$ and $d_{G'}(x_0) = r$.

In this case it suffices to prove that $Q_r(S,T) = 0$. To see that this holds, note that applying Lemma 2.3.16 with $H = G\backslash M$, $W = V(G)$ and $B = \{x_0\}$ implies that $G' - \{x_0\}$ is connected. But the unique component C of $G' - \{x_0\}$ satisfies $r|C| + e_{G'}(C, \{x_0\}) = r(n-1) + r = rn$, and since r is even C cannot be odd.

Case 2: $|S| > 0$ and $\rho \leqslant 1/2$. Since $e_{G'}(S,T) \leqslant \overline{d}_{G'}(T)|T|$ and $|S| \geqslant 2|T|$, we have

$$R_r(S,T) \geqslant r(|S| - |T|) \geqslant \frac{r}{3}(|S| + |T|) \geqslant |S| + |T| \geqslant Q_r(S,T),$$

where the last inequality follows by Corollary 2.3.17.

Case 3: $|S| > 0$, $\rho \geqslant 1/2$ and $\left(\frac{1}{s} + \frac{1}{\rho s}\right)\frac{\log n}{p} \geqslant \frac{7}{2}$. In this case we have that $e_{G'}(S,T) \leqslant e_G(S,T) + (\rho+1)s \leqslant 2(\rho+1)s\log n + (\rho+1)s$, where the last inequality follows from Definition 2.3.5(b). So

$$R_r(S,T) - Q_r(S,T) \geqslant \rho s(\overline{d}_{G'}(T) - 2\log n - r - 2) + s(r - 2\log n - 2), \qquad (2.3.19)$$

and it suffices to prove that the right-hand side of (2.3.19) is non-negative. Now observe that if $\rho \leqslant \frac{r - 2\log n - 2}{2\log n + 2}$, then this is immediate since $\overline{d}_{G'}(T) \geqslant r$. On the other hand, if $\rho \geqslant \frac{r-2\log n - 2}{2\log n + 2}$, then $|T| \geqslant \frac{r - 2\log n - 2}{2\log n + 2} \geqslant \frac{r}{3\log n} \geqslant 3\log n$. So by Definition 2.3.5(d) we have

$\overline{d}_{G'}(T) \geqslant r + 3\log n - 1 \geqslant r + 2\log n + 2$, and the result follows.

Case 4: $|S| > 0$, $\rho \geqslant 1/2$, $\left(\frac{1}{s} + \frac{1}{\rho s}\right)\frac{\log n}{p} \leqslant \frac{7}{2}$ and $\rho s \leqslant n/30$. In this case Definition 2.3.5(b) implies that $e_{G'}(S,T) \leqslant 7\rho s^2 p + u$. So

$$R_r(S,T) - Q_r(S,T) \geqslant \rho s(\overline{d}_{G'}(T) - r - 1) + s(r - 7\rho sp - u/s - 1),$$

and it suffices to prove that the right-hand side of this inequality is non-negative. Recall that $\overline{d}_{G'}(T) \geqslant r+1$ as $|T| = \rho s \geqslant 2(\rho+1)\log n/7p \geqslant 2$, and so the first bracket is non-negative. Also $7\rho sp \leqslant np/4 < r/3$ by Definition 2.3.5(c), and $u/s < u < r/3$. Hence the second bracket is positive.

Case 5: $|S| > 0$, $\rho \geqslant \frac{1}{2}$ and $\rho s \geqslant n/30$. Since G is $(p, 2\sqrt{np(1-p)})$-jumbled, (2.3.2) implies that

$$e_G(S,T) \leqslant \rho s^2 p + 4\sqrt{np(1-p)}s(\rho+1).$$

Note that $\overline{d}_{G'}(T)|T| - e_{G'}(S,T) \geqslant \overline{d}_G(T)|T| - e_G(S,T) - u$. Now we claim that $\overline{d}_G(T)|T| - u \geqslant (np - 130\sqrt{np(1-p)})|T|$. Indeed,

$$\begin{aligned}
\overline{d}_G(T)|T| &= 2e_G(T) + e_G(T, V(G)\backslash T) \\
&\overset{(2.3.2)}{\geqslant} 2p\binom{|T|}{2} + p|T|(n - |T|) - 4\sqrt{np(1-p)}|T| - 4\sqrt{np(1-p)}n \\
&\geqslant np|T| - p|T| - (4n + 4|T|)\sqrt{np(1-p)} \geqslant (np - 130\sqrt{np(1-p)})|T| + u.
\end{aligned}$$

Hence

$$\begin{aligned}
R_r(S,T)/s &\geqslant \rho(np - 130\sqrt{np(1-p)}) - \rho sp - 4\sqrt{np(1-p)}(\rho+1) + r(1-\rho) \\
&\geqslant \rho(np - r - 142\sqrt{np(1-p)}) + r - \rho sp.
\end{aligned}$$

But

$$\delta(G) - 1 \leqslant r \leqslant \delta(G) + u \leqslant np - 200\sqrt{np(1-p)} + 4\sqrt{np(1-p)}$$
$$= np - 196\sqrt{np(1-p)}.$$

Hence

$$R_r(S,T)/s \geqslant 54\rho\sqrt{np(1-p)} + \delta(G) - 1 - \rho sp \geqslant 50\rho\sqrt{np(1-p)} + \delta(G) - \rho sp.$$

Further as $(\rho+1)s \leqslant n$, we have

$$\delta(G) - \rho sp \geqslant np - 2\sqrt{np\log n} - np\left(\frac{\rho}{\rho+1}\right) = \frac{np}{\rho+1} - 2\sqrt{np\log n}.$$

Now if $\rho \leqslant \frac{1}{6}\sqrt{\frac{np}{\log n}}$ then $\frac{np}{\rho+1} \geqslant \frac{np}{3\rho} \geqslant 2\sqrt{np\log n}$ and so $R_r(S,T)/s \geqslant 50\rho\sqrt{np(1-p)} \geqslant \rho + 1$.

On the other hand if $\rho \geqslant \frac{1}{6}\sqrt{\frac{np}{\log n}} \geqslant \sqrt{\frac{\log n}{1-p}}$ then $2\rho\sqrt{np(1-p)} \geqslant 2\sqrt{np\log n}$ and so $R_r(S,T)/s \geqslant 48\rho\sqrt{np(1-p)} + np/(\rho+1) \geqslant \rho + 1$. □

If we only require a $2\lfloor\delta(G_{n,p})/2\rfloor$-factor in $G_{n,p}$, then a less restrictive assumption on p will suffice, as per the following corollary.

Corollary 2.3.20 *Let $p = p(n)$ be such that $np/\log^6 n \to \infty$ and $n(1-p)/\log^6 n \to \infty$. Then whp, $G_{n,p}$ contains a regular spanning subgraph of degree $2\lfloor\delta(G_{n,p})/2\rfloor$.*

Proof. It follows from Lemmas 2.3.7–2.2.2, Lemma 2.3.14 and Corollary 3.13 of [15] that $G_{n,p}$ is p-pseudorandom for the given range of p. Thus we may apply Lemma 2.3.18 with $u = 0$ to obtain the desired subgraph. □

We remark that the results of [90] and [84] imply that Corollary 2.3.20 in fact holds for the entire range of p. Indeed, as described in Section 1.3 these results even give a collection of $\lfloor \delta(G_{n,p})/2 \rfloor$ Hamilton cycles, whose union is the desired subgraph.

2.3.4 Splitting into 2-factors

Our aim in this section is show that under certain conditions, an even-regular graph H can be decomposed into 2-factors so that the sum of the number of cycles in these 2-factors is not too large. We would like such a result to hold for arbitrary even-regular graphs H, but this does not seem feasible for the densities that we consider. So we first show that it suffices for H to be a spanning subgraph of a pseudorandom graph G (see Corollary 2.3.26), and then that it suffices for H to be 'partly' pseudorandom (see Corollary 2.3.30). More precisely, we show that for any even-regular graph H' of degree $r_{H'}$, the union of H' with an even-regular graph H which is close to being p-pseudorandom may be decomposed into 2-factors with few cycles in total, provided only that p is somewhat larger than $r_{H'}/n$.

In order to bound the number of cycles in each 2-factor we will transfer the problem to a bipartite setting and use the Egorychev-Falikman-Waerden theorem (see Theorem 2.2.19). Given a bipartite graph B on vertex classes $V = \{v_1, \dots, v_n\}$ and $V' = \{v'_1, \dots, v'_n\}$, and a matching M in B, let $D(M)$ be the digraph on $[n]$ formed by including an edge from a to b if and only if there is an edge of M between v_a and v'_b. Note that if P is a perfect matching then $D(P)$ is 1-regular. For an integer C, let $\mathcal{P}_C(B)$ be the set of perfect matchings P of B such that $D(P)$ has at least C cycles. Let $\mathcal{P}_{k,\ell}(B)$ be the set of perfect matchings P such that $D(P)$ has at least k cycles of length ℓ. We now use a counting argument to show that $\mathcal{P}_C(B)$ is small whenever C is large and B satisfies a very weak pseudorandomness condition. The proof builds on ideas from [42] and later developments in [81], as well as from Section 2.2 of this thesis.

Lemma 2.3.21 *Let B be an r_B-regular bipartite graph on vertex classes $V = \{v_1, \dots, v_n\}$ and $V' = \{v'_1, \dots, v'_n\}$. Let $\log^2 n \leqslant r_G \leqslant n$ and $\sqrt{r_G}\log n \leqslant r_B \leqslant r_G/2$, and let $C = 2n\sqrt{r_G \log n}/r_B$. Suppose that all $S \subseteq V$ and $T \subseteq V'$ such that $|S|, |T| \geqslant n/\sqrt{r_G}$*

satisfy $e_B(S,T) \leqslant \frac{5r_G(|S|+|T|)^2}{2n}$. *Let* $X = (\frac{r_B}{n})^n n!$. *Then* $|\mathcal{P}_C(B)| < X$.

A similar result was proved in [42] and also in Section 2.2 (see Lemma 2.2.17). In the proof of Lemma 2.3.21 we use the following result, which is part of the proof of Lemma 2 in [42] (see inequality (5) there).

Lemma 2.3.22 (Frieze and Krivelevich, [42]) *Let B be an r-regular bipartite graph on vertex classes V and V' of size n. Let $m < n$ and let $S \subseteq V$ and $T \subseteq V'$ each have size $n - m$. Let B' be the bipartite subgraph of B induced by S and T, and let Y_m be the number of perfect matchings of B'. Suppose that $e_B(S,T) \geqslant rn/2$. Then*

$$Y_m \leqslant \left(\frac{(n-2m)r + e_B(V\backslash S, V'\backslash T)}{n-m}\right)^{n-m} e^{-n+m}(3n)^{10n/r}.$$

Note that Lemma 2.3.22 is stated in [42] with an estimate m^2 for $e_B(V\backslash S, V'\backslash T)$, rather than explicitly including the term $e_B(V\backslash S, V'\backslash T)$. Since the graphs we consider are subgraphs of random graphs we will have $e_B(V\backslash S, V'\backslash T) \ll m^2$. This will allow us to obtain nontrivial bounds on $|\mathcal{P}_C(B)|$ even for very sparse graphs.

Proof. [Proof of Lemma 2.3.21] Let $k = n\log n/r_B$. For each ℓ, let $X_{k,\ell} = |\mathcal{P}_{k,\ell}(B)|$, and let $Y_{k,\ell}$ be the maximum number of perfect matchings between two subsets of the vertex classes of B, each of size $n - k\ell$. Let $\ell_0 = r_B/\sqrt{r_G \log n}$. Since $\ell_0 \geqslant \sqrt{\log n}$ we may take ℓ_0 to be an integer. Note that

$$k\ell_0 = \frac{n\sqrt{\log n}}{\sqrt{r_G}} = o(n). \tag{2.3.23}$$

We first derive a bound on $Y_{k,\ell}$, for $\ell \leqslant \ell_0$. Consider any $S \subseteq V$ and $T \subseteq V'$, each of size $n - k\ell$. We claim that $e_B(V\backslash S, V'\backslash T) \leqslant 10r_G k^2\ell_0^2/n$. Indeed, let $S' \subseteq V$ and $T' \subseteq V'$

be sets of size $k\ell_0$ such that $V\backslash S \subseteq S'$ and $V'\backslash T \subseteq T'$. Then

$$e_B(V\backslash S, V'\backslash T) \leqslant e_B(S',T') \leqslant \frac{5r_G(|S'|+|T'|)^2}{2n} = \frac{10r_Gk^2\ell_0^2}{n}.$$

Moreover, $|S| = |T| = n - o(n)$ by (2.3.23) and so $e_B(S,T) \geqslant r_Bn/2$. Hence by Lemma 2.3.22,

$$Y_{k,\ell} \leqslant \left(\frac{(n-2k\ell)r_B + 10r_Gk^2\ell_0^2/n}{n-k\ell}\right)^{n-k\ell} e^{-n+k\ell}(3n)^{10n/r_B}.$$

We can further estimate this as follows: Observe that

$$\begin{aligned}\frac{(n-2k\ell)r_B + 10r_Gk^2\ell_0^2/n}{n-k\ell} &= r_B\left(1 - \frac{k\ell}{n-k\ell} + \frac{10r_Gk^2\ell_0^2}{nr_B(n-k\ell)}\right)\\ &\leqslant r_B \exp\left[-\frac{k\ell}{n-k\ell} + \frac{10r_Gk^2\ell_0^2}{nr_B(n-k\ell)}\right],\end{aligned}$$

and hence

$$\begin{aligned}Y_{k,\ell} &\leqslant r_B^{n-k\ell}e^{-k\ell}\exp\left[\frac{10r_Gk^2\ell_0^2}{nr_B}\right]e^{-n+k\ell}(3n)^{10n/r_B}\\ &= r_B^{n-k\ell}e^{10k}e^{-n}(3n)^{10n/r_B},\end{aligned} \tag{2.3.24}$$

where the last line follows from the fact that

$$\frac{r_Gk\ell_0^2}{nr_B} \overset{(2.3.23)}{=} \frac{r_G}{nr_B}\cdot\frac{n\sqrt{\log n}}{\sqrt{r_G}}\cdot\frac{r_B}{\sqrt{r_G\log n}} = 1.$$

Now we proceed to bound $X_{k,\ell}$. Note that if M_1 and M_2 are matchings such that $D(M_1)$ and $D(M_2)$ are cycles, then M_1 and M_2 are vertex-disjoint if and only if $D(M_1)$ and $D(M_2)$ are vertex-disjoint. We claim that there are at most $\binom{n}{k}r_B^{k(\ell-1)}\ell^{-k}$ ways of choosing k vertex-disjoint matchings M of size ℓ in B, such that $D(M)$ is a cycle of length ℓ for each M. Indeed, we can construct these matchings as follows: First choose an initial vertex in V for each matching ($\binom{n}{k}$ choices). For an initial vertex v_{i_1} of M, select an unused neighbour $v'_{i_2} \in V'$ of v_{i_1} (at most r_B choices) and add $v_{i_1}v'_{i_2}$ to M. Then select an unused neighbour v'_{i_3} of v_{i_2} (recall that v_{i_2} is the vertex of V corresponding to v'_{i_2}), etc., until we have selected v'_{i_ℓ}. Now if v'_{i_1} is a neighbour of v_{i_ℓ} then add $v_{i_\ell}v'_{i_1}$ to M and note that $D(M)$ is a cycle of length ℓ. Repeat this process for each initial vertex, in each case using only vertices we have not used before. This process described above constructs every collection of k vertex-disjoint matchings of size ℓ at least ℓ^k times, since we choose the initial vertex of each matching arbitrarily. The claim follows immediately.

Since $Y_{k,\ell}$ bounds the number of perfect matchings on the remaining vertices (i.e., those which are not contained in the above matchings), we have

$$X_{k,\ell} \leqslant \binom{n}{k} r_B^{k(\ell-1)} \ell^{-k} Y_{k,\ell}.$$

Estimating $\binom{n}{k} \leqslant \left(\frac{ne}{k}\right)^k$ and $n! \geqslant \left(\frac{n}{e}\right)^n$, we have

$$\begin{aligned} X_{k,\ell}X^{-1} &\overset{(2.3.24)}{\leqslant} \left(\frac{ne}{k}\right)^k r_B^{k(\ell-1)} \ell^{-k} r_B^{n-k\ell} e^{10k} e^{-n} (3n)^{10n/r_B} \left(\frac{n}{r_B}\right)^n \left(\frac{e}{n}\right)^n \\ &= \left(\frac{ne^{11}}{k}\right)^k r_B^{k\ell-k} \ell^{-k} r_B^{n-k\ell} (3n)^{10n/r_B} r_B^{-n} = \left(\frac{ne^{11}}{k\ell r_B}\right)^k (3n)^{10n/r_B} \\ &= \left(\frac{e^{11}}{\ell \log n}\right)^k (3n)^{10n/r_B}. \end{aligned}$$

Hence

$$X^{-1} \sum_{\ell=3}^{\ell_0} X_{k,\ell} \leqslant \left(\left(\frac{e^{11}}{\log n}\right)^{\log n/10} 3n \right)^{10n/r_B} \sum_{\ell=3}^{\ell_0} \ell^{-k}.$$

But $((e^{11}/\log n)^{\log n/10} 3n)^{10n/r_B} < 2^{-10n/r_B} < 1$ and $\ell^{-k} \leqslant 3^{-\log n} < 1/n$ for $\ell \geqslant 3$. Hence

$$X^{-1} \sum_{\ell=3}^{\ell_0} X_{k,\ell} < \sum_{\ell=3}^{\ell_0} 1/n < \ell_0/n < 1.$$

Note that since $k\ell_0 \leqslant C/2$ and $n/\ell_0 = C/2$, we have $C \geqslant k\ell_0 + n/\ell_0$. But any 1-regular digraph which, for all $\ell \leqslant \ell_0$, contains fewer than k cycles of length ℓ has fewer than $k\ell_0 + n/\ell_0$ cycles in total. Hence $|\mathcal{P}_C(B)| \leqslant \sum_{\ell=3}^{\ell_0} X_{k,\ell} < X$. $\square$

The following lemma and corollary accomplish our aim of decomposing an even-regular graph H which is close to being pseudorandom into 2-factors with few cycles in total.

Lemma 2.3.25 *Let $r_G \geqslant \log^2 n$, and let G be an r_G/n-pseudorandom graph on n vertices. Let $r_H \geqslant 2\sqrt{r_G}\log n$ be even and let H be an r_H-regular spanning subgraph of G. Then H contains a 2-factor with at most $C = 4n\sqrt{r_G \log n}/r_H$ cycles.*

Proof. Recall that Petersen's theorem [100] states that every even-regular graph can be decomposed into 2-factors. Therefore H can be decomposed into a collection $F_1, \dots, F_{r_B}$ of 2-factors, where $r_B = r_H/2$. For each of these 2-factors we orient the edges so that each cycle is an oriented cycle, thus forming a collection of 1-regular digraphs on $V(H)$. Now taking the union of these digraphs yields an orientation D of H which is r_B-regular.

Label the vertices of $V(H)$ as $v_1, \dots, v_n$. Form a bipartite graph B on vertex classes $V = \{v_1, \dots, v_n\}$ and $V' = \{v'_1, \dots, v'_n\}$ by joining $v_i \in V$ to $v'_j \in V'$ if and only if there is an edge of D from v_i to v_j. Now for any sets $S \subseteq V$ and $T \subseteq V'$ with $|S|, |T| \geqslant n/\sqrt{r_G}$, let S', T' be the corresponding subsets of $V(H)$. Then

$$\begin{aligned} e_B(S,T) \leqslant e_H(S' \cup T') \leqslant e_G(S' \cup T') &\leqslant \frac{r_G(|S|+|T|)^2}{2n} + 2\sqrt{r_G}(|S|+|T|) \\ &\leqslant \frac{3r_G(|S|+|T|)^2}{2n}, \end{aligned}$$

where the third inequality follows since G is $(r_G/n, 2\sqrt{r_G})$-jumbled by Definition 2.3.5(a), and the fourth since $2 \leqslant \sqrt{r_G}(|S|+|T|)/n$. Now Lemma 2.3.21 implies that $|\mathcal{P}_C(B)| < \left(\frac{r_B}{n}\right)^n n!$. But by Theorem 2.2.19 the total number of perfect matchings of B is at least $\left(\frac{r_B}{n}\right)^n n!$. So there exists a perfect matching P of B such that $D(P)$ has at most C cycles. Now ignoring the orientation of $D(P)$ yields the desired 2-factor of H. $\square$

For a 2-regular graph F, let $c(F)$ be the number of cycles of F, and for a collection $\mathcal{F}$ of 2-factors let $c(\mathcal{F}) = \sum_{F\in\mathcal{F}} c(F)$. Repeated application of Lemma 2.3.25 gives us the following result.

Corollary 2.3.26 *Let G be an r_G/n-pseudorandom graph on n vertices with $r_G \geqslant \log^2 n$. Let H be an r_H-regular spanning subgraph of G, such that r_H is even. Then H can be decomposed into a collection of 2-factors $\mathcal{F} = \{F_1, \ldots, F_{r_H/2}\}$, such that $c(\mathcal{F}) \leqslant 3n\sqrt{r_G \log^3 n}$.*

Proof. If $r_H \leqslant 8\sqrt{r_G}\log n$ then the result follows from Petersen's theorem [100], noting that trivially $c(F_i) \leqslant n/3$ for each i. So suppose $r_H \geqslant 8\sqrt{r_G}\log n$. Let $i_0 = r_H/2 - \sqrt{r_G}\log n$ and note that since $i_0 \geqslant r_H/4 \geqslant \log^2 n$, we may assume that i_0 is an integer. By repeatedly applying Lemma 2.3.25 we have

$$\begin{aligned} c(\mathcal{F}) &\leqslant \sum_{i=1}^{i_0} \frac{2n\sqrt{r_G \log n}}{r_H/2 - i + 1} + \sum_{i=i_0+1}^{r_H/2} n/3 \leqslant 2n\sqrt{r_G \log n} \sum_{i=\sqrt{r_G}\log n}^{r_H/2} \frac{1}{i} + n\sqrt{r_G}\log n \\ &\leqslant 2n\sqrt{r_G \log n}(\log(r_H/2) + \sqrt{\log n}) \leqslant 3n\sqrt{r_G \log^3 n}. \end{aligned}$$

□

Now we consider an arbitrary even-regular graph H' and aim to show that if we take an even-regular graph H on $V(H')$ which is close to being pseudorandom and edge-disjoint from H', then we can decompose $H \cup H'$ into 2-factors with a similar bound on the total number of cycles as in Corollary 2.3.26. Our strategy will be to first split H' up into matchings. We then extend each of these matchings into two perfect matchings using edges of H (see Lemma 2.3.27), and apply Lemma 2.3.21 to transform each of these perfect matchings into a 2-factor with few cycles (see Lemma 2.3.28).

We start by considering the case of a single matching. If n is odd, then call a graph F on n vertices a *pseudomatching* if it has a unique vertex of degree 2 and all other vertices are of degree 1. A perfect pseudomatching in a graph G is a pseudomatching covering every vertex of G.

Lemma 2.3.27 *Let G be an r_G/n-pseudorandom graph on n vertices with $r_G \geqslant \log^2 n$ and let H be a spanning subgraph of G such that $\delta(H) \geqslant \frac{549}{550} r_G + 5$. Let M be a matching with $V(M) \subseteq V(G)$ and suppose M and H are edge-disjoint. Then*

(i) *if n is even, then there exists a partition of M into submatchings M_1, M_2, such that each of M_1, M_2 can be completed to a perfect matching on $V(G)$ using disjoint sets of edges of H, and*

(ii) *if n is odd, then there exists a partition of M into submatchings M_1, M_2, such that each of M_1, M_2 can be completed to a perfect pseudomatching on $V(G)$ using disjoint sets of edges of H.*

It would be more convenient to extend M directly into a perfect matching rather than splitting it first, but in general this may not be possible. Indeed, in the case in which n is even and M has exactly $n/2 - 1$ edges, then M can be extended to a perfect matching only if the two vertices of $V(G)\backslash V(M)$ form an edge of H.

Proof. Let $r_H = \frac{549}{550} r_G$. We partition M randomly by placing each edge into either M_1 or M_2, with equal probability. For a vertex $v \in V(G)$, let Y_v be the set of edges ww' of M such that both w and w' lie in $N_H(v)$, and Z_v the set such that exactly one of w and w' lies in $N_H(v)$. Note that $2|Y_v| + |Z_v| \leqslant d_H(v) \leqslant \Delta(G) \leqslant 2r_H$, and that

$$|N_H(v)\backslash V(M_1)| \sim 2Bin(|Y_v|, 1/2) + Bin(|Z_v|, 1/2) + |N_H(v)\backslash V(M)|.$$

(Here we use that M and H are edge-disjoint and so the unique neighbour of v in M is not a neighbour of v in H.)

Suppose $|Y_v| \geqslant 8\log n$. Let $\varepsilon = 4\sqrt{\log n/|Y_v|} \leqslant 3/2$ and note that $2\sqrt{r_H \log n} \geqslant \varepsilon|Y_v|/2$, since $r_H \geqslant |Y_v|$. By Lemma 1.8.1(i) we have that

$$\mathbb{P}\left[Bin(|Y_v|, 1/2) \leqslant |Y_v|/2 - 2\sqrt{r_H \log n}\right] < 2e^{-\varepsilon^2|Y_v|/6} \leqslant 4e^{-16\log n/6} < \frac{1}{n^2}.$$

On the other hand, if $|Y_v| \leqslant 8\log n$ then $|Y_v|/2 \leqslant 2\sqrt{r_H \log n}$ and so

$$\mathbb{P}\left[Bin(|Y_v|, 1/2) \leqslant |Y_v|/2 - 2\sqrt{r_H \log n}\right] = 0.$$

By the same argument we can show that

$$\mathbb{P}\left[Bin(|Z_v|, 1/2) \leqslant |Z_v|/2 - 4\sqrt{r_H \log n}\right] < \frac{1}{n^2}.$$

So whp, every $v \in V(G)$ satisfies

$$\begin{aligned}|N_H(v)\backslash V(M_1)| &\geqslant |Y_v| + |Z_v|/2 - 8\sqrt{r_H \log n} + |N_H(v)\backslash V(M)| \\ &= \frac{1}{2}|N_H(v) \cap V(M)| + |N_H(v)\backslash V(M)| - 8\sqrt{r_H \log n} \geqslant r_H/3 + 7,\end{aligned}$$

and hence $\delta(H - V(M_1)) \geqslant r_H/3 + 7$. By a similar argument the same holds for M_2. Now choose a partition of M into M_1 and M_2 such that the above bounds on $|N_H(v)\backslash V(M_1)|$ and $|N_H(v)\backslash V(M_2)|$ hold for all vertices $v \in V(G)$.

If n is odd, then we first find a path $v_1v_2v_3$ of length 2 in $H - V(M_1)$ and add it to M_1 to form a pseudomatching, so that $V(H)\backslash V(M_1)$ has an even number of vertices. We also remove v_1v_2 and v_2v_3 from H. (If n is even then we simply omit this step.) Henceforth we proceed identically in cases (i) and (ii).

Let $W = V(H)\backslash V(M_1)$ and note that we still have $\delta(H) \geqslant r_H + 3$ and $\delta(H[W]) \geqslant r_H/3 + 4$. We now use Theorem 2.3.15 to show that $H[W]$ contains a perfect matching.

For this it suffices to prove that for any set $S \subseteq W$, the number of components of $H[W \backslash S]$ which have an odd number of vertices is at most $|S|$. If S is nonempty then this follows from Lemma 2.3.16. On the other hand if $S = \emptyset$ then by Lemma 2.3.16 $H[W \backslash S] = H[W]$ has exactly one component, namely W. But this component has an even number of vertices, and so by Theorem 2.3.15 $H[W]$ contains a perfect matching M_1'. So $M_1 \cup M_1'$ is a perfect matching or pseudomatching on $V(G)$. We delete the edges in M_1' from H.

We now repeat the process for M_2. First we extend M_2 to a pseudomatching (if n is odd). Now we still have $\delta(H) \geqslant r_H$ and $\delta(H - V(M_2)) \geqslant r_H/3$, so the conditions of Lemma 2.3.16 still hold and hence we can complete M_2 to a perfect matching or pseudomatching on $V(G)$. □

Lemma 2.3.28 *Let G be an r_G/n-pseudorandom graph on n vertices with $r_G \geqslant 2\log^2 n$, and let H be a spanning subgraph of G such that $\delta(H) \geqslant \frac{29}{30} r_G$. Let P be a perfect matching (if n is even) or a perfect pseudomatching (if n is odd) on $V(G)$, edge-disjoint from H. Then H contains a matching P' such that $P \cup P'$ forms a 2-regular graph with at most $\frac{3n\sqrt{r_G \log n}}{\delta(H)}$ cycles.*

We will apply Lemma 2.3.28 in the proof of Lemma 2.3.29 with P being one of the perfect matchings M_i obtained from Lemma 2.3.27.

Proof. Let $r_H = \delta(H)$ and note that $r_H \geqslant \log^2 n$ by Definition 2.3.5(c). If n is odd, then let v_1 be the vertex of P of degree 2 and v_2, v_3 its neighbours. For the remainder of the proof we treat v_2v_3 as if it were an edge of P, and v_1v_2, v_1v_3 as if they were not edges of P. If v_2v_3 is also an edge of H, then we delete it from H (so that P and H remain edge-disjoint). The proof then proceeds identically whether n is odd or even.

Label the edges of P as $e_1, e_2, \ldots, e_{\lfloor n/2 \rfloor}$. For each edge e_i, label one of the vertices a_i (chosen at random and independently from the other choices) and the other b_i. Let

$B_1 = \{a_1, \ldots, a_{\lfloor n/2 \rfloor}\}$ and $B_2 = \{b_1, \ldots, b_{\lfloor n/2 \rfloor}\}$, and let B be the bipartite graph on vertex classes B_1, B_2 with edges $E_H(B_1, B_2)$.

Now for each $v \in V(B)$, let Y_v be the set of edges $ab \in P$ such that both a and b lie in $N_H(v)$, and let Z_v be the set of edges $ab \in P$ such that exactly one of a and b lies in $N_H(v)$. Note that

$$d_H(v) - 1 \leqslant 2|Y_v| + |Z_v| \leqslant d_H(v) \leqslant \Delta(H) \leqslant \Delta(G) \leqslant r_G + 2\sqrt{r_G \log n} \leqslant \frac{11r_H}{10},$$

where the second-last inequality follows from Definition 2.3.5(c). Now $d_B(v) \sim |Y_v| + Bin(|Z_v|, 1/2)$. (Here we use the condition that P and H are edge-disjoint and so the unique neighbour of v in P is not a neighbour of v in H.)

Suppose $|Z_v| \geqslant 8 \log n$. Let $\varepsilon = 4\sqrt{\log n / |Z_v|} \leqslant 3/2$ and note that $4\sqrt{r_H \log n} \geqslant \varepsilon |Z_v|/2$. By Lemma 1.8.1(i) we have that

$$\mathbb{P}\left[Bin(|Z_v|, 1/2) \leqslant |Z_v|/2 - 4\sqrt{r_H \log n}\right] < 2e^{-\varepsilon^2 |Z_v|/6} \leqslant 4e^{-16 \log n/6} < \frac{1}{n^2}.$$

On the other hand if $|Z_v| \leqslant 8 \log n$ then $|Z_v|/2 \leqslant 4\sqrt{r_H \log n}$ and so

$$\mathbb{P}\left[Bin(|Z_v|, 1/2) \leqslant |Z_v|/2 - 4\sqrt{r_H \log n}\right] = 0.$$

So whp every $v \in V(B)$ satisfies $d_B(v) \geqslant |Y_v| + |Z_v|/2 - 4\sqrt{r_H \log n} \geqslant r_H/2 - 5\sqrt{r_H \log n}$. Similarly, whp every $v \in V(B)$ satisfies $d_B(v) \leqslant 11r_H/20 + 5\sqrt{r_H \log n}$. Now we choose B such that $\delta(B) \geqslant r_H/2 - 5\sqrt{r_H \log n}$ and $\Delta(B) \leqslant 11r_H/20 + 5\sqrt{r_H \log n}$. Let $r_B = 2r_H/5$ and note that $r_B \geqslant r_G/3$.

Claim: *B contains a regular spanning subgraph B' of degree r_B.*

To prove the claim, we use the Max-flow Min-cut theorem. Let each edge of B have capacity 1. Add a source σ joined to each vertex of B_1 by an edge of capacity r_B and a sink τ joined to each vertex of B_2 by an edge of capacity r_B. Let $n' = |B_1| = \lfloor n/2 \rfloor$. Now we show that the minimum cut must have capacity at least $n'r_B$ (indeed exactly $n'r_B$, since one could cut all of the edges incident to σ). It follows that the maximum flow from σ to τ is $n'r_B$. Further there is some flow which achieves this maximum and such that the flow along each edge is an integer. Taking the edges of B which have flow 1 immediately yields the desired r_B-regular spanning subgraph.

So consider a cut $\mathcal{C}$. Let T be the set of vertices $v \in B_1$ such that σv is not part of $\mathcal{C}$. Similarly let S be the set of vertices $v \in B_2$ such that $v\tau$ is not part of $\mathcal{C}$. Now since every edge in $E_B(T, S)$ must be part of $\mathcal{C}$, the capacity of $\mathcal{C}$ is at least $e_B(T, S)+r_B(2n'-|S|-|T|)$. So noting that $e_B(T, S) = e_H(T, S)$, it suffices to prove that

$$e_H(T, S) \geqslant n'r_B - r_B(2n' - |S| - |T|) = r_B(|S| + |T| - n').$$

Let $S' = B_2 \backslash S$. Then an equivalent statement is that

$$e_H(T, B_2) - e_H(T, S') + r_B(|S'| - |T|) \geqslant 0.$$

We can assume without loss of generality that $|T| + |S'| \leqslant n' \leqslant n/2$; otherwise we can rearrange the inequality as $e_H(B_1, S) - e_H(T', S) + r_B(|T'| - |S|) \geqslant 0$ where $T' = B_1 \backslash T$ and the proof proceeds analogously. We now consider the following cases:

Case 1: $|T| \geqslant 2|S'|$. We can estimate $e_H(T, S') \leqslant |S'|\Delta(B)$ and $e_H(T, B_2) \geqslant |T|\delta(B)$. Now

$$e_H(T, B_2) - e_H(T, S') + r_B(|S'| - |T|) \geqslant |T|(\delta(B) - r_B) - |S'|(\Delta(B) - r_B),$$

and the right-hand side is positive since $\delta(B) - r_B \geqslant r_H/10 - 5\sqrt{r_H \log n} > r_H/12$ and $\Delta(B) - r_B \leqslant 3r_H/20 + 5\sqrt{r_H \log n} < r_H/6$.

Case 2: $|S'| \geqslant |T|$. Then the inequality holds trivially.

Case 3: $|S'| \leqslant |T| \leqslant 2|S'|$. Note that $|T| \leqslant n/3$ (otherwise $|T| + |S'| > n/2$). So since G is $(r_G/n, 2\sqrt{r_G})$-jumbled by Definition 2.3.5(c), (2.3.2) implies that

$$\begin{aligned} r_B|S'| - e_H(T, S') &\geqslant r_B|S'| - r_G|T||S'|/n - 4\sqrt{r_G}(|T| + |S'|) \\ &\geqslant |S'|(r_B - r_G|T|/n - 12\sqrt{r_G}) \geqslant |S'|(r_G/3 - r_G/3 - 12\sqrt{r_G}) \\ &= -12|S'|\sqrt{r_G} \geqslant -12|T|\sqrt{r_G}. \end{aligned}$$

Hence $e_H(T, B_2) - e_H(T, S') + r_B(|S'| - |T|) \geqslant |T|\delta(B) - r_B|T| - 12|T|\sqrt{r_G} \geqslant 0$, which completes the proof of our claim.

Now Theorem 2.2.19 implies that the number of perfect matchings of B' is at least $X = (\frac{r_B}{n'})^{n'}(n')!$. Note that for any $S \subseteq B_1$ and $T \subseteq B_2$ with $|S|, |T| \geqslant n'/\sqrt{r_G}$,

$$\begin{aligned} e_{B'}(S, T) \leqslant e_G(S, T) &\overset{(2.3.2)}{\leqslant} r_G|S||T|/n + 4\sqrt{r_G}(|S| + |T|) \\ &\leqslant r_G|S||T|/n' + 2r_G(|S| + |T|)^2/n' \\ &\leqslant 5r_G(|S| + |T|)^2/2n', \end{aligned}$$

where in the third inequality we use that $2 \leqslant \sqrt{r_G}(|S| + |T|)/n'$ and in the last inequality we use that $2|S||T| \leqslant (|S|+|T|)^2$. Set $C = 2n'\sqrt{r_G \log n'}/r_B$. Now applying Lemma 2.3.21 with $V = B_1$, $V' = B_2$, and $v_i = a_i$ and $v'_i = b_i$ for each $1 \leqslant i \leqslant n'$ implies that $|\mathcal{P}_C(B')| < X$. So there exists a perfect matching P' in B' such that $D(P')$ has at most C cycles. (Recall that $D(P')$ was defined in the paragraph before Lemma 2.3.21.) But we have a one-to-one correspondence between cycles $i_1 i_2 \dots i_\ell i_1$ of $D(P')$ and cycles $b_{i_1} a_{i_1} b_{i_2} a_{i_2} \dots b_{i_\ell} a_{i_\ell} b_{i_1}$ of $P \cup P'$. Hence $P \cup P'$ has at most C cycles. Now note that $C \leqslant 3n\sqrt{r_G \log n}/r_H$. Finally, if n is odd then on the cycle of $P \cup P'$ containing the edge $v_2 v_3$, we replace $v_2 v_3$ by the path $v_2 v_1 v_3$, so that $P \cup P'$ is 2-regular. $\square$

We now combine Lemmas 2.3.27 and 2.3.28 to show that under suitable conditions, the union of an arbitrary even-regular graph H' and a graph H which is close to being pseudorandom contains a collection of edge-disjoint 2-factors which together cover the edges of H'.

Lemma 2.3.29 *Let G be an r_G/n-pseudorandom graph on n vertices with $r_G \geqslant 2\log^2 n$, and let H be a spanning subgraph of G such that $\delta(H) \geqslant (1 - 1/1100)r_G$. Let H' be an arbitrary $r_{H'}$-regular graph on the same vertex set, and let $E_{bad} \subseteq E(H')$. Suppose that H' is edge-disjoint from H and that $r_{H'} + 1 + 10^6|E_{bad}|/n \leqslant r_G/5000$. Then there exists $t \leqslant r_G/5000$ and a collection $F_1, \ldots, F_{2t}$ of edge-disjoint 2-factors in $H \cup H'$ whose union covers all of the edges of H', such that each F_i has at most $4n\sqrt{\log n}/\sqrt{r_G}$ cycles and each set $E(F_i) \cap E_{bad}$ is a matching of size at most $n/10^6$.*

We will apply Lemma 2.3.29 with E_{bad} being a set of 'bad' edges which we will want to avoid when merging the cycles of each F_i into a Hamilton cycle. The purpose of the restrictions on $F_i \cap E_{bad}$ is to spread the edges in E_{bad} out among the F_i's and thus make them easier to avoid.

Proof. Let $r_H = (1 - 1/550)r_G + 5$. By Vizing's theorem, we can decompose $E(H')$ into edge-disjoint matchings $M'_1, M'_2, \ldots, M'_{r_{H'}+1}$. For each i, split M'_i into $\lfloor 10^6|E(M'_i) \cap E_{bad}|/n + 1\rfloor$ matchings M_j such that $|E(M_j) \cap E_{bad}| \leqslant n/10^6$. Let $M_1, \ldots, M_m$ be the resulting collection of matchings and note that $m \leqslant r_{H'} + 1 + 10^6|E_{bad}|/n \leqslant r_G/5000$. Now for each matching M_j, we will find a pair of edge-disjoint 2-factors in $H \cup M_j$, each with at most $4n\sqrt{\log n}/\sqrt{r_G}$ cycles, which together cover the edges of M_j. Thus the total number of 2-factors will be $2m$.

Firstly observe that for any $v \in V(H)$, no more than $4t - r_{H'} \leqslant r_G/1200$ edges of H incident to v will be used during this process to construct our 2-factors (as each 2-factor uses up at most 2 edges of H incident to v). Hence after deleting all the edges of H lying in the 2-factors found so far, we still have $\delta(H) \geqslant (1-1/1100-1/1200)r_G \geqslant r_H$. For each M_j, we use Lemma 2.3.27 to decompose M_j into two matchings and complete each one to a perfect matching or pseudomatching using edges of H. Now by Lemma 2.3.28 we can complete each of these perfect matchings or pseudomatchings to a 2-factor using edges of H, such that each 2-factor produced has at most $3n\sqrt{r_G \log n}/r_H \leqslant 4n\sqrt{\log n}/\sqrt{r_G}$ cycles. $\square$

Finally we combine Corollory 2.3.26 and Lemma 2.3.29 to fully decompose $H \cup H'$ into 2-factors with few cycles in total.

Corollary 2.3.30 *Let G be an r_G/n-pseudorandom graph with $r_G \geqslant 2\log^2 n$ and let H be an even-regular spanning subgraph of G of degree r_H, with $\delta(G)-1 \leqslant r_H \leqslant \delta(G)$. Let H' be an arbitrary even-regular graph of degree $r_{H'}$ on the same vertex set and let $E_{bad} \subseteq E(H')$. Suppose that H' is edge-disjoint from H and that $r_{H'} + 1 + 10^6|E_{bad}|/n \leqslant r_G/5000$. Let $t = (r_H + r_{H'})/2$. Then there exists a decomposition $\mathcal{F} = \{F_1, \dots, F_t\}$ of $H \cup H'$ into 2-factors such that $c(\mathcal{F}) \leqslant 4n\sqrt{r_G \log^3 n}$ and $E(F_i) \cap E_{bad}$ is a matching of size at most $n/10^6$ for each i.*

Proof. Apply Lemma 2.3.29 to construct a collection $\mathcal{F}'$ of at most $r_G/2500$ edge-disjoint 2-factors in $H \cup H'$, such that $\bigcup \mathcal{F}'$ covers H',

$$c(\mathcal{F}') \leqslant \frac{4n\sqrt{\log n}r_G}{2500\sqrt{r_G}} \leqslant n\sqrt{r_G \log n},$$

and the set $E(F) \cap E_{bad}$ is a matching of size at most $n/10^6$ for each $F \in \mathcal{F}'$. Let $H'' = (H \cup H') \setminus \bigcup \mathcal{F}'$ and note that H'' is an even-regular spanning subgraph of G. So by Corollary 2.3.26 with $H = H''$ we can decompose H'' into a collection $\mathcal{F}''$ of 2-factors such that $c(\mathcal{F}'') \leqslant 3n\sqrt{r_G \log^3 n}$. Taking $\mathcal{F} = \mathcal{F}' \cup \mathcal{F}''$, we have $c(\mathcal{F}) \leqslant 4n\sqrt{r_G \log^3 n}$. $\square$

2.3.5 Merging cycles

So far we have the necessary tools to find large collections of disjoint 2-factors in a pseudorandom graph G, by first finding a regular spanning subgraph and then decomposing this subgraph into a collection $\mathcal{F}$ of 2-factors. The aim of this section is to show that we can transform these 2-factors into Hamilton cycles, by 'merging' the cycles of each 2-factor together using edges which are taken from a pseudorandom graph G'. The crucial observation here is that G' need not be as dense as G originally was; in fact, under certain conditions G' may be taken to be significantly sparser. To establish this we make use of the bounds on $c(\mathcal{F})$ proved in Section 2.3.4.

Recall that $N(S)$ denotes the external neighbourhood of S, i.e. $N(S) = \bigcup_{s\in S} N(s)\backslash S$. All paths P are considered to have a 'direction' in the sense that the first and last endpoints are distinguished (so if $P = v_1 \dots v_\ell$ and $P' = v_\ell \dots v_1$ then we view P and P' as different paths). If $P = x \dots y$ then we call x the *first endpoint* and y the *last endpoint* of P. We define the *reverse* of a path $P = x \dots y$ to be the path $P' = y \dots x$ which has the same vertices and edges as P.

We will use the rotation-extension method. For this we need to recall Definitions 2.2.23 and 2.2.30, and also make the following additional definitions. Let

$$\tau_0 = \frac{\log n}{\log \log n} + 3. \tag{2.3.31}$$

In what follows, the sequences of rotations we consider will generally have length at most τ_0.

Definition 2.3.32 *Let $P = v_1 v_2 \dots v_\ell$ be a path, and let $C = w_1 w_2 \dots w_m w_1$ be a cycle which is vertex-disjoint from P. An* extension of P to incorporate C, *with* join vertex w_i *and* broken edge $w_{i-1}w_i$ *is the operation of deleting the edge $w_{i-1}w_i$ from C and adding the edge $v_1 w_i$ to form a new path $w_{i-1}w_{i-2} \dots w_{i+1} w_i v_1 \dots v_{\ell-1} v_\ell$. Call $v_1 w_i$ the* new edge *of the extension.*

Note that in effect we performed extensions in Section 2.2 as well. We make a formal definition here since we need to be quite careful about which edges we break.

Given a 2-regular graph F and an edge-disjoint graph H on the same vertex set, we say that a Hamilton cycle C is formed by *merging the cycles of F using edges of H* if $E(C) \subseteq E(F) \cup E(H)$. Our strategy will be to first merge two of the cycles of F together to form a long path P. We then show that if $Q \subseteq V(H)$ is 'large', the set of (H, Q, τ_0)-reachable vertices of P must in general be large (see Corollary 2.3.41). This will allow us to either extend the path by incorporating another cycle, or close P to a cycle, thus reducing the number of cycles in F. Repeating this process will eventually produce a Hamilton cycle.

Recall that Lemma 2.2.31 gives us many reachable vertices provided that H 'expands', i.e., that $N_H(S) \cap Q$ is large (compared to S) for any $S \subseteq Q$ which is not itself too large. The fact that we have the required expansion property will follow from Lemma 2.3.35. Lemma 2.3.35 relies on Corollary 2.3.34, which in turn relies on Lemma 2.3.33. Lemma 2.3.33 is a simple consequence of pseudorandomness.

Lemma 2.3.33 *Let G be a p-pseudorandom graph on n vertices, and let $S, T \subseteq V(G)$ be disjoint with $s = |S|$ and $t = |T|$. Let*

$$g(s,t) = \begin{cases} 2(s+t)\log n & \text{if } \frac{\log n}{sp} \geqslant \frac{7}{2}; \\ 7s(s+t)p & \text{otherwise,} \end{cases}$$

and

$$h(s) = \begin{cases} 2s\log n & \text{if } \frac{\log n}{sp} \geqslant \frac{7}{2}; \\ 7s^2 p & \text{otherwise.} \end{cases}$$

Then $e_G(S,T) \leqslant g(s,t)$ and $e_G(S) \leqslant h(s)$.

Proof. Suppose first that $\frac{\log n}{sp} \geqslant \frac{7}{2}$. Then Definition 2.3.5(b)(i) implies that $e_G(S,T) \leqslant 2(s+t)\log n = g(s,t)$. If $\frac{\log n}{sp} \leqslant \frac{7}{2}$ and $\left(\frac{1}{s}+\frac{1}{t}\right)\frac{\log n}{p} \geqslant \frac{7}{2}$ then Definition 2.3.5(b)(i) implies that $e_G(S,T) \leqslant 2(s+t)\log n \leqslant 7s(s+t)p = g(s,t)$. Finally if $\left(\frac{1}{s}+\frac{1}{t}\right)\frac{\log n}{p} \leqslant \frac{7}{2}$ then Definition 2.3.5(b)(ii) implies that $e_G(S,T) \leqslant 7stp \leqslant 7s(s+t)p = g(s,t)$.

For the second part, if $\frac{7}{4} \leqslant \frac{\log n}{sp} \leqslant \frac{7}{2}$, then by Definition 2.3.5(b)(iii) we have $e_G(S) \leqslant 2s\log n \leqslant 7s^2p = h(s)$, and otherwise the result follows immediately from Definition 2.3.5(b)(iii) and (iv). $\square$

Corollary 2.3.34 *Let G be a p-pseudorandom graph on n vertices. Let $S \subseteq V(G)$ with $|S| = s$, and for each vertex $x \in S$, let $A_x \subseteq N_G(x)$. Let $T = \bigcup_{x\in S} A_x \backslash S$, and $t = |T|$. Then the following properties hold:*

(i) *If $s \leqslant \frac{2\log n}{7p}$ and $\sum_{x\in S}|A_x| \geqslant 12s\log n$, then $t \geqslant \frac{\sum_{x\in S}|A_x|}{4\log n}$.*

(ii) *If $s \geqslant \frac{2\log n}{7p}$, then $t + 3s \geqslant \frac{\sum_{x\in S}|A_x|}{7sp}$.*

Proof. (i) By Lemma 2.3.33 we have

$$\sum_{x\in S}|A_x| \leqslant e_G(S,T) + 2e_G(S) \leqslant g(s,t) + 2h(s) = 2(s+t)\log n + 4s\log n.$$

Hence $t \geqslant (\sum_{x\in S}|A_x| - 6s\log n)/2\log n \geqslant \sum_{x\in S}|A_x|/4\log n$.

(ii) We have

$$\sum_{x\in S}|A_x| \leqslant e_G(S,T) + 2e_G(S) \leqslant g(s,t) + 2h(s) = 7sp(s+t) + 14s^2p$$

and the result follows immediately. $\square$

Lemma 2.3.35 *Let $0 < \varepsilon < 1$ be a constant. Let G' be an $r_{G'}/n$-pseudorandom graph on n vertices and let G be an r_G/n-pseudorandom spanning subgraph of G', with $r_G \leqslant r_{G'}$*

and $\varepsilon r_G \geqslant 16\log^2 n$. *Let* H *be a spanning subgraph of* G *such that*

$$|E(G)\backslash E(H)| \leqslant 2n\sqrt{r_G \log n}. \tag{2.3.36}$$

Let H' *be a spanning subgraph of* G', *such that*

$$|E(H)\backslash E(H')| \leqslant \frac{\varepsilon^2 n r_G^2}{10^4 r_{G'}} \text{ and } |E(H')\backslash E(H)| \leqslant \frac{\varepsilon^2 n r_G^2}{10^4 r_{G'}}. \tag{2.3.37}$$

Let $Q', S \subseteq V(G)$ *and suppose that* $|N_{H'}(x) \cap Q'| \geqslant \varepsilon r_G$ *for every vertex* $x \in S$. *Then at least one of the following holds:*

(i) $\frac{1}{2}|N_{H'}(S) \cap Q'| - |S| \geqslant |S| \log n$,

(ii) $|S| \leqslant \frac{\varepsilon n r_G}{50 r_{G'}}$ *and* $\frac{1}{2}|N_{H'}(S) \cap Q'| - |S| \geqslant \frac{\varepsilon n r_G}{49 r_{G'}}$,

(iii) $|S| \leqslant \varepsilon n/90$ *and* $\frac{1}{2}|N_{H'}(S) \cap Q'| - |S| \geqslant \varepsilon n/45$,

(iv) $|S| > |Q'|/6$,

(v) $\frac{1}{2}|N_{H'}(S) \cap Q'| - |S| \geqslant |Q'|/6$.

Note that if r_G is much smaller than $r_{G'}$ then (2.3.37) is more restrictive than simply requiring the symmetric difference of $E(H)$ and $E(H')$ to be $o(e(H))$. We need this more restrictive bound in Cases 3 and 4 below to obtain good expansion for small sets. (See also the remark after Corollary 2.3.56.)

Proof. Let $T = N_{H'}(S) \cap Q'$, $s = |S|$ and $t = |T|$. For each vertex $x \in V(G)$, let $B_x = N_{H'}(x) \cap Q'$ and note that

$$\sum_{x \in S} |B_x| \geqslant \varepsilon r_G s \geqslant 16 s \log^2 n. \tag{2.3.38}$$

We consider the following cases:

Case 1: $s \leqslant \frac{2n\log n}{7r_{G'}}$. In this case we apply Corollary 2.3.34(i) with $G = G'$ and $A_x = B_x$ to obtain $t \geqslant \frac{\sum_{x\in S}|A_x|}{4\log n} \geqslant \frac{16s\log^2 n}{4\log n} = 4s\log n$. Hence (i) holds.

Case 2: $\frac{2n\log n}{7r_{G'}} \leqslant s \leqslant \frac{\varepsilon n r_G}{50r_{G'}}$. Apply Corollary 2.3.34(ii) with $G = G'$ and $A_x = B_x$ to obtain

$$t + 3s \geqslant \frac{n\sum_{x\in S}|A_x|}{7sr_{G'}} \geqslant \frac{\varepsilon n r_G}{7r_{G'}}.$$

Hence $t \geqslant \frac{4\varepsilon n r_G}{49r_{G'}}$ and so $\frac{1}{2}t - s \geqslant \frac{\varepsilon n r_G}{49r_{G'}}$, i.e., (ii) holds.

Case 3: $\frac{\varepsilon n r_G}{50r_{G'}} \leqslant s \leqslant \frac{2n\log n}{7r_G}$. Let $A_x = B_x \cap N_H(x)$ and note that

$$\sum_{x\in S}|B_x| \overset{(2.3.38)}{\geqslant} \varepsilon r_G s \geqslant \varepsilon r_G \frac{\varepsilon n r_G}{50r_{G'}} \overset{(2.3.37)}{\geqslant} 8|E(H')\backslash E(H)|.$$

Hence $\sum_{x\in S}|A_x| \geqslant \sum_{x\in S}|B_x| - 2|E(H')\backslash E(H)| \geqslant \frac{3}{4}\sum_{x\in S}|B_x| \geqslant 12s\log^2 n$ by (2.3.38). So Corollary 2.3.34(i) with $G = G$ implies that $t \geqslant \frac{\sum_{x\in S}|A_x|}{4\log n} \geqslant 3s\log n$, and hence (i) holds.

Case 4: $\max\{\frac{\varepsilon n r_G}{50r_{G'}}, \frac{2n\log n}{7r_G}\} \leqslant s \leqslant \varepsilon n/90$. Let $A_x = B_x \cap N_H(x)$, and note similarly to Case 3 that $\sum_{x\in S}|A_x| \geqslant \frac{3}{4}\sum_{x\in S}|B_x|$. Hence Corollary 2.3.34(ii) with $G = G$ implies that

$$t + 3s \geqslant \frac{n\sum_{x\in S}|A_x|}{7sr_G} \geqslant \frac{n\sum_{x\in S}|B_x|}{10sr_G} \overset{(2.3.38)}{\geqslant} \frac{\varepsilon n}{10}.$$

So $t \geqslant \varepsilon n/15$, and $t/2 - s \geqslant \varepsilon n/45$. Hence (iii) holds.

Case 5: $s \geqslant \varepsilon n/90$. We may assume without loss of generality that $s \leqslant |Q'|/6$ (otherwise (iv) holds). In this case we must have $|Q'| \geqslant \varepsilon n/15$.

Claim: $|N_{H'}(S) \cap Q'| \geqslant 2|Q'|/3$.

Let $Q'' = Q'\backslash(N_{H'}(S) \cup S)$. To prove the claim, suppose for a contradiction that $|Q'\backslash N_{H'}(S)| \geqslant |Q'|/3$. Then

$$|Q''| \geqslant |Q'|/3 - |S| \geqslant |Q'|/6 \geqslant s \geqslant \varepsilon n/90.$$

Now since $e_{H'}(S,Q'')=0$, we have

$$e_G(S,Q'') \leqslant |E(G)\backslash E(H')| \leqslant |E(G)\backslash E(H)| + |E(H)\backslash E(H')| \\ \overset{(2.3.36),(2.3.37)}{\leqslant} 2n\sqrt{r_G \log n} + \frac{\varepsilon^2 n r_G^2}{10^4 r_{G'}} \leqslant \frac{\varepsilon^2 n r_G}{9000}. \qquad (2.3.39)$$

But on the other hand Definition 2.3.5(a) and (2.3.2) imply that

$$e_G(S,Q'') \geqslant \frac{r_G s|Q''|}{n} - 4\sqrt{r_G}(s+|Q''|) \geqslant \frac{\varepsilon^2 n r_G}{90^2} - 4n\sqrt{r_G} \geqslant \frac{\varepsilon^2 n r_G}{8500},$$

contradicting (2.3.39). This proves the claim. Now $\frac{1}{2}|N_{H'}(S)\cap Q'|-|S| \geqslant |Q'|/3-|Q'|/6 = |Q'|/6$ and so (v) holds. □

For a graph H and a set $S \subseteq V(H)$, let $Int_H(S)$ be the set of vertices $x \in V(H)$ such that $x \in S$ and $N_H(x) \subseteq S$. If S is a set of vertices such that $S \nsubseteq V(H)$, then we take $Int_H(S)$ to mean $Int_H(S \cap V(H))$. Further let $Cl_H(S) = S \cup N_H(S)$. Note that $V(H)\backslash Int_H(S) = Cl(V(H)\backslash S)$, and that if P is a path then it is possible for an endpoint x of P to be in $Int_P(S)$.

Roughly speaking, the following lemma states that rotations (and extensions) of a path P do not have much effect on $Int_P(S)$, provided that the pivots (or join vertices) themselves lie in $Int_P(S)$.

Lemma 2.3.40 *Let $P = x\dots y$ be a path, and let Q be a set of vertices.*

(i) *Let $z \in Int_P(Q)$. Suppose we perform a rotation of P with pivot z, and let P' be the resulting path. Then $Int_P(Q)\backslash\{z\} \subseteq Int_{P'}(Q) \subseteq Int_P(Q)$. Further, if $x \in Q$ then $Int_{P'}(Q) = Int_P(Q)$.*

(ii) *Let C be a cycle vertex-disjoint from P. Let zz^- be an edge of C and suppose that $z \in Int_C(Q)$. Suppose we perform an extension of P to incorporate C with join* *vertex z and broken edge zz^-, and let P' be the resulting path. Then $(Int_P(Q) \cup Int_C(Q))\backslash\{z\} \subseteq Int_{P'}(Q) \subseteq Int_P(Q) \cup Int_C(Q)$. Further, if $x \in Q$ then $Int_{P'}(Q) = Int_P(Q) \cup Int_C(Q)$.*

Proof. (i) Let z^- be the predecessor of z along P, z^{--} the predecessor of z^-, and x^+ the successor of x. The only vertices whose neighbourhoods change as a result of the rotation are x, z and z^-. However, if $x \in Int_P(Q)$ then $x^+ \in Q$ and thus $x \in Int_{P'}(Q)$ and vice versa (here we use that $z \in Q$). Similarly, $z^- \in Int_P(Q)$ and $z^- \in Int_{P'}(Q)$ each hold if and only if $z^{--} \in Q$. Since $z \in Int_P(Q)$, the first part of (i) follows. If $x \in Q$ then we have $z \in Int_{P'}(Q)$ and hence $Int_{P'}(Q) = Int_P(Q)$.

(ii) The proof of (ii) is similar. $\square$

The following corollary gives, under fairly weak conditions, a lower bound on the number of (H', Q, τ_0)-reachable vertices of a long path P. This allows us to make any one of a large number of vertices of P into the first endpoint of P via a short sequence of rotations. Further, it allows us to 'avoid' a specified set (namely $V(P)\backslash Q$) while doing so. We will use this second property for two main purposes: Firstly in order to make sure that certain edges of our 2-factor F which we want to keep are *not* broken during the process of transforming F into a Hamilton cycle, and secondly when we want to prevent one half of P from being affected by the rotations at all. In each of these cases we need to construct sets satisfying the conditions on Q and Q' in Corollary 2.3.41; Lemmas 2.3.42 and 2.3.45 accomplish this.

Corollary 2.3.41 *Let $0 < \varepsilon < 1$ be a constant. Let G' be an $r_{G'}/n$-pseudorandom graph on n vertices and let G be an r_G/n-pseudorandom spanning subgraph of G', with $r_G \leqslant r_{G'}$ and $\varepsilon r_G \geqslant 16 \log^2 n$. Let H be a spanning subgraph of G such that (2.3.36) holds, and let H' be a spanning subgraph of G' such that (2.3.37) holds.*

Let $P = x \dots y$ be a path in G' such that P and H' are edge-disjoint. Let $Q \subseteq V(P)$

with $x \in Q$ and let $Q' \subseteq V(G')$ be such that $Q' \cap V(P) \subseteq Int_P(Q)$. Suppose that $|N_{H'}(v) \cap Q'| \geqslant \varepsilon r_G$ for every vertex $v \in Q$.

Then either

(i) *there exists a (H', Q', τ_0)-reachable vertex v of P which has a neighbour in H' lying in $Q' \backslash V(P)$, or*

(ii) *the set of (H', Q', τ_0)-reachable vertices of P has size at least $|Q'|/6$.*

Proof. Suppose that (i) does not hold, i.e., that $N_{H'}(v) \cap Q' \subseteq V(P)$ for every (H', Q', τ_0)-reachable vertex v of P. For each $\tau \leqslant \tau_0$, let U_τ be the set of (H', Q', τ)-reachable vertices of P. Assume for contradiction that (ii) does not hold either, i.e., that $|U_\tau| < |Q'|/6$ for all $\tau \leqslant \tau_0$. By Lemma 2.2.31 with $Q = Q' \cap V(P)$ we have that $|U_{\tau+1}| \geqslant \frac{1}{2}|N_{H'}(U_\tau) \cap Q'| - |U_\tau|$ for each $\tau < \tau_0$, since $N_{H'}(U_\tau) \cap Q' \cap V(P) = N_{H'}(U_\tau) \cap Q'$.

Note that any rotation of P whose pivot lies in $Q' \cap V(P) \subseteq Int_P(Q)$ produces a path P' whose first endpoint lies in Q. Thus $Int_{P'}(Q) = Int_P(Q)$ by Lemma 2.3.40(i) and the same holds for all paths obtained by further rotations with pivots in $Q' \cap V(P)$. So $U_\tau \subseteq Q$ and hence $|N_{H'}(v) \cap Q'| \geqslant \varepsilon r_G$ for each $v \in U_\tau$. Now Lemma 2.3.35 with $S = U_\tau$ implies that $|U_{\tau+1}| \geqslant \min\{\frac{\varepsilon n r_G}{49 r_{G'}}, \frac{\varepsilon n}{45}, |U_\tau| \log n\}$ for any $\tau < \tau_0$. (Here we use that conclusions (iv) and (v) of Lemma 2.3.35 cannot hold since $|U_\tau| < |Q'|/6$ for all $\tau \leqslant \tau_0$.) Hence there must exist some

$$\tau_1 \leqslant \frac{\log\left(\min\left\{\frac{\varepsilon n r_G}{49 r_{G'}}, \frac{\varepsilon n}{45}\right\}\right)}{\log\log n} + 1 \leqslant \tau_0 - 1,$$

such that $|U_{\tau_1}| \geqslant \min\{\frac{\varepsilon n r_G}{49 r_{G'}}, \frac{\varepsilon n}{45}\}$.

Suppose first that $\frac{\varepsilon n}{45} \geqslant \frac{\varepsilon n r_G}{49 r_{G'}}$. Then $|U_{\tau_1}| \geqslant \frac{\varepsilon n r_G}{49 r_{G'}}$, and hence Lemma 2.3.35 with $S = U_\tau$ implies that $|U_{\tau+1}| \geqslant \min\{\frac{\varepsilon n}{45}, |U_\tau| \log n\}$ for each τ such that $\tau_1 < \tau < \tau_0$. Hence there exists $\tau_2 < \tau_0$ such that $|U_{\tau_2}| \geqslant \frac{\varepsilon n}{45}$. But now setting $S = U_{\tau_2}$, none of the conclusions of Lemma 2.3.35 can hold. So we obtain the desired contradiction.

On the other hand, if $\frac{\varepsilon n}{45} \leqslant \frac{\varepsilon n r_G}{49 r_{G'}}$ then we already have $|U_{\tau_1}| \geqslant \frac{\varepsilon n}{45}$. So we derive a contradiction in a similar way. $\square$

The following lemma will be used to obtain a set of vertices in which the expansion property we require still holds, and which also contains no endpoints of any 'bad' edges that we want to avoid while rotating. (W will be the set of these endpoints.)

Lemma 2.3.42 *Let G' be an $r_{G'}/n$-pseudorandom graph on n vertices and let G be an r_G/n-pseudorandom spanning subgraph of G', with $300\log^3 n \leqslant r_G \leqslant r_{G'}$. Let H be an even-regular spanning subgraph of G with degree r_H such that $\delta(G)-1 \leqslant r_H \leqslant \delta(G)$. Let H' be an r_H-regular spanning subgraph of G', such that*

$$|E(H)\backslash E(H')| \leqslant \frac{nr_G^2}{2500 r_{G'}\log^2 n},$$

and let F be a 2-factor of G' which is edge-disjoint from H'. Let $W \subseteq V(G)$ with $|W| \leqslant n/400$.

Then there exist sets $V' \subseteq V(G)$ and $V'' \subseteq V'\backslash W$, such that

- $|Int_F(V'')| \geqslant n - 6|W|$,
- $|Int_F(V')| \geqslant n - |W|/\log^2 n$, *and*
- $|N_{H'}(v) \cap Int_F(V'')| \geqslant r_H/2$ *for every* $v \in V'$.

Proof. Note that by Definition 2.3.5(c),

$$\delta(G) \geqslant r_G - 2\sqrt{r_G \log n} \geqslant r_G\left(1 - \frac{2}{\sqrt{300\log n}}\right) \geqslant 290\log^3 n,$$

and hence $r_H \geqslant 288\log^3 n$.

If $W = \emptyset$ then we can take $V'' = V' = V(G)$, so we assume hereafter that $|W| \geqslant 1$. We define V'' as follows: Initially $V'' = V(G)\backslash W$. As long as there exists a vertex $v \in V''$ such that $|N_{H'}(v) \cap Int_F(V'')| \leqslant r_H/2$, we remove v from V''. Let W' be the set of removed vertices and note that $|N_{H'}(v) \cap Cl_F(W \cup W')| \geqslant r_H/2$ for every $v \in W'$. Suppose that this process continues until $|W'| = |W|$. Then for each $x \in W'$, let $B_x = N_{H'}(x) \cap Cl_F(W \cup W')$. Thus we have $|B_x| \geqslant r_H/2$ for each $x \in W'$.

Note that

$$\sum_{x \in W'} |B_x| \geqslant |W'| r_H/2 \geqslant 144|W'| \log^3 n, \tag{2.3.43}$$

and also that $|Cl_F(W \cup W')| \leqslant 3(|W| + |W'|) = 6|W|$. Now we separate into two cases:

Case 1: $|W'| = |W| \leqslant nr_H/150r_{G'}$. Applying Corollary 2.3.34 with $G = G'$, $S = W'$ and $A_x = B_x$, we obtain either

$$|Cl_F(W \cup W')| \geqslant \left| \bigcup_{x \in W'} B_x \right| \geqslant \frac{|W| r_H}{8 \log n} \geqslant 7|W|$$

or

$$|Cl_F(W \cup W')| \geqslant \left| \bigcup_{x \in W'} B_x \right| \geqslant \frac{|W| r_H n}{14 r_{G'} |W|} - 3|W| \geqslant 7|W|,$$

either of which yields an immediate contradiction.

Case 2: $|W'| = |W| \geqslant nr_H/150r_{G'}$. Let $A_x = B_x \cap N_H(x)$. Note that

$$\sum_{x \in W'} |B_x| \geqslant \frac{|W'| r_H}{2} \geqslant \frac{r_H}{2} \cdot \frac{n r_H}{150 r_{G'}} \geqslant 4|E(H) \backslash E(H')|,$$

and that

$$\sum_{x \in W'} |A_x| \geqslant \sum_{x \in W'} |B_x| - 2|E(H) \backslash E(H')| \geqslant \frac{1}{2} \sum_{x \in W'} |B_x| \overset{(2.3.43)}{\geqslant} 72|W| \log^3 n.$$

So we can apply Corollary 2.3.34 with $G = G$ and $S = W'$ to obtain either

$$|Cl_F(W \cup W')| \geqslant \left| \bigcup_{x \in W'} A_x \right| \geqslant \frac{|W| r_H}{16 \log n} \geqslant 7|W|$$

or

$$|Cl_F(W \cup W')| \geqslant \left| \bigcup_{x \in W'} A_x \right| \geqslant \frac{|W| r_H n}{28 r_G |W|} - 3|W| \geqslant \frac{n}{30} - 3|W| \geqslant 7|W|,$$

either of which again yields an immediate contradiction.

So the process must terminate before $|W'| = |W|$. Fix V'' in its state at the point when the process terminates. Note that $|V(G) \backslash V''| \leqslant 2|W|$ and hence $|Cl_F(V(G) \backslash V'')| \leqslant 6|W|$, i.e., $|Int_F(V'')| \geqslant n - 6|W|$.

Let $W'' = \{v \in V(G) \mid |N_{H'}(v) \cap Cl_F(W \cup W')| \geqslant r_H/2\}$ and $V' = V(G) \backslash W''$. Since $W'' \subseteq W \cup W'$ we have $|W''| \leqslant 2|W|$. We will show that V'' and V' satisfy the assertions of the lemma. Since $W'' \subseteq W' \cup W = V(G) \backslash V''$, we have $V'' \subseteq V' \backslash W$ and so it remains to prove that $|Int_F(V')| \geqslant n - |W|/\log^2 n$. This holds provided that

$$|Cl(W'')| \leqslant |W|/\log^2 n. \tag{2.3.44}$$

We now claim that $|Cl_F(W \cup W')| \geqslant 18|W''| \log^2 n$. Again we consider two cases:

Case 1: $|W''| \leqslant n r_H / 300 r_{G'} \log^2 n$. Then applying Corollary 2.3.34 with $G = G'$, $S = W''$ and $A_x = B_x$, we have either

$$|Cl_F(W \cup W')| \geqslant \frac{\sum_{x \in W''} |B_x|}{4 \log n} \geqslant \frac{|W''| r_H}{8 \log n} \geqslant 18|W''| \log^2 n$$

or

$$|Cl_F(W \cup W')| \geqslant \frac{|W''| n r_H}{14 r_{G'} |W''|} - 3|W''| \geqslant |W''|(19 \log^2 n - 3) \geqslant 18|W''| \log^2 n,$$

as desired.

Case 2: $|W''| \geqslant nr_H/300r_{G'}\log^2 n$. Note that

$$\sum_{x\in W''} |B_x| \geqslant \frac{r_H}{2} \cdot \frac{nr_H}{300r_{G'}\log^2 n} \geqslant 4|E(H)\backslash E(H')|,$$

and so setting $A_x = N_H(x) \cap B_x$ again we have

$$\sum_{x\in W''} |A_x| \geqslant \sum_{x\in W''} |B_x| - 2|E(H)\backslash E(H')| \geqslant \frac{1}{2}\sum_{x\in W''} |B_x| \geqslant 72|W''|\log^2 n.$$

Now again applying Corollary 2.3.34, we have either

$$|Cl_F(W\cup W')| \geqslant \frac{|W''|r_H}{16\log n} \geqslant 18|W''|\log^2 n$$

or

$$|Cl_F(W\cup W')| \geqslant \frac{|W''|r_H n}{28r_G|W''|} - 3|W''| \geqslant \frac{n}{30} - 3|W''| \geqslant 13|W| - 6|W| \geqslant 7|W|.$$

But the latter case cannot occur since $|Cl_F(W\cup W')| \leqslant 3|W\cup W'| \leqslant 6|W|$. This proves the claim. Hence $|W''| \leqslant \frac{|Cl_F(W\cup W')|}{18\log^2 n} \leqslant |W|/3\log^2 n$ and so $|Cl_F(W'')| \leqslant |W|/\log^2 n$, which proves (2.3.44). $\square$

Roughly speaking, the following lemma states that if we have a graph H' which is close to being pseudorandom and a long path P subdivided into $\log n$ segments, then most of the vertices will have many neighbours in most of the segments. This will enable us to carry out rotations involving only the initial half of a long path in the proof of Lemma 2.3.47. The proof of Lemma 2.3.45 is similar to that of Lemma 2.2.29.

Lemma 2.3.45 *Let G be an r_G/n-pseudorandom graph with $r_G \geqslant 10^5\log^2 n$. Let $\varepsilon \leqslant 1/5$ be a positive constant and let n' be an integer such that $n/10 \leqslant n' \leqslant n$. Let $U \subseteq V(G)$ be* *such that* $|V(G)\backslash U| \leqslant \varepsilon^2 n'/8$*, and let* H' *be a graph on* $V(G)$ *such that*

$$|E(G)\backslash E(H')| \leqslant \frac{\varepsilon^3 r_G n}{32000}. \tag{2.3.46}$$

Let P *be a path which is edge-disjoint from* H'*, with* $V(P) \subseteq V(G)$ *and* $|P| = n'$*, divided into* $\log n$ *segments* $J_1, \ldots, J_{\log n}$ *whose lengths are as equal as possible. Then there exists a set* $I \subseteq [\log n]$ *such that* $|I| \geqslant (1-\varepsilon)\log n$ *and for every* $i \in I$*, there exists* $J_i' \subseteq J_i \cap U$ *such that* $|Int_P(J_i')| \geqslant (1-\varepsilon)n'/\log n$*, which satisfies the following condition: For every* $v \in J_i'$*, and for all but at most* $\varepsilon \log n$ *indices* $j \in I$*,* $|N_{H'}(v) \cap Int_P(J_j')| \geqslant \frac{r_G}{25\log n}$*.*

Proof. We define I and $\{J_i' \mid i \in I\}$ as follows: Initially $I = [\log n]$ and $J_i' = J_i \cap U$. For a vertex v and for $j \in I$, call v *weakly connected* to J_j' if $|N_{H'}(v) \cap Int_P(J_j')| \leqslant \frac{r_G}{25\log n}$. As long as there exists $i \in I$ and a vertex $v \in J_i'$, such that v is weakly connected to J_j' for more than $\varepsilon \log n$ values of $j \in I$, we remove v from J_i'. Further, if at any stage there exists $i \in I$ such that $|Int_P(J_i')| \leqslant (1-\varepsilon)n'/\log n$, then we remove i from I.

We claim that this process must terminate before $\varepsilon^2 n'/4 - |V(P)\backslash U|$ vertices are removed. Indeed, suppose we have removed $\varepsilon^2 n'/4 - |V(P)\backslash U|$ vertices and let R be the set of removed vertices. Note that no vertices of $V(P)\backslash U$ are included in R, and that $|R| \geqslant \varepsilon^2 n'/8$ since $|V(P)\backslash U| \leqslant |V(G)\backslash U| \leqslant \varepsilon^2 n'/8$. Now $|R \cup (V(P)\backslash U)| = \varepsilon^2 n'/4$, and so $\sum_{i=1}^{\log n} |Int_P(J_i')| \geqslant (1 - 3\varepsilon^2/4)n'$. Hence we have $|Int_P(J_i')| \geqslant (1-\varepsilon)n'/\log n$ for at least $(1 - 3\varepsilon/4)\log n$ values of i, i.e., at most $3\varepsilon \log n/4$ indices have been removed from I. So for each vertex $v \in R$, there remain more than $\varepsilon \log n/4$ indices $i \in I$ for which v is weakly connected to J_i'. For each $i \in I$, let $WC(i)$ be the set of vertices $v \in R$ which are weakly connected to J_i'. Let $I_0 = \{i \in I \mid |WC(i)| \geqslant \varepsilon^3 n'/64\}$.

Claim: *For each* $i \in I_0$*, there are at least* $\frac{|WC(i)|r_G}{50\log n}$ *edges in* $E(G)\backslash E(H')$ *between* $Int_P(J_i')$ *and* $WC(i)$*.*

Let $S = Int_P(J_i')$. To prove the claim, note that $S \cap WC(i) = \emptyset$ since $WC(i) \subseteq R$, and hence (2.3.2) implies that

$$\begin{aligned}
e_G(S, WC(i)) &\geqslant \frac{|S||WC(i)|r_G}{n} - 4\sqrt{r_G}(|S| + |WC(i)|) \\
&\geqslant \frac{(1-\varepsilon)n'|WC(i)|r_G}{n \log n} - 5\sqrt{r_G}|WC(i)| \\
&\geqslant \frac{|WC(i)|r_G}{50 \log n}\left(4 - \frac{250 \log n}{\sqrt{r_G}}\right) \geqslant \frac{3|WC(i)|r_G}{50 \log n},
\end{aligned}$$

where the last inequality follows from the condition $r_G \geqslant 10^5 \log^2 n$. But the definition of $WC(i)$ implies that $e_{H'}(S, WC(i)) \leqslant \frac{|WC(i)|r_G}{25 \log n}$, and the claim follows immediately.

Observe that $\sum_{i \in I} |WC(i)| > |R|\varepsilon \log n/4 \geqslant \varepsilon^3 n' \log n/32$, since each $v \in R$ is weakly connected to J_i' for more than $\varepsilon \log n/4$ values of $i \in I$. But now

$$\sum_{i \in I \setminus I_0} |WC(i)| \leqslant |I \setminus I_0| \frac{\varepsilon^3 n'}{64} \leqslant \frac{\varepsilon^3 n' \log n}{64},$$

and hence $\sum_{i \in I_0} |WC(i)| > \varepsilon^3 n' \log n/64$. Together with the claim, this implies that

$$|E(G) \setminus E(H')| \geqslant \frac{r_G}{50 \log n} \sum_{i \in I_0} |WC(i)| > \frac{\varepsilon^3 r_G n' \log n}{3200 \log n} \geqslant \frac{\varepsilon^3 r_G n}{32000},$$

which contradicts (2.3.46). This proves that the process must terminate before $\varepsilon^2 n'/4 - |V(P) \setminus U|$ vertices are removed. Now $|I| \geqslant (1-\varepsilon) \log n$ at the point at which the process terminates, since as before at most $3\varepsilon \log n/4$ indices have been removed, and so at this point I and $\{J_i' \mid i \in I\}$ satisfy the assertions of the lemma. $\square$

In the next lemma we exhibit a method for merging cycles of a 2-factor F together. The basic idea is fairly simple: given F, we first use rotation-extension to transform some of the cycles in F into a long path. When we can no longer extend the path, we apply Lemma 2.3.45 to show that there is a large set Q_1 on the first half of the path whose vertices are reachable (by rotations which only use vertices in the first half of the path as pivots), and a similar set Q_2 in the last half of the path. Since Q_1 and Q_2 are large we can find an edge between them and close the path to a cycle, thus forming a new 2-factor F' with fewer cycles. However, the fact that we need to avoid bad edges forces us to be very careful.

Lemma 2.3.47 *Let G' be an $r_{G'}/n$-pseudorandom graph on n vertices, and let G be an r_G/n-pseudorandom spanning subgraph of G' with $300\log^3 n \leqslant r_G \leqslant r_{G'}$. Let F be a 2-factor of G' and let H be an even-regular spanning subgraph of G of degree r_H such that $\delta(G) - 1 \leqslant r_H \leqslant \delta(G)$. Let H' be an r_H-regular spanning subgraph of G', such that*

$$|E(H)\backslash E(H')| = |E(H')\backslash E(H)| \leqslant \frac{nr_G^2}{2500 r_{G'}\log^2 n}. \tag{2.3.48}$$

Let $E_{bad} \subseteq E(F)$ be a matching of size at most $n/220000$. Suppose that F and H' are edge-disjoint and that $2|E_{bad}|/\log^2 n \leqslant c(F) \leqslant \frac{4nr_G^2}{r_{G'}\log^3 n}$.

Then unless F is a Hamilton cycle, we can obtain a new 2-factor F' with the following properties:

- $E(F') \subseteq E(F) \cup E(H')$,
- $c(F') < c(F)$,
- $|E(F') \cap E(H')| \leqslant \frac{5\log n}{\log\log n}(c(F) - c(F'))$, *and*
- $E_{bad} \subseteq E(F')$.

Proof. Let W be the set of endpoints of edges in E_{bad}. Note that $|W| \leqslant n/110000$ and so by Lemma 2.3.42 there exist sets $V' \subseteq V(G)$ and $V'' \subseteq V'\backslash W$, with $|Int_F(V'')| \geqslant n - 6|W|$ and $|Int_F(V')| \geqslant n - |W|/\log^2 n$ and such that

$$|N_{H'}(v) \cap Int_F(V'')| \geqslant r_H/2. \tag{2.3.49}$$

for every vertex $v \in V'$. Note that

$$c(F) \geqslant 2|E_{bad}|/\log^2 n = |W|/\log^2 n \geqslant n - |Int_F(V')|. \tag{2.3.50}$$

We will obtain F' by deleting an edge yy' in a cycle of F to form a path P, performing a sequence of rotations and extensions to incorporate other cycles of F into P, and finally closing P to a cycle, thus reducing the number of cycles. The new edges for these rotations and extensions will be taken from a graph H'', which is defined as follows: Initially $H'' = H' \cup \{yy'\}$. Whenever a rotation or extension with new edge $e \in H''$ is performed, we remove e from H'' and call e a *used* edge. For each edge of H'' used we will add back to H'' the edge of F broken during the corresponding rotation or extension.

We can view H'' as the 'current' version of H'; note that we always have $|E(H'')| = |E(H')|$ or $|E(H'')| = |E(H')| + 1$. Similarly, let F' be the current version of F (so F' is either a 2-factor of $H' \cup F$, or consists of a path and various cycles). Note that since we do not change F itself during the proof, the function Int_F will not change either.

As long as we use on average at most $5 \log n/ \log\log n$ edges per cycle of F merged, we have

$$\begin{aligned} |E(H'')\backslash E(H)| &\leqslant |E(H)\backslash E(H'')| + 1 \\ &\leqslant |E(H)\backslash E(H')| + |E(H')\backslash E(H'')| + 1 \\ &\overset{(2.3.48)}{\leqslant} \frac{nr_G^2}{2500 r_{G'} \log^2 n} + \frac{5c(F)\log n}{\log\log n} + 1 \leqslant \frac{2nr_G^2}{r_{G'}\log^2 n}. \end{aligned} \tag{2.3.51}$$

This fulfils condition (2.3.37) of Lemma 2.3.35 and Corollary 2.3.41 whenever they are applied with $H' = H''$. We also have that $e(G) \leqslant nr_G/2 + 2n\sqrt{r_G}$ since G is $(r_G/n, 2\sqrt{r_G})$-jumbled by Definition 2.3.5(a), and

$$e(H) = \frac{nr_H}{2} \geqslant \frac{n(\delta(G)-1)}{2} \geqslant \frac{nr_G}{2} - n\sqrt{r_G \log n} - \frac{n}{2}$$

by Definition 2.3.5(c). So since $E(H) \subseteq E(G)$,

$$|E(G)\backslash E(H)| \leqslant 2n\sqrt{r_G} + n\sqrt{r_G \log n} + \frac{n}{2} \leqslant 2n\sqrt{r_G \log n}. \tag{2.3.52}$$

This fulfils condition (2.3.36) of Lemma 2.3.35 (and Corollary 2.3.41). Further, still assuming that on average at most $5\log n/\log\log n$ edges are used per cycle merged, we have

$$\begin{aligned} |E(G)\backslash E(H'')| &\leqslant |E(G)\backslash E(H)| + |E(H)\backslash E(H'')| \\ &\overset{(2.3.51),(2.3.52)}{\leqslant} 2n\sqrt{r_G \log n} + \frac{2nr_G^2}{r_{G'}\log^2 n} \leqslant \left(\frac{1}{15}\right)^3 \frac{r_G n}{32000}. \end{aligned} \tag{2.3.53}$$

Claim: $H'[Int_F(V')]$ *is connected.*

To prove the claim, suppose for contradiction that $H'[Int_F(V')]$ has two components S and T. Since (2.3.49) implies that $|N_{H'}(v) \cap Int_F(V')| \geqslant r_H/2$ for any $v \in Int_F(V')$, we can apply Lemma 2.3.35 with $\varepsilon = 1/3$, $Q' = Int_F(V')$ and $S = S$. But $N_{H'}(S) \cap Q' = \emptyset$, and so of the possible conclusions of Lemma 2.3.35 only (iv) can hold. This implies that $|S| \geqslant |Q'|/6 \geqslant n/7$, and similarly $|T| \geqslant |Q'|/6 \geqslant n/7$. Now since G is $(r_G/n, 2\sqrt{r_G})$-jumbled, (2.3.2) implies that

$$e_G(S,T) \geqslant r_G|S||T|/n - 4\sqrt{r_G}(|S| + |T|) \geqslant nr_G/50.$$

But by (2.3.53) this implies that $e_{H'}(S,T) > 0$. Hence $H'[Int_F(V')]$ has only one component, which proves the claim.

Our first aim is to find a path $P_1 = x'' \dots y''$ whose vertices span at least two cycles of F and so that x'' and y'' lie in $Int_F(V'')$. Let C_1 be a cycle of F containing a vertex of

$Int_F(V')$. It is easy to see that $Int_F(V')\backslash V(C_1)$ is nonempty. Indeed, if $V' = V(G)$ then it is sufficient to recall that F is not a Hamilton cycle. Otherwise $|Int_F(V')| \leqslant n-3$ and so by (2.3.50) the number of vertices which are not part of C_1 is at least $3(c(F)-1) \geqslant 3(n - |Int_F(V')| - 1) > n - |Int_F(V')|$. Hence one of these vertices must be in $Int_F(V')$. (This is the only place in the proof where we use the lower bound on $c(F)$.)

Since $H'[Int_F(V')]$ is connected, there exists an edge xy of H' with $x \in Int_F(V') \cap V(C_1)$ and $y \in Int_F(V')\backslash V(C_1)$. Let C_2 be the cycle of F in which y lies. Since E_{bad} is a matching, there will be at least one edge $yy' \notin E_{bad}$ in C_2 incident to y. Delete yy' from C_2 to form a path $C_2' = y \dots y'$. Note that $Int_F(V'') \cap V(C_2') \subseteq Int_{C_2'}(V'')$.

Again since E_{bad} is a matching we can find $xx' \notin E_{bad}$ in C_1 incident to x. Perform an extension of C_2' to incorporate C_1, with join vertex x and broken edge xx'. Let $P_0 = x' \dots y'$ denote the resulting path and note that $(Int_F(V'') \cap V(P_0))\backslash\{x, y\} \subseteq Int_{P_0}(V'')$.

Now by (2.3.49), x' has a neighbour z_1 in H'', which lies in $Int_F(V'')\backslash\{x, y\}$. If $z_1 \in V(P_0)$ then we perform a rotation with pivot z_1; note that the first endpoint x'' of the resulting path lies in V''. Otherwise, let C be the cycle containing z_1 and let $z_1x'' \notin E_{bad}$ be an edge of C. Perform an extension of P_0 to incorporate C and again note that the new first endpoint x'' will lie in V''. Call the resulting path P_0' and let $P_0'' = y' \dots x''$ be the reverse of P_0'. (Recall that the reverse of a path was defined in the second paragraph of Section 2.3.5.) Now similarly we can perform a rotation (or extension) of P_0'' with some pivot z_2 (or join vertex z_2), so that the new first endpoint y'' also lies in V''. Call the resulting path P_0''' and let $P_1 = x'' \dots y''$ be the reverse of P_0'''.

Applying Lemma 2.3.40 twice with $P = P_0$ and P_0'' respectively, $Q = V''$ and $z = z_1$ and z_2 respectively implies that

$$(Int_F(V'') \cap V(P_1))\backslash\{z_1, z_2, x, y\} \subseteq Int_{P_1}(V''). \tag{2.3.54}$$

All rotations performed during the remainder of the proof will have pivots in $(Int_F(V'') \cap V(P_1)) \backslash \{z_1, z_2, x, y\}$ (where P_1 is the current path). Thus by (2.3.54) the pivots lie in $Int_{P_1}(V'')$. Also the join vertices of all extensions performed during the remainder of the proof will lie in $Int_F(V'')$. Hence no edges in E_{bad} will be broken. Moreover, the endpoints of P_1 will always lie in V'' and hence Lemma 2.3.40 implies that (2.3.54) will always hold.

Note that for each $v \in V'$, we have

$$|N_{H''}(v) \cap Int_F(V'')| \geqslant |N_{H'}(v) \cap Int_F(V'')| - 2$$

since $H' \subseteq H'' \cup F'$ and $\Delta(F') = 2$. So (2.3.49) implies that

$$|N_{H''}(v) \cap Int_F(V'')| \geqslant r_G/3 + 4$$

with room to spare. We now incorporate additional cycles into P_1 by an iterative procedure. Each iteration proceeds as follows:

Apply Corollary 2.3.41 with $\varepsilon = 1/3$, $H' = H''$, $P = P_1$, $Q = V'' \cap V(P_1)$ and $Q' = Int_F(V'') \backslash \{z_1, z_2, x, y\}$. We obtain one of two possible conclusions:

Case 1: *There exists an (H'', Q', τ_0)-reachable vertex v of P_1 which has a neighbour $z \in Int_F(V'') \backslash V(P_1)$.* Let C be the cycle of F on which z lies. Now we perform the necessary rotations to make v the first endpoint of P_1, and then an extension with join vertex z to incorporate C into P_1. We then redefine P_1 to be the resulting path and begin the next iteration.

Case 2: *The number of (H'', Q', τ_0)-reachable vertices of P_1 is at least $|Q'|/6 \geqslant 3n/20$.* Note that this immediately implies that $|P_1| \geqslant |Q'/6| \geqslant n/10$. Let P_1 be divided into $\log n$ segments J_i whose lengths are as equal as possible. Noting that $|V(G) \backslash Q'| \leqslant 6|W| + 4 \leqslant n/18000$, we may apply Lemma 2.3.45 with $\varepsilon = 1/15$, $H' = H''$, $U = Q'$ and $n' = |P_1|$.

Note that (2.3.46) is satisfied due to (2.3.53). Thus we obtain a set $I \subseteq [\log n]$ of size at least $14 \log n/15$ and sets $J_i' \subseteq J_i \cap Q'$ for each $i \in I$, such that $|Int_{P_1}(J_i')| \geqslant 14|P_1|/15 \log n$ and the following holds: For every $v \in J_i'$, and for all but at most $\log n/15$ indices $j \in I$, $|N_{H''}(v) \cap Int_{P_1}(J_j')| \geqslant \frac{r_G}{25 \log n}$.

Note that $|\bigcup_{i \in I} Int_{P_1}(J_i')| \geqslant (1 - 2/15)|P_1|$. Hence for some $i \in I$, J_i' contains a (H'', Q', τ_0)-reachable vertex of P_1. Perform the necessary rotations to make one of these vertices into the first endpoint, breaking at most τ_0 edges in the process. Redefine P_1 to be the resulting path, and let P_1' be the reverse of P_1. Now apply Corollary 2.3.41 with $\varepsilon = 1/3$, $H' = H''$, $P = P_1'$, $Q = V'' \cap V(P_1')$ and $Q' = Int_F(V'') \backslash \{z_1, z_2, x, y\}$. Again we obtain one of two possible conclusions:

Case 2a: *There exists an (H'', Q', τ_0)-reachable vertex v of P_1' which has a neighbour $z \in Int_F(V'') \backslash V(P_1')$.* In this case we extend P_1' as in Case 1, redefine P_1 to be the resulting path and begin the next iteration. Note that in future instances of Case 2, the sets I, J_i and J_i' will be redefined for each new path P_1'.

Case 2b: *The number of (H'', Q', τ_0)-reachable vertices of P_1' is at least $|Q'|/6 \geqslant 3n/20$.* Note that any instance of Case 2b is immediately preceded by an instance of Case 2, and so the segments J_i have been defined (possibly redefined) so that they partition $V(P_1')$. Call a segment J_i *broken* if one of its edges was used as a broken edge in one of the rotations in Case 2 (or, for later uses, in Case 2b), and let $I' \subseteq I$ be the set of all indices $i \in I$ such that J_i is an unbroken segment. Note that since each rotation breaks at most one segment, $|\bigcup_{i \in I'} Int_{P_1'}(J_i')| \geqslant (1 - 2/15 - \tau_0/\log n)|P_1'| > (1 - 3/20)|P_1'|$. Hence for some $i \in I'$, J_i' contains a (H'', Q', τ_0)-reachable vertex of P_1'. Perform the necessary rotations to make one of these vertices into the first endpoint, breaking at most τ_0 edges in the process. Call the resulting path $P_2 = x_2 \dots y_2$.

Eventually, the iterative procedure has to conclude by entering Case 2b. Let $I'' \subseteq I$ be the set of all indices $i \in I$ such that J_i was broken in neither Case 2 nor Case 2b. Let

I_1, I_2 be subsets of I'' such that $|I_1| = |I_2| = |I''|/2$ and for every $i_1 \in I_1$, $i_2 \in I_2$ the segment J_{i_1} precedes J'_{i_2} where the ordering is taken along P_2. Now

$$|I_1|, |I_2| \geqslant \frac{1}{2}((1-1/15)\log n - 2\tau_0) \geqslant \frac{1}{2}(1-1/10)\log n.$$

Let $Q_1 = \{x_2\} \cup \bigcup_{i \in I_1} J'_i$, $Q_2 = \{y_2\} \cup \bigcup_{i \in I_2} J'_i$ and note that

$$|Int_{P_2}(Q_1)|, |Int_{P_2}(Q_2)| \geqslant \frac{1}{2}\left(1-\frac{1}{10}\right)\left(1-\frac{1}{15}\right)|P_1| \geqslant \frac{2|P_1|}{5} \geqslant \frac{n}{25}. \tag{2.3.55}$$

Further, for each vertex v of Q_1, and for at least $|I_1| - \log n/15 \geqslant \log n/3$ indices $i \in I_1$, we have that $|N_{H''}(v) \cap Int_{P_2}(J'_i)| \geqslant \frac{r_G}{25\log n}$ (noting that $Int_{P_2}(J'_i) = Int_{P_1}(J'_i)$ since J'_i is unbroken). Hence

$$|N_{H''}(v) \cap Int_{P_2}(Q_1)| \geqslant r_G/75$$

for all $v \in Q_1$, and the corresponding statement holds for vertices of Q_2.

Now applying Corollary 2.3.41 with $P = P_2$, $Q = Q_1$, $Q' = Int_{P_2}(Q_1)$, $H' = H''$ and $\varepsilon = 1/75$ implies that there exists a set A of $|Int_{P_2}(Q_1)|/6$ $(H'', Int_{P_2}(Q_1), \tau_0)$-reachable vertices of P_2 (since $Q' \subseteq V(P_2)$, it is impossible for the first conclusion of Corollary 2.3.41 to hold). Let P'_2 be the reverse of P_2 and apply Corollary 2.3.41 again with $P = P'_2$, $Q = Q_2$ and $Q' = Int_{P'_2}(Q_2)$ to obtain a set B of $|Int_{P'_2}(Q_2)|/6$ $(H'', Int_{P'_2}(Q_2), \tau_0)$-reachable vertices of P'_2. Now $|A|, |B| \geqslant n/150$ by (2.3.55), and so using (2.3.2) we have

$$e_G(A,B) \geqslant \frac{|A||B|r_G}{n} - 4\sqrt{r_G}(|A|+|B|) \geqslant \frac{nr_G}{22500} - 4n\sqrt{r_G} \geqslant \frac{nr_G}{30000}.$$

Hence

$$e_{H''}(A,B) \geqslant \frac{nr_G}{30000} - |E(G)\backslash E(H'')| \overset{(2.3.53)}{>} 0.$$

Now let $x_3y_3 \in E_{H''}(A,B)$. Noting that x_3 is $(H'', Int_{P_2}(Q_1), \tau_0)$-reachable, perform a set of at most τ_0 rotations of P_2 with pivots in $Int_{P_2}(Q_1)$, to form a new path P_2'' whose first endpoint is x_3. Note that $Int_{P_2''}(Q_2) = Int_{P_2}(Q_2)$. Thus setting P_2''' to be the reverse of P_2'', we have that y_3 is still $(H'', Int_{P_2'''}(Q_2), \tau_0)$-reachable. Perform a set of at most τ_0 rotations to make y_3 the first endpoint of P_2'''. Note that all the pivots of these rotations lie in $Q_1 \cup Q_2 \subseteq \bigcup_{i \in I} J_i' \subseteq Q'$, and so as discussed after (2.3.54) no edge in E_{bad} is broken. Now use x_3y_3 to close P_2''' to a cycle.

It remains to estimate the number of edges broken during the process. In the initial formation of P_1 we break only 4 edges (two while forming P_0 and one for each of the two subsequent rotations or extensions). During each instance of Case 1 or Case 2a we break at most $\tau_0 + 1$ edges, during Case 2 we break at most τ_0 edges, during Case 2b we break at most τ_0 edges, and during the remainder of the proof we break at most $2\tau_0$ edges. Each time we merge a new cycle into P_1, we need either to go through Case 1 once, or through both Cases 2 and 2a once. Case 2b occurs only once (at the end, after Case 2).

Let k_1 be the number of times we repeat Case 1 and k_2 be the number of times we repeat Cases 2 and 2a. Then the number of cycles of F decreases by $k_1 + k_2 + 1$. Further the total number of edges broken is at most

$$4 + k_1(\tau_0 + 1) + k_2(2\tau_0 + 1) + 4\tau_0 \leqslant 4(k_1 + k_2 + 1)\tau_0 + 4$$

So on average the number of edges broken per cycle merged is at most $4\tau_0 + 4 \leqslant 5 \log n / \log \log n$. $\square$

We can now apply Lemma 2.3.47 repeatedly to transform 2-factors into Hamilton cycles. Given a 2-regular graph F and an edge-disjoint graph H on the same vertex set, we say that a Hamilton cycle C is formed by *merging* the cycles of F using edges of H if $E(C) \subseteq E(F) \cup E(H)$. $E(H) \cap E(C)$ is the set of *used edges*, and $E(F) \backslash E(C)$ is the set of *broken edges.*

Corollary 2.3.56 *Let G' be an $r_{G'}/n$-pseudorandom graph on n vertices, and let G be an r_G/n-pseudorandom spanning subgraph of G' with $300\log^3 n \leqslant r_G \leqslant r_{G'}$. Let H be an even-regular spanning subgraph of G with degree r_H, such that $\delta(G) - 1 \leqslant r_H \leqslant \delta(G)$. Let $\mathcal{F}$ be a collection of edge-disjoint 2-factors $F_1, F_2, \ldots, F_m$ of G', such that each F_i is edge-disjoint from H.*

Let $E_{bad} \subseteq E(\bigcup \mathcal{F})$ be such that $E_{bad} \cap E(F_i)$ is a matching and $|E_{bad} \cap E(F_i)| \leqslant n/10^6$ for each F_i. Suppose that $c(\mathcal{F}) r_{G'} \log^3 n \leqslant 4nr_G^2$. Then we can merge the cycles of each F_i into a Hamilton cycle C_i using the edges of H. Further, we can ensure that the number of edges of E_{bad} broken during the process is at most $|E_{bad}|/\log n$, and that all of the C_i's are pairwise edge-disjoint.

When r_G is much smaller than $r_{G'}$ the bound on $c(\mathcal{F})$ is more restrictive than simply requiring that $c(\mathcal{F})$ is small compared to $e(H)$ as one might at first expect. The assumption is necessary due to (2.3.48) which in turn arises from (2.3.37). In the proof of Lemma 2.3.60 (and thus of Theorem 2.3.62) this assumption will be the limiting factor in determining how small we can make r_G compared to $r_{G'}$, and thus how many iterations we need to use.

Proof. We merge cycles by repeatedly applying Lemma 2.3.47. During this process we will remove certain edges from H (namely those which lie in the new 2-factor obtained by Lemma 2.3.47) and add certain edges to H (namely those edges which are removed from the old 2-factor in Lemma 2.3.47 to obtain the new one). Let H' denote the 'current' version of H (so H always denotes the original version).

We use Lemma 2.3.47 repeatedly to reduce $c(F_i)$ until $c(F_i) = 1$, i.e., F_i is a Hamilton cycle for each i. We make use of the fact that on average at most $5\log n/\log\log n$ edges of H' are used by Lemma 2.3.47 for each cycle that needs to be merged. Hence we have

$$|E(H)\backslash E(H')| = |E(H')\backslash E(H)| \leqslant \frac{5c(\mathcal{F})\log n}{\log\log n} \leqslant \frac{r_G^2 n}{2500 r_{G'}\log^2 n}$$

throughout the process. So (2.3.48) is satisfied.

More precisely, we proceed as follows: Given $1 \leqslant i \leqslant m$, suppose that the number of cycles of F_i is at least $2|E_{bad} \cap E(F_i)|/\log^2 n$. Then we can apply Lemma 2.3.47 with $E_{bad} = E_{bad} \cap E(F_i)$ to merge one or more cycles of F_i together, without breaking any edges of E_{bad} and using on average at most $5\log n/\log\log n$ edges of H' per cycle merged. (Note that the required upper bound on $c(F_i)$ holds since $c(F_i) \leqslant c(\mathcal{F})$.)

On the other hand if the number of cycles is fewer than $2|E_{bad} \cap E(F_i)|/\log^2 n$, then we can apply Lemma 2.3.47 with $E_{bad} = \emptyset$ repeatedly to merge all of the remaining cycles of F_i together, thus forming a Hamilton cycle. Again we use at most $5\log n/\log\log n$ edges of H' per cycle and thus the total number of edges of E_{bad} which are broken during this process is at most

$$\frac{2|E_{bad} \cap E(F_i)|}{\log^2 n} \cdot \frac{5\log n}{\log\log n} \leqslant \frac{|E_{bad} \cap E(F_i)|}{\log n}.$$

The bound on the total number of edges of E_{bad} broken follows immediately. □

For a given application of Corollary 2.3.56 during the remainder of the chapter, define the *leftover graph* to be the final state of H'. Thus H' is obtained from H by deleting the edges of the Hamilton cycles produced by Corollary 2.3.56, and adding those edges of $\bigcup \mathcal{F}$ which do not lie in any of these Hamilton cycles.

We now prove variants of Lemma 2.3.47 and Corollary 2.3.56. These will be used in cases where instead of needing to avoid a set of bad edges when rotating, we only want to avoid a vertex x_0 (when we apply the lemmas x_0 will be the vertex of minimum degree in

$G_{n,p}$). For clarity we prove these as separate lemmas as the conditions are different, but the proofs proceed along similar lines in each case.

Lemma 2.3.57 *Let G' be an $r_{G'}/n$-pseudorandom graph on n vertices, and let G be an r_G/n-pseudorandom spanning subgraph of G' with $10^5 \log^2 n \leqslant r_G \leqslant r_{G'}$. Let x_0 be a vertex of G and let H be formed from G by removing all of the edges of G incident to x_0. Let H' be a spanning subgraph of G', such that (2.3.48) holds and $d_H(v) = d_{H'}(v)$ for each $v \in V(G)$.*

Let F be a 2-factor of G', edge-disjoint from H', such that $c(F) \leqslant \frac{4nr_G^2}{r_{G'} \log^3 n}$. Then unless F is a Hamilton cycle, we can obtain a new 2-factor F' such that the following hold:

- $E(F') \subseteq E(F) \cup E(H')$,
- $c(F') < c(F)$, *and*
- $|E(F') \cap E(H')| \leqslant \frac{5 \log n}{\log \log n}(c(F) - c(F'))$.

Proof. The proof is similar to that of Lemma 2.3.47. We will again obtain F' from F by performing a sequence of rotations and extensions, with all the new edges for these rotations and extensions taken from a graph H'' which is defined as in the proof of Lemma 2.3.47. Again, we say that an edge e of H'' is *used* if it is the new edge of some rotation or extension performed during the proof. Define the current version F' of F as in the proof of Lemma 2.3.47. Observe that $e(H) \geqslant e(G) - n$, and so assuming we use on average at most $\frac{5 \log n}{\log \log n}$ edges for each cycle of F merged, the bounds (2.3.51), (2.3.52) and (2.3.53) still hold. Moreover,

$$\begin{aligned}\delta(H'' - \{x_0\}) &\geqslant \delta(H' - \{x_0\}) = \delta(H - \{x_0\}) \\ &\geqslant \delta(G) - 1 \geqslant r_G - 2\sqrt{r_G \log n} - 1 \geqslant \frac{r_G}{2}.\end{aligned} \tag{2.3.58}$$

Claim: $H' - \{x_0\}$ *is connected.*

To prove the claim, suppose for a contradiction that $H' - \{x_0\}$ has two components, S and T. By (2.3.58) we can apply Lemma 2.3.35 with $\varepsilon = 1/3$, $Q' = V(G)\backslash\{x_0\}$ and $S = S$. But now $N_{H'-\{x_0\}}(S) = \emptyset$, and so of the possible conclusions of Lemma 2.3.35 only (iv) can hold. This implies that $|S| \geqslant (n-1)/6 > n/7$. Similarly $|T| \geqslant (n-1)/6 > n/7$. Now since G is $(r_G/n, 2\sqrt{r_G})$-jumbled we have that

$$e_G(S,T) \geqslant r_G|S||T|/n - 4\sqrt{r_G}(|S| + |T|) \geqslant nr_G/50.$$

But by (2.3.53) this implies that $e_{H'}(S,T) > 0$, which contradicts our assumption that S and T were components. Hence $H' - \{x_0\}$ has only one component, which proves the claim.

Let C_1 be a cycle of F. Since $H' - \{x_0\}$ is connected, there exists an edge xy of H' joining two distinct cycles C_1 and C_2 of F, with x on C_1 and y on C_2 and such that $x, y \neq x_0$. Let yy' be an edge of C_2 incident to y such that $y' \neq x_0$. Delete yy' from C_2 to form a path C_2'. Let xx' be an edge of C_1 such that $x' \neq x_0$, and perform an extension of C_2' with join vertex x and broken edge xx' to incorporate C_1. Let $P_1 = x' \ldots y'$ denote the resulting path. Note that x_0 cannot be an endpoint of P_1.

Let $V'' = V(G)\backslash\{x_0\}$ and $Q' = Int_F(V'')$. Note that $Int_{P_1}(V'') = Q' \cap V(P_1)$. All rotations performed during the remainder of the proof will have pivots in $Int_F(V'') \cap V(P_1)$, where P_1 is the current path. Similarly, all extensions performed during the remainder of the proof will have join vertices in Q'. Thus by applying Lemma 2.3.40 with $Q = V''$, we always have that

$$Int_{P_1}(V'') = Int_F(V'') \cap V(P_1) = Q' \cap V(P_1).$$

So in particular x_0 will never be an endpoint of P_1.

By (2.3.58) we can apply Corollary 2.3.41 with $\varepsilon = 1/3$, $H' = H''$, $P = P_1$, $Q = V'' \cap V(P_1)$ and $Q' = Q'$ and use the same case analysis as in Lemma 2.3.47. The remainder of the argument is also identical to that in the proof of Lemma 2.3.47. (Note that $x_0 \notin J_i'$ for any i since $J_i' \subseteq J_i \cap Q'$, and hence $x_0 \notin Q_1$ and $x_0 \notin Q_2$.) □

Corollary 2.3.59 *Let G' be an $r_{G'}/n$-pseudorandom graph on n vertices, and let G be an r_G/n-pseudorandom spanning subgraph of G' with $10^5 \log^2 n \leqslant r_G \leqslant r_{G'}$. Let x_0 be a vertex of G and let H be formed from G by removing all of the edges of G incident to x_0. Let $\mathcal{F}$ be a collection of edge-disjoint 2-factors $F_1, F_2, \ldots, F_m$ of G', such that each F_i is edge-disjoint from H and $c(\mathcal{F}) \leqslant \frac{4nr_G^2}{r_{G'} \log^3 n}$. Then we can merge the cycles of each F_i into a Hamilton cycle C_i using the edges of H, such that the C_i's are pairwise edge-disjoint. (Recall that merging was defined in the paragraph before Corollary 2.3.56.)*

Proof. We merge cycles by repeatedly applying Lemma 2.3.57. During this process we will remove certain edges from H (namely those which lie in the new 2-factor obtained by Lemma 2.3.57) and add certain edges to H (namely those edges which are removed from the old 2-factor in Lemma 2.3.57 to obtain the new one). Let H' denote the 'current' version of H (so H always denotes the original version).

We use Lemma 2.3.57 repeatedly to reduce $c(F_i)$ until $c(F_i) = 1$, i.e., F_i is a Hamilton cycle for each i. We make use of the fact that on average at most $5 \log n / \log \log n$ edges of H' are used by Lemma 2.3.57 for each cycle that needs to be merged. So

$$|E(H) \backslash E(H')| = |E(H') \backslash E(H)| \leqslant \frac{5c(\mathcal{F}) \log n}{\log \log n} \leqslant \frac{r_G^2 n}{2500 r_{G'} \log^2 n},$$

and hence (2.3.48) is satisfied throughout the process. □

2.3.6 Completing the proof

In this section we combine our results to prove Theorem 1.3.5. Roughly speaking, the following lemma states that given a graph H_0 which is close to being pseudorandom and given about $\log n$ pseudorandom graphs $H_1, \dots, H_{2t+1}$, we can find a set of edge-disjoint Hamilton cycles in the union of these graphs which cover all of the edges of H_0. While the H_i cannot be too sparse, they need not be as dense as H_0. The point is that the remaining 'uncovered' graph is much sparser than H_0 and is also not too far from being pseudorandom. In the proof of Theorem 2.3.62 we apply this lemma three times in succession to obtain an uncovered graph which is very sparse.

Lemma 2.3.60 *Let $p_0 \geqslant \log^{14} n/n$, and let $p_1 \geqslant ((np_0)^3 \log^{10} n)^{1/4}/n$. Let $t = \frac{\log(n^2 p_1)}{\log\log n}$ and let $p_2, \dots, p_{2t+1}$ be positive reals such that $p_i = p_1$ for odd i and $p_i = 10^{10} p_1$ for even i. Let G_0 be a p_0-pseudorandom graph on n vertices. Suppose that G_i is a p_i-pseudorandom spanning subgraph of G_0, and let H_i be an even-regular spanning subgraph of G_i such that $\delta(G_i) - 1 \leqslant \delta(H_i) \leqslant \delta(G_i)$, for each $1 \leqslant i \leqslant 2t+1$. Suppose that the graphs G_i are pairwise edge-disjoint for $1 \leqslant i \leqslant 2t+1$ and let H_0 be an even-regular spanning subgraph of G_0 which is edge-disjoint from $\bigcup_{i=1}^{2t+1} H_i$. Then there exists a collection $\mathcal{HC}$ of edge-disjoint Hamilton cycles such that $H_0 \subseteq \bigcup \mathcal{HC} \subseteq \bigcup_{i=0}^{2t+1} H_i$.*

Formally the assumption $p_0 \geqslant \log^{14} n/n$ can be omitted. It is included for clarity since if p_0 is significantly smaller then $p_0 \leqslant \sum_{i=1}^{2t+1} p_i$ and so the lemma becomes vacuous.

To prove Lemma 2.3.60 we first decompose H_0 into 2-factors which on average have few cycles. We then use edges of H_1 to transform these 2-factors into Hamilton cycles. Because edges are exchanged between H_1 and the 2-factors there will still be some edges of H_0 left uncovered (the 'bad' edges). We decompose $H_1' \cup H_2$ (where H_1' is the leftover of H_1 and H_0) into 2-factors with few cycles and then use H_3 to transform them into Hamilton cycles (we cannot decompose H_1' on its own since it is no longer close to being pseudorandom). Again some edges of H_0 will be left uncovered, but we can guarantee that the number of such edges will be reduced (by a factor of about $\log n$). After about $\log n/\log\log n$ iterations we arrive at a leftover graph which contains no edges of H_0, i.e., all of the edges of H_0 are covered.

Proof. [Proof of Lemma 2.3.60] Note that $p_1 \geqslant \log^{13} n/n$. Corollary 2.3.26 with $G = G_0$ and $H = H_0$ implies that H_0 can be decomposed into a collection $\mathcal{F}_1$ of 2-factors such that $c(\mathcal{F}_1) \leqslant 3n\sqrt{np_0 \log^3 n}$. Since

$$c(\mathcal{F}_1) \cdot np_0 \log^3 n \leqslant 4n(np_0 \log^3 n)^{3/2} \leqslant 4n(np_1)^2, \tag{2.3.61}$$

we may apply Corollary 2.3.56 with $G' = G_0$, $G = G_1$, $H = H_1$, $\mathcal{F} = \mathcal{F}_1$ and $E_{bad} = \emptyset$. This allows us to merge the cycles of each 2-factor of $\mathcal{F}_1$ to form a collection of edge-disjoint Hamilton cycles. Let $H_1' \subseteq H_0 \cup H_1$ be the leftover graph from Corollary 2.3.56 (as defined after the proof of Corollary 2.3.56), and let $E_{bad} = E(H_1') \cap E(H_0)$. Note that $|E_{bad}| \leqslant |E(H_1')| = |E(H_1)| \leqslant n^2p_1/2$.

Now H_1' is even-regular with degree at most np_1, and $np_1 + 1 + 10^6 np_1/2 \leqslant np_2/5000$. Hence we may apply Corollary 2.3.30 with $G = G_2$, $H = H_2$ and $H' = H_1'$ to obtain a decomposition $\mathcal{F}_2$ of $H_1' \cup H_2$ into 2-factors, such that $c(\mathcal{F}_2) \leqslant 4n\sqrt{np_2 \log^3 n}$ and $E(F) \cap E_{bad}$ is a matching of size at most $n/10^6$ for each $F \in \mathcal{F}_2$.

By (2.3.61) (noting that $p_1 = p_3$) we may apply Corollary 2.3.56 with $G' = G_0$, $G = G_3$, $H = H_3$, $\mathcal{F} = \mathcal{F}_2$ and $E_{bad} = E_{bad}$. This yields another collection of edge-disjoint Hamilton cycles and a leftover graph H_3'. Furthermore, if we redefine $E_{bad} = E(H_3') \cap E(H_0)$, then $|E_{bad}|$ is reduced by a factor of $\log n$.

We now repeat this process a further $t-1$ times, using up the graphs $H_4, H_5, \dots,$ H_{2t+1}. Now we have $|E(H_{2t+1}') \cap E(H_0)| \leqslant \frac{n^2 p_1}{2(\log n)^t} < 1$, i.e., $E(H_{2t+1}') \cap E(H_0) = \emptyset$. Let $\mathcal{HC}$ be the union of all the collections of Hamilton cycles produced by this process and note that $H_0 \subseteq \bigcup \mathcal{HC} \subseteq \bigcup_{i=0}^{2t+1} H_i$. $\square$

Before proving Theorem 1.3.5, we state a 'pseudorandom' version of the theorem. The conditions in this version are significantly more complicated than those of Theorem 1.3.5; however, it has the advantage of being entirely deterministic and we believe it to be of independent interest.

Theorem 2.3.62 *Let $\log^{50} n/n \leqslant p_0 \leqslant 1 - n^{-1/4}\log^9 n$. Let*

$$\begin{aligned} p_2 &= \frac{(np_0)^{3/4}\log^{7/2} n}{n}, \; p_3 = \frac{(np_2)^{3/4}\log^{7/2} n}{n}, \\ p_4 &= \frac{(np_3)^{3/4}\log^{7/2} n}{n}, \; p_5 = \frac{\sqrt{np_0(1-p_0)}}{n\log^{3/4} n} \end{aligned} \tag{2.3.63}$$

and $p_1 = p_0 - p_2 - p_3 - p_4 - p_5$. For $i = 2, 3, 4$ let

$$m_i = \frac{2\log(n^2 p_i)}{\log\log n}.$$

For $1 \leqslant j \leqslant 2t_i + 1$ let $p_{(i,j)} = \frac{p_i}{(10^{10}+1)m_i+1}$ if j is odd and $p_{(i,j)} = \frac{10^{10} p_i}{(10^{10}+1)m_i+1}$ if j is even.

Let G_0 be a p_0-pseudorandom graph on n vertices. Suppose that G_0 has a decomposition into graphs G_1, G_2, G_3, G_4, G_5, and that G_i has a decomposition into graphs $G_{(i,1)}, G_{(i,2)}, \ldots, G_{(i,2t_i+1)}$ for $i = 2, 3, 4$, such that the following conditions hold:

(i) *G_i is p_i-pseudorandom for each $1 \leqslant i \leqslant 5$,*

(ii) *$G_{(i,j)}$ is $p_{(i,j)}$-pseudorandom for $i = 2, 3, 4$ and $1 \leqslant j \leqslant 2t_i + 1$,*

(iii) *$G_i \cup G_{i+1}$ is $(p_i + p_{i+1})$-pseudorandom for $i = 2, 3, 4$, and*

(iv) *G_0 is $4u$-jumping where*

$$u = 2np_5 = \frac{2\sqrt{np_0(1-p_0)}}{\log^{3/4} n}. \tag{2.3.64}$$

Then G_0 has property $\mathcal{H}$.

For the moment we will assume the truth of Theorem 2.3.62 and use it to prove Theorem 1.3.5.

Proof. [Proof of Theorem 1.3.5] Let $p_0 = p$ and let $G_0 \sim G_{n,p}$. Define as in the statement of Theorem 2.3.62 the real numbers p_i for $1 \leqslant i \leqslant 5$, integers t_i for $i = 2, 3, 4$, and reals $p_{(i,j)}$ for $i = 2, 3, 4$ and $1 \leqslant j \leqslant 2t_i + 1$.

Form graphs G_1, G_5 and $G_{(i,j)}$ for $i = 2, 3, 4$ and $1 \leqslant j \leqslant 2t_i + 1$ as follows: For each edge e of G_0, place e in G_1 with probability p_1/p_0, in G_5 with probability p_5/p_0, and in $G_{(i,j)}$ with probability $p_{(i,j)}/p_0$ for each $i = 2, 3, 4$ and $1 \leqslant j \leqslant 2t_i + 1$. Let $G_i = \bigcup_{j=1}^{2t_i+1} G_{(i,j)}$ for $i = 2, 3, 4$.

Note that $p_i = o(p_0)$ for $i = 2, 3, 4, 5$ and that

$$np_5 \geqslant \log^{24} n,\ np_i \geqslant \log^{14} n \text{ and } np_{(i,j)} \geqslant \log^{13} n \tag{2.3.65}$$

for $i = 2, 3, 4$, where the second inequality holds since $x \geqslant \log^{14} n$ implies that $x^{3/4} \log^{7/2} n \geqslant \log^{14} n$. Thus the bounds on each p_i and $p_{(i,j)}$ in Lemma 2.3.13 hold, and hence Lemma 2.3.13 implies that G_0 is p_0-pseudorandom and that conditions (i), (ii) and (iii) of Theorem 2.3.62 hold whp. Moreover, condition (iv) holds whp by Lemma 2.3.14. Hence Theorem 2.3.62 implies that G_0 has property $\mathcal{H}$. $\square$

It remains to prove Theorem 2.3.62.

Proof. [Proof of Theorem 2.3.62] Recall from the proof of Theorem 1.3.5 that $p_i = o(p_0)$ for $i = 2, 3, 4, 5$. Thus

$$p_1 = (1 - o(1))p_0. \tag{2.3.66}$$

Note also that (2.3.65) holds and

$$np_4 = (np_0)^{27/64} \log^{\frac{7}{2}\left(1+\frac{3}{4}+\frac{9}{16}\right)} \leqslant (np_0)^{27/64} \log^{49/6} n. \tag{2.3.67}$$

Let x_0 be the vertex of G_0 of minimum degree. If $\delta(G_0)$ is odd, then at this point we use Lemma 2.3.18(i) to remove an optimal matching M_{Opt} which covers x_0 from G_1, and let G_1' be the remainder. If $\delta(G_0)$ is even then let $G_1' = G_1$ and $M_{Opt} = \emptyset$.

Form H_5 from G_5 by removing all of the edges incident to x_0, and add the removed edges to G_1'. For each $G_{(i,j)}$, apply Lemma 2.3.18(ii) with $u = 0$ to form a regular spanning subgraph $H_{(i,j)}$ whose degree is either $\delta(G_{(i,j)})$ (if $\delta(G_{(i,j)})$ is even) or $\delta(G_{(i,j)})-1$ (otherwise). If there are edges of $G_{(i,j)} \backslash H_{(i,j)}$ which are incident to x_0, move these edges into G_1'. Let $H_i = \bigcup_{j=1}^{2t_i+1} H_{(i,j)}$ for $i = 2, 3, 4$. Now all edges of G_0 which are incident to x_0 lie in $G_1' \cup H_2 \cup H_3 \cup H_4 \cup M_{Opt}$, i.e.,

$$d_{G_0}(x_0) = d_{G_1'}(x_0) + d_{H_2}(x_0) + d_{H_3}(x_0) + d_{H_4}(x_0) + d_{M_{Opt}}(x_0). \tag{2.3.68}$$

Let $\sum_{i,j}$ denote the summation $\sum_{i=2}^{4} \sum_{j=1}^{2t_i+1}$. Note that (2.3.68) implies that

$$d_{G_1}(x_0) - 1 \leqslant d_{G_1'}(x_0) \leqslant \delta(G_0) - 2\sum_{i,j} \lfloor \delta(G_{(i,j)})/2 \rfloor \leqslant \delta(G_0) - \sum_{i,j} (\delta(G_{(i,j)}) - 1). \tag{2.3.69}$$

The next claim shows that the number of edges incident to x_0 which we added to G_1 to form G_1' is at most u.

Claim 1:

$$\Delta(G_5) + \sum_{i,j} \big(\Delta(G_{(i,j)}) - \delta(G_{(i,j)}) + 1\big) \leqslant u. \tag{2.3.70}$$

Indeed, by Definition 2.3.5(c) we have that

$$\Delta(G_5) \leqslant np_5 + 2\sqrt{np_5 \log n} \leqslant \frac{4np_5}{3} \overset{(2.3.64)}{=} \frac{2u}{3}.$$

Further, note that

$$(5 \log^4 n)^8 \leqslant np_0(1-p_0)^4. \tag{2.3.71}$$

Also, Definition 2.3.5(c) implies that

$$\Delta(G_{(2,j)}) - \delta(G_{(2,j)}) \leqslant 4\sqrt{np_{(2,j)} \log n} \leqslant 4\sqrt{np_2 \log n}$$

for each $1 \leqslant j \leqslant 2t_2 + 1$. Hence using $t_2 = o(\log n)$, we have

$$\begin{aligned} \sum_{j=1}^{2t_2+1} \left(\Delta(G_{(2,j)}) - \delta(G_{(2,j)}) + 1\right) &\leqslant 4(2t_2+1)\left(\sqrt{np_2 \log n} + 1\right) \\ &\leqslant \log n \sqrt{np_2 \log n} = \log^{13/4} n (np_0)^{3/8} \\ &\overset{(2.3.71)}{\leqslant} \frac{\sqrt{np_0(1-p_0)}}{5 \log^{3/4} n} = \frac{np_5}{5} \overset{(2.3.64)}{=} \frac{u}{10}. \end{aligned}$$

Similarly

$$\sum_{j=1}^{2t_i+1} \left(\Delta(G_{(i,j)}) - \delta(G_{(i,j)}) + 1\right) \leqslant \log n \sqrt{np_i \log n} \leqslant \frac{u}{10}$$

for $i = 3, 4$. Thus the left-hand side of (2.3.70) is at most $2u/3 + 3(u/10) \leqslant u$, which proves the claim.

Claim 2: *Each $x \neq x_0$ satisfies $d_{G_1}(x) \geqslant d_{G_1}(x_0) + 2u$.*

In other words, x_0 is the vertex of minimum degree in G_1 and G_1 is $2u$-jumping. To prove Claim 2, recall our assumption that G_0 is $4u$-jumping. Hence we have

$$\begin{aligned} d_{G_1}(x) &\geqslant \delta(G_0) + 4u - \Delta(G_5) - \sum_{i,j} \Delta(G_{(i,j)}) \\ &\overset{(2.3.70)}{\geqslant} \delta(G_0) - \sum_{i,j} (\delta(G_{(i,j)}) - 1) + 3u \overset{(2.3.69)}{\geqslant} d_{G_1}(x_0) - 1 + 3u \geqslant d_{G_1}(x_0) + 2u, \end{aligned}$$

which proves the claim.

Since each $H_{(i,j)}$ is even-regular and since M_{Opt} covers x_0 if and only if $d_{G_0}(x_0)$ is odd, (2.3.68) implies that $d_{G_1'}(x_0)$ is even. Also recall that the number of edges added to G_1 at the vertex x_0 (after removing M_{Opt}) to form G_1' is at most the left-hand side of (2.3.70), and hence is at most u. So x_0 is the vertex of minimum degree in G_1'. Moreover, since $1 - p_0 < 1 - p_1$,

$$u \overset{(2.3.64),(2.3.66)}{\leqslant} \frac{2\sqrt{n \cdot 2p_1(1-p_1)}}{\log^{3/4} n} \leqslant 4\sqrt{np_1(1-p_1)}.$$

Hence we may apply Lemma 2.3.18 with $G = G_1$ and $G' = G_1'$ to form a regular spanning subgraph H_1 of G_1' with degree $\delta(G_1') = d_{G_1'}(x_0)$. Note that H_1 contains every edge of G_1' incident to x_0.

By Lemma 2.3.60 where $p_0 = p_0$, $p_j = p_{(2,j)}$ for $1 \leqslant j \leqslant 2t_2 + 1$, $G_0 = G_0$, $H_0 = H_1$, $G_j = G_{(2,j)}$, and $H_j = H_{(2,j)}$ for $1 \leqslant j \leqslant 2t_2 + 1$, there exists a collection $\mathcal{HC}_1$ of edge-disjoint Hamilton cycles in $H_1 \cup H_2$ which cover the edges of H_1. Let H_2' be the graph formed by the edges of H_2 which are not contained in one of these Hamilton cycles.

Applying Lemma 2.3.60 again with $p_0 = p_2 + p_3$, $p_j = p_{(3,j)}$ for $1 \leqslant j \leqslant 2t_3 + 1$, $G_0 = G_2 \cup G_3$, $H_0 = H_2'$, $G_j = G_{(3,j)}$, and $H_j = H_{(3,j)}$, we obtain a collection $\mathcal{HC}_2$ of edge-disjoint Hamilton cycles in $H_2' \cup H_3$ which cover H_2'. Let H_3' be the graph formed by the edges of H_3 which are not covered by one of the Hamilton cycles.

Applying Lemma 2.3.60 again with $p_0 = p_3 + p_4$, $p_j = p_{(4,j)}$ for $1 \leqslant j \leqslant 2t_4 + 1$, $G_0 = G_3 \cup G_4$, $H_0 = H_3'$, $G_j = G_{(4,j)}$, and $H_j = H_{(4,j)}$, we obtain a collection $\mathcal{HC}_3$ of edge-disjoint Hamilton cycles in $H_3' \cup H_4$ which cover H_3'. Let H_4' be the graph formed by the edges of H_4 which are not covered by one of the Hamilton cycles.

Note that H_4' is a subgraph of G_4. Hence by Corollary 2.3.26 we can find a decomposition $\mathcal{F}$ of H_4' into 2-factors such that $c(\mathcal{F}) \leqslant 3n\sqrt{np_4 \log^3 n}$. Now we claim that

$$36(np_4)^3 \log^9 n \leqslant (np_5)^4. \tag{2.3.72}$$

To prove (2.3.72), note that $(np_5)^4 = (np_0(1-p_0))^2/\log^3 n$. So by (2.3.67) it suffices to prove that $36(np_0)^{81/64} \log^{67/2} n \leqslant (np_0(1-p_0))^2/\log^3 n$, or equivalently that $(np_0)^{47/64}(1-p_0)^2 \geqslant 36 \log^{73/2} n$. But if $p_0 \leqslant 1/2$ then we have

$$(np_0)^{47/64}(1-p_0)^2 \geqslant (np_0)^{47/64}/4 \geqslant (\log^{50} n)^{47/64}/4 \geqslant 36 \log^{73/2} n,$$

and if $p_0 \geqslant 1/2$ then

$$(np_0)^{47/64}(1-p_0)^2 \geqslant n^{47/64}(n^{-1/4})^2/2 \geqslant 36 \log^{73/2} n,$$

with room to spare, which proves (2.3.72).

It follows immediately from (2.3.72) that if $p_5 \leqslant p_4$, then

$$c(\mathcal{F}) \leqslant 3n\sqrt{np_4 \log^3 n} \leqslant \frac{n(np_5)^2}{2np_4 \log^3 n} \leqslant \frac{n(np_5)^2}{n(p_4+p_5)\log^3 n}.$$

On the other hand if $p_5 \geqslant p_4$, then

$$c(\mathcal{F}) \leqslant 3n\sqrt{np_4 \log^3 n} \leqslant 3n\sqrt{np_5 \log^3 n} \leqslant \frac{n(np_5)}{2\log^3 n} \leqslant \frac{n(np_5)^2}{n(p_4+p_5)\log^3 n}.$$

Hence in either case we can apply Corollary 2.3.59 with $G' = G_4 \cup G_5$, $G = G_5$, $H = H_5$ and $\mathcal{F} = \mathcal{F}$ to obtain a collection $\mathcal{HC}_4$ of edge-disjoint Hamilton cycles in $H'_4 \cup H_5$ which cover H'_4. Now let $\mathcal{HC} = \mathcal{HC}_1 \cup \mathcal{HC}_2 \cup \mathcal{HC}_3 \cup \mathcal{HC}_4$. Observe that $\mathcal{HC}$ covers every edge of H_1, H_2, H_3 and H_4. But recall that every edge of G_0 incident to x_0 is contained in either H_1, H_2, H_3, H_4 or M_{Opt}. Hence $\mathcal{HC}$ contains exactly $\lfloor d_{G_0}(x_0)/2 \rfloor = \lfloor \delta(G_0)/2 \rfloor$ Hamilton cycles. $\square$

Note that the only place where we use the full strength of the condition on p_0 is in the proof of (2.3.71) and (2.3.72). Also note that if we omit one of the iterations (i.e., if instead of defining p_4, G_4 and H_4 we simply use H_5 to finish the decomposition of H'_3) then the proof of Theorem 1.3.5 still works as long as $\log^{125} n/n \leqslant p \leqslant 1 - n^{-1/7}$ (say). On the other hand, we could have improved the lower bound on p in Theorem 1.3.5 somewhat by adding extra iterations. However, even a large number of iterations will only reduce the lower bound to approximately $\log^{30} n/n$, and in view of the result of [90], it is not necessary to reduce the exponent further in any case.

Chapter 3

Embedding spanning bipartite graphs of small bandwidth

3.1 Preliminaries

3.1.1 Notation

In this chapter, we use slightly different notation to that of Chapter 2. Specifically, given $S \subseteq V(G)$ we define $N(S) := \bigcup_{v \in S} N(v)$ (and not the 'external neighbourhood', as in Chapter 2).

3.1.2 Degree sequence and Ore-type conditions forcing robust expansion

In this section we show that Theorem 1.4.8 implies Theorem 1.4.4 and that Theorem 1.4.4 implies Theorem 1.4.5. The remainder of the chapter is devoted to proving Theorem 1.4.8 directly.

The following result is an immediate consequence of Lemma 13 from [87].

Lemma 3.1.1 ([87]) *Given positive constants $\tau \ll \eta < 1$ there exists an integer n_0 such* *that whenever G is a graph on $n \geq n_0$ vertices with*

$$d_i \geqslant i + \eta n \text{ or } d_{n-i-\eta n} \geqslant n - i \text{ for all } i < n/2,$$

then $\delta(G) \geq \eta n$ and G is a robust (τ^2, τ)-expander. □

Notice that Lemma 3.1.1 together with Theorem 1.4.8 implies Theorem 1.4.4. We now show that Theorem 1.4.4 implies Theorem 1.4.5.

Lemma 3.1.2 *Let $0 < \gamma < 1/2$. Suppose G is a graph on n vertices such that, for all non-adjacent $x \neq y \in V(G)$,*

$$d(x) + d(y) \geqslant (1 + 2\gamma)n.$$

Let $d_1 \leqslant \ldots \leqslant d_n$ denote the degree sequence of G. Then

$$d_i \geqslant i + \gamma n \text{ or } d_{n-i-\gamma n} \geqslant n - i \text{ for all } i < n/2.$$

Proof. Firstly note that for $(1-\gamma)n/2 \leqslant i < n/2$ we wish to show that either $d_{n-i'-\gamma n} \geqslant n - i'$ or $d_{i'} \geqslant i' + \gamma n$, where $i' := n - i - \gamma n$. Notice that $n/2 - \gamma n < i' \leqslant n/2 - \gamma n/2$. Thus, it suffices to only consider i such that $1 \leqslant i \leqslant (1-\gamma)n/2$.

Suppose there is some $1 \leqslant i \leqslant (1-\gamma)n/2$ such that the statement does not hold. Then there is a set A of i vertices, each of degree less than $i + \gamma n \leqslant n/2 + \gamma n/2$. So for any $x, y \in A$, $d(x) + d(y) < (1 + 2\gamma)n$ and hence $G[A]$ is a clique. Set $B := V(G) \backslash A$. Note that $e_G(A, B) < (\gamma n + 1)i$. Hence, there is a vertex $x \in B$ that receives less than $\min\{\gamma n + 1, i\}$ edges from A. Therefore, there is a vertex $y \in A$ such that $xy \notin E(G)$. Thus, $d(x) + d(y) < (n - i - 1 + \gamma n + 1) + (i + \gamma n) \leqslant (1 + 2\gamma)n$, contradicting our assumption. □

3.2 Outline of the proof of Theorem 1.4.8

3.2.1 Proof overview

The overall strategy is similar to that of the proof of Theorem 1.4.1 in [22]. Indeed, as in [22] the proof is split into two main lemmas; the *Lemma for G* and the *Lemma for H*. However, many of the methods used in [22] break down in our setting, as detailed in the remainder of this section, so our argument proceeds somewhat differently.

The role of the Lemma for G (Lemma 3.6.1) is to obtain some special structure within G so that it will be suitable for embedding H into; By applying Theorem 1.4.7, we show that G contains a spanning subgraph G' which 'looks' like the blow-up of a cycle $C = V_1V_2 \dots V_{2k}V_1$. (This structure is somewhat simpler than in the corresponding lemma in [22].) More precisely, there is a partition $V_1, \dots, V_{2k}$ of $V(G)$ such that:

(i) $(V_{2i-1}, V_{2i})_{G'}$ is a 'super-regular' pair of density at least $d > 0$ for each $1 \leqslant i \leqslant k$;

(ii) $(V_{2i}, V_{2i+1})_{G'}$ is an 'ε-regular' pair of density at least d for each $1 \leqslant i \leqslant k$,

where the density of a pair (A, B) is $e(A, B)/|A||B|$. Furthermore, there are even integers $1 \leqslant i_1 \neq j_1 \leqslant 2k$ such that:

(iii) $(V_{i_1}, V_{j_1})_{G'}$ is 'ε-regular' with density at least d.

(So $V_{i_1}V_{j_1}$ can be thought of as a chord of C.) Crucially, this partition is 'robust' in the sense that one can modify the sizes of each partition class V_i somewhat without essentially destroying the properties (i)–(iii). In [22] the high minimum degree of G allowed this to be done relatively easily. Since in our case G may not have high minimum degree, we rely on the *Mobility Lemma* (Lemma 3.5.1), which is given in Section 3.5.)

Set $c := V_{i_1}V_{j_1}$. The role of the Lemma for H (Lemma 3.7.1) is to construct a graph homomorphism f from H to $C \cup \{c\}$ in such a way that 'most' of the edges of H are mapped to edges of the form $V_{2i-1}V_{2i}$ for some i. (Recall that these are the edges which correspond to super-regular pairs in G'.) The homomorphism f is such that every $V_i \in C$ receives roughly $|V_i|$ vertices of H. So f can be viewed as a 'guide' as to which vertex class $V_i \subseteq V(G)$ each vertex from H is embedded into. In [22], an argument using colourings was used to obtain a different structure, which would not be compatible with our Lemma for G. Here we instead use a randomised algorithm to obtain our desired homomorphism. In particular, since the partition $V_1, \dots, V_{2k}$ is 'robust', we can alter the sizes of the classes V_i such that (i)–(iii) still hold and so that now $|f^{-1}(V_i)| = |V_i|$ for all i. Properties (i)–(iii) then allow us to apply the Blow-up lemma [76] to embed H into G' and thus G. (Actually we apply a result from [18] which is a consequence of the Blow-up lemma.)

3.2.2 Techniques for the Lemma for G

In order to obtain the partition $V_1, \dots, V_{2k}$ of $V(G)$ we modify a partition $V'_0, V'_1, \dots, V'_{2k}$ obtained by applying Szemerédi's Regularity lemma [110] to G. Roughly speaking, $V'_1, \dots, V'_{2k}$ will satisfy (i)–(iii). Thus, we need to redistribute the vertices from V'_0 into the other vertex classes whilst retaining these properties. We also require our partition $V_1, \dots, V_{2k}$ to satisfy $|V_{2i-1}| \approx |V_{2i}|$ for each $1 \leqslant i \leqslant k$. So we need to redistribute vertices in a 'balanced' way. In the Lemma for G in [22], the minimum degree condition of Theorem 1.4.1 is heavily relied on to achieve this.

However, our graph G may have very small minimum degree. So instead we introduced the notion of a 'shifted M-walk' to help us redistribute vertices: Given a perfect matching M in a graph R a shifted M-walk is a walk whose edges alternate between edges of M and edges of $R \backslash M$ (see Section 3.4 for the precise definition). Since G is a robust expander, we can find short shifted M-walks in a reduced graph R of G. (Here, M will be the perfect matching in R that corresponds to the super-regular pairs from (i) above.) These walks act as a 'guide' as to how we redistribute vertices amongst the vertex classes.

3.2.3 Techniques for the Lemma for H

In [22] the techniques used are actually strong enough to prove a more general result than Theorem 1.4.1 (and so Theorem 1.4.1 is not proved directly). For example, in the case when $r = 2$, their result concerns not only bipartite H but also a special class of 3-colourable graphs H where the third colour class is very small (see Theorem 2 in [22] for precise details). One example of such a graph H is a Hamilton cycle C' with a chord between two vertices of distance 2 on C'. H is 3-colourable and has bounded bandwidth. However, H cannot be embedded into *every* graph G satisfying the hypothesis of Theorem 1.4.8. Indeed, consider the graph G defined at the end of Section 1.4.3.

In particular, this means we have to approach the proof of the Lemma for H differently: Since H has bandwidth $o(n)$ we can chop $V(H)$ into small linear sized segments $A_1, B_1, \dots, A_m, B_m$ where all the edges of H lie in pairs of the form $(A_i, B_i)_H$ and $(B_i, A_{i+1})_H$ and such that $A := \cup_{i=1}^m A_i$ and $B := \cup_{i=1}^m B_i$ are the colour classes of H. Ideally we would want to construct f to map the vertices of A_1 into V_1, the vertices of B_1 into V_2 and so on, continuing around C many times until all the vertices have been assigned. However, since $|A|$ and $|B|$ may vary widely, this would map vertices in an unbalanced way. That is, the total number of vertices mapped to 'odd' classes V_{2i-1} would differ widely from the total number of vertices mapped to 'even' classes V_{2i}. We get around this problem by using the chord $c = V_{i_1}V_{j_1}$ to 'flip' halfway in the process. So after this, vertices from the B_i are mapped to 'odd' classes V_{2i-1} and vertices from the A_i are mapped to the 'even' classes V_{2i}. We also 'randomise' part of the mapping procedure to ensure that the number of vertices of H assigned to each V_i is approximately $|V_i|$. (A randomisation technique of a similar flavour was used in [89].)

3.3 The Regularity Lemma

In the proof of the Lemma for G (Lemma 3.6.1) we will use Szemerédi's Regularity lemma [110]. In this section we will introduce all the information we require about this result. To do this, we firstly introduce some more notation. Recall that the *density* of a bipartite graph G with vertex classes A and B is defined to be

$$d(A,B) := \frac{e(A,B)}{|A||B|}.$$

Given any $\varepsilon, d > 0$, we say that G is (ε, d)*-regular* if $d(A,B) \geqslant d$ and, for all sets $X \subseteq A$ and $Y \subseteq B$ with $|X| \geq \varepsilon|A|$ and $|Y| \geq \varepsilon|B|$, we have $|d(A,B) - d(X,Y)| < \varepsilon$. We say that G is (ε, d)*-super-regular* if additionally every vertex $a \in A$ has at least $d|B|$ neighbours in B and every vertex $b \in B$ has at least $d|A|$ neighbours in A. We also say that (A,B) is an (ε, d)*-(super-)regular pair*. We will frequently use the following simple fact.

Fact 3.3.1 *Let $\varepsilon, d > 0$. Suppose that $G = (A,B)$ is an (ε, d)-regular pair. Let $B' \subseteq B$ be such that $|B'| \geqslant \varepsilon|B|$. Then there are at most $\varepsilon|A|$ vertices in A with fewer than $(d-\varepsilon)|B'|$ neighbours in B'.* $\square$

We will also require the next simple proposition which allows us to modify a (super-)regular pair without destroying its (super-)regularity (see e.g., [21, Proposition 8]).

Proposition 3.3.2 *Let $\alpha, \beta, \varepsilon, \delta > 0$ and suppose that (A,B) is an (ε, d)-regular pair. Let A' and B' be vertex sets with $|A'\Delta A| \leqslant \alpha|A|$ and $|B'\Delta B| \leqslant \beta|B|$. Then (A', B') is an (ε', d')-regular pair where*

$$\varepsilon' := \varepsilon + 3(\sqrt{\alpha} + \sqrt{\beta}) \text{ and } d' := d - 2(\alpha + \beta).$$

If, moreover, (A,B) is (ε, d)-super-regular and each vertex in A' has at least $d|B'|$ neigh-

bours in B' and each vertex in B' has at least $d|A'|$ neighbours in A', then (A', B') is (ε', d')-super-regular. □

We will use the following degree form of Szemerédi's Regularity lemma [110] which can be easily derived from the classical version.

Lemma 3.3.3 (Regularity lemma) *For every $\varepsilon > 0$ and $k_0 \in \mathbb{N}$ there exists $K_0 = K_0(\varepsilon, k_0)$ such that for every $d \in [0, 1]$ and for every graph G on $n \geqslant K_0$ vertices there exists a partition $V_0, V_1, \ldots, V_k$ of $V(G)$ and a spanning subgraph G' of G, such that the following conditions hold:*

(i) $k_0 \leqslant k \leqslant K_0$,

(ii) $d_{G'}(x) \geqslant d_G(x) - (d + \varepsilon)n$ *for every* $x \in V(G)$,

(iii) *the subgraph* $G'[V_i]$ *is empty for all* $1 \leqslant i \leqslant k$,

(iv) $|V_0| \leqslant \varepsilon n$,

(v) $|V_1| = |V_2| = \ldots = |V_k|$,

(vi) *for all* $1 \leqslant i < j \leqslant k$ *either* $(V_i, V_j)_{G'}$ *is an* (ε, d)*-regular pair or* $G'[V_i, V_j]$ *is empty.*

□

We call $V_1, \ldots, V_k$ *clusters*, V_0 the *exceptional set* and the vertices in V_0 *exceptional vertices.* We refer to G' as the *pure graph.* The *reduced graph R of G with parameters ε, d and k_0* is the graph whose vertices are $V_1, \ldots, V_k$ and in which V_iV_j is an edge precisely when $(V_i, V_j)_{G'}$ is (ε, d)-regular.

The following result implies that the property of a graph G being a robust expander is 'inherited' by the reduced graph R of G. It is an immediate consequence of Lemma 14 from [87].

Lemma 3.3.4 ([87]) *Let k_0, n_0 be positive integers and let $\varepsilon, d, \eta, \nu, \tau$ be positive constants such that $1/n_0 \ll \varepsilon \ll d \ll \nu, \tau, \eta < 1$ and such that $k_0 \ll n_0$. Let G be a graph on $n \geq n_0$ vertices with $\delta(G) \geq \eta n$ and such that G is a robust (ν, τ)-expander. Let R be the reduced graph of G with parameters ε, d and k_0. Then $\delta(R) \geq \eta|R|/2$ and R is a robust $(\nu/2, 2\tau)$-expander.* □

3.4 Shifted walks and robust expanders

Let G be a graph containing a perfect matching M. A *shifted M-walk* in G with endpoints $a = v_1$ and $b = v_{2\ell}$ is a walk $v_1 v_2 \dots v_{2\ell}$ in G such that $v_{2i}v_{2i+1} \in M$ for every $1 \leqslant i \leqslant \ell - 1$ and $v_{2i-1}v_{2i} \notin M$ for any $1 \leqslant i \leqslant \ell$. A shifted M-walk is *simple* if it contains each edge of M at most twice. Note that a path containing a single edge $v_1v_2 \notin M$ is a (simple) shifted M-walk for any perfect matching M.

Lemma 3.4.1 *Let G be a graph containing a perfect matching M, and let W be a shifted M-walk in G with endpoints a and b. Then W contains a simple shifted M-walk W' with endpoints a and b.*

Proof. We proceed by induction. Let $W = v_1 \dots v_{2\ell}$. If W is already simple then we set $W' := W$; otherwise, there exists an edge $xy \in M$ which appears at least three times in W. Let the first three appearances of xy be $v_{i_1}v_{i_1+1}$, $v_{i_2}v_{i_2+1}$ and $v_{i_3}v_{i_3+1}$ (in order). Now each of v_{i_1}, v_{i_2} and v_{i_3} is either x or y, and so without loss of generality we can assume that $v_{i_1} = v_{i_2} = x$. This yields a shorter shifted M-walk $v_1 \dots v_{i_1} v_{i_2+1} \dots v_{2\ell}$ whose endpoints are the same as those of W. We repeat this process until we obtain a walk in which every edge appears at most twice, and set W' to be the walk thus obtained. □

Lemma 3.4.2 *Let M be a perfect matching in a graph G and let $A \subseteq V(M)$ be a set containing at most one vertex from each edge of M. Suppose that W is a shifted M-walk* *both of whose endpoints lie in A. Then W contains a shifted M-walk W' such that the endpoints of W' both lie in A and no other vertices of W' lie in A.*

Proof. Let $W = v_1 \dots v_{2\ell}$. Let $1 \leqslant i_1 \leqslant \ell$ be minimal such that $v_{2i_1} \in A$, and let $1 \leqslant i_2 \leqslant i_1$ be maximal such that $v_{2i_2-1} \in A$. Now $v_{2i_2-1}v_{2i_2} \dots v_{2i_1-1}v_{2i_1}$ is the desired shifted M-walk. □

In the proof of the Lemma for G we will use shifted walks in the reduced graph R of G as a "guide" as to how to redistribute vertices in G. Since the reduced graph R will be a robust expander, the following result ensures we can find our desired shifted walks.

Let G be a graph containing a perfect matching M, and let $A \subseteq V(G)$. For each $v \in V(G)$, let $v' \in V(G)$ be the unique vertex such that $vv' \in M$. The *shifted M-neighbourhood* of A is the set $SN_M(A) = \{v' \mid v \in N(A)\}$. $SN^r_M(A)$ is defined recursively by $SN^1_M(A) := SN_M(A)$ and $SN^r_M(A) := SN_M(SN^{r-1}_M(A))$ for $r \geqslant 2$.

Lemma 3.4.3 *Let $0 < \nu \leqslant \tau < \eta \ll 1$ be constants. Suppose G is a graph on n vertices with $\delta(G) \geqslant \eta n$ which is a robust (ν, τ)-expander, and let M be a perfect matching in G. Then for any $a \in V(G)$, G contains a shifted M-walk of length at most $3/\nu$ which both starts and finishes at a.*

Proof. The minimum degree condition implies that $|SN_M(a)| = |N(a)| \geqslant \eta n \geqslant \tau n$. Since G is a robust (ν, τ)-expander,

$$|SN^r_M(a)| = |N(SN^{r-1}_M(a))| \geqslant \min\left\{|SN^{r-1}_M(a)| + \nu n, (1 - \tau + \nu)n\right\},$$

for all $r \geqslant 2$. Hence $|SN^{1/2\nu}_M(a)| \geqslant (\tau + 1/2 - \nu)n$ and so

$$|N(SN^{1/2\nu}_M(a))| \geqslant (1/2 + \tau)n.$$

Thus, there exists some edge $vv' \in M$ such that both v and v' lie in $N(SN_M^{1/2\nu}(a))$. This implies that there exists a shifted M-walk P with endpoints a and v and a shifted M-walk P' with endpoints a and v', each of length at most $1/\nu + 1$. Now $P \cup vv' \cup P'$ forms a shifted M-walk of length at most $2/\nu + 3 \leqslant 3/\nu$ which starts and finishes at a. $\square$

The next lemma allows us to delete a small number of vertices from a robust expander without essentially destroying this property.

Lemma 3.4.4 *Let $0 < \alpha < \nu \leqslant \tau \ll 1$ be constants. Suppose that G is a graph on n vertices which is a robust (ν, τ)-expander and let $S \subseteq V(G)$ be a set of size αn. Then $G - S$ is a robust $(\nu - \alpha, \tau + \alpha)$-expander.*

Proof. Let $G' := G - S$ and $n' := |G'|$. Consider any $A \subseteq V(G')$ such that $(\tau + \alpha)n' \leqslant |A| \leqslant (1 - \tau - \alpha)n'$. Set $A' := A \cup S$. Then

$$\tau n \leqslant (\tau + \alpha)n' + \alpha n \leqslant |A'| \leqslant (1 - \tau - \alpha)n' + \alpha n \leqslant (1 - \tau)n.$$

So $|RN_{\nu,G}(A')| \geqslant |A'| + \nu n$. Now every vertex of $RN_{\nu,G}(A')$ has at least νn neighbours in A' and since $|S| = \alpha n$, at least $(\nu - \alpha)n \geqslant (\nu - \alpha)n'$ of these must lie in A. Hence every vertex of $RN_{\nu,G}(A') \backslash S$ lies in $RN_{\nu-\alpha,G'}(A)$, and so $|RN_{\nu-\alpha,G'}(A)| \geqslant |A'| + \nu n - |S| \geqslant |A| + (\nu - \alpha)n'$, as desired. $\square$

3.5 The Mobility Lemma

In order to state our next result we first introduce a slight variant of the notion of a reduced graph, as obtained via Lemma 3.3.3. Let $\varepsilon, \varepsilon', d, d' > 0$. Suppose that G is a graph and $V_1, \dots, V_k$ is a partition of $V(G)$. We say that a graph R is an (ε, d)*-reduced graph of G on $V_1, \dots, V_k$* if the following holds:

- $V(R) = \{V_1, \dots, V_k\}$;

- If $V_iV_j \in E(R)$ then $(V_i, V_j)_G$ is an (ε, d)-regular pair (for all $1 \leqslant i \neq j \leqslant k$).

Suppose $V'_1, \dots, V'_k$ is another partition of $V(G)$ and R is as above (in particular, $V(R) = \{V_1, \dots, V_k\}$). We also say that R is an (ε', d')*-reduced graph of G on* $V'_1, \dots, V'_k$ if the following holds:

- If $V_iV_j \in E(R)$ then $(V'_i, V'_j)_G$ is an (ε', d')-regular pair (for all $1 \leqslant i \neq j \leqslant k$).

Suppose that $V_1, \dots, V_k$ and $V'_1, \dots, V'_k$ are both partitions of the vertex set of a graph G. Given a cluster $V = V_i$ for some $1 \leqslant i \leqslant k$, we will often denote by V' the cluster V'_i. The above definition allows us, after modifying a partition $\{V_1, \dots, V_k\}$ to obtain $\{V'_1, \dots, V'_k\}$, to continue to refer to the same reduced graph R, even though strictly speaking the clusters of R and the parts of the partition are no longer identical.

We will apply the next result in the proof of the Lemma for G (Lemma 3.6.1) so that we can alter a particular partition of a graph G somewhat without destroying the structure of our reduced graph R.

Lemma 3.5.1 (Mobility lemma) *Let $k \in \mathbb{N}$, and let $\xi, \varepsilon, \varepsilon', d', d$ be positive constants such that*

$$0 < \xi \ll 1/k \ll \varepsilon \ll \varepsilon' \ll d' \ll d \ll 1.$$

Suppose G is a graph on n vertices, $A_1, B_1, A_2, B_2, \dots, A_k, B_k$ is a partition of $V(G)$ such that $|A_i|, |B_i| \geqslant n/3k$ for all $1 \leqslant i \leqslant k$ and R is an (ε, d)-reduced graph on $A_1, B_1, \dots, A_k, B_k$. Let $(a_i)_{i=1}^k$ and $(b_i)_{i=1}^k$ be integers. Suppose that the following conditions hold:

(i) *R contains the Hamilton cycle $C = A_1B_1A_2B_2 \dots A_kB_kA_1$;*

(ii) *R contains an edge $A_{i_1}A_{j_1}$ for some $i_1 \neq j_1$;*

(iii) *R contains an edge $B_{i_2}B_{j_2}$ for some $i_2 \neq j_2$;*

(iv) *The pair* $(A_i, B_i)_G$ *is* (ε, d)*-super-regular for all* $1 \leqslant i \leqslant k$*;*

(v) $|a_i|, |b_i| < \xi n$ *for each* $1 \leqslant i \leqslant k$*;*

(vi) $\sum_{i=1}^k a_i + \sum_{i=1}^k b_i = 0$*;*

(vii) $|\sum_{i=1}^k a_i| = |\sum_{i=1}^k b_i| \leqslant \xi n$*.*

Then there exists a partition $A'_1, B'_1, A'_2, B'_2, \dots, A'_k, B'_k$ *of* $V(G)$ *such that* $|A'_i| = |A_i| + a_i$ *and* $|B'_i| = |B_i| + b_i$ *for each* $1 \leqslant i \leqslant k$*,* R *is an* (ε', d')*-reduced graph of* G *on* $A'_1, B'_1, A'_2, B'_2, \dots, A'_k, B'_k$*, and* $(A'_i, B'_i)_G$ *is* (ε', d')*-super-regular for each* $1 \leqslant i \leqslant k$*.*

Proof. Without loss of generality we may assume that $\sum_{i=1}^k a_i \geqslant 0$. (As a consequence of this assumption we will in fact only need the edge $B_{i_2}B_{j_2}$, and not the edge $A_{i_1}A_{j_1}$.) Note that by (iii) and Fact 3.3.1 there are at least $(1-\varepsilon)|B_{i_2}| \gg \xi n$ vertices in B_{i_2} with at least $(d-\varepsilon)|B_{j_2}|$ neighbours in B_{j_2}. Pick $\sum_{i=1}^k a_i \leqslant \xi n$ of these vertices and move them from B_{i_2} into A_{j_2}. Call the resulting sets $B^*_{i_2}$ and $A^*_{j_2}$ respectively.

We now perform an iterative procedure which will reassign vertices among the vertex classes $(A_i)_{i=1}^k$ and, separately, $(B_i)_{i=1}^k$. Initially we define the classes $A^*_i = A_i$ for each $i \neq j_2$ and $B^*_i = B_i$ for each $i \neq i_2$. Roughly speaking, A^*_i (or B^*_i) will be the current version of A_i (or B_i). The choice of how we defined $B^*_{i_2}$ and $A^*_{j_2}$ is such that, initially,

$$\sum_{i=1}^k |A^*_i| = \sum_{i=1}^k |A_i| + \sum_{i=1}^k a_i \quad \text{and} \quad \sum_{i=1}^k |B^*_i| = \sum_{i=1}^k |B_i| + \sum_{i=1}^k b_i. \tag{3.5.2}$$

Throughout the procedure we will ensure that (3.5.2) holds. Furthermore, throughout we will ensure that

$$|A^*_i \Delta A_i|, |B^*_i \Delta B_i| \leqslant 5k\xi n \leqslant \varepsilon |A_i|, \varepsilon |B_i| \tag{3.5.3}$$

for every $1 \leqslant i \leqslant k$. We will also ensure that whenever a vertex v is moved to a cluster A^*_i, v has at least $(d-\varepsilon)|B_i|$ neighbours in B_i, and vice versa. We will terminate the procedure when $|A_i^*| = |A_i| + a_i$ and $|B_i^*| = |B_i| + b_i$, and then set $A_i' := A_i^*$ and $B_i' := B_i^*$ for each $1 \leqslant i \leqslant k$.

Each iteration proceeds as follows: Let $1 \leqslant i \leqslant k$ be such that $|A_i^*| < |A_i| + a_i$ and let $j \neq i$ be such that $|A_j^*| > |A_j| + a_j$. (Such i and j exist by (3.5.2).) Suppose that $i < j$. Note that (i) implies that $(B_{j-1}, A_j)_G$ is an (ε, d)-regular pair. So by (3.5.3) and Fact 3.3.1 there is a vertex v in A_j^* which has at least $(d-\varepsilon)|B_{j-1}|$ neighbours in B_{j-1}. We move v from A_j^* to A_{j-1}^*. Similarly we move one vertex (which need not be v) from A_{j-1}^* to A_{j-2}^*, and so on until we move one vertex from A_{i+1}^* to A_i^*. On the other hand, if $j < i$ we perform the same procedure moving vertices in the *same* direction as before. That is, we move a vertex from A_j^* to A_{j-1}^* and so on until we move a vertex A_2^* to A_1^*. Then we move a vertex A_1^* to A_k^* and continue until we move a vertex from A_{i+1}^* to A_i^*.

Since in each step of the process we only move vertices between the A_i^*, certainly (3.5.2) holds throughout. Now when the procedure terminates we have $|A_i^*| = |A_i| + a_i$ for all $1 \leqslant i \leqslant k$. It remains to show that (3.5.3) holds. Note that in each step of the iteration we add at most one vertex to each A_i^* and remove at most one vertex from each A_i^*. Further, in total we need to perform the iterative procedure at most

$$\sum_{i=1}^{k} |a_i| + \sum_{i=1}^{k} a_i \leqslant (k+1)\xi n \leqslant 2k\xi n$$

times. (The $\sum_{i=1}^{k} a_i$ here comes from the fact that, at the start, we moved $\sum_{i=1}^{k} a_i$ vertices from B_{i_2} to A_{j_2}.) Thus, at the end of the procedure $|A_{j_2}^* \Delta A_{j_2}| \leqslant 5k\xi n$ and $|A_i^* \Delta A_i| \leqslant 4k\xi n$ for all $i \neq j_2$. We now set $A_i' := A_i^*$ for each $1 \leqslant i \leqslant k$.

We apply an identical iterative procedure to the B_i^*. However, we now move vertices in the *opposite* direction to before (so vertices are moved from B_j^* to B_{j+1}^*, etc.). Therefore we obtain sets B_i' such that $|B_i'| = |B_i| + b_i$ and $|B_i' \Delta B_i| \leqslant 5k\xi n$ for all $1 \leqslant i \leqslant k$.

Given any $VW \in E(R)$, $(V, W)_G$ is an (ε, d)-regular pair. Since by (3.5.3), $|V'\Delta V|$, $|W'\Delta W| \leqslant \varepsilon|V|, \varepsilon|W|$, Proposition 3.3.2 implies that $(V', W')_G$ is an (ε', d')-regular pair. So indeed, R is an (ε', d')-reduced graph of G on $A'_1, B'_1, A'_2, B'_2, \dots, A'_k, B'_k$. It remains to show that the pair $(A'_i, B'_i)_G$ is (ε', d')-super-regular for every $1 \leqslant i \leqslant k$. By (iv) and (3.5.3), every vertex $v \in A_i$ has at least $(d - \varepsilon)|B_i| \geqslant d'|B'_i|$ neighbours in B'_i. Further, during our iterative procedure we ensured that every vertex $v \in A'_i \backslash A_i$ has at least $(d - \varepsilon)|B_i|$ neighbours in B_i. Hence (3.5.3) implies that every $v \in A'_i$ has at least

$$(d - \varepsilon)|B_i| - \varepsilon|B_i| \geqslant d'|B'_i|$$

neighbours in B'_i. Similarly each $w \in B'_i$ has at least $d'|A'_i|$ neighbours in A'_i. So $(A'_i, B'_i)_G$ is an (ε', d')-super-regular pair for all $1 \leqslant i \leqslant k$, as desired. □

3.6 The Lemma for G

Lemma 3.6.1 (Lemma for G) *Let $n_0 \in \mathbb{N}$ and let $\lambda, \xi, \varepsilon, d, \nu, \tau, \eta$ be positive constants such that*

$$0 < 1/n_0 \ll \lambda \ll \xi \ll \varepsilon \ll d \ll \nu \leqslant \tau \ll \eta \ll 1.$$

Suppose G is a graph on $n \geqslant n_0$ vertices with $\delta(G) \geqslant \eta n$ which is a robust (ν, τ)-expander. Then for some $\xi \ll 1/k \ll \varepsilon$, there exist integers $1 \leqslant i_1 \neq j_1, i_2 \neq j_2 \leqslant k$ and a partition $(n_i)_{i=1}^{2k}$ of n with $n_i > n/3k$ for all $1 \leqslant i \leqslant 2k$ and $|n_{2i-1} - n_{2i}| \leqslant \lambda n$ for all $1 \leqslant i \leqslant k$ such that the following holds: For every partition $(n'_i)_{i=1}^{2k}$ of n satisfying $n'_i \leqslant n_i + \xi n$ for all $1 \leqslant i \leqslant 2k$, there exists a partition $A'_1, B'_1, A'_2, B'_2, \dots, A'_k, B'_k$ of $V(G)$ and a spanning subgraph G' of G such that the following properties are satisfied.

(α_1) $|A'_i| = n'_{2i-1}$ *and* $|B'_i| = n'_{2i}$ *for all* $1 \leqslant i \leqslant k$;

(α_2) $(A'_i, B'_i)_{G'}$ *is* (ε, d)*-super-regular for all* $1 \leqslant i \leqslant k$;

(α_3) $(B'_i, A'_{i+1})_{G'}$ *is* (ε, d)*-regular for all* $1 \leqslant i \leqslant k$ *(where* $A'_{k+1} := A'_1$*);*

(α_4) $(A'_{i_1}, A'_{j_1})_{G'}$ *is* (ε, d)*-regular;*

(α_5) $(B'_{i_2}, B'_{j_2})_{G'}$ *is* (ε, d)*-regular.*

Proof. Choose additional constants ε' and d' such that

$$\xi \ll \varepsilon' \ll \varepsilon \ll d \ll d' \ll \nu.$$

Apply the Regularity Lemma (Lemma 3.3.3) with parameters ε', d' and $k_0 := 1/\varepsilon'$ to obtain clusters $V_1, \dots, V_{k'}$ of size m (where $(1-\varepsilon')n/k' \leqslant m \leqslant n/k'$), an exceptional set V_0, a pure graph $G' \subseteq G$ (recall that the pure graph was defined after Lemma 3.3.3) and the reduced graph R of G with parameters ε', d' and k_0. Since $\xi \ll \varepsilon'$ we may assume that

$$\xi \ll 1/k' \leqslant \varepsilon'.$$

If k' is odd then we delete $V_{k'}$ from R and add all of the vertices of $V_{k'}$ to V_0. So $|V_0| \leqslant \varepsilon' n + m \leqslant 2\varepsilon' n$. We now refer to this modified reduced graph as R and redefine $k' = |R|$. By Lemma 3.3.4, R originally had minimum degree at least $\eta k'/2$ and was a robust $(\nu/2, 2\tau)$-expander. So R still has minimum degree at least $\eta k'/3$ and by Lemma 3.4.4, R is still a robust $(\nu/3, 3\tau)$-expander.

Set $k := k'/2$. Since $1/k' \ll \nu \leqslant \tau \ll \eta < 1$, Theorem 1.4.7 implies that R contains a Hamilton cycle $C = A_1B_1 \dots A_kB_kA_1$. Since $|C| = 2k$ is even, C contains a perfect matching $M = \{A_1B_1, \dots, A_kB_k\}$. Notice that R contains an edge $A_{i_1}A_{j_1}$ for some $1 \leqslant i_1 \neq j_1 \leqslant k$ and an edge $B_{i_2}B_{j_2}$ for some $1 \leqslant i_2 \neq j_2 \leqslant k$. Indeed, let $A := \{A_i\}_{i=1}^k$ and note that since R is a robust $(\nu/3, 3\tau)$-expander we have $|RN_{\nu,R}(A)| \geqslant k + \nu k'$. This implies that $A \cap RN_{\nu,R}(A) \neq \emptyset$ and hence that R contains some edge $A_{i_1}A_{j_1}$. Similarly R contains an edge $B_{i_2}B_{j_2}$.

Fact 3.3.1 implies that we can replace each cluster in $V(R)$ with a subcluster of size $m' := (1-\varepsilon')m$ such that for every edge $A_jB_j \in M$ the chosen subclusters of A_j and B_j form a $(2\varepsilon', d'/2)$-super-regular pair in G'. We add all of the vertices not in these subclusters to V_0, and from now on we refer to the subclusters as the clusters of R. So $(V, W)_{G'}$ is still a $(2\varepsilon', d'/2)$-regular pair for all $VW \in E(R)$. Note that $|V_0| \leqslant 2\varepsilon' n + \varepsilon' n = 3\varepsilon' n$.

Our next task is to incorporate the vertices of V_0 into the clusters $V_1, \dots, V_{k'}$ so that the pairs $(A_j, B_j)_{G'}$ remain super-regular and such that the pairs $(V_i, V_j)_{G'}$ remain regular for all $V_iV_j \in E(R)$ (with somewhat weaker constants in each case). Let $V_0 = \{x_1, \dots, x_t\}$ where $t \leqslant 3\varepsilon' n$. We will assign the vertices of V_0 in such a way that:

(a) At most $8\varepsilon' m'/\eta$ vertices are assigned to each cluster $V \in V(R)$;

(b) Whenever a vertex $x_i \in V_0$ is assigned to a cluster A_j, x_i has at least $\eta m'/4$ neighbours in B_j. Similarly any vertex from V_0 assigned to B_j has at least $\eta m'/4$ neighbours in A_j.

Suppose we have assigned $x_1, \dots, x_{i-1}$ to clusters in $V(R)$ such that (a) and (b) are satisfied. Call a cluster $V \in V(R)$ *full* if it has already been assigned $8\varepsilon' m'/\eta$ vertices of V_0. Let F be the set of full clusters. Since $|V_0| \leqslant 3\varepsilon' n$ we have $|F| \leqslant (3\varepsilon' \eta n)/(8\varepsilon' m') \leqslant \eta k$. Thus, as $\delta(G) \geqslant \eta n$,

$$\left| N_G(x_i) \backslash \left(V_0 \cup \bigcup_{V \in F} V \right) \right| \geqslant \eta n - 3\varepsilon' n - (\eta k) m' \geqslant \eta n/3.$$

Hence, by the pigeonhole principle there exists some $V \in V(R) \backslash F$ such that $|N_G(x_i) \cap V| \geqslant \eta n/3k' \geqslant \eta m'/4$. Now if $V = A_j$ for some $1 \leqslant j \leqslant k$ then we add x_i to B_j; otherwise, $V = B_j$ for some $1 \leqslant j \leqslant k$ and we add x_i to A_j. Repeating this process for each x_i we indeed assign all of the vertices of V_0 to the clusters of R in such a way that (a) and (b) are satisfied. We now incorporate all of the assigned vertices into their respective clusters. Further, we add all those edges from G with endpoints in V_0 to G'. Note that

(c) $m' \leqslant |V| \leqslant m' + 8\varepsilon' m'/\eta \leqslant (1+\sqrt{\varepsilon'})m'$ for all $V \in V(R)$,

(d) $(V, W)_{G'}$ is a $((\varepsilon')^{1/3}, d'/4)$-regular pair for every edge $VW \in E(R)$ and

(e) $(V, W)_{G'}$ is a $((\varepsilon')^{1/3}, d'/4)$-super-regular pair for every edge $VW \in E(M)$.

(Conditions (d) and (e) follow by Proposition 3.3.2.)

Next we will perform an algorithm which redistributes vertices among the clusters in R in such a way that $||A_i| - |B_i|| \leqslant \lambda n$ for each $1 \leqslant i \leqslant k$. We define $\{A_i^*, B_i^*\}_{i=1}^k$, R^* and M^* as follows: Initially we set $A_i^* := A_i$ and $B_i^* := B_i$ for all $1 \leqslant i \leqslant k$, $R^* := R$ and $M^* := M$. At each step we will redefine each A_i^* and B_i^*, R^* and M^* and reassign vertices so that the quantity

$$\Sigma^* = \sum_{1 \leqslant i \leqslant k, ||A_i^*| - |B_i^*|| > \lambda n} ||A_i^*| - |B_i^*||$$

decreases by at least λn. The algorithm will terminate when $\Sigma^* = 0$, i.e., when $||A_i^*| - |B_i^*|| \leqslant \lambda n$ for all $1 \leqslant i \leqslant k$. Initially $\Sigma^* \leqslant 8\varepsilon' m' k/\eta \leqslant 4\varepsilon' n/\eta$ by (c), and hence we need at most $4\varepsilon'/\eta\lambda$ steps to complete the process. R^* will always be an *induced* subgraph of R and at each step we set M^* to be the submatching of M induced by $V(R^*)$. (Note that $V(R^*)$ is a subset of $V(R) = \{A_i, B_i\}_{i=1}^k$ throughout the algorithm.)

We will ensure that the inequality

$$|R^*| \geqslant (1 - \nu/12)\, k' \tag{3.6.2}$$

holds throughout, and that M^* is a perfect matching in R^*. Further we will ensure that

$$|A_i^* \backslash A_i| \leqslant (\varepsilon')^{1/3} m' \text{ and } |B_i^* \backslash B_i| \leqslant (\varepsilon')^{1/3} m'; \tag{3.6.3}$$

$$|A_i \backslash A_i^*| \leqslant (\varepsilon')^{1/3} m' \text{ and } |B_i \backslash B_i^*| \leqslant (\varepsilon')^{1/3} m' \tag{3.6.4}$$

for all $1 \leqslant i \leqslant k$.

Each step proceeds as follows: Call a vertex v *well-connected* to a cluster $V \in V(R)$ if v has at least $d'm'/8$ neighbours in V. Recall that if $VW \in E(R)$ then $(V, W)_{G'}$ is a $((\varepsilon')^{1/3}, d'/4)$-regular pair and so V contains at least $m'/2$ vertices v which are well-connected to W. In what follows we will ensure that every vertex we redistribute to a cluster A_i^* is well-connected to B_i and vice versa. Since (3.6.4) holds throughout the process, given any $VW \in E(R^*)$, V^* will always contain at least $m'/2 - (\varepsilon')^{1/3}m' \geqslant m'/3 \gg \lambda n$ vertices that are well-connected to W (where $V^* := A_i^*$ if $V = A_i$ for some i and $V^* := B_i^*$ if $V = B_i$ for some i). Thus, at any point during the algorithm we may choose a set of λn well-connected vertices from any of the A_i^* and B_i^*. (When it is clear from the context, we will not explicitly specify which cluster a vertex v is well-connected to.)

Let S be the set of clusters $V^* \in V(R^*)$ such that either $V^* = A_i$ where $|A_i^*| > |B_i^*| + \lambda n$ or $V^* = B_i$ where $|B_i^*| > |A_i^*| + \lambda n$. If S is empty then the algorithm terminates. (We shall see later that in this case we must have that $\Sigma^* = 0$.) Otherwise, choose $V^* \in S$ arbitrarily. Since R is a robust $(\nu/3, 3\tau)$-expander and $\delta(R) \geqslant \eta k'/3$, (3.6.2) implies that $\delta(R^*) \geqslant \eta |R^*|/4$ and Lemma 3.4.4 implies that R^* is a robust $(\nu/4, 4\tau)$-expander. Hence Lemma 3.4.3 implies that R^* contains a shifted M^*-walk P' of length at most $12/\nu$ which starts and finishes at V^*. By Lemma 3.4.1, P' contains a simple shifted M^*-walk P'' which also starts and finishes at V^*. Now apply Lemma 3.4.2 to P'' to obtain a simple shifted M^*-walk P of length at most $12/\nu$, such that the endpoints of P both lie in S and no other vertices of P lie in S. We call P the *active walk* of this step of the algorithm.

Let $P = U_1 W_2 U_2 \dots W_{\ell-1} U_{\ell-1} W_\ell$ such that $W_i U_i \in E(M^*)$ for each $2 \leqslant i \leqslant \ell - 1$. Let W_1 and U_ℓ denote the clusters such that $W_1 U_1, W_\ell U_\ell \in E(M^*)$. Given any $1 \leqslant i \leqslant \ell$, if $U_i = A_j$ for some $1 \leqslant j \leqslant k$, set $U_i^* := A_j^*$; otherwise $U_i = B_j$ for some $1 \leqslant j \leqslant k$, so set

$U_i^* := B_j^*$. Define W_i^* analogously for each $1 \leqslant i \leqslant \ell$. Move $\lambda n/2$ well-connected vertices from U_1^* into U_2^*, $\lambda n/2$ well-connected vertices from U_2^* into U_3^*, and so on until we have moved $\lambda n/2$ well-connected vertices from $U_{\ell-1}^*$ to U_ℓ^*. Then move $\lambda n/2$ well-connected vertices from W_2^* into W_1^*, $\lambda n/2$ well-connected vertices from W_3^* into W_2^*, and so on until we have moved $\lambda n/2$ well-connected vertices from W_ℓ^* to $W_{\ell-1}^*$. Note that since P is simple, each cluster loses at most λn vertices and gains at most λn vertices. Further for each $1 < i < \ell$ the quantity $||W_i^*| - |U_i^*||$ remains unchanged (in fact, $|W_i^*|$ and $|U_i^*|$ remain unchanged). For $i = 1, \ell$, $||W_i^*| - |U_i^*||$ decreases by precisely λn (or $2\lambda n$ if $U_1 = W_\ell$).

In order to ensure that (3.6.3) holds, we remove from R^* every pair $\{A_i, B_i\}$ of clusters such that $|A_i^* \backslash A_i| \geqslant (\varepsilon')^{1/3}m' - \lambda n$ or $|B_i^* \backslash B_i| \geqslant (\varepsilon')^{1/3}m' - \lambda n$. Since each cluster gains at most λn vertices in each step, any clusters which are not removed will still satisfy (3.6.3) at the end of the next step. Similarly, to ensure that (3.6.4) holds, we remove from R^* every pair $\{A_i, B_i\}$ of clusters such that $|A_i \backslash A_i^*| \geqslant (\varepsilon')^{1/3}m' - \lambda n$ or $|B_i \backslash B_i^*| \geqslant (\varepsilon')^{1/3}m' - \lambda n$.

Claim 3.6.5 *For each pair $\{A_i, B_i\}$ of clusters which is removed from R^* we have that $||A_i^*| - |B_i^*|| \leqslant \lambda n$.*

Proof. To prove the claim, suppose for a contradiction that some pair $\{A_i, B_i\}$ of clusters is removed from R^* and that $||A_i^*| - |B_i^*|| > \lambda n$. In order for $\{A_i, B_i\}$ to be removed we must have that $|A_i^* \backslash A_i| \geqslant (\varepsilon')^{1/3}m' - \lambda n$, $|B_i^* \backslash B_i| \geqslant (\varepsilon')^{1/3}m' - \lambda n$, $|A_i \backslash A_i^*| \geqslant (\varepsilon')^{1/3}m' - \lambda n$ or $|B_i \backslash B_i^*| \geqslant (\varepsilon')^{1/3}m' - \lambda n$. Without loss of generality assume that $|A_i^* \backslash A_i| \geqslant (\varepsilon')^{1/3}m' - \lambda n$. Since in each step we add at most λn vertices to A_i^*, there must have been at least $(\varepsilon')^{1/3}m'/2\lambda n$ steps in the algorithm so far, such that A_i is contained in the active walk P of each step. By the definition of P, either A_i or B_i must be an endpoint of P. So $||A_i^*| - |B_i^*||$ is reduced by at least λn during each such step, and hence by at least $(\varepsilon')^{1/3}m'/2$ during the algorithm so far. But this is a contradiction since initially $||A_i^*| - |B_i^*|| \leqslant \sqrt{\varepsilon'}m' \leqslant (\varepsilon')^{1/3}m'/2$. □

It remains to show that (3.6.2) holds throughout the process. Suppose for a contradiction that at some point more than $\nu k'/12$ clusters have been removed from R^*. Then at least

$$((\varepsilon')^{1/3}m' - \lambda n)\left(\frac{\nu k'}{48}\right) > \sqrt{\varepsilon'}n \tag{3.6.6}$$

vertices of G must have been redistributed during the process so far. But at most $12\lambda n/\nu$ vertices were redistributed during each step, and at most $4\varepsilon'/\eta\lambda$ steps were performed during the process. Hence the number of redistributed vertices is at most

$$\frac{12\lambda n}{\nu} \cdot \frac{4\varepsilon'}{\eta\lambda} < \sqrt{\varepsilon'}n,$$

which contradicts (3.6.6). This proves that (3.6.2) holds throughout.

By construction, when the algorithm terminates we have that $V(R^*)$ does not contain any A_i and B_i such that $||A_i^*| - |B_i^*|| > \lambda n$. Further, by Claim 3.6.5, for those clusters $A_i, B_i \notin V(R^*)$ we have that $||A_i^*| - |B_i^*|| \leqslant \lambda n$. Hence, we indeed obtain clusters $\{A_i^*, B_i^*\}_{i=1}^k$ such that $\Sigma^* = 0$ and (3.6.3) and (3.6.4) hold.

We now set $n_{2i-1} := |A_i^*|$ and $n_{2i} := |B_i^*|$ for each $1 \leqslant i \leqslant k$. Notice that $n_j \geqslant (1-(\varepsilon')^{1/3})m' > n/3k$ for each $1 \leqslant j \leqslant 2k$. We now relabel the clusters of R in the natural way so that $V(R) = \{A_i^*, B_i^*\}_{i=1}^k$. Note that by (3.6.3) and (3.6.4) we have

$$|A_i \Delta A_i^*| \leqslant 2(\varepsilon')^{1/3}m' \text{ and } |B_i \Delta B_i^*| \leqslant 2(\varepsilon')^{1/3}m'$$

for each $1 \leqslant i \leqslant k$. Hence by Proposition 3.3.2 the pair $(V, W)_{G'}$ is $((\varepsilon')^{1/10}, d'/10)$-regular for every edge $VW \in E(R)$. Further, the pair $(A_i^*, B_i^*)_{G'}$ is $((\varepsilon')^{1/10}, d'/10)$-super-regular for every $1 \leqslant i \leqslant k$. Indeed, we ensured that every vertex v which was redistributed to A_i^* had at least $d'm'/8$ neighbours in B_i. Since $|B_i^* \Delta B_i| \leq 2(\varepsilon')^{1/3}m'$, v has at least $d'|B_i^*|/10$ neighbours in B_i^*. Similarly every vertex $w \in B_i^*$ has at least $d'|A_i^*|/10$ neighbours in A_i^*.

Now suppose we are given a partition $(n'_i)_{i=1}^{2k}$ of n such that $n'_i \leqslant n_i+\xi n$ for each $1 \leqslant i \leqslant 2k$. Set $a_i := n'_{2i-1}-n_{2i-1}$ and $b_i := n'_{2i}-n_{2i}$ for all $1 \leqslant i \leqslant k$. Notice that $|a_i|, |b_i| \leqslant 2k\xi n$ for each $1 \leqslant i \leqslant k$ and $|\sum_{i=1}^k a_i| = |\sum_{i=1}^k b_i| \leqslant 2k\xi n$. Recall that R contains the edges $A^*_{i_1}A^*_{j_1}$ and $B^*_{i_2}B^*_{j_2}$. Thus, we can apply the Mobility lemma (Lemma 3.5.1) with parameters $k, 2k\xi, (\varepsilon')^{1/10}, \varepsilon, d$ and $d'/10$ to obtain a partition $A'_1, B'_1, \ldots, A'_k, B'_k$ of $V(G)$ which satisfies conditions (α_1)–(α_5). $\square$

3.7 The Lemma for H

Lemma 3.7.1 (Lemma for H) *For any $\Delta, k \in \mathbb{N}$ and $\xi > 0$, there exist $\beta > 0$ and $n_0 \in \mathbb{N}$ such that the following holds: Let H be a bipartite graph on $n \geqslant n_0$ vertices with bandwidth at most βn and such that $\Delta(H) \leqslant \Delta$. Let $n_1, n_2, \ldots, n_{2k}$ be an integer partition of n such that $n_i > n/(3k)$ for all $1 \leqslant i \leqslant 2k$ and $|n_{2i-1} - n_{2i}|/n \ll \xi$ for all $1 \leqslant i \leqslant k$. Suppose C is the cycle $12\ldots(2k)1$ on $[2k]$, and let $c = \{2i_1, 2i_2\}$ be a chord of C (for some distinct $1 \leqslant i_1, i_2 \leqslant k$). Then there exists a set $S \subseteq V(H)$ and a graph homomorphism $f : H \to C \cup \{c\}$, such that*

(β_1) $|S| \leqslant \xi n$;

(β_2) $|f^{-1}(i)| \leqslant n_i + \xi n$ for all $1 \leqslant i \leqslant 2k$;

(β_3) Every edge which is not in $H[S]$ is mapped to an edge $\{2i-1, 2i\}$, for some $1 \leqslant i \leqslant k$.

Proof. Choose $\beta > 0$ and integers n_0, m_1, m_2 and k_1 such that

$$1/n_0 \ll \beta \ll 1/m_1 \ll 1/m_2 \ll 1/k_1 \ll 1/\Delta, 1/k, \xi.$$

Further, we may assume that m_2 divides m_1. We begin by defining a new cycle C' with chord c' which will act as an intermediate stage between C and H, i.e., we will construct homomorphisms $f_1 : C' \cup \{c'\} \to C \cup \{c\}$ and $f_2 : H \to C' \cup \{c'\}$ such that $f = f_1 \circ f_2$ is our desired homomorphism. The homomorphism f_2 will be constructed to map roughly the same number of vertices to each vertex in C'. Notice however, that our desired homorphism f may not map vertices in an 'equal' way (since, in general, the n_i may be far from equal). Thus, the role of f_1 is to ensure f maps the 'correct' proportion of vertices to each vertex in C.

Let C' be the cycle $12\dots(2k')1$ on $[2k']$, where $k' := \sum_{i\in[k]}\lceil(n_{2i-1}+n_{2i})k_1/n\rceil$. Note that $k_1 \leqslant k' \leqslant k_1+k$. We define f_1 as follows: For each $1 \leqslant j \leqslant k'$, let $g(j) \in \mathbb{N}$ be such that $\sum_{i=1}^{g(j)-1}\lceil(n_{2i-1}+n_{2i})k_1/n\rceil < j \leqslant \sum_{i=1}^{g(j)}\lceil(n_{2i-1}+n_{2i})k_1/n\rceil$. Then set $f_1(2j-1) = 2g(j)-1$ and $f_1(2j) = 2g(j)$ for each $1 \leqslant j \leqslant k'$.

Recall that $c = \{2i_1, 2i_2\}$ is a chord of C. Suppose that i_1', i_2' are such that $f_1(2i_1') = 2i_1$ and $f_1(2i_2') = 2i_2$. Notice that as $i_1 \neq i_2$, we have that $i_1' \neq i_2'$. Thus, set $c' := \{2i_1', 2i_2'\}$ to be the chord of C'.

By construction $f_1(c') = c$. Given any edge $c_1 = \{2j-1, 2j\}$ on C', we have that $f_1(c_1) = \{2g(j)-1, 2g(j)\}$. Further, consider any edge $c_2 = \{2j, 2j+1\} = \{2j, 2(j+1)-1\}$ on C'. Then $f_1(2j) = 2g(j)$ and $f_1(2(j+1)-1) = 2g(j+1)-1$. By definition of f_1, either $g(j+1) = g(j)$ or $g(j+1) = g(j)+1$. But both $\{2g(j), 2g(j)-1\}$ and $\{2g(j), 2g(j)+1\}$ are edges of C. So in either case f_1 maps c_2 to an edge of C. Therefore, indeed f_1 is a graph homomorphism.

Roughly speaking, we will construct f_2 as follows: Initially we split H up into *small segments* $A_1, B_1, \dots, A_{m_1}, B_{m_1}$ in such a way that almost all of the edges of H lie in the pairs $(A_i, B_i)_{i=1}^{m_1}$ and the remainder lie in the pairs $(B_i, A_{i+1})_{i=1}^{m_1}$. Our ideal strategy would be to map all of the vertices of A_1 onto vertex 1 of C', the vertices of B_1 onto vertex 2, the vertices of A_2 onto vertex 3, etc. This ensures that f_2 is a homomorphism and that almost all of the edges of H are mapped onto an edge of the form $\{2i-1, 2i\}$ for some i. However the number of vertices mapped onto each vertex of C' may vary widely. To solve this problem we introduce 'drunken' segments in which the assignment of the vertices is random, and use a probabilistic argument to show that with positive probability each vertex of C' receives approximately the same number of vertices of H. We also use the chord c' to 'turn around' at some point during the process, in order to eliminate the possible inequality between the number of vertices of H assigned to odd and even vertices of C'.

Chopping H up into segments. Since H has bandwidth at most βn, there exists an ordering $x_1, x_2, \ldots, x_n$ of $V(H)$ such that for every edge $x_i x_j$ of H, $|i-j| \leqslant \beta n$. Let (A, B) be a bipartition of $V(H)$. We define $\{A_i, B_i\}_{i=1}^{m_1}$ as follows: for each vertex $x_s \in A$ there exists $1 \leqslant i \leqslant m_1$ such that $(i-1)n/m_1 - \beta n < s \leqslant in/m_1 - \beta n$ (unless $s > n - \beta n$). We assign x_s to A_i (or to A_{m_1} if $s > n - \beta n$). Similarly for each vertex $x_t \in B$ there exists $1 \leqslant j \leqslant m_1$ such that $(j-1)n/m_1 < t \leqslant jn/m_1$, and we assign x_t to B_j. Let S be the set of vertices x_s such that $in/m_1 - 2\beta n < s \leqslant in/m_1 + \beta n$ for some $1 \leqslant i \leqslant m_1$. Note that the following properties hold:

(a) $n/m_1 - \beta n \leqslant |A_i| + |B_i| \leqslant n/m_1 + \beta n$ for each $1 \leqslant i \leqslant m_1$;

(b) $|S| \leqslant 3m_1 \beta n \ll \xi n$;

(c) Every edge of H which is not in $H[S]$ lies in one of the pairs (A_i, B_i) for some $1 \leqslant i \leqslant m_1$;

(d) Every edge of $H[S]$ lies in one of the pairs (A_i, B_i) or one of the pairs (B_i, A_{i+1}) for some $1 \leqslant i \leqslant m_1$ (where $A_{m_1+1} := A_1$).

Properties (c) and (d) follow from the fact that H has bandwidth at most βn and that $n/m_1 \gg \beta n$. We now modify the small segments so that properties (a)–(d) are still satisfied and so that every small segment has size at least $n/(4\Delta m_1)$. Suppose a small segment A_i has size smaller than $n/(4\Delta m_1)$. Note that $|N_H(A_i)| \leqslant n/(4m_1)$ and so

$$|B_i \backslash (S \cup N_H(A_i))| \overset{(a)}{\geqslant} (n/m_1 - \beta n - n/(4\Delta m_1)) - 3\beta n - n/(4m_1) \geqslant n/(4m_1).$$

But (c) implies that any vertex in $B_i \backslash (S \cup N_H(A_i))$ must be isolated in H and so may be reassigned to A_i without affecting properties (a)–(d). Hence we may reassign sufficiently many vertices so that $|A_i|, |B_i| \geqslant n/(4\Delta m_1)$. For any segment B_i which has size smaller than $n/(4\Delta m_1)$ we proceed in an identical way. From now on we denote by A the union of small segments $\bigcup_{i=1}^{m_1} A_i$ and by B the union $\bigcup_{i=1}^{m_1} B_i$.

We now group the small segments together to form *large segments* $\{L_j\}_{j=1}^{m_2}$, which are defined as

$$L_j := \bigcup_{\frac{(j-1)m_1}{m_2} < t \leqslant \frac{jm_1}{m_2}} (A_t \cup B_t) \,.$$

Note that since $\beta \ll 1/m_1$, (a) implies that

$$\frac{n}{m_2} - \sqrt{\beta} n \leqslant |L_j| \leqslant \frac{n}{m_2} + \sqrt{\beta} n \tag{3.7.2}$$

for each $1 \leqslant j \leqslant m_2$. In order to eliminate any inequality between the number of vertices of H assigned to odd and even vertices of C' we need to partition $\{L_j\}_{j=1}^{m_2}$ into two parts. We will assign the vertices in each part separately. For each $1 \leqslant j \leqslant m_2$, set $s_j := |L_j \cap A| - |L_j \cap B|$. Note that $|s_j| \leqslant n/m_2 + \sqrt{\beta} n - n/(4\Delta m_2) \leqslant n/m_2$ for each $1 \leqslant j \leqslant m_2$, and that $\sum_{j=1}^{m_2} s_j = |A| - |B|$. Suppose without loss of generality that $|A| - |B| \geqslant 0$. Then since $\beta, 1/m_2 \ll \xi$ there exists an integer m_3 so that $\xi m_2/20 \leqslant m_3 \leqslant (1 - \xi/20)m_2$ and

$$\frac{(|A| - |B|)}{2} - \frac{\xi n}{20} \leqslant \sum_{i=1}^{m_3} s_j \leqslant \frac{(|A| - |B|)}{2} + \frac{\xi n}{20}. \tag{3.7.3}$$

We will embed separately the large segments $\{L_j\}_{j=1}^{m_3}$ and the segments $\{L_j\}_{j=m_3+1}^{m_2}$.

For each $1 \leqslant j \leqslant m_3$, let the *drunken segment* D_j be the union of the last $k_2 := \xi m_1/(6k' m_2)$ pairs of small segments in L_j and let the *sober segment* S_j be the union of the rest of the small segments in L_j.

Defining our algorithms. We now define three different algorithms for assigning the vertices of a segment to vertices of C', given an *initial vertex* $2i'_0 - 1 \in C'$. In each case we work mod $2k'$ when dealing with vertices of C'. The *sober algorithm* is a deterministic process which proceeds as follows: Let S_j be a sober segment whose first small segments are A_{i_0} and B_{i_0}. For every i such that A_i and B_i are small segments of S_j, assign every vertex of A_i to the vertex $2i'-1$ of C' and every vertex of B_i to the vertex $2i'$ of C' where $i' \equiv i'_0 + i - i_0 \mod k'$. So whenever B_i is assigned to $2i'$, A_{i+1} is assigned to $2i'+1$ $(\mathrm{mod}\ 2k')$. We call the vertex $2i^*$ of C' to which the vertices of the last small segment of S_j are assigned the *final vertex* of the algorithm and define this term in a similar way for the remaining two algorithms.

The *drunken algorithm* is a randomised algorithm which proceeds as follows: Given a drunken segment D_j whose first small segment is A_{i_0}, assign every vertex of A_{i_0} to the vertex $2i'_0 - 1$ of C' and every vertex of B_{i_0} to the vertex $2i'_0$ of C'. Then for every pair A_{i+1}, B_{i+1} of small segments in D_j, let $2i'$ be the vertex to which the vertices of B_i were assigned and let

$$i'' = \begin{cases} i' & \text{with probability } \frac{1}{2}; \\ i'+1 & \text{with probability } \frac{1}{2}. \end{cases}$$

(All random choices are made independently.) Assign every vertex of A_{i+1} to $2i''-1$ and every vertex of B_{i+1} to $2i''$.

Claim 3.7.4 *Suppose that the vertices of D_j are assigned using the drunken algorithm with initial vertex $2i'_0 - 1$, and let $i'_1 \in [k']$ be arbitrary. Let the random variable I be the final vertex of the drunken algorithm. Then*

$$\mathbb{P}[I = 2i'_1 \mid i'_0] \leqslant \frac{1 + \xi/20}{k'}.$$

Proof. To prove the claim, note that $I \sim 2(i_0' + Bin(k_2, 1/2))$, that $k_2 \gg (k')^3/6$ and that $1/k' \ll \xi/20$. So Lemma 1.8.5 with $\varepsilon := \xi/20$ implies that $\mathbb{P}[I - 2i_0' = 2i_1' - 2i_0'] \leqslant (1 + \xi/20)/k'$, and Claim 3.7.4 follows immediately. $\square$

The $2i_1'$-*seeking algorithm* is a deterministic algorithm which proceeds as follows: Given a drunken segment D_j whose first small segment is A_{i_0}, assign every vertex of A_{i_0} to the vertex $2i_0' - 1$ of C' and every vertex of B_{i_0} to the vertex $2i_0'$ of C'. Then for every pair A_{i+1}, B_{i+1} of small segments in D_j, let $2i'$ be the vertex of C' to which the vertices of B_i were assigned and let

$$i'' = \begin{cases} i' & \text{if } i' = i_1; \\ i' + 1 & \text{otherwise.} \end{cases}$$

Assign every vertex of A_{i+1} to $2i'' - 1$ and every vertex of B_{i+1} to $2i''$. Note that the final vertex of this algorithm is always $2i_1'$, since $k' \leqslant \xi m_1/(6k'm_2)$.

Applying the algorithms. We use these algorithms to assign small segments to vertices of C' as follows: Choose $i_0' \in [k']$ randomly and let $2i_0' - 1$ be the initial vertex for S_1. For each $1 \leqslant j \leqslant m_3 - 1$, use the sober algorithm to assign the segments of S_j and then use the drunken algorithm to assign the vertices of D_j, where in each case the initial vertex of each segment is the successor of the final vertex of the previous segment. (So for example, if the final vertex of the drunken algorithm, when applied to D_j, is $2i^*$, then the initial vertex of the sober algorithm, when applied to S_{j+1}, is $2i^* + 1$.) Then assign the vertices of S_{m_3} using the sober algorithm and assign the vertices of D_{m_3} using the $2i_1'$-seeking algorithm. (Recall that $2i_1'$ was a vertex of the chord c'.) We explain how we assign the small segments from $\bigcup_{j=m_3+1}^{m_2} L_j$ later.

Claim 3.7.5 *For each $1 \leqslant i \leqslant k'$, let X_i be the number of vertices of $\bigcup_{j=1}^{m_3} S_j$ assigned to* *the vertex* $2i-1$ *of* C'. *Then*

$$\mathbb{P}\left[X_i > \frac{1}{2k'}\left(\frac{m_3 n}{m_2} + \frac{|A|-|B|}{2}\right) + \frac{\xi n}{6k'}\right] \leqslant \frac{1}{3k'}.$$

Proof. For each $1 \leqslant j \leqslant m_3$, let $Y_j = |S_j \cap A|$ and let $X_{i,j}$ be the number of vertices of S_j which are assigned to $2i-1$. To prove the claim, we first use Claim 3.7.4 to bound $\mathbb{E}[X_{i,j} \mid X_{i,j-1}, \dots, X_{i,1}]$. Let r_j be the initial vertex of S_j for each j. Let $B_{x_{i,1},\dots,x_{i,j-1}}$ be the event $X_{i,j-1} = x_{i,j-1}, \dots, X_{i,1} = x_{i,1}$ for some $x_{i,1}, \dots x_{i,j-1}$. Now for $1 \leqslant i' \leqslant k'$ and any integer x we have $\mathbb{P}[X_{i,j} = x \mid B_{x_{i,1},\dots,x_{i,j-1}} \cap (r_{j-1} = 2i'-1)] = \mathbb{P}[X_{i,j} = x \mid r_{j-1} = 2i'-1]$. Hence Lemma 1.8.6 implies that $\mathbb{E}[X_{i,j} \mid X_{i,j-1}, \dots, X_{i,1}] \leqslant \max_{i'=1}^{k'} \mathbb{E}[X_{i,j} \mid r_{j-1} = 2i'-1]$. But Claim 3.7.4 implies that

$$\begin{aligned}\mathbb{E}[X_{i,j} \mid r_{j-1} = 2i'-1] &= \sum_{i''=1}^{k'} \mathbb{E}[X_{i,j} \mid r_j = 2i''-1]\mathbb{P}[r_j = 2i''-1 \mid r_{j-1} = 2i'-1] \\ &\leqslant \frac{1+\xi/20}{k'} \sum_{i''=1}^{k'} \mathbb{E}[X_{i,j} \mid r_j = 2i''-1] = \frac{(1+\xi/20)Y_j}{k'}.\end{aligned}$$

Hence $\mathbb{E}[X_{i,j} \mid X_{i,j-1}, \dots, X_{i,1}] \leqslant (1+\xi/20)Y_j/k'$. Set $X'_{i,j} := X_{i,j}m_2/n$. Since

$$|S_j| \overset{(3.7.2)}{\leqslant} \left(\frac{n}{m_2} + \sqrt{\beta}n\right) - \left(\frac{\xi m_1}{6k'm_2}\right)\left(\frac{n}{4\Delta m_1}\right) \leqslant \frac{n}{m_2},$$

we have that $X'_{i,j} \in [0,1]$, for each $1 \leqslant j \leqslant m_3$. Let

$$\mu = \sum_{j=1}^{m_3} \frac{(1+\xi/20)Y_j m_2}{k'n}, \tag{3.7.6}$$

and note that

$$\sum_{j=1}^{m_3} Y_j \leqslant \sum_{j=1}^{m_3} |A \cap L_j| = \frac{1}{2}\left(\sum_{j=1}^{m_3} |L_j| + s_j\right)$$
$$\overset{(3.7.2),(3.7.3)}{\leqslant} \left(\frac{m_3 n}{2m_2} + \frac{m_3\sqrt{\beta}n}{2}\right) + \left(\frac{|A|-|B|}{4} + \frac{\xi n}{40}\right) \leqslant \frac{m_3 n}{2m_2} + \frac{|A|-|B|}{4} + \frac{\xi n}{20}. \tag{3.7.7}$$

Note also that $Y_j \geqslant (n/(4\Delta m_1)) \times (m_1/2m_2) = n/(8\Delta m_2)$ for each $1 \leqslant j \leqslant m_1$. Thus we have $\mu \geqslant m_3/(8\Delta k') \gg (\log k')/\xi^2$. We now apply Lemma 1.8.4 with $\delta := \xi/20$ to obtain

$$\mathbb{P}\left[\sum_{j=1}^{m_3} X'_{i,j} > (1+\xi/20)\mu\right] \leqslant e^{-\frac{\xi^2\mu}{1200}} \leqslant \frac{1}{3k'}.$$

It follows that with probability at least $1 - 1/3k'$,

$$X_i \overset{(3.7.6)}{\leqslant} \frac{(1+\xi/20)^2}{k'} \sum_{i=1}^{m_3} Y_j \overset{(3.7.7)}{\leqslant} \frac{(1+\xi/20)^2}{k'}\left(\frac{m_3 n}{2m_2} + \frac{|A|-|B|}{4} + \frac{\xi n}{20}\right)$$
$$\leqslant \frac{1}{2k'}\left(\frac{m_3 n}{m_2} + \frac{|A|-|B|}{2}\right) + \frac{\xi n}{6k'},$$

which proves Claim 3.7.5. □

By a similar argument we have that if X'_i is the number of vertices of $\bigcup_{j=1}^{m_3} S_j$ assigned to $2i$, then

$$\mathbb{P}\left[X'_i > \frac{1}{2k'}\left(\frac{m_3 n}{m_2} + \frac{|B|-|A|}{2}\right) + \frac{\xi n}{6k'}\right] \leqslant \frac{1}{3k'}$$

for every $1 \leqslant i \leqslant k'$. Taken together with Claim 3.7.5 this implies that with probability at least 1/3,

$$X_i \leqslant \frac{1}{2k'}\left(\frac{m_3 n}{m_2} + \frac{|A| - |B|}{2}\right) + \frac{\xi n}{6k'} \qquad (3.7.8)$$
$$\text{and } X'_i \leqslant \frac{1}{2k'}\left(\frac{m_3 n}{m_2} + \frac{|B| - |A|}{2}\right) + \frac{\xi n}{6k'}$$

for every $1 \leqslant i \leqslant k'$, and hence there exists an assignment such that (3.7.8) holds.

For each $m_3 < j \leqslant m_2$, let D_j be the union of the *first* k_2 small segments of L_j and S_j the union of the remaining small segments. We now assign the vertices of $\bigcup_{j=m_3+1}^{m_2} L_j$ using an algorithm similar to that for $\bigcup_{j=1}^{m_3} L_j$, but in reverse order. That is, we first choose $1 \leqslant i''_0 \leqslant k'$ randomly and assign the vertices of S_{m_2} using the sober algorithm, but with the roles of A_i and B_i exchanged for each i. Thus we assign the vertices of B_{m_1} to $2i''_0 - 1$, the vertices of A_{m_1} to $2i''_0$, etc. Similarly we use the drunken algorithm to assign the vertices of D_{m_2} (again with the roles of A_i and B_i exchanged for each i), and so on until we have assigned all the vertices up to S_{m_3+1}. (As before, the initial vertex of any application of an algorithm is the successor of the final vertex of the previous application of an algorithm.) Finally we use the $2i'_2$-seeking algorithm to assign the vertices of the last drunken segment D_{m_3+1}. (Recall that the final vertex of the $2i'_2$-seeking algorithm is always $2i'_2$.) Let $\overline{X_i}$ be the number of vertices of $\bigcup_{j=m_3+1}^{m_2} S_j$ assigned to to $2i-1$ and $\overline{X'_i}$ the number assigned to $2i$. By using a proof analogous to that of Claim 3.7.5, we can ensure that

$$\overline{X_i} \leqslant \frac{1}{2k'}\left(\frac{(m_2 - m_3)n}{m_2} + \frac{|B| - |A|}{2}\right) + \frac{\xi n}{6k'} \qquad (3.7.9)$$
$$\text{and } \overline{X'_i} \leqslant \frac{1}{2k'}\left(\frac{(m_2 - m_3)n}{m_2} + \frac{|A| - |B|}{2}\right) + \frac{\xi n}{6k'}$$

for each $1 \leqslant i \leqslant k'$.

Note that

$$|\bigcup_{j=1}^{m_2} D_j| \leqslant m_2 \times \frac{\xi m_1}{6k'm_2} \times (n/m_1 + \beta n) \leqslant \xi n/3k',$$

and hence in total we assign at most

$$\begin{aligned}
&X_i + X_i' + |\bigcup_{j=1}^{m_2} D_j| \\
\overset{(3.7.8),(3.7.9)}{\leqslant} &\frac{1}{2k'}\left(\frac{m_3 n}{m_2} + \frac{|A|-|B|}{2}\right) + \frac{1}{2k'}\left(\frac{(m_2-m_3)n}{m_2} + \frac{|B|-|A|}{2}\right) + \frac{2\xi n}{3k'} \\
= &\frac{1}{2k'}(1+4\xi/3)n
\end{aligned}$$

vertices to $2i-1$ and at most $(1+4\xi/3)n/2k'$ vertices to $2i$, for each $1 \leqslant i \leqslant k'$.

Verifying that f has the desired properties. This completes our definition of f_2. We now check that f_2 is a homomorphism. By properties (c) and (d) it suffices to show for each i that whenever A_i and B_i (or B_i and A_{i+1}) are assigned to vertices $1 \leqslant i', j' \leqslant 2k'$ of C', then $i'j'$ is an edge of $C' \cup \{c'\}$. Observe first that the sober, drunken and seeking algorithms all assign vertices in such a way that $i'j'$ is an edge of C'. Further, recall that the initial vertex of any application of an algorithm is the successor of the final vertex of the previous application of an algorithm. So if, for example, the vertices of B_i are assigned to the final vertex $2i^*$ where B_i is the final segment assigned in an application one of the algorithms, then the vertices of A_{i+1} will be assigned to the initial vertex $2i^*+1$ in the next application of an algorithm.

The only pair this argument does not deal with is the pair (B_j, A_{j+1}), where B_j is the last small segment of D_{m_3} (and hence A_{j+1} is the first small segment of D_{m_3+1}, and therefore the last to be assigned). Now by the definition of the seeking algorithm, the vertices of B_j are assigned to the vertex $2i_1'$ of C' and the vertices of A_{j+1} are assigned to the vertex $2i_2'$ of C'. Recalling that $c' = \{2i_1', 2i_2'\}$ we have that f_2 is indeed a homomorphism.

Now consider $f = f_1 \circ f_2$. Since f_1 and f_2 are both homomorphisms we have that f is a homomorphism. Property (b) implies that condition (β_1) holds. By (c), every edge xy not in $H[S]$ lies in a pair (A_i, B_i) for some i. Thus, by definition of our three algorithms, xy is mapped to an edge $\{2j-1, 2j\}$ by f_2 for some j. By definition of f_1, $\{2j-1, 2j\}$ is mapped to $\{2j'-1, 2j'\}$ by f_1 for some j'. Therefore, f satisfies (β_3).

To see that condition (β_2) also holds, recall that f_1 assigns to each vertex $2i-1$ of C (and also to $2i$) exactly $\lceil (n_{2i-1} + n_{2i})k_1/n \rceil$ vertices of C'. Hence f assigns at most

$$\begin{aligned} \frac{1}{2k'}(1+4\xi/3)n\lceil (n_{2i-1}+n_{2i})k_1/n\rceil &\leqslant (1+4\xi/3)(n_{2i-1}+\xi n/10) + \frac{1}{2k'}(1+4\xi/3)n \\ &\leqslant n_{2i-1} + \xi n \end{aligned}$$

vertices of H to the vertex $2i-1$ of C, for each $1 \leqslant i \leqslant k$. Similarly f assigns at most $n_{2i} + \xi n$ vertices of H to the vertex $2i$. □

3.8 Completing the proof

In this section we use Lemmas 3.6.1 and 3.7.1 to prove Theorem 1.4.8. We use the following definition and lemma from [18]; these allow us to prove that H embeds into G by checking some relatively simple conditions.

Definition 3.8.1 *Let H be a graph on n vertices, let R be a graph on $[k]$, and let $R' \subseteq R$. We say that a vertex partition $V(H) = (W_i)_{i\in[k]}$ of H is* ε-compatible *with an integer partition $(n_i)_{i\in[k]}$ of n and $R' \subseteq R$ if the following holds. For $i \in [k]$ let S_i be the set of vertices in W_i with neighbours in some W_j with $ij \notin E(R')$ and $i \neq j$. Set $S := \bigcup_{i\in[k]} S_i$ and $T_i := N_H(S) \cap (W_i \backslash S)$. Then for all $i, j \in [k]$ we have that*

(γ_1) $|W_i| = n_i$;

(γ_2) *$xy \in E(H)$ for $x \in W_i$ and $y \in W_j$ implies that $ij \in E(R)$;*

(γ_3) $|S_i| \leqslant \varepsilon n_i$ *and* $|T_i| \leqslant \varepsilon \cdot \min\{n_j \mid i$ *and* j *are in the same component of* $R'\}$.

The partition $V(H) = (W_i)_{i\in[k]}$ *is* ε-compatible *with a partition* $V(G) = (V_i)_{i\in[k]}$ *of a graph* G *and* $R' \subseteq R$ *if* $V(H) = (W_i)_{i\in[k]}$ *is* ε-compatible *with* $(|V_i|)_{i\in[k]}$ *and* $R' \subseteq R$.

Lemma 3.8.2 ([18], Lemma 3.12) *For all* $d, \Delta, r > 0$ *there is a constant* $\varepsilon = \varepsilon(d, \Delta, r)$ *such that the following holds. Let* G *be a graph on* n *vertices and suppose that* $(V_i)_{i\in[k]}$ *is a partition of* $V(G)$. *Suppose* R *is an* (ε, d)*-reduced graph of* G *on* $V_1, \dots, V_k$ *and that* R' *is a subgraph of* R *whose connected components have size at most* r. *Assume that* $(V_i, V_j)_G$ *is an* (ε, d)*-super-regular pair for every edge* $V_iV_j \in E(R')$. *Further, let* H *be a graph on* n *vertices with maximum degree* $\Delta(H) \leqslant \Delta$ *that has a vertex partition* $V(H) = (W_i)_{i\in k}$ *which is* ε*-compatible with* $V(G) = (V_i)_{i\in[k]}$ *and* $R' \subseteq R$. *Then* $H \subseteq G$. □

Lemma 3.8.2 is a consequence of the Blow-up lemma of Komlós, Sárközy and Szemerédi [76]. We now prove Theorem 1.4.8.

Proof. [Proof of Theorem 1.4.8] Firstly, note that it suffices to prove Theorem 1.4.8 under the additional assumption that $\eta \ll 1$. We choose β, n_0 as well as additional constants $d, \varepsilon, \xi, \lambda$ as follows: Choose $d \ll \nu$ as required by Lemma 3.6.1 also ensuring that $d \ll 1/\Delta$. Then take $\varepsilon \leqslant \varepsilon(d, \Delta, 2)$ as in Lemma 3.8.2, ensuring also that $\varepsilon \ll d$. Finally choose

$$1/n_0 \ll \beta \ll \lambda \ll \xi \ll \varepsilon,$$

as required by Lemmas 3.6.1 and 3.7.1.

Apply Lemma 3.6.1 to G to obtain an integer k such that $\xi \ll 1/k \ll \varepsilon$, a partition $(n_i)_{i=1}^{2k}$ of n and integers $1 \leqslant i_1 \neq j_1, i_2 \neq j_2 \leqslant k$. Suppose C is the cycle $12\dots(2k)1$ with the chord $c = \{2i_2, 2j_2\}$. Next we apply Lemma 3.7.1 with the partition $(n_i)_{i=1}^{2k}$ of n as input to obtain a set $S \subseteq V(H)$ and a homomorphism $f : H \to C \cup \{c\}$, such that

(i) $|S| \leqslant \xi n$;

(ii) $|f^{-1}(i)| \leqslant n_i + \xi n$ for all $1 \leqslant i \leqslant 2k$;

(iii) Every edge which is not in $H[S]$ is mapped to the edge $\{2i-1, 2i\}$, for some $1 \leqslant i \leqslant k$.

Let $W_i := f^{-1}(i)$ and $n'_i := |f^{-1}(i)|$ for each i, and note that $(n'_i)_{i=1}^k$ is a partition of n. Condition (ii) together with Lemma 3.6.1 imply that there is a partition $A'_1, B'_1, A'_2, B'_2, \dots, A'_k, B'_k$ of $V(G)$ and a spanning subgraph G' of G which satisfy conditions (α_1)–(α_5).

Relabel these clusters $V_1, \dots, V_k$ such that $V_{2i-1} := A'_i$ and $V_{2i} := B'_i$ for all $1 \leqslant i \leqslant k$. So $|V_i| = n'_i$ for all $1 \leqslant i \leqslant 2k$. Let R be the (ε, d)-reduced graph of G' on $V_1, \dots, V_k$ with the maximal number of edges. Hence (α_2)–(α_5) imply that R contains the Hamilton cycle $C' = V_1V_2 \dots V_{2k}V_1$ and the chord $c' := V_{2i_2}V_{2j_2}$ (we view $C' \cup \{c'\}$ as a copy of $C \cup \{c\}$ in R). Let R' be the spanning subgraph of R containing precisely the edges $V_{2i-1}V_{2i}$ for $1 \leqslant i \leqslant k$. Note that (α_2) implies that $(V_{2i-1}V_{2i})_{G'}$ is an (ε, d)-super-regular pair for all $1 \leqslant i \leqslant k$.

We now check that the partition $V(H) = (W_i)_{i=1}^{2k}$ is ε-compatible with the partition $(V_i)_{i=1}^{2k}$ and $R' \subseteq R$. We defined $(V_i)_{i=1}^{2k}$ so that $|V_i| = |W_i|$ for each $1 \leqslant i \leqslant 2k$ and hence condition (γ_1) of Definition 3.8.1 holds. Condition (γ_2) holds since $f : H \to C \cup \{c\}$ is a homomorphism and $C \cup \{c\}$ is a subgraph of R. Note that for all $1 \leqslant i \leqslant 2k$, Lemma 3.6.1 implies that

$$\varepsilon n'_i \geqslant \varepsilon(n_i - 2k\xi n) \geqslant \varepsilon(n/3k - 2k\xi n) \geqslant \varepsilon n/4k \gg \xi n \overset{(i)}{\geqslant} |S|,$$

where the first inequality follows from $n'_{i'} \leqslant n_{i'} + \xi n$ for every $i' \neq i$ in $[2k]$. Furthermore, $|N_H(S) \cap (W_i \backslash S)| \leqslant \Delta|S| \leqslant \Delta\xi n \ll \varepsilon n/4k \leqslant \varepsilon n'_j$ for all $1 \leqslant i, j \leqslant 2k$. Thus, condition (γ_3) holds. Hence, Lemma 3.8.2 implies that G' (and therefore G) contains H, as desired. □

Chapter 4

Perfect matchings in hypergraphs

4.1 Outline

The main goal of this chapter is to prove Theorems 1.5.1 and 1.5.2. In Section 4.2 we deduce Theorem 1.5.1 from Theorem 1.5.2 by showing that Procedure DeterminePM has the requisite properties. The rest of the chapter is devoted to the proof of Theorem 1.5.2, beginning with Section 4.3 in which we briefly sketch how this theorem is proved and outline the key lemmas of the proof (Lemmas 4.6.1 and 4.6.8). In Sections 4.4 and 4.5 we introduce a number of necessary preliminaries for the proof. Section 4.4 details the results from [65] which we use, theory regarding k-partite k-graphs, the Weak Regularity Lemma, convex geometry and some well-known probabilistic tools. Section 4.5 focuses on a key definition, that of 'robust maximality', which allows our proof to run much more smoothly.

Having established these preliminaries, we proceed to state and prove Lemmas 4.6.1 and 4.6.8 in Section 4.6. Finally, in Section 4.7 we deduce Theorem 1.5.2 as a consequence of Lemma 4.6.8.

4.2 Algorithmic analysis

In this section we prove Theorem 1.5.1 by demonstrating that Procedure DeterminePM has the requisite properties. We begin with the least evident claim, namely that the list of partitions $\mathcal{P}$ which is the range for the innermost *for* loop can be generated in polynomial time.

Lemma 4.2.1 *Fix $\delta > 0$. Let H be a k-graph on a vertex set V of size n with $\delta_{k-1}(H) \geqslant \delta n$. Let $2 \leqslant d < k$ and let $L \subseteq \mathbb{Z}^d$ be a $(1, -1)$-free edge-lattice. Then there are at most $d^{k-2+1/\delta}$ partitions $\mathcal{P}$ of V such that $\mathbf{i}_{\mathcal{P}}(e) \in L$ for every $e \in H$, and these partitions can be listed in time $O(n^{k+1})$.*

Proof. We prove Lemma 4.2.1 by considering Procedure ListPartitions. For this, we imagine each vertex class V_j to be a 'bin' to which vertices may be assigned, and keep track of a set *Unassigned* of vertices yet to be assigned to a vertex class. So initially we take each V_j to be empty and $Unassigned = V(H)$. The procedure operates as a *search tree*; at certain points the instruction is to branch over a range of possibilities. This means to select one of these possibilities and continue with this choice, then, when the algorithm halts, to return to the branch point, select the next possibility, and so forth. Each branch may produce an output partition; the output of the procedure consists of all output partitions.

Note that since L is assumed to be $(1, -1)$-free, for any partition $\mathcal{P}$ of the assigned vertices and any set S of $k-1$ assigned vertices, there is at most one $j \in [k]$ such that $\mathbf{i}_{\mathcal{P}}(S) + \mathbf{u}_j \in L$. So the instruction 'Choose such a j arbitrarily' will actually only ever have at most one choice.

The final line of the procedure ensures that any partition $\mathcal{P}$ of $V(H)$ which is output has the property that $\mathbf{i}_{\mathcal{P}}(e) \in L$ for every $e \in H$. The converse is also true: any partition $\mathcal{P}$ of $V(H)$ such that $\mathbf{i}_{\mathcal{P}}(e) \in L$ for every $e \in H$ will be output by some branch of the

Procedure ListPartitions

Input: A k-graph H and a $(1,-1)$-free edge-lattice $L \subseteq \mathbb{Z}^d$.
Output: Outputs all partitions of $V(H)$ with $\mathbf{i}_{\mathcal{P}}(e) \in L$ for every $e \in H$.

Set each $V_j = \emptyset$ and $Unassigned = V(H)$.;
while $Unassigned \neq \emptyset$ **do**
  if $xy_1 \dots y_{k-1} \in H$ *for some vertices* $x \in Unassigned$ *and* $y_1, \dots, y_{k-1} \notin Unassigned$ **then**
    if $\mathbf{i}_{\mathcal{P}}(y_1 \dots y_{k-1}) + \mathbf{u}_j \in L$ *for some* $j \in [k]$ **then**
      Choose such a j arbitrarily;
      Assign x to V_j and remove x from $Unassigned$.
    else
      Halt with no output.
  else
    Choose $x \in Unassigned$;
    Branch over all possible assignments of x.
if $\mathbf{i}_{\mathcal{P}}(e) \in L$ *for every* $e \in H$ **then** halt with output $\mathcal{P}$.

procedure. To see this, consider the branch of the procedure in which, at each branch point, the vertex x under consideration is assigned to the vertex class in which it lies in $\mathcal{P}$. By the above remark, every other vertex of H must also be assigned to the vertex class in which it lies in $\mathcal{P}$. We conclude that Procedure ListPartitions indeed runs correctly, returning all partitions $\mathcal{P}$ of V such that $\mathbf{i}_{\mathcal{P}}(e) \in L$ for every $e \in H$.

To bound the number of such partitions, observe that the procedure will branch over the assignment of each of the first $k-2$ vertices considered. Let x be a vertex considered by the algorithm which is *not* one of these first $k-2$ vertices. The procedure will branch over the assignment of x only if there is no edge $xy_1 \dots y_{k-1}$ of H where $y_1, \dots, y_{k-1}$ have all previously been assigned. Suppose that this is the case, and let $y_1 y_2 \dots y_{k-2}$ be an arbitrary set of $k-2$ distinct previously-assigned vertices. Since $\delta_{k-1}(H) \geqslant \delta n$ there are at least δn vertices v such that $xvy_1 \dots y_{k-2}$ is an edge of H; none of these v can have been previously assigned, and each will be assigned before the next branch of the procedure. Hence, except for the first $k-2$ vertices considered, after any branch in the procedure at least δn vertices are assigned before the next branch. We conclude that the search tree has depth at most $k-2+1/\delta$, and its degree is the number of vertex classes d, and so it has at most $d^{k-2+1/\delta}$ leaf nodes. At most one partition is output at each leaf node, so at most $d^{k-2+1/\delta}$ partitions $\mathcal{P}$ will be output by the algorithm, as required. Furthermore, over all branches there will be at most $d^{k-2+1/\delta}n$ iterations of the *while* loop, and the condition of the first *if* statement takes $O(n^k)$ operations to check, and so the overall running time is $O(n^{k+1})$. $\square$

Proof of Theorem 1.5.1. Fix $k \geqslant 3$ and $\gamma > 0$, and let H be a k-graph on n vertices with $\delta_{k-1}(H) \geqslant (1/k+\gamma)n$. We begin by demonstrating that Procedure DeterminePM will determine correctly whether or not H has a perfect matching. This is trivial if $n < n_0 = n_0(k, \gamma, k^{k^k+1})$, so assume without loss of generality that $n \geqslant n_0$.

Suppose first that H does not contain a perfect matching. Then by Theorem 1.5.2 applied with $C = k^{k^k}$ there exist

(i) a partition $\mathcal{P}$ of $V(H)$ into d parts, where $1 \leqslant d < k$, and

(ii) a $(1,-1)$-free edge-lattice $L \subseteq \mathbb{Z}^d$ such that any matching M in H formed of edges $e \in H$ with $\mathbf{i}_{\mathcal{P}}(e) \notin L$ has size less than k^{k^k}

such that no matching M' in H of size at most $k-2$ has $\mathbf{i}_{\mathcal{P}}(V \setminus V(M')) \in L$. Let M be a maximal matching in H formed of edges $e \in H$ with $\mathbf{i}_{\mathcal{P}}(e) \notin L$, so $|M| < k^{k^k}$. By maximality of M any edge $e \in H$ which does not intersect $V(M)$ has $\mathbf{i}_{\mathcal{P}}(e) \in L$. Therefore, when Procedure DeterminePM considers these M, d, and L, this partition $\mathcal{P}$ of V will be considered in some iteration of the *for* loop. Procedure DeterminePM will then find that no matching $M' \subseteq H$ of size at most $k-2$ has $\mathbf{i}_{\mathcal{P}}(V \setminus V(M')) \in L$ and output that H does not contain a perfect matching, as required.

Now suppose that H does contain a perfect matching, and suppose for a contradiction that Procedure DeterminePM incorrectly claims that this is not the case. This can only arise if Procedure DeterminePM considers a matching M in H of size at most k^{k^k}, an integer $1 \leqslant d < k$, a $(1,-1)$-free edge-lattice $L \subseteq \mathbb{Z}^d$ and a partition $\mathcal{P}$ of V into d parts so that any edge $e \in H$ which does not intersect $V(M)$ has $\mathbf{i}_{\mathcal{P}}(e) \in L$, and finds that no matching $M' \subseteq H$ of size at most $k-2$ has $\mathbf{i}_{\mathcal{P}}(V(H) \setminus V(M')) \in L$. Since any edge $e \in H$ with $\mathbf{i}_{\mathcal{P}}(e) \notin L$ must intersect $V(M)$, any matching M_0 in H formed of edges $e \in H$ with $\mathbf{i}_{\mathcal{P}}(e) \notin L$ has size at most $k|M| \leqslant k^{k^k+1}$. So $\mathcal{P}$, d and L are as in condition (ii) of Theorem 1.5.2 with $C = k^{k^k+1}$; since H has a perfect matching this is a contradiction.

It remains only to show that Procedure DeterminePM must terminate in polynomial time. To see this, note that the range of the first *for* loop consists of at most $n^{k \cdot k^{k^k}}$ matchings M, and these can be generated in time $O(n^{k^{k^k+1}})$ (by considering each set of at most k^{k^k} edges in turn). The range of the next *for* loop (over d and L) has a constant number of elements and can be generated in constant time. Indeed, since L must be an edge-lattice, it must be generated by vectors with non-negative co-ordinates which sum to k; the number of such generating sets is bounded by a function of k. Lemma 4.2.1 applied to $H - V(M)$ shows that the range of the innermost *for* loop also has constant size, and the list can be generated in time $O(n^{k+1})$. Finally, similarly to above it takes time $O(n^{k(k-2)})$ to evaluate the truth of the condition of the central *if* statement. We conclude that in any case this procedure will run in time $O(n^{k^{k^k+1}+k(k-2)})$, as required. $\square$

4.3 Outlines of the proofs

In this section we briefly sketch the proof of Theorem 1.5.2. To do this we will use informal descriptions of theory which will be rigorously detailed in later sections. The argument showing that (i) implies (ii) in Theorem 1.5.2 is fairly straightforward. Indeed, if a k-graph H as in the theorem contains a perfect matching M^*, then for any $\mathcal{P}$, d, L and M as in condition (ii) of Theorem 1.5.2 we have $\mathbf{i}_{\mathcal{P}}(V(H) - V(M^*)) = \mathbf{i}_{\mathcal{P}}(\emptyset) = 0 \in L$. Whilst we cannot simply take $M' = M^*$ since M^* is too large, a simple argument (Lemma 4.7.6) shows that since $\mathbf{i}_{\mathcal{P}}(V(H) - V(M^*)) \in L$ there must be a submatching $M' \subseteq M^*$ of size at most $k-2$ in H such that $\mathbf{i}_{\mathcal{P}}(V(H) - V(M')) \in L$, as required. So the main difficulty in proving Theorem 1.5.2 lies in proving the converse, that (ii) implies (i).

For much of this argument it is simpler to work with k-partite k-graphs H. These are graphs whose vertex set is partitioned into vertex classes $V_1, \dots, V_k$ so that every edge of H contains one vertex from each V_j. The central result in the proof of Theorem 1.5.2, namely Lemma 4.6.1, is expressed in terms of k-partite k-graphs; we then translate this result into the non-partite setting (Lemma 4.6.8) by using a random k-partition of H. When working with k-partite k-graphs our codegree conditions will apply only to *partite* $(k-1)$-sets of vertices, i.e. sets which contain at most one vertex from any vertex class.

A key result on which we rely is Theorem 4.4.7, which is covered in detail in the next section. This theorem allows us to restrict our attention to the case in which H is close to a divisiblity barrier. We combine this with the following observation, which is formalised in the proof of Lemma 4.6.1: Suppose that H has the property that almost all partite $(k-1)$-sets have codegree at least $n/\ell + \gamma n$ in H, and is close to a divisiblity barrier. Then H can be divided into k-partite subgraphs $H^1, \dots, H^r$ whose vertex sets partition $V(H)$ and each of which satisfies a stronger codegree condition: almost all partite $(k-1)$-sets in H^j have codegree at least $(1/(\ell-1) + \gamma/4)n_j$ in H^j, where n_j is the size of each vertex class of H^j. This allow us to prove, by induction on ℓ, that if in addition H satifies a minimum vertex degree condition as well as a further condition relating to the divisibility barrier, then H does contain a perfect matching. The induction is detailed in Lemma 4.6.1, which constitutes the core of our proof.

We will frequently speak of a partition $\mathcal{Q}$ of a vertex set V. We use this term in a slightly non-standard way to mean a family of pairwise-disjoint subsets of V whose union is V (so it is possible for a part to be empty, though this will rarely be the case). Furthemore, we implicitly fix an order on the parts of the partition, so we may consistently speak of, for example, the ith part of $\mathcal{Q}$. Also, in a slight abuse of notation we will sometimes refer to $\mathcal{Q}$ as also being partition of a subset $V' \subseteq V$; the partition referred to in this way is the natural restriction of $\mathcal{Q}$ to V'.

Let V be a (finite) set of vertices, and let $\mathcal{Q}$ partition V into d parts. Then for any set $S \subseteq V$, the *index vector* $\mathbf{i}_{\mathcal{Q}}(S)$ of S with respect to $\mathcal{Q}$ is the vector in $\mathbb{Z}^d$ whose ith co-ordinate is the size of the intersection of S with the ith part of $\mathcal{Q}$ (note that $\mathbf{i}_{\mathcal{Q}}(S)$ is well-defined by the implicit order of the parts of $\mathcal{Q}$). When $\mathcal{Q}$ is clear from the context we write simply $\mathbf{i}(S)$ for $\mathbf{i}_{\mathcal{Q}}(S)$. This notation gives us a convenient way of describing where the edges of a k-graph lie with regard to a given partition of the vertex set, as follows.

Definition 4.3.1 *Let H be a k-graph on vertex set V, and let $\mathcal{Q}$ be a partition of V into d parts. Then for any $\mu > 0$,*

(i) $I^{\mu}_{\mathcal{Q}}(H)$ denotes the set of all $\mathbf{i} \in \mathbb{Z}^d$ such that at least $\mu|V(H)|^k$ edges $e \in G$ have $\mathbf{i}(e) = \mathbf{i}$, that is, those vectors which are the index vector of many edges of G.

(ii) $I_{\mathcal{Q}}(H)$ denotes the set of all $\mathbf{i} \in \mathbb{Z}^d$ such that at least one edge $e \in H$ has $\mathbf{i}(e) = \mathbf{i}$.

(iii) $L^{\mu}_{\mathcal{Q}}(H)$ denotes the lattice in $\mathbb{Z}^d$ generated by $I^{\mu}_{\mathcal{Q}}(H)$, and $L_{\mathcal{Q}}(H)$ denotes the lattice in $\mathbb{Z}^d$ generated by $I_{\mathcal{Q}}(H)$.

(iv) If $\mathcal{Q}$ is the trivial partition in which every part contains one vertex, then we write $L(H)$ and $I(H)$ for $L_{\mathcal{Q}}(H)$ and $I_{\mathcal{Q}}(H)$ respectively.

4.3.1 An analogue for tripartite 3-graphs

To illustrate the ideas of the proof better, we now outline the the proof of an analogous result to Theorem 1.5.3 for 3-partite 3-graphs. Indeed, let H be a 3-partite 3-graph with vertex classes V_1, V_2 and V_3 each of size n. Suppose that any partite pair in H has codegree at least $(1/3 + \gamma)n$. We shall prove that either H contains a perfect matching, or $V(H)$ can be partitioned into two parts A and B such that every edge of H has an even number of vertices in A, but $|A|$ is odd (in particular, H cannot contain a perfect matching in this case).

We begin by applying Theorem 4.4.7 to H. This implies that either H contains a perfect matching (in which case we are done) or that there is a partition $\mathcal{P}$ of $V(H)$ into parts of size at least $(1/3 + \gamma/2)n$ which refines the k-partition of $V(H)$ and is such that $L^{\mu}_{\mathcal{P}}(H)$ is incomplete. Note that this means that $\mathcal{P}$ refines each vertex class V_i into at most two parts; in fact, it is not hard to see that $\mathcal{P}$ must partition each vertex class V_i into precisely two parts, W_i^1 and W_i^2, and additionally that $L^{\mu}_{\mathcal{P}}(H)$ is $(1, -1)$-free. (Recall that $(1, -1)$-free lattices were defined in Section 1.5.3.)

We now decompose H into eight 3-partite subgraphs $(H^{ijk})_{i,j,k\in[2]}$, where H^{ijk} consists of all edges of H which contain one vertex from each of W_1^i, W_2^j and W_3^k. So every edge of H lies in precisely one subgraph H^{ijk}. Since $L^{\mu}_{\mathcal{P}}(H)$ is $(1, -1)$-free, we may assume without loss of generality that almost all of the edges of H lie in either $H^{111}, H^{122}, H^{212}$ or H^{221} (the remaining subgraphs have very low density). As described earlier, we now observe that almost all partite pairs in each subgraph have codegree at least $(1/2+\gamma/2)n$. Indeed, our assumption on H tells us that any pair of vertices xy with $x \in W_1^1$ and $y \in W_2^1$ has at least $(1/3 + \gamma)n$ neighbours in V_3 (i.e. vertices $z \in V_3$ such that $xyz \in H$). But since H_{112} has very low density, very few such pairs xy can have many neighbours in W_3^2. So almost all of these pairs must have $(1/3 + \gamma/2)n \geqslant (1/2 + \gamma/2)|W_{31}|$ neighbours in W_{31}. So H^{111} satisfies a significantly stronger codegree condition than that satisfied by H, ignoring the fact that a small number of pairs, which we will call *bad* pairs, fail this codegree condition.

At this point we delete a (small) matching M in H to achieve two aims. Firstly, M will cover all vertices which lie in many bad pairs. Since there are few bad pairs there will only be a small number of such vertices. Secondly, after deleting the vertices covered by M there will be an *even* number of vertices remaining in $W^2 := W_1^2 \cup W_2^2 \cup W_3^2$. This can be done provided that at least one edge of H intersects W^2 in an odd number of vertices; if there is no such edge then we are done, and H contains no perfect matching.

Delete the vertices covered by M from H; this only slightly weakens the codegree condition on H^{111}, H^{122}, H^{212} and H^{221} since M does not contain many edges. To complete the proof, we choose a random partition of the remaining vertices of H into subsets $S_{111}, S_{122}, S_{212}$ and S_{221} under the constraints that (i) each subset contains equally many vertices from each vertex class V_j, and (ii) $S_{ijk} \subseteq W_1^i \cup W_2^j \cup W_3^k$. Such partitions exist since each part W_j^i has size greater than $n/3$ and $|W^2|$ is even (this is a special case of Proposition 4.5.4). Let H'_{111} be the 3-partite subgraph of H induced by S_{111}, and define H'_{122}, H'_{212}, and H'_{221} similarly. The fact that the partition was chosen randomly implies that, in each subgraph, almost every partite pair has codegree at least $(1/2 + \gamma/2)n'$, where n' is the number of vertices in each part, and no vertex lies in many bad pairs. It follows (by Theorem 4.4.7) that each subgraph contains a perfect matching; together with the deleted matching this yields a perfect matching in H.

4.3.2 The general case

For the general case of Theorem 1.5.2, we use the ideas outlined above, but there are several additional complications. Firstly, there may now be many possibilities for the partition $\mathcal{P}$ returned by applications of Theorem 4.4.7. Also, we will need to use Theorem 4.4.7 several times to repeatedly strengthen the codegree condition under consideration. Both of these complications are handled in the core of the proof, Lemma 4.6.1. Loosely speaking, this says the following statement holds for $\ell \leqslant k$. Let H be a k-partite k-graph with vertex classes of size n in which almost all partite $(k-1)$-sets have codegree at least $(1/\ell + \gamma)n$. Let $\mathcal{P}$ be a partition of each vertex class of H into parts of size at least n/k, which is maximal with the property that $L^\mu_\mathcal{P}(H)$ is $(1, -1)$-free (for some fixed small μ). That is, no refinement of $\mathcal{P}$ has this property. Provided that every vertex lies in many edges with index in $L^\mu_{\mathcal{P}}(H)$, and also $\mathbf{i}_{\mathcal{P}}(V(H)) \in L^\mu_{\mathcal{P}}(H)$ then H contains a perfect matching.

We prove this statement by induction on ℓ by a similar argument to that in the previous section (roughly speaking, there we reduced the $\ell = 3$ case to the $\ell = 2$ case, for $k = 3$). That is, we apply Theorem 4.4.7 to yield a perfect matching (in which case we are done), or a partition $\mathcal{P}''$ of $V(H)$ into parts of size at least $(1/\ell + \gamma)n$ such that $L^\mu_{\mathcal{P}''}(H)$ is incomplete. For the base case of the induction we observe that the latter outcome is impossible for $\ell = 2$. We then consider the non-trivial subgraphs $H_{\mathbf{i}}$ for $\mathbf{i} \in L^\mu_{\mathcal{P}''}(H)$. By a similar argument to the previous section (formalised in Proposition 4.4.5), we find that each $H_{\mathbf{i}}$ satisfies a codegree condition similar to H but with $\ell - 1$ in place of ℓ. Our aim is then to find vertex-disjoint subgraphs $H'_{\mathbf{i}} \subseteq H_{\mathbf{i}}$ whose vertex sets partition $V(H)$, each of which satisfies the conditions of Lemma 4.6.1 with a stronger codegree condition (i.e. $\ell - 1$ in place of ℓ). We can then apply the inductive hypothesis to find a perfect matching in each $H'_{\mathbf{i}}$; together these form a perfect matching in H.

To do this, for each $\mathbf{i}$ we take a partition $\mathcal{Q}_{\mathbf{i}}$ of $V(H_{\mathbf{i}})$ which is maximal with the property that $L^\mu_{\mathcal{Q}_{\mathbf{i}}}(H_{\mathbf{i}})$ is $(1, -1)$-free. To apply the inductive hypothesis we then need to ensure that our partition into subgraphs $H'_{\mathbf{i}}$ satisfies $\mathbf{i}_{\mathcal{Q}_{\mathbf{i}}}(V(H'_{\mathbf{i}})) \in L^\mu_{\mathcal{Q}_{\mathbf{i}}}(H_{\mathbf{i}})$ for each $\mathbf{i}$. We write $\mathcal{Q}^\cap$ for the common refinement of all the partitions $\mathcal{Q}_{\mathbf{i}}$, and determine how many vertices each $H'_{\mathbf{i}}$ must include from each part of $\mathcal{Q}^\cap$ for all of these conditions to be satisfied. This is the most technical part of the proof, and is accomplished by proving Claims 4.6.2, 4.6.3 and 4.6.5 in succession. Finally, we choose the vertex set of each $H'_{\mathbf{i}}$ at random with the given number of vertices from each part of $\mathcal{Q}^\cap$. In Claim 4.6.7 we demonstrate that with high probability this random selection does indeed give subgraphs $H'_{\mathbf{i}}$ to which we can apply the inductive hypothesis, completing the proof.

However, the condition above that '$\mathcal{P}$ is maximal with the property that $L^\mu_{\mathcal{P}}(H)$ is $(1, -1)$-free' turns out to be insufficient for the proof. Essentially this is because a slight change to μ or H might result in a refinement of $\mathcal{P}$ having this property also. We instead use the notion of 'robust maximality': a partition $\mathcal{P}$ of $V(H)$ is robustly maximal with respect to H if $L^{\mu}_{\mathcal{P}}(H)$ is $(1,-1)$-free but $L^{\mu'}_{\mathcal{Q}}(H)$ is not $(1,-1)$-free for any refinement $\mathcal{Q}$ of $\mathcal{P}$, even if μ' is much larger than μ. Section 4.5 is devoted to the study of this property, showing that any k-graph admits a robustly maximal partition, and that the property of being robustly maximal is preserved after small changes to H or by taking a random subgraph of H.

Having proved Lemma 4.6.1, we deduce the non-partite equivalent Lemma 4.6.8 by starting with a (non-partite) k-graph, taking a random k-partition and applying Lemma 4.6.1. Finally, in Section 4.7 we complete the proof of Theorem 1.5.2 by a fairly straightforward argument, showing that if condition (ii) of the theorem holds, then we can delete the vertices covered by some small matching from H so that after this deletion some partition $\mathcal{P}$ which is robustly maximal with respect to H has the property that $\mathbf{i}_{\mathcal{P}}(V) \in L^{\mu}_{\mathcal{P}}(H)$. We can then apply Lemma 4.6.8 to find a perfect matching in H, as required.

4.4 Hypergraph Theory and Geometry

A *hypergraph* consists of a vertex set V and an edge set E, where every $e \in E$ is a subset of V. We frequently identify a hypergraph with its edge set, writing $|H|$ for the number of edges of H, and $e \in H$ to mean that e is an edge of H. A *k-graph* is a hypergraph in which every edge has size precisely k. Let H be a k-graph with vertex set V, and let S be a set of $k-1$ vertices of H. Then the *codegree* $d_H(S)$ of S is the number of edges in H which contain S as a subset, or equivalently, the number of vertices $v \in V$ such that $S \cup \{v\} \in H$. When H is clear from the context we write simply $d(S)$.

4.4.1 Partitions, index vectors and lattices

Recall Definition 4.3.1. We shall make extensive use of the following elementary proposition.

Proposition 4.4.1 *Let G and H be k-graphs sharing a common vertex set V of size n, let $\mathcal{P}$ be a partition of V, and let μ and α be positive constants. If $|G \triangle H| \leqslant \alpha n^k$, then $I^{\mu}_{\mathcal{P}}(G) \subseteq I^{\mu-\alpha}_{\mathcal{P}}(H)$, and consequently $L^{\mu}_{\mathcal{P}}(G) \subseteq L^{\mu-\alpha}_{\mathcal{P}}(H)$.*

Proof. If $\mathbf{i} \in I^{\mu}_{\mathcal{P}}(G)$, then at least μn^k edges $e \in G$ have $\mathbf{i}(e) = \mathbf{i}$, and so at least $(\mu - \alpha)n^k$ edges $e \in H$ have $\mathbf{i}(e) = \mathbf{i}$. So $\mathbf{i} \in I^{\mu-\alpha}_{\mathcal{P}}(G)$, as required for the first conclusion. The second conclusion follows immediately. □

Let V be a vertex set and $\mathcal{Q}$ and $\mathcal{Q}'$ be partitions of V. We say that $\mathcal{Q}'$ *refines* $\mathcal{Q}$, which we denote by $\mathcal{Q} \prec \mathcal{Q}'$, if every part of $\mathcal{Q}'$ is a subset of a part of $\mathcal{Q}$ (possibly the whole of the part, so that in particular every partition refines itself). If $\mathcal{Q}'$ refines $\mathcal{Q}$, then for any set $S \subseteq V$ the index vectors $\mathbf{i}_{\mathcal{Q}}(S)$ and $\mathbf{i}_{\mathcal{Q}'}(S)$ are linked by the equation $(\mathbf{i}_{\mathcal{Q}}(S))_X = \sum (\mathbf{i}_{\mathcal{Q}'}(S))_Y$ for each $X \in \mathcal{Q}$, where the sum is taken over all parts Y of $\mathcal{Q}'$ which refine X. This leads to the following definition.

Definition 4.4.2 *Let V be a finite set and let $\mathcal{P}$ and $\mathcal{P}'$ be partitions of V such that $\mathcal{P}$ refines $\mathcal{P}'$. Let $\mathbf{i}$ be an index vector with respect to $\mathcal{P}$. Then the* restriction $(\mathbf{i} \mid \mathcal{P}')$ *of $\mathbf{i}$ to $\mathcal{P}'$ is defined by*

$$(\mathbf{i} \mid \mathcal{P}')_X = \sum_{Y \in \mathcal{P}, Y \subseteq X} i_X$$

for each $X \in \mathcal{P}'$.

If there are many edges of a given index vector with respect to $\mathcal{Q}$, then there must be many edges of some index vector with respect to $\mathcal{Q}'$ which restricts to $\mathcal{Q}'$. This is the essence of the following proposition.

Proposition 4.4.3 *Let V be a vertex set, and let $\mathcal{P}$ and $\mathcal{Q}$ be partitions of V such that $\mathcal{Q}$ refines $\mathcal{P}$. Let r be the number of parts of $\mathcal{P}$, and for each $j \in [r]$ let m_j be the number of parts into which the jth part of $\mathcal{P}$ is refined by $\mathcal{Q}$. Finally let H be any k-graph on V; then*

(i) if $\mathbf{i} \in L^{\mu}_{\mathcal{Q}}(H)$ then $(\mathbf{i} \mid \mathcal{P}) \in L^{\mu}_{\mathcal{P}}(H)$, and

(ii) for any $\mathbf{i} \in L^{\mu}_{\mathcal{P}}(H)$ there exists $\mathbf{i}' \in L^{\mu/m}_{\mathcal{Q}}(H)$ such that $(\mathbf{i}' \mid \mathcal{P}) = \mathbf{i}$, where $m := \prod_{j\in[r]} \binom{m_j}{i_j}$.

Proof. Let $n := |V|$. For (i), let $\mathbf{i} \in I^{\mu}_{\mathcal{Q}}(H)$, so at least μn^k edges $e \in H$ have $\mathbf{i}_{\mathcal{Q}}(e) = \mathbf{i}$. Each such edge has $\mathbf{i}_{\mathcal{P}}(e) = (\mathbf{i} \mid \mathcal{P})$, so $(\mathbf{i} \mid \mathcal{P}) \in I^{\mu}_{\mathcal{P}}(H)$. Since restriction is a linear operation the result follows.

Likewise, for (ii), fix $\mathbf{i} \in I^{\mu}_{\mathcal{P}}(H)$; then at least μn^k edges $e \in H$ have $\mathbf{i}_{\mathcal{P}}(e) = \mathbf{i}$. Note that there are precisely m index vectors $\mathbf{i}'$ with respect to $\mathcal{Q}$ which satisfy $(\mathbf{i}' \mid \mathcal{P}) = \mathbf{i}$. So by the pigeonhole principle there is some $\mathbf{i}'$ with $(\mathbf{i}' \mid \mathcal{P}) = \mathbf{i}$ for which at least $\mu n^k/m$ edges $e \in H$ have $\mathbf{i}_{\mathcal{Q}}(e) = \mathbf{i}'$, that is, $\mathbf{i}' \in I^{\mu/m}_{\mathcal{Q}}(H)$. Again the result follows by linearity of the restriction operator. □

For any partition $\mathcal{P}$ of a vertex set V and any part $X \in \mathcal{P}$, recall that the *unit vector* $\mathbf{u}_X$ is the index vector with respect to $\mathcal{P}$ which is 1 on X and zero on every other part of $\mathcal{P}$. Let $L \subseteq \mathbb{Z}^d$ be a lattice. Recall that L is $(1, -1)$-*free* if it does not contain any difference of unit vectors $\mathbf{u}_i - \mathbf{u}_j$ with $i, j \in [d]$ and $i \neq j$. Such lattices play a crucial role in this chapter.

4.4.2 Partite hypergraphs

Many of our results will apply specifically to partite k-graphs. Let V be a set of vertices, and let $\mathcal{P}$ partition V. Then we say that a set $S \subseteq V$ is $\mathcal{P}$-*partite* if S has at most one vertex in any part of $\mathcal{P}$. We say that a hypergraph H on V is $\mathcal{P}$-*partite* if every edge of

H is $\mathcal{P}$-partite. In this case we refer to the parts of $\mathcal{P}$ as the *vertex classes* of H.

Usually we will consider k-graphs H which are $\mathcal{P}$-partite for some partition $\mathcal{P}$ of $V(H)$ into k parts. The following definition effectively complements the definition of $I^\mu_{\mathcal{Q}}(H)$ for such k-graphs. Let V be a set of vertices, let $\mathcal{P}$ partition V into k parts, and let $\mathcal{Q}$ refine $\mathcal{P}$. For a $\mathcal{P}$-partite k-graph H on vertex set V and a constant μ, we form the *auxiliary k-graph* $\mathrm{Aux}^\mu_{\mathcal{Q}}(H)$ as follows. First, take a single vertex corresponding to each part of $\mathcal{Q}$. Next, add an edge whenever the corresponding k parts of $\mathcal{Q}$ induce at least $\mu|V(H)|^k$ edges of H. We naturally obtain a partition $\mathcal{P}_{\mathrm{Aux}}$ of the vertices of $\mathrm{Aux}^\mu_{\mathcal{Q}}(H)$, where two vertices lie in the same part of $\mathcal{P}_{\mathrm{Aux}}$ if and only if the corresponding parts of $\mathcal{Q}$ lie in the same part of $\mathcal{P}$. Then $\mathrm{Aux}^\mu_{\mathcal{Q}}(H)$ is $\mathcal{P}_{\mathrm{Aux}}$-partite. The key observation is that by definition of $\mathrm{Aux}^\mu_{\mathcal{Q}}(H)$ we have $I(\mathrm{Aux}^\mu_{\mathcal{Q}}(H)) = I^\mu_{\mathcal{Q}}(H)$, and so $L(\mathrm{Aux}^\mu_{\mathcal{Q}}(H)) = L^\mu_{\mathcal{Q}}(H)$.

For this reason it will often be convenient to work with $\mathrm{Aux}^\mu_{\mathcal{Q}}(H)$. One result which we will apply to $\mathrm{Aux}^\mu_{\mathcal{Q}}(H)$ to obtain information about H is the following proposition. For each $0 \leqslant j \leqslant k-1$ the *partite minimum j-degree* $\delta^*_j(J)$ is defined to be the largest m such that any j-edge e has at least m extensions to a $(j+1)$-edge in any part not used by e, i.e.

$$\delta^*_j(J) := \min_{e \in J_j} \min_{i : e \cap V_i = \emptyset} |\{v \in V_i : e \cup \{v\} \in J\}|.$$

Proposition 4.4.4 *Let $k \geqslant 3$ and let H be a k-partite k-graph with vertex classes $V_1, \dots, V_k$ of size r and $\delta^*(H) \geqslant 1$. Let $\mathbf{i} \in \mathbb{Z}^{rk}$ be such that $\sum_{v \in V_j} i_v$ is constant over $j \in [k]$. Then for any $v \in V(H)$, there exists $v' \in V(H)$ such that $\mathbf{i} - \mathbf{u}_v + \mathbf{u}_{v'} \in L(H)$.*

Proof. The proof proceeds in three stages. We first show that for any $j, j' \in [k]$ and any vertices $x, x' \in V_j$ and $y \in V_{j'}$, there is some $y' \in V_{j'}$ such that $\mathbf{u}_x + \mathbf{u}_y - \mathbf{u}_{x'} - \mathbf{u}_{y'} \in L(H)$. We then show that for any $\mathbf{i} \in \mathbb{Z}^{rk}$, there exist $w, w' \in V$ such that $\mathbf{i} - \mathbf{u}_w + \mathbf{u}_{w'} \in L(H)$. Finally, the statement of the proposition follows by combining these two facts.

The proof of the first part is as follows: If $j \neq j'$ then take e to be any edge of H which contains x and y. So $S = \{x'\} \cup e \setminus \{x, y\}$ is a $(k-1)$-tuple of vertices of H, and so some edge $e' \in H$ contains S; say $e' = \{y'\} \cup S$. Then $\mathbf{i}(e) - \mathbf{i}(e') = \mathbf{u}_x + \mathbf{u}_y - \mathbf{u}_{x'} - \mathbf{u}_{y'} \in L(H)$ by definition. If instead $j = j'$ then choose $\ell \neq j$ and $z \in V_\ell$, whereupon applying the previous argument twice gives first $z' \in V_\ell$ then $y' \in V_{j'}$ such that $\mathbf{u}_x + \mathbf{u}_z - \mathbf{u}_{x'} - \mathbf{u}_{z'} \in L(H)$ and $\mathbf{u}_y + \mathbf{u}_z - \mathbf{u}_{y'} - \mathbf{u}_{z'} \in L(H)$, so $\mathbf{u}_x + \mathbf{u}_y - \mathbf{u}_{x'} - \mathbf{u}_{y'} \in L(H)$ as required.

Now, let e be any edge of of H. Then $\mathbf{i}' := \left(\sum_{w \in V_1} i_w\right) \mathbf{i}(e)$ is a member of $L(H)$ with the property that $((\mathbf{i} - \mathbf{i}') \mid \mathcal{P}') = \mathbf{0}$. Now choose $\mathbf{i}' \in L(H)$ with this property such that $||\mathbf{i} - \mathbf{i}'||_1$ is minimised. The observation above then implies that $||\mathbf{i} - \mathbf{i}'||_1 \leqslant 2$. To see this, suppose for a contradiction that $||\mathbf{i} - \mathbf{i}'||_1 > 3$. Then we may choose $x, y \in V_j$ and $x' \in V_{j'}$ such that $i_x - i'_x > 0$, $i_{x'} - i'_{x'} > 0$ and $i_y - i'_y < 0$. The observation above yields y' such that $\mathbf{i}^* := \mathbf{u}_x + \mathbf{u}_{x'} - \mathbf{u}_y - \mathbf{u}_{y'} \in L(H)$. But then $\mathbf{i}' + \mathbf{i}^* \in L(H)$, and $||\mathbf{i} - (\mathbf{i}' + \mathbf{i}^*)||_1 \leqslant ||\mathbf{i} - \mathbf{i}'||_1 - 3 + 1 < ||\mathbf{i} - \mathbf{i}'||_1$, contradicting the assumed minimality of $\mathbf{i}'$. So $||\mathbf{i} - \mathbf{i}'||_1 \leqslant 2$, and it follows that $\mathbf{i} - \mathbf{i}' = \mathbf{u}_w - \mathbf{u}_{w'}$ for some $w, w' \in V$. Then $\mathbf{i} - \mathbf{u}_w + \mathbf{u}_{w'} = \mathbf{i}' \in L(H)$.

Finally, for any $v \in V$ we can find $v' \in V$ such that $\mathbf{u}_v + \mathbf{u}_w - \mathbf{u}_{v'} - \mathbf{u}_{w'} \in L(H)$. Now $\mathbf{i} - \mathbf{u}_v + \mathbf{u}_{v'} = \mathbf{i}' - (\mathbf{u}_v + \mathbf{u}_w - \mathbf{u}_{v'} - \mathbf{u}_{w'}) \in L(H)$. □

We are now ready to establish some key properties of $\mathcal{P}'$-partite hypergraphs and partitions $\mathcal{P}$ of their vertex sets which refine $\mathcal{P}'$. Specifically, we consider hypergraphs H in which all but a small proportion of $\mathcal{P}'$-partite $(k-1)$-sets in $V(H)$ have codegree at least dn, and $L^\mu_{\mathcal{P}}(H)$ is $(1, -1)$-free (for appropriate values of d and μ): from these conditions we may deduce a significant amount of information about the structure of H, $\mathcal{P}$ and $L^\mu_{\mathcal{P}}(H)$. In particular, property (v) of the following proposition shows that the subgraph $H_{\mathbf{i}}$ of H satisfies a similar codegree condition (with slightly weaker constants), and so we may then apply Proposition 4.4.5 to this subgraph also. This 'inheritance' would not be the case if Proposition 4.4.5 instead assumed the stronger condition that H had minimum codegree at least dn; this is the reason why for much of this chapter we work with the weaker type of codegree condition, which applies to almost all edges of H.

Proposition 4.4.5 *Suppose that $1/n_0 \ll \mu, \varepsilon \ll \psi \leqslant d, c, 1/k$. Let $n \geqslant n_0$ and let $\mathcal{P}'$ partition a vertex set V into k parts $V_1, \dots, V_k$, where $cn \leqslant |V_i| \leqslant n$ for each i. Let H be a $\mathcal{P}'$-partite k-graph on vertex set V in which at most εn^{k-1} $\mathcal{P}'$-partite sets S of $k-1$ vertices have $d(S) < dn$. Finally, let $\mathcal{P}$ be a partition of V which refines $\mathcal{P}'$, has parts each of size at least cn, and is such that $L^\mu_{\mathcal{P}}(H)$ is $(1, -1)$-free. Then the following properties hold for some integer r.*

(i) Any $\mathcal{P}'_{\mathrm{Aux}}$-partite set S of $k-1$ vertices of $\mathrm{Aux}^\mu_{\mathcal{P}}(H)$ has $d(S) = 1$.

(ii) Each part of $\mathcal{P}'$ is refined into exactly r parts by $\mathcal{P}$.

(iii) Every part of $\mathcal{P}$ is indexed by exactly r^{k-2} vectors $\mathbf{i} \in I^\mu_{\mathcal{P}}(H)$.

(iv) $I^{dc^{k-1}/2}_{\mathcal{P}}(H) = I^\mu_{\mathcal{P}}(H)$.

Furthermore, for any $\mathbf{i} \in I^\mu_{\mathcal{P}}(H)$ let $S_{\mathbf{i}}$ denote the parts of $\mathcal{P}$ indexed by $\mathbf{i}$; let $\mathcal{P}_{\mathbf{i}}$ be the natural k-partition of $S_{\mathbf{i}}$, and let $H_{\mathbf{i}}$ be the $\mathcal{P}_{\mathbf{i}}$-partite k-graph on vertex set $S_{\mathbf{i}}$ whose edges are all edges of H of index $\mathbf{i}$. Then the following properties hold.

(v) For any $\mathbf{i} \in I^\mu_{\mathcal{P}}(H)$ at most ψn^{k-1} $\mathcal{P}_{\mathbf{i}}$-partite sets $S \subseteq S_{\mathbf{i}}$ of $k-1$ vertices have $d_{H_{\mathbf{i}}}(S) < (d - \psi)n$.

(vi) For any $\mathbf{i} \in I^\mu_{\mathcal{P}}(H)$ at most ψn vertices of $S_{\mathbf{i}}$ lie in fewer than $c^{k-2} d n^{k-1}/2$ edges of $H_{\mathbf{i}}$.

(vii) For any $\mathbf{i} \in \mathbb{Z}^{kr}$ such that $(\mathbf{i} \mid \mathcal{P}')$ is a multiple of $\mathbf{1}$ and any $X \in \mathcal{P}$, there exists a part $X' \in \mathcal{P}$ such that $\mathbf{i} - \mathbf{u}_X + \mathbf{u}_{X'} \in L^\mu_{\mathcal{P}}(H)$.

Finally, each part of $\mathcal{P}$ has size at least $dn - \psi n$.

Proof. Introduce a new constant ψ' with $\mu, \varepsilon \ll \psi' \ll \psi, d, c, 1/k$. First, note that since each part V_i of $\mathcal{P}'$ is refined into at most $1/c$ parts of $\mathcal{P}$, there are at most $1/c^k$ possible values of $\mathbf{i}_{\mathcal{P}}(e)$ for an edge e of H.

For (i), note first that $d(S) \leqslant 1$, since $L(\mathrm{Aux}^{\mu}_{\mathcal{P}}(H)) = L^{\mu}_{\mathcal{P}}(H)$ is $(1,-1)$-free. Fix parts $X_1, \dots X_{k-1}$ of $\mathcal{P}$, where $X_j \subseteq V_j$ for each $j \in [k-1]$; then by symmetry it is sufficient to show that there is some part X_k of $\mathcal{P}$ with $X_k \subseteq V_k$ such that $H[X_1 \cup \dots \cup X_k]$ contains at least $\mu|V|^k$ edges. To do this, note that $\prod_{j \in [k-1]} |X_j| \geqslant (cn)^{k-1}$ $(k-1)$-tuples $S = (x_1, \dots, x_{k-1})$ have $x_j \in X_j$ for each j, and so at least $(c^{k-1} - \varepsilon)n^{k-1}$ such $(k-1)$-tuples have $d(S) \geqslant dn$. Thus at least $(c^{k-1} - \varepsilon)dn^k \geqslant \mu|V|^k/c$ edges of H contain one vertex from each X_j. Since $\mathcal{P}$ refines V_k into at most $1/c$ parts, some part $X_k \subseteq V_k$ of $\mathcal{P}$ must be as required. Next recall that $I(\mathrm{Aux}^{\mu}_{\mathcal{P}}(H)) = I^{\mu}_{\mathcal{P}}(H)$; then (ii) and (iii) follow immediately from (i).

For (v), fix $\mathbf{i} \in I = I^{\mu}_{\mathcal{P}}(H)$, and let $W_1 \subseteq V_1, \dots, W_k \subseteq V_k$ be the parts of $\mathcal{P}$ indexed by $\mathbf{i}$ (so $S_{\mathbf{i}} = \bigcup W_j$). Consider the sets $S = \{x_1, \dots, x_{k-1}\}$ with $x_i \in W_i$ for each i. By assumption, at most εn^{k-1} such S have $d^*_H(S) < dn$. Furthermore, if S satisfies $d^*_H(S) \geqslant dn$ but $d^*_{H_{\mathbf{i}}}(S) < dn - \psi' n$, then S is contained in at least $\psi' n$ edges $e \in H$ with $\mathbf{i}(e) \notin I$. Since there are at most $c^k \mu (kn)^k$ such edges of H, we conclude that at most $c^k \mu (kn)^k / \psi' n \leqslant \psi' n^{k-1}$ edges of H have this form. This proves a stronger form of (v), with ψ' in place of ψ. In particular this implies that $H_{\mathbf{i}}$ contains at least $((cn)^{k-1} - \psi' n^{k-1})(d - \psi')n \geqslant c^{k-1} dn^k/2$ edges, and so we have (iv).

Now let $v \in S_{\mathbf{i}}$; then v lies in at least $(cn)^{k-2}$, $\mathcal{P}_{\mathbf{i}}$-partite sets S of $k-1$ vertices of $S_{\mathbf{i}}$. So either v lies in at least $2(cn)^{k-2}(d - \psi)n/3 \geqslant dc^{k-2}n^{k-1}/2$ edges of $H_{\mathbf{i}}$, or v lies in at least $(cn)^{k-2}/3$ of the sets S counted in (v). Since there are at most $\psi' n^{k-1}$ such sets, there are at most $\psi' n^{k-1}/((cn)^{k-2}/3) \leqslant \psi n$ vertices of the latter type in $S_{\mathbf{i}}$, which proves (vi).

One further consequence of (iii) and (v) is the 'Finally' part of the statement, since each part $W \in \mathcal{P}$ is indexed by some $\mathbf{i}$, and then some $(k-1)$-tuple in $H_{\mathbf{i}}$ has at least $(d-\psi)n$ neighbours in W. Finally, (vii) follows by applying Proposition 4.4.4 to $\mathrm{Aux}^{\mu}_{\mathcal{P}}(H)$. □

Although for the most part we deal with k-partite hypergraphs in the main lemmas, we will also need analogues of the properties in Proposition 4.4.5 in the main body of the proof of Theorem 1.5.2. For this we use the following proposition.

Proposition 4.4.6 *Suppose that $1/n_0 \ll \mu, \varepsilon \ll \psi, d, c, 1/k$, and let V be a vertex set of size $n \geqslant n_0$. Let H be a k-graph V in which at most εn^{k-1} sets S of $k-1$ vertices have $d(S) < dn$. Let $\mathcal{P}$ be a partition of V which has r parts each of size at least cn and such that $L^{\mu}_{\mathcal{P}}(H)$ is $(1,-1)$-free. Then the following properties hold.*

(i) For any non-negative index vector $\mathbf{i}$ with respect to $\mathcal{P}$ such that $||\mathbf{i}||_1 = k-1$, there exists $X \in \mathcal{P}$ such that $\mathbf{i} + \mathbf{u}_X \in L^{\mu}_{\mathcal{P}}(H)$.

(ii) $I^{dc^{k-1}/3(k-1)!}_{\mathcal{P}}(H) = I^{\mu}_{\mathcal{P}}(H)$.

Furthermore, for any $\mathbf{i} \in I^{\mu}_{\mathcal{P}}(H)$ let $S_{\mathbf{i}} = \bigcup_{X \in \mathcal{P}, i_X > 0} X$ and let $H_{\mathbf{i}}$ be the k-graph whose edges are all edges of H of index $\mathbf{i}$. Then the following properties hold.

(iii) For any $\mathbf{i} \in I^{\mu}_{\mathcal{P}}(H)$ at most ψn^{k-1} sets $S \subseteq S_{\mathbf{i}}$ of $k-1$ vertices such that $||\mathbf{i}_{\mathcal{P}}(S) - \mathbf{i}||_1 = 1$ have $d_{H_{\mathbf{i}}}(S) < (d-\psi)n$.

(iv) For any $\mathbf{i} \in I^{\mu}_{\mathcal{P}}(H)$ at most ψn vertices of $S_{\mathbf{i}}$ lie in fewer than $c^{k-2}dn^{k-1}/4(k-1)!$ edges of $H_{\mathbf{i}}$.

(v) For any $\mathbf{i} \in \mathbb{Z}^r$ such that $\sum_{X \in \mathcal{P}} i_X$ is a multiple of k and any $X \in \mathcal{P}$, there exists a part $X' \in \mathcal{P}$ such that $\mathbf{i} - \mathbf{u}_X + \mathbf{u}_{X'} \in L^{\mu}_{\mathcal{P}}(H)$.

Finally, each part of $\mathcal{P}$ has size at least $dn - \psi n$.

Proof. The proof proceeds analogously to that of Proposition 4.4.5. Introduce a new constant ψ' with $\mu, \varepsilon \ll \psi' \ll \psi, d, c, 1/k$. First, note that since $\mathcal{P}$ has $r \leqslant 1/c$, there are at most $1/c^k$ possible values of $\mathbf{i}_{\mathcal{P}}(e)$ for an edge e of H.

For (i), let $\{X_1, \dots X_{k-1}\}$ be a multiset in which every part $X \in \mathcal{P}$ is included with multiplicity i_X. Note that at least $\prod_{j\in[k-1]}(|X_j| - k)/(k-1)! \geqslant \psi'(cn)^{k-1}$ $(k-1)$-tuples $S = (x_1, \dots, x_{k-1})$ have $x_j \in X_j$ for each j, and so at least $(\psi' c^{k-1} - \varepsilon)n^{k-1}$ such $(k-1)$-tuples have $d(S) \geqslant dn$. Thus at least $(\psi' c^{k-1} - \varepsilon)dn^k \geqslant \mu n^k/c$ edges of H contain at least i_X vertices from each $X \in \mathcal{P}$, i.e., they each have index vector $\mathbf{i} + \mathbf{u}_{X'}$ for some $X' \in \mathcal{P}$. Since $\mathcal{P}$ has at most $1/c$ parts, there exists $X \in \mathcal{P}$ such that at least μn^k of these edges have index vector $\mathbf{i} + \mathbf{u}_X$ with respect to $\mathcal{P}$.

For (iii), fix $\mathbf{i} \in I = I^{\mu}_{\mathcal{P}}(H)$ and choose $X \in \mathcal{P}$ such that $i_X > 0$. Consider the sets S such that $\mathbf{i}_{\mathcal{P}}(S) = \mathbf{i} - \mathbf{u}_X$. By assumption, at most εn^{k-1} such S have $d_H(S) < dn$. Furthermore, if S satisfies $d_H(S) \geqslant dn$ but $d_{H_{\mathbf{i}}}(S) < dn - \psi' n$, then since I is $(1,-1)$-free S is contained in at least $\psi' n$ edges $e \in H$ with $\mathbf{i}(e) \notin I$. Since there are at most $c^k \mu n^k$ such edges of H, we conclude that at most $c^k \mu n^k/\psi' n \leqslant \psi' n^{k-1}$ edges of H have this form, where the inequality follows from $\mu \ll \psi', c$. This proves a stronger form of (iii), with ψ' in place of ψ. In particular this implies that $H_{\mathbf{i}}$ contains at least $((cn)^{k-1}/2(k-1)! - \psi' n^{k-1})(d - \psi')n \geqslant c^{k-1}dn^k/3(k-1)!$ edges, and so we have (ii).

Now let $v \in S_{\mathbf{i}}$; then v lies in at least $(cn-k)^{k-2}/(k-1)! \geqslant (cn)^{k-2}/2(k-1)!$, sets S of $k-1$ vertices H with $||\mathbf{i}_{\mathcal{P}}(S) - \mathbf{i}||_1 = 1$. So either v lies in at least $(cn)^{k-2}(d - \psi)n/3(k-1)! \geqslant dc^{k-2}n^{k-1}/4(k-1)!$ edges of $H_{\mathbf{i}}$, or v lies in at least $(cn)^{k-2}/6(k-1)!$ of the sets S counted in (iii). Since there are at most $\psi' n^{k-1}$ such sets, there are at most $\psi' n^{k-1}/((cn)^{k-2}/6(k-1)!) \leqslant \psi n$ vertices of the latter type in $S_{\mathbf{i}}$, which proves (iv). One further consequence of (i) and (iii) is the 'Finally' part of the statement, since each part $W \in \mathcal{P}$ is indexed by some $\mathbf{i}$, and then some $(k-1)$-tuple in $H_{\mathbf{i}}$ has at least $(d-\psi)n$ neighbours in W.

It remains to prove (v), which we do in a similar way to Proposition 4.4.4. We first show that for any $X_1, X'_1, X_2 \in \mathcal{P}$, there exists $X'_2 \in \mathcal{P}$ such that $\mathbf{u}_{X_1} + \mathbf{u}_{X_2} - \mathbf{u}_{X'_1} - \mathbf{u}_{X'_2} \in L^{\mu}_{\mathcal{P}}(H)$. Choose any non-negative index vector $\mathbf{i}'$ such that $||\mathbf{i}'||_1 = k-3$, and (using (i)) let $Y \in \mathcal{P}$ be such that $\mathbf{i}' + \mathbf{u}_{X_1} + \mathbf{u}_{X_2} + \mathbf{u}_Y \in L^{\mu}_{\mathcal{P}}(H)$ and let X'_2 be such that $\mathbf{i}' + \mathbf{u}_{X'_1} + \mathbf{u}_Y + \mathbf{u}_{X'_2} \in L^{\mu}_{\mathcal{P}}(H)$. Then $\mathbf{u}_{X_1} + \mathbf{u}_{X_2} - \mathbf{u}_{X'_1} - \mathbf{u}_{X'_2}$ is the difference of these two index vectors and hence lies in $L^{\mu}_{\mathcal{P}}(H)$.

Now for any $\mathbf{i} \in \mathbb{Z}^r$ with $\sum_{X \in \mathcal{P}} i_X = km$, choose $\mathbf{i}' \in L^{\mu}_{\mathcal{P}}(H)$ such that $\sum_{X \in \mathcal{P}} i'_X = km$ and which minimises $||\mathbf{i} - \mathbf{i}'||_1$ subject to this condition. The above observation implies that $||\mathbf{i} - \mathbf{i}'||_1 \leqslant 2$. To see this, suppose for a contradiction that $||\mathbf{i} - \mathbf{i}'||_1 \geqslant 3$. Then we can find X_1, X'_1, X_2 such that $i_{X_1} - i'_{X_1} > 0$, $i_{X_2} - i'_{X_2} > 0$ and $i_{X'_1} - i'_{X'_1} < 0$. Let $X'_2 \in \mathcal{P}$ be such that $\mathbf{i}^* = \mathbf{u}_{X_1} + \mathbf{u}_{X_2} - \mathbf{u}_{X'_1} - \mathbf{u}_{X'_2} \in L^{\mu}_{\mathcal{P}}(H)$; then $||\mathbf{i} - (\mathbf{i}' + \mathbf{i}^*)||_1 < ||\mathbf{i} - \mathbf{i}'||_1$ and $\sum_{X \in \mathcal{P}} i^*_X = km$, contradicting our choice of $\mathbf{i}'$. It follows that $\mathbf{i}' = \mathbf{i} - \mathbf{u}_Y + \mathbf{u}_{Y'}$ for some $Y, Y' \in \mathcal{P}$. Now applying the observation once more we obtain X' such that $\mathbf{i}^{**} = \mathbf{u}_X + \mathbf{u}_Y - \mathbf{u}_{X'} - \mathbf{u}_{Y'} \in L^{\mu}_{\mathcal{P}}(H)$ and hence $\mathbf{i} + \mathbf{u}_X - \mathbf{u}_{X'} = \mathbf{i}' + \mathbf{i}^{**} \in L^{\mu}_{\mathcal{P}}(H)$. $\square$

4.4.3 Hypergraph matching theory

A central theorem from [65] will play a key role in this chapter. For this we make the following definitions. A *k-complex* J is a hypergraph such that $\emptyset \in J$, every edge of J has size at most k, and J is closed downwards, that is, if $e \in J$ and $e' \subseteq e$ then $e' \in J$. We write J_r to denote the r-graph formed by edges of size r in J. Also, we use the following notion of partite degree in a multipartite k-complex from [65]. Let V be a set of vertices, let $\mathcal{P}$ partition of V into k parts $V_1, \dots, V_k$, and let J be a $\mathcal{P}$-partite k-complex on V. The *partite degree sequence* $\delta^*(J) = (\delta^*_0(J), \dots, \delta^*_{k-1}(J))$ (recall that $\delta^*_j(J)$ is the largest m such that any j-edge e has at least m extensions to a $(j+1)$-edge in any part not used by e). Note that we suppress the dependence on $\mathcal{P}$ in our notation: this will always be clear from the context. Minimum degree sequence conditions on a k-complex will therefore take the form $\delta^*(J) \geqslant (a_0, \ldots, a_{k-1})$, where the inequality is to be interpreted pointwise. Finally, observe that if J is a $\mathcal{P}$-partite k-complex on V, and $\mathcal{Q}$ is a partition of V which refines $\mathcal{P}$ into d parts, then every edge $e \in J_k$ has $(\mathbf{i}_{\mathcal{Q}}(e) \mid \mathcal{P}) = \mathbf{i}_{\mathcal{P}}(e) = \mathbf{1}$. So we say that $L \subseteq \mathbb{Z}^d$ is *complete with respect to* $\mathcal{P}'$ if $\mathbf{i} \in L$ for every $\mathbf{i} \in \mathbb{Z}^d$ with $(\mathbf{i} \mid \mathcal{P}') = \mathbf{1}$, and *incomplete with respect to* $\mathcal{P}'$ otherwise.

We can now state the theorem from [65] which will play a key role in our proof (the form stated here is actually slightly weaker, as we assume that $\mathcal{P}'$ has exactly k parts rather than at least k).

Theorem 4.4.7 (Theorem 2.9, [65]) *Suppose that $1/n \ll \mu \ll \gamma \ll 1/k$. Let $\mathcal{P}'$ partition a set V into parts $V_1, \ldots, V_k$ each of size n, and suppose that J is a $\mathcal{P}'$-partite k-complex on V with*

$$\delta^*(J) \geqslant \left(n, \left(\frac{k-1}{k} + \gamma\right) n, \left(\frac{k-2}{k} + \gamma\right) n, \ldots, \left(\frac{1}{k} + \gamma\right) n\right).$$

Then either J_k contains a perfect matching, or there exists a 'divisibility barrier' to such a matching. That is, there is some partition $\mathcal{P}$ of V into $d \leq k^2$ parts of size at least $\delta^*_{k-1}(J) - \mu n$ such that $\mathcal{P}$ refines $\mathcal{P}'$ and $L^{\mu}_{\mathcal{P}}(J_k)$ is incomplete with respect to $\mathcal{P}'$.

We will want to apply Theorem 4.4.7 to a k-partite k-graph H in which most (but not all) partite sets of size $k-1$ have sufficiently high degree. For this we will use the following lemma, which yields a k-complex J so that J_k is a subgraph of H and which will satisfy the minimum degree condition of Theorem 4.4.7 after the deletion of a small number of vertices (assuming D is sufficiently large).

Lemma 4.4.8 *Suppose that $1/n \ll \varepsilon \ll d, 1/k$ and let $\alpha = \varepsilon^{1/3k}$. Let $\mathcal{P}$ partition a vertex set V into parts $V_1, \ldots, V_k$ of size n, and let H be a $\mathcal{P}$-partite k-graph on V such that at most εn^{k-1} $\mathcal{P}$-partite sets $S \subseteq V$ of size $k-1$ have $d(S) < dn$. Then there exists a k-complex J on V such that $J_k \subseteq H$ and $\delta^*(J) \geqslant ((1-\sqrt{\varepsilon})n, (1-\alpha)n, \ldots, (1-\alpha)n, dn - \alpha n)$.*

Proof. Let $\beta = \alpha/k!$ and note that $\varepsilon \leqslant (\beta/k^k)^{2k}$. Call a $\mathcal{P}'$-partite set $A \subseteq V$ of size $k-1$ *bad* if $d_H(A) < dn$, and *good* otherwise. We define *bad* and *good* $\mathcal{P}'$-partite sets A of size i recursively for $1 < i < k$ by saying that A is bad if it there are more than βn vertices x in some part of $\mathcal{P}$ such that $A \cup \{x\}$ is bad, and good otherwise. By a trivial induction the number of bad i-sets is at most $\varepsilon k^{k^2}\binom{n}{k-i}/\beta^{i-1} \leqslant \sqrt{\varepsilon}\binom{n}{k-i}$ for every i. Call a vertex v good if $\{v\}$ is a good 1-set.

We now define J_1 to be the set of good vertices of H; note that $\delta_0(J) = |J_1| \geqslant (1-\sqrt{\varepsilon})n$. We define J_i recursively for $1 < i < k$ as the family of good i-sets A such that every $(i-1)$-subset of A is an element of J_{i-1}. We now prove by induction that $\delta^*_i(J) \geqslant (1-(i+1)!\beta)n \geqslant (1-\alpha)n$ for each $0 \leqslant i \leqslant k-2$. The case $i = 0$ is already done.

Suppose that $\delta^*_i(J) \geqslant (1-(i+1)!\beta)n$. Suppose for a contradiction that there exists a part $W \in \mathcal{P}$ and $e \in J_{i+1}$ with $|N_J(e) \cap W| < (1-(i+1)!\beta)n$. Then there exists a family $F \subseteq \binom{V}{i+2}$, such that $\mathbf{i}_{\mathcal{P}}(e') = \mathbf{i}_{\mathcal{P}}(e) + \mathbf{u}_W$ for each $e' \in F$, with $|F| > (i+2)!\beta n - i - 1 \geqslant (i+1)(i+1)!\beta n + \beta n$ and $F \cap J_{i+2} = \emptyset$. Since $e \in J_{i+1}$ we know that F contains at most βn bad sets, and hence more than $(i+1)(i+1)!\beta n$ good ones. Now for every good $e_i \in F$, there exists $e'_i \in \binom{V}{i+1} \setminus J_{i+1}$ such that $e'_i \subseteq e_i$. Further knowing e and e'_i uniquely determines $e_i = e \cup e'_i$, and so there is a family $F' \subseteq \binom{V}{i+1}$ such that $|F'| > (i+1)(i+1)!\beta n$ and each $e' \in F'$ intersects e in some set e^- of i vertices. Since there are only $i+1$ possible choices for e^-, there is some $(i-1)$-subset e^- of e, which is contained in more than $(i+1)!\beta n$ $i+1$-sets $e' \in \binom{V}{i+1} \setminus J_{i+1}$. But since $e \in J_{i+1}$ we have $e^- \in J_i$, and so $\delta^*_i(J) < (1-(i+1)!\beta)n$, contradicting our assumption.

Similarly $\delta^*_{k-1}(J) \geqslant dn - k!\beta n \geqslant dn - \alpha n$. □

4.4.4 Geometry

Given points $\mathbf{v}_1, \ldots, \mathbf{v}_r \in \mathbb{R}^d$, we define their *positive cone* as

$$PC(\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}) := \{\sum_{j \in [r]} \lambda_j \mathbf{v}_j : \lambda_1, \ldots, \lambda_r \geqslant 0\}.$$

The well-known Farkas' Lemma shows that any point $\mathbf{v}$ outside of the positive cone of these points can be separated from them by a separating hyperplane.

Theorem 4.4.9 (Farkas' Lemma) *Suppose $\mathbf{v} \in \mathbb{R}^d \setminus PC(Y)$ for some finite set $Y \subseteq \mathbb{R}^d$. Then there is some $\mathbf{a} \in \mathbb{R}^d$ such that $\mathbf{a} \cdot \mathbf{y} \geq 0$ for every $\mathbf{y} \in Y$ and $\mathbf{a} \cdot \mathbf{v} < 0$.*

We will apply this result, via the following proposition, to show that the index vector $\mathbf{i}_{\mathcal{P}}(V)$ of the k-graphs we consider must lie within the positive cone of the well-represented index vectors.

Proposition 4.4.10 *Let $\mathcal{P}'$ partition a set V into parts $V_1, \ldots, V_k$ each of size n, and let $\mathcal{P}$ be a partition refining $\mathcal{P}'$ into parts of size at least n/k. Also let I be a set of index vectors with respect to $\mathcal{P}$ such that for any $i \in [k]$ and parts $X_j \subseteq V_j$ of $\mathcal{P}$ for each $j \neq i$ there is some part $X_i \subseteq V_i$ of $\mathcal{P}$ with $\sum_{i \in [k]} \mathbf{u}_{X_i} \in I$. Then $\mathbf{i}_{\mathcal{P}}(V) \in PC(I)$.*

Proof. Suppose for a contradiction that $\mathbf{i}_{\mathcal{P}}(V) \notin PC(I)$. Then by Theorem 4.4.9 we may fix $\mathbf{a} \in \mathbb{R}^d$ such that $\mathbf{a} \cdot \mathbf{i} \geq 0$ for every $\mathbf{i} \in I$ and $\mathbf{a} \cdot \mathbf{i}_{\mathcal{P}}(V) < 0$. For each $i \in [k]$, let $X_i^1, \ldots, X_i^{b_i}$ be the parts of $\mathcal{P}$ which are subsets of V_i, and let $a_i^1, \ldots, a_i^{b_i}$ be the corresponding coordinates of $\mathbf{a}$, with the labels chosen so that $a_i^1 \leqslant \ldots \leqslant a_i^{b_i}$. Fix $i \in [k]$ for which $a_i^{b_i} - a_i^1$ is minimised, so in particular $a_i^{b_i} - a_i^1 \leqslant \frac{1}{k} \sum_{j \in [k]} a_j^{b_j} - a_j^1$. By assumption we may choose $\mathbf{i} \in I$ such that $\mathbf{i} = \mathbf{u}_{X_i^s} + \sum_{j \neq i} \mathbf{u}_{X_j^1}$ for some $s \in [b_i]$. Then

$$\begin{aligned} 0 > \mathbf{a} \cdot \mathbf{i}_{\mathcal{P}}(V) &\geqslant \sum_{j \in [k]} n a_j^1 + \frac{n}{k}(a_j^{b_j} - a_j^1) \\ &\geqslant n \left(\sum_{j \in [k]} a_j^1 + (a_i^{b_i} - a_i^1) \right) \geqslant n \mathbf{a} \cdot \mathbf{i} \geqslant 0, \end{aligned}$$

a contradiction. So $\mathbf{i}_{\mathcal{P}}(V) \in PC(I)$. $\square$

Finally, we shall need a slightly rephrased version of a simple proposition from [65]. Here $B^d(\mathbf{0}, r)$ is the closed ball of radius r, centred at $\mathbf{0}$.

Proposition 4.4.11 ([65], Proposition 4.8) *Suppose that $1/s \ll 1/r, 1/d$. Let $X \subseteq \mathbb{Z}^d \cap B^d(\mathbf{0}, r)$, and let L_X be the sublattice of $\mathbb{Z}^d$ generated by X. Then for any vector $\mathbf{x} \in L_X \cap B^d(\mathbf{0}, r)$ we may choose integers $a_{\mathbf{i}}$ with $|a_{\mathbf{i}}| \leqslant s$ for each $\mathbf{i} \in X$ such that $\mathbf{x} = \sum_{\mathbf{i} \in X} a_{\mathbf{i}} \mathbf{i}$.*

4.5 Robust maximality

Recall that if G is a k-graph on a vertex set V, and $\mathcal{Q}$ partitions V into d parts, then $L^{\mu}_{\mathcal{Q}}(G)$ is generated by those vectors $\mathbf{i} \in \mathbb{Z}^d$ for which at least μn^k edges of G have $\mathbf{i}(e) = \mathbf{i}$. So the constant μ can be viewed as the 'detection threshold'; it specifies the number of edges of a given index which are required for this index to contribute to $L^{\mu}_{\mathcal{Q}}(G)$.

The nature of a divisibility barrier is that $L^{\mu}_{\mathcal{Q}}(G)$ is $(1, -1)$-free. For our methods we will need the partition $\mathcal{Q}$ to be maximal with this property; that is, for any refinement $\mathcal{Q}'$ of $\mathcal{Q}$ we need $L^{\mu}_{\mathcal{Q}'}(G)$ to be $(1, -1)$-free. More then this, we need the same to hold even with a much higher detection threshold. This leads to the following key definition of a robustly maximal partition.

Definition 4.5.1 *Let H be a k-graph on a vertex set V of size n. We say that a partition $\mathcal{P}$ of V is (c, μ, μ')-robustly maximal with respect to H if* *(i)* $L^{\mu}_{\mathcal{P}}(H)$ *is* $(1,-1)$*-free, and*

(ii) for any partition $\mathcal{P}^+$ *of* V *which strictly refines* $\mathcal{P}$ *into parts of size at least* cn*, the lattice* $L^{\mu'}_{\mathcal{P}^+}(H)$ *is not* $(1,-1)$*-free.*

The next lemma allows us to obtain a robustly maximal partition of the vertex set of any k-graph. In addition, we may do this so as to refine an existing partition.

Proposition 4.5.2 *Let* $k \geqslant 2$ *be an integer and* $c > 0$ *be a constant. Let* $s = \lfloor 1/c \rfloor$ *and fix constants* $0 < \mu_1 < \cdots < \mu_{s+1}$*. Suppose that* H *is a* k*-graph on a vertex set* V *of size* n*, and* $\mathcal{P}$ *is a partition of* V *with parts of size at least* cn *such that* $L^{\mu_1}_{\mathcal{P}}(H)$ *is* $(1,-1)$*-free. Then there exists* $t \in [s]$ *and a partition* $\mathcal{P}'$ *of* V *with parts of size at least* cn *which refines* $\mathcal{P}$ *and is* (c, μ_t, μ_{t+1})*-robustly maximal with respect to* H*.*

Proof. Let $\mathcal{P}^{(1)} = \mathcal{P}$; then we proceed iteratively. At step t, if there is a partition $\mathcal{P}^+$ which strictly refines $\mathcal{P}^{(t)}$ into parts of size at least cn such that the lattice $L^{\mu_{t+1}}_{\mathcal{P}^+}(H)$ is $(1,-1)$-free, then let $\mathcal{P}^{(t+1)} = \mathcal{P}^+$. Otherwise terminate with output $\mathcal{P}^{(t)}$. By definition each $\mathcal{P}^{(t)}$ refines $\mathcal{P}$ and has parts of size at least cn. Furthermore, if the algorithm terminates at time t, then $L^{\mu_t}_{\mathcal{P}^{(t)}}$ is $(1,-1)$-free by choice of $\mathcal{P}^{(t)}$, but for any $\mathcal{P}^+$ which refines $\mathcal{P}^{(t)}$ into parts of size at least cn the lattice $L^{\mu_{t+1}}_{\mathcal{P}^+}(H)$ is not $(1,-1)$-free. That is, $\mathcal{P}^{(t)}$ is (c, μ_t, μ_{t+1})-robustly maximal with respect to H. So it remains only to show that this algorithm must terminate with $t \leqslant 1/c$; for this we observe that the number of parts of each $\mathcal{P}^{(i)}$ is strictly greater than the number of parts of $\mathcal{P}^{(i-1)}$, but $\mathcal{P}^{(t)}$ can have at most $1/c$ parts since each part has size at most cn. $\square$

It is intuitive that if $\mathcal{Q}$ is a robustly maximal partition of the vertex set of a k-graph G, and H is a k-graph which is close to G, then $\mathcal{Q}$ should also be a robustly maximal partition of the vertex set of H (allowing weaker parameters of robust maximality). This intuition is captured in the next lemma.

Proposition 4.5.3 *Suppose that $1/n_0 \ll \mu, \mu', \alpha \ll c, 1/k$. Let H be a k-graph on $n \geqslant n_0$ vertices with m edges, and let H' be a subgraph of H with at least $(1-\alpha)n$ vertices and at least $m - \alpha n^k$ edges. Suppose that $\mathcal{P}$ is a (c, μ, μ')-robustly maximal partition of $V(H)$ with respect to H, with parts of size at least cn. Then $\mathcal{P}$ is a $(c + 2\alpha, \mu + 2\alpha, \mu' - \alpha)$-robustly maximal partition of $V(H')$ with respect to H'.*

Proof. Let H'' be the k-graph with the vertex set of H but the edge set of H'. We first show that $I^{\mu+2\alpha}_{\mathcal{P}}(H') \subseteq I^{\mu+\alpha}_{\mathcal{P}}(H'') \subseteq I^{\mu}_{\mathcal{P}}(H)$. Indeed, since $|H \triangle H''| \leqslant \alpha n^k$ we have $I^{\mu+\alpha}_{\mathcal{P}}(H'') \subseteq I^{\mu}_{\mathcal{P}}(H)$ by Proposition 4.4.1. Since $(\mu + 2\alpha)|V(H')|^k \geqslant (\mu+2\alpha)((1-\alpha)n)^k \geqslant (\mu + \alpha)n^k$, we also have $I^{\mu+2\alpha}_{\mathcal{P}}(H') \subseteq I^{\mu+\alpha}_{\mathcal{P}}(H'')$. So $L^{\mu+2\alpha}_{\mathcal{P}}(H') \subseteq L^{\mu}_{\mathcal{P}}(H)$, and so $L^{\mu}_{\mathcal{P}}(H)$ being $(1, -1)$-free implies that $L^{\mu+2\alpha}_{\mathcal{P}}$ is $(1, -1)$-free.

Now suppose that there is some partition $\mathcal{P}^+$ of $V(H')$ which strictly refines $\mathcal{P}$ into parts of size at least $(c+2\alpha)|V(H')|$ such that the lattice $L^{\mu'-\alpha}_{\mathcal{P}^+}(H')$ is $(1, -1)$-free. Form a partition $\mathcal{P}^{++}$ of $V(H)$ by assigning the vertices of $V(H) \setminus V(H')$ to the parts of $\mathcal{P}^+$ arbitrarily. Similarly to above we prove $I^{\mu'-\alpha}_{\mathcal{P}^+}(H') \supseteq I^{\mu'-\alpha}_{\mathcal{P}^{++}}(H'') \supseteq I^{\mu'}_{\mathcal{P}^{++}}(H)$. Indeed, the second inequality holds by Proposition 4.4.1, and the first holds since $(\mu' - \alpha)|V(H')|^k \leqslant (\mu' - \alpha)n^k$. So $L^{\mu'}_{\mathcal{P}^{++}}(H) \subseteq L^{\mu'-\alpha}_{\mathcal{P}^+}(H')$, and so $\mathcal{P}^{++}$ is a partition of $V(H)$ which strictly refines $\mathcal{P}$ into parts of size at least $(c + 2\alpha)|V(H')| \geqslant cn$ such that the lattice $L^{\mu'}_{\mathcal{P}^{++}}(H)$ is $(1, -1)$-free, contradicting the robust maximality of H. We conclude that $\mathcal{P}$ is $(c + 2\alpha, \mu + 2\alpha, \mu' - \alpha)$-robustly maximal with respect to H'. $\square$

Recall that Proposition 4.4.3 showed that if $\mathcal{P}$ partitions a vertex set V and $\mathcal{Q}$ refines $\mathcal{P}$, then for any $\mathbf{i} \in L^{\mu}_{\mathcal{P}}(H)$ there is some $\mu' \ll \mu$ and $\mathbf{i}' \in L^{\mu'}_{\mathcal{Q}}(H)$ such that $(\mathbf{i}' \mid \mathcal{P}) = \mathbf{i}$. The next proposition shows that if $\mathcal{P}$ is robustly maximal with respect to H then we have a much stronger result, namely that $\mathbf{i}' \in L^{\mu'}_{\mathcal{Q}}(H)$ for *any* $\mathbf{i}'$ such that $(\mathbf{i}' \mid \mathcal{P}) = \mathbf{i}$.

Proposition 4.5.4 *Suppose that $1/n \ll \mu \ll \mu' \ll c, 1/k$. Let H be a k-graph on a vertex set V of size n, let $\mathcal{P}$ be a partition of V with parts of size at least cn which is*

(c, μ, μ')-robustly maximal with respect to H, and let $\mathcal{Q}$ be a partition of V which refines $\mathcal{P}$ into parts of size at least cn. Suppose that $\mathbf{i} \in L_{\mathcal{P}}^{\mu'}(H)$. Then $\mathbf{i}' \in L_{\mathcal{Q}}^{\mu}(H)$ for any index vector $\mathbf{i}'$ with respect to $\mathcal{Q}$ such that $(\mathbf{i}' \mid \mathcal{P}) = \mathbf{i}$.

Proof. Let p be the number of parts of $\mathcal{P}$. We prove the following statement by induction on q: if $\mathcal{Q}$ is a refinement of $\mathcal{P}$ into q parts of size at least cn and $\mathbf{i}'$ is an index vector with respect to $\mathcal{Q}$ such that $(\mathbf{i}' \mid \mathcal{P}) = \mathbf{i}$ then $\mathbf{i}' \in L_{\mathcal{Q}}^{\mu'/2^{q-p}}(H)$. The base case $q = p$ is trivial, since then we must have $\mathcal{Q} = \mathcal{P}$, so $\mathbf{i}' = \mathbf{i} \in L_{\mathcal{Q}}^{\mu'}(V)$ by assumption. Assume therefore that we have proved the proposition for any refinement $\mathcal{Q}'$ of $\mathcal{P}$ into $q-1$ parts, and that $\mathcal{Q}$ refines $\mathcal{P}$ into q parts. Since $\mathcal{P}$ is (c, μ, μ')-robustly maximal, there are parts $X, X' \in \mathcal{Q}$ such that $\mathbf{u}_X - \mathbf{u}_{X'} \in L_{\mathcal{Q}}^{\mu'}(H)$. Let $\mathcal{Q}'$ be formed by merging X and X' into a single part; then $\mathcal{Q}'$ is a refinement of $\mathcal{P}$ into $q-1$ parts. So $(\mathbf{i}' \mid \mathcal{Q}' \in L_{\mathcal{Q}'}^{\mu^*/2^{q-1-p}}(V)$ by our inductive hypothesis. By Proposition 4.4.3 there exists $\mathbf{i}^* \in L_{\mathcal{Q}}^{\mu'/2^{q-p}}(V)$ with $(\mathbf{i}^* \mid \mathcal{Q}') = (\mathbf{i}' \mid \mathcal{Q}')$. Now $\mathbf{i}^*$ differs from $\mathbf{i}'$ only in the co-ordinates corresponding to X and X'; since $\mathbf{u}_X - \mathbf{u}_{X'} \in L_{\mathcal{Q}}^{\mu'}(H) \subseteq L_{\mathcal{Q}}^{\mu'/2^{q-p}}(V)$ we have $\mathbf{i}' \in L_{\mathcal{Q}}^{\mu'/2^{q-p}}(V)$, completing the induction. Since $\mathcal{Q}$ as in the lemma can have at most $1/c$ parts, we conclude that $\mathbf{i}_{\mathcal{Q}}(V) \in L_{\mathcal{Q}}^{\mu'/2^{1/c}}(V) \subseteq L_{\mathcal{Q}}^{\mu}(V)$. □

Our goal for the rest of this section is to prove Lemma 4.5.6 that robust maximality is preserved by random selection. That is, if H is a k-graph on V and some partition $\mathcal{P}$ is robustly maximal with respect to H, then with high probability $\mathcal{P}$ is also robustly maximal (with slightly weaker parameters) with respect to the subgraph H' induced by a randomly chosen subset $V' \subset V$. It is easy to prove that the first part of the definition of robust maximality is preserved by such a random selection. Instead the difficulties arise with the second part, where we must consider all refinements of $\mathcal{P}$: there are too many possible refinements to simply take a union bound. We therefore make use of weak hypergraph regularity to prove this lemma. This necessitates the following definitions.

Suppose that $\mathcal{P}$ partitions a set V into r parts $V_1, \dots, V_r$ and G is a $\mathcal{P}$-partite k-graph on V. For $\varepsilon > 0$ and $A \in \binom{[r]}{k}$, we say that the k-partite sub-k-graph G_A is (ε, d)*-vertex-regular* if for any sets $V_i' \subseteq V_i$ with $|V_i'| \geq \varepsilon|V_i|$ for $i \in A$, writing $V' = \bigcup_{i \in A} V_i'$, we have $d(G_A[V']) = d \pm \varepsilon$. We say that G_A is ε*-vertex-regular* if it is (ε, d)-vertex-regular for some d, and a partition $\mathcal{P}$ of $V(G)$ is ε-regular if all but at most εn^k edges of G belong to ε-vertex-regular k-partite sub-k-graphs. The following lemma has essentially the same proof as that of the Szemerédi Regularity Lemma [110], namely iteratively refining a partition until an 'energy function' does not increase by much; in this case the energy function is the sum of the mean square densities of the k-graphs with respect to the partition.

Theorem 4.5.5 *(Weak Regularity Lemma) Suppose that $1/n \ll 1/m_0 \ll \varepsilon \ll 1/r \leq 1/k$. Suppose that G is a k-graph on n vertices and that $\mathcal{P}'$ is a partition of $V = V(G)$. Then there is a refinement $\mathcal{P}$ of $\mathcal{P}'$ into $m \leq m_0$ parts which is ε-regular in which each part has size $\lfloor n/m \rfloor$ or $\lceil n/m \rceil$.*

Lemma 4.5.6 *Suppose that $1/n_0 \ll \mu \ll \mu' \ll c \ll \lambda \leqslant 1/k$. Let H be a k-graph on vertex set V of size $n \geqslant n_0$, and let $\mathcal{P}$ be a partition of V which is (c, μ, μ')-robustly maximal with respect to H. Let $\mathcal{Q}^\cap$ be a partition with at most $1/\lambda$ parts which refines $\mathcal{P}$. Let $\mathbf{n}$ be an index vector with respect to $\mathcal{Q}^\cap$ such that $n_Z \geqslant \lambda|Z| - cn$ for $Z \in \mathcal{Q}^\cap$. Suppose that a set S_Z of size n_Z is chosen uniformly at random from each Z, with all choices being independent. Let $S = \bigcup_{Z \in \mathcal{Q}^\cap} S_Z$, and let $\widehat{\mathcal{P}}$ be the restriction of $\mathcal{P}$ to S. Then with probability $1 - 1/k^k$, $\widehat{\mathcal{P}}$ is $(2c, \mu/c, (\mu')^3)$-robustly maximal with respect to $H[S]$.*

Proof. In this proof we omit floor and ceiling symbols where they do not affect the argument. Introduce a new constant ε with $1/n_0 \ll \varepsilon \ll \mu$. We apply Theorem 4.5.5 to H to obtain an integer $n_R \geqslant 3k^{k+1}/c$, and an ε-regular partition $\mathcal{Q}$ of $V(H)$ which refines $\mathcal{P}$ into n_R parts of almost equal size. Let $n_0 = n/n_R$. Let R be the $\mu'/3$-reduced graph on $\mathcal{Q}$ - that is, R contains an edge between any parts $X_1, \dots, X_k$ of $\mathcal{Q}$ which form an (ε, d)-regular tuple for some $d \geqslant \mu'/3$.

Let $H' = H[S]$. We first show that $L^{\mu/c}_{\widehat{\mathcal{P}}}(H')$ is $(1,-1)$-free. It suffices to prove that $I^{\mu/c}_{\widehat{\mathcal{P}}}(H') \subseteq I^{\mu}_{\mathcal{P}}(H)$. Note that $n' := \sum_{Z \in \mathcal{Q}^\cap} n_Z \geqslant \lambda n - cn/\lambda \geqslant \lambda n/2$. If $\mathbf{i} \in I^{\mu/c}_{\widehat{\mathcal{P}}}(H')$ then there are at least $(\mu/c)(\lambda n/2)^k \geqslant (2^k\mu/\lambda^k)(\lambda n/2)^k = \mu n^k$ edges $e \in H'$ with $\mathbf{i}_{\widehat{\mathcal{P}}}(e) = \mathbf{i}$, where the inequality holds since $c \ll \lambda$. But all of these edges are also edges of H and hence $\mathbf{i} \in I^{\mu}_{\mathcal{P}}(H)$. Further, each part $\widehat{X} = X \cap S$ of $\widehat{\mathcal{P}}$ (where X is the corresponding part of $\mathcal{P}$) has size $\sum_Z n_Z \geqslant \sum_Z(\lambda|Z| - cn) \geqslant \lambda|X| - cn/\lambda \geqslant 3cn$, where the sum is taken over all $Z \in \mathcal{Q}^\cap$ with $Z \subseteq X$ and the last inequality holds since $c \ll \lambda$.

Let $\mathcal{P}_R$ be the partition of $V(R) = \mathcal{Q}$ induced by $\mathcal{P}$ (i.e., a part of $\mathcal{P}_R$ contains of all parts of $\mathcal{Q}$ which are subsets of a given part of $\mathcal{P}$). For any $Y \in \mathcal{Q}$, note that $|Y \cap S|$ is a sum of independent hypergeometric random variables and that

$$\begin{aligned}\mathbb{E}[|Y \cap S|] &= \sum_{Z \in \mathcal{Q}^\cap} \mathbb{E}[|Y \cap Z \cap S|] = \sum_{Z \in \mathcal{Q}^\cap} n_Z|Y \cap Z|/|Z| \\ &\geqslant \sum_{Z \in \mathcal{Q}^\cap} \lambda|Y \cap Z| - \sum_{Z \in \mathcal{Q}^\cap} cn|Y \cap Z|/|Z| \geqslant \lambda|Y| - cn/\lambda \qquad \geqslant 2cn_0,\end{aligned}$$

where the last inequality follows from $c \ll \lambda$. Hence Corollary 1.8.2 implies that $\mathbb{P}[|Y \cap S| < cn_0] < 2\exp[-cn_0/12] \leqslant 1/n^2$, and a union bound implies that with high probability $|Y \cap S| \geqslant cn_0$ for every $Y \in \mathcal{Q}$.

Now suppose for a contradiction that there is a refinement $\widehat{\mathcal{P}}^*$ of $\widehat{\mathcal{P}}$ with parts of size at least $2cn$, such that $L^{(\mu')^3}_{\widehat{\mathcal{P}}^*}(H')$ is $(1,-1)$-free. Define a partition $\mathcal{S} = \{X_{\mathcal{S}} \mid X \in \widehat{\mathcal{P}}^*\}$ of $V(R)$, whose parts correspond to those of $\widehat{\mathcal{P}}^*$, by assigning clusters in $V(R)$ as follows: For each $Y \in V(R)$, assign Y to a part $X_{\mathcal{S}}$ such that $|S \cap Y \cap X| \geqslant c|S \cap Y|$ with probability proportional to $|S \cap Y \cap X|$, all choices being made independently. (So if either $|S \cap Y \cap X| \geqslant c|S \cap Y|$ or $S \cap Y \cap X = \emptyset$ for every X then the probability is $|S \cap Y \cap X|/|S \cap Y|$.)

Note that

$$\sum_{Y \in V(R)} \frac{|S \cap Y \cap X|}{|S \cap Y|} \geqslant \sum_{Y \in V(R)} \frac{|S \cap Y \cap X|}{n_0} \geqslant \frac{3cn}{n_0} = 3cn_R$$

and hence

$$\begin{aligned}
\mathbb{E}[|X_{\mathcal{S}}|] = \sum_{Y \in V(R)} \mathbb{P}[Y \in X_R] &\geqslant \sum_{Y \in V(R), |S \cap Y \cap X| \geqslant c|S \cap Y|} \frac{|S \cap Y \cap X|}{|S \cap Y|} \\
&= \sum_{Y \in V(R)} \frac{|S \cap Y \cap X|}{|S \cap Y|} - \sum_{Y \in V(R), |S \cap Y \cap X| < c|S \cap Y|} \frac{|S \cap Y \cap X|}{|S \cap Y|} \\
&\geqslant 3cn_R - cn_R \geqslant 2cn_R
\end{aligned}$$

for each $X_{\mathcal{S}} \in \mathcal{S}$. Now let N_Y be the random variable which is 1 if $Y \in X_{\mathcal{S}}$ and 0 otherwise for each $Y \in V(R)$. We have $|X_{\mathcal{S}}| = \sum_{Y \in V(R)} N_Y$. By applying Lemma 1.8.1(i) to the variables N_Y we obtain

$$\mathbb{P}[|X_{\mathcal{S}}| < cn_R] \leqslant \exp\left[-\frac{2(cn_R)^2}{n_R}\right] \leqslant \exp[-2c^2 n_R] \leqslant 1/n_R \leqslant c/3k^{k+1},$$

and taking a union bound on the error probabilities, we obtain that with probability at least $1 - 1/3k^k$ each part $X_{\mathcal{S}} \in \mathcal{S}$ has size at least cn_R. Note that whenever $Y \in X_{\mathcal{S}}$ we have $|S \cap X \cap Y| \geqslant c|S \cap Y| \geqslant c^2 n_0 \geqslant \varepsilon n_0$. For each $\mathbf{i} \in I_{\mathcal{S}}^{\mu'/3}(R)$ there are at least $\mu' n_R^k/3$ edges $e \in R$ with $\mathbf{i}_{\mathcal{S}}(e) = \mathbf{i}$. For each such e, order the clusters in e as $(Y_1, \dots, Y_k)$, let $(X_j)_{\mathcal{S}}$ be the part of $\mathcal{S}$ containing Y_j for each $j \in [k]$ and let F be the k-partite subgraph of H with vertex classes $(S \cap X_j \cap Y_j)_{j \in [k]}$. Now every edge of F has index vector $\mathbf{i}$ with respect to $\widehat{\mathcal{P}}^*$ and $|F| \geqslant (\mu'/3 - \varepsilon)(c^2 n_0)^k \geqslant 3(\mu')^2(n'/n_R)^k$, where the second inequality follows from $\mu' \ll c$. Hence there are at least $(\mu')^3(n')^k$ edges $e' \in H'$ with $\mathbf{i}_{\widehat{\mathcal{P}}^*}(e') = \mathbf{i}$. Thus $I_{\mathcal{S}}^{\mu'/3}(R) \subseteq I_{\widehat{\mathcal{P}}^*}^{(\mu')^3}(H')$ and so $L_{\mathcal{S}}^{\mu'/3}(R)$ is $(1,-1)$-free.

Define a refinement $\mathcal{P}^*$ of $\mathcal{P}$ by $\mathcal{P}^* = \{\bigcup_{Y \in X} Y \mid X \in \mathcal{S}\}$. Note that each part of $\mathcal{P}^*$

has size at least $cn_R n_0 = cn$. For any $\mathbf{i} \in I^{\mu'/3}_{\mathcal{P}^*}(H)$ there are at most $\varepsilon n^k + n_R^k \cdot (\mu'/3) n_0^k \leqslant \mu' n^k/2$ edges $e \in H$ with $\mathbf{i}_{\mathcal{P}}(e) = \mathbf{i}$ which do not lie in k-graphs corresponding to edges of R. Hence there are at least $\mu' n^k/2$ edges which do lie in such k-graphs, and so there are at least $\mu' n^k / 2n_0^k = \mu' n_R^k/2$ edges $e \in R$ such that $\mathbf{i}_{\mathcal{P}_R}(e) = \mathbf{i}$. Hence $\mathbf{i} \in I^{\mu'/3}_{\mathcal{P}_R}$, and so $I^{\mu'}_{\mathcal{P}^*}(H) \subseteq I^{\mu'/3}_{\mathcal{S}}(R)$ and $L^{\mu'}_{\mathcal{P}^*}(H) \subseteq L^{\mu'/3}_{\mathcal{P}_R}(R)$. But now $L^{\mu'}_{\mathcal{P}^*}(H)$ is $(1,-1)$-free, which contradicts our assumption that $\mathcal{P}$ is (c, μ, μ')-robustly maximal. □

4.6 The key lemmas and their proofs

The goal of this section is to state and prove the two key lemmas in the proof of Theorem 1.5.2, Lemma 4.6.1, which applies to k-partite k-graphs, and Lemma 4.6.8, which applies to general k-graphs. The k-partite form is simpler to prove, while in fact we deduce the general form from the k-partite form by a randomisation argument. Therefore we first present the proof of Lemma 4.6.1.

Lemma 4.6.1 *Let k and ℓ be integers with $k \geqslant \ell \geqslant 2$, and suppose that constants $n, \varepsilon, \mu, \mu', c, d, \gamma$ satisfy $1/n \ll \varepsilon, \mu \ll \mu' \ll c, d \ll \gamma, 1/k$. Let $\mathcal{P}'$ partition a set V into vertex classes $V_1, \dots, V_k$ of size n, and let H be a $\mathcal{P}'$-partite k-graph on V in which at most εn^{k-1} $\mathcal{P}'$-partite subsets $S \subseteq V$ of size $k-1$ have $d(S) < (1/\ell + \gamma)n$.*

Let $\mathcal{P}$ be a partition of V which refines $\mathcal{P}'$ into parts of size at least cn and is (c, μ, μ')-robustly maximal with respect to H. Suppose that

(i) for any vertex $x \in V$ there are at least dn^{k-1} edges $e \in H$ with $x \in e$ and $\mathbf{i}_{\mathcal{P}}(e) \in L^{\mu}_{\mathcal{P}}(H)$, and

(ii) $\mathbf{i}_{\mathcal{P}}(V) \in L^{\mu}_{\mathcal{P}}(H)$.

Then H contains a perfect matching.

Proof. We fix k and proceed by induction on ℓ, using the following hierarchy of constants.

$$1/n \ll 1/D \ll \varepsilon, \mu \ll \varepsilon', \mu'_0 \ll \mu'_1 \ll \dots \mu'_{k^{k^k}} \ll \mu' \ll c, d \ll \eta \ll \gamma, 1/k.$$

Step 1: The case $\mathcal{P} = \mathcal{P}'$. Suppose first that $\mathcal{P} = \mathcal{P}'$. By applying Lemma 4.4.8 with $\mathcal{P}'$ and $(1/\ell + \gamma)n$ in place of $\mathcal{P}$ and D respectively we obtain a $\mathcal{P}'$-partite k-complex J on V with $J_k \subseteq H$ and

$$\delta^*(J) \geqslant \left((1-\sqrt{\varepsilon})n, (1-\varepsilon')n, \dots, (1-\varepsilon')n, (1/\ell + \gamma - \varepsilon')n\right).$$

We now greedily construct a matching M^* in H of size at most $k\sqrt{\varepsilon}n$ which includes all of the at most $k\sqrt{\varepsilon}n$ vertices $x \in V$ for which $\{x\}$ is not an edge of J. We can do this greedily since by assumption every vertex lies in at least $dn^{k-1} \geqslant k^2\sqrt{\varepsilon}n^{k-1}$ edges of H. Let $V' = V \setminus V(M^*)$ and let $J' := J[V']$. Then

$$\delta^*(J') \geqslant (n', (1-2\varepsilon')n', \dots, (1-2\varepsilon')n', (1/\ell + \gamma - 2\varepsilon')n'),$$

where $n' := n - |M^*|$ is the size of each part of V'. So J' satisfies the degree sequence condition of Theorem 4.4.7 with $n', \mu/2$ and $\gamma/2$ in place of n, μ and γ respectively. Furthermore $\mathcal{P}$ is $(c, 2\mu, \mu'/2)$-robustly maximal with respect to J'_k by Lemma 4.5.3.

Now suppose that $\mathcal{P}''$ is a partition of V' into parts of size at least $\delta^*_{k-1}(J) - \mu n/2 \geqslant n/\ell \geqslant cn$ which refines $\mathcal{P} = \mathcal{P}'$. Since J'_k is $\mathcal{P}$-partite every edge $e \in J'_k$ has $\mathbf{i}_{\mathcal{P}}(e) = \mathbf{1}$; in particular we have $\mathbf{1} \in L^{\mu'/2}_{\mathcal{P}}(J'_k)$. So by Proposition 4.5.4 we have $\mathbf{i} \in L^{2\mu}_{\mathcal{P}''}(J'_k)$ for any index vector $\mathbf{i}$ with respect to $\mathcal{P}''$ such that $(\mathbf{i} \mid \mathcal{P}) = \mathbf{1}$. That is, $L^{2\mu}_{\mathcal{P}''}(J'_k)$ is complete with respect to $\mathcal{P}$. We conclude that there is no 'divisibility barrier' to a perfect matching in J'_k, so there exists a perfect matching in J'_k by Theorem 4.4.7. Together with M^* this forms a perfect matching in H.

In particular, this gives the base case $\ell = 2$ of the induction, since in this case the condition that each part of $\mathcal{P}$ must be of size at least $n/\ell + \gamma n > n/2$ implies that we must have $\mathcal{P} = \mathcal{P}'$.

Step 2: Defining subpartitions and removing bad vertices. For the rest of the proof we assume that $\mathcal{P} \neq \mathcal{P}'$, that $3 \leqslant \ell \leqslant k$ and that the lemma holds with $\ell - 1$ in place of ℓ. We also write $I := I^\mu_{\mathcal{P}}(H)$, $L := L^\mu_{\mathcal{P}}(H)$ and $\mathrm{Aux} = \mathrm{Aux}^\mu_{\mathcal{P}}(H)$. Recall that we then have $I(\mathrm{Aux}) = I$ and $L(\mathrm{Aux}) = L$, and that $\mathcal{P}'_{\mathrm{Aux}}$ is the natural partition of $V(\mathrm{Aux})$ inherited from $\mathcal{P}'$

Our strategy will be to split H up randomly into a number of vertex-disjoint k-partite subgraphs, each of which satisfies the conditions of the lemma (with weaker constants) when ℓ is replaced by $\ell - 1$. The inductive hypothesis then implies the existence of a perfect matching in each subhypergraph, and taking the union of these matchings gives a perfect matching in H.

First, note that H, $\mathcal{P}'$ and $\mathcal{P}$ meet the conditions of Proposition 4.4.5 with $1/\ell + \gamma$ in place of d. So by Proposition 4.4.5(ii) we may fix r such that each part V_j of $\mathcal{P}'$ is further partitioned into r parts by $\mathcal{P}$, so the k parts of $\mathcal{P}_{\mathrm{Aux}}$ each consist of r vertices. Next Proposition 4.4.5(i) tells us that $d(S) = 1$ for any $\mathcal{P}'_{\mathrm{Aux}}$-partite set of $k-1$ vertices of Aux. Also, by Proposition 4.4.5(iii) each part of $\mathcal{P}$ is indexed by precisely r^{k-2} vectors $\mathbf{i} \in I$, so in particular we have $|I| = r^{k-1}$. The final statement of Proposition 4.4.5 implies that every part of $\mathcal{P}$ has size at least $(1/\ell + \gamma/2)n$, a significantly stronger bound than that assumed in the statement of the lemma. Together with $\mathcal{P} \neq \mathcal{P}'$ this implies that $2 \leqslant r < \ell$.

For each $\mathbf{i} \in I$, let the set $S_{\mathbf{i}}$ be the union of the parts $\mathcal{P}$ indexed by $\mathbf{i}$, and let $\mathcal{P}_{\mathbf{i}}$ be the restriction of $\mathcal{P}'$ to $S_{\mathbf{i}}$. Define $H_{\mathbf{i}}$ to be the $\mathcal{P}_{\mathbf{i}}$-partite k-graph on vertex set $S_{\mathbf{i}}$ whose edge set consists of all edges of H with index $\mathbf{i}$. By repeated application of Proposition 4.5.2

there exist $i \in [s]$ and a partition $\mathcal{Q}_{\mathbf{i}}$ of $S_{\mathbf{i}}$ for each $\mathbf{i} \in I$ such that the partition $\mathcal{Q}_{\mathbf{i}}$ refines $\mathcal{P}_{\mathbf{i}}$, has parts of size at least $cn/2$ and is $(c/2, \mu'_i, \mu'_{i+1})$-robustly maximal with respect to $H_{\mathbf{i}}$. For simplicity of notation we now relabel μ'_i as μ_0 and μ'_{i+1} as μ_9. Since $\mu, \varepsilon \ll \mu'_1 \leqslant \mu_0 = \mu'_i \ll \mu_9 = \mu'_{i+1} \ll \mu'$ we may introduce new constants $\mu_1, \dots, \mu_8$ such that

$$\mu, \varepsilon \ll \mu_0 \ll \mu_1 \ll \cdots \ll \mu_9 \ll \mu'.$$

Also, it will sometimes be convenient to regard $H_{\mathbf{i}}$ as a k-graph on V rather than on $S_{\mathbf{i}}$ (but with the same edge set). For this purpose define $\mathcal{Q}^0_{\mathbf{i}}$ to be the partition of V whose parts are the parts of $\mathcal{Q}_{\mathbf{i}}$ and the parts of $\mathcal{P}$ not indexed by $\mathbf{i}$.

Fix some $\mathbf{i} \in I$, and let $I_{\mathbf{i}} := I^{\mu_0}_{\mathcal{Q}_{\mathbf{i}}}(H_{\mathbf{i}})$ and $L_{\mathbf{i}} := L^{\mu_0}_{\mathcal{Q}_{\mathbf{i}}}(H_{\mathbf{i}})$. The fact that $\mathcal{Q}_{\mathbf{i}}$ is $(c/2, \mu_0, \mu_9)$-robustly maximal with respect to $H_{\mathbf{i}}$ implies that $L_{\mathbf{i}}$ is $(1, -1)$-free. Furthermore, by applying Proposition 4.4.5(v) with μ_0 in place of ψ we find that at most $\mu_0 n^{k-1}$ $\mathcal{P}'$-partite $(k-1)$-sets $S \subseteq S_{\mathbf{i}}$ of $k-1$ vertices have $d_{H_{\mathbf{i}}}(S) < n/\ell + 3\gamma n/4$. So we conclude that $H_{\mathbf{i}}$ also meets the conditions of Proposition 4.4.5 with $\mathcal{P}_{\mathbf{i}}$ in place of $\mathcal{P}'$, $\mathcal{Q}_{\mathbf{i}}$ in place of $\mathcal{P}$, $1/\ell + 3\gamma/4$ in place of d and μ_0 in place of both μ and ε. Every part of $\mathcal{Q}_{\mathbf{i}}$ therefore has size at least $n/\ell + 2\gamma n/3$; in particular we may apply Proposition 4.4.5 with $1/\ell$ in place of c also. By Proposition 4.4.5(iv) we deduce that $I^{1/2\ell^k}_{\mathcal{Q}_{\mathbf{i}}}(H_{\mathbf{i}}) = I_{\mathbf{i}}$. Also, Proposition 4.4.5(vi) with μ_1/r^{k-1} in place of ψ implies that at most $\mu_1 n/r^{k-1}$ vertices of $S_{\mathbf{i}}$ lie in fewer than $n^{k-1}/2\ell^{k-1} \geqslant 5dn^{k-1}$ edges $e \in H_{\mathbf{i}}$ with $\mathbf{i}(e) \in I_{\mathbf{i}}$. We refer to such vertices as being *bad for* $\mathbf{i}$. Since $\mathbf{i}$ was arbitrary the comments of this paragraph apply to any $\mathbf{i} \in I$; we say that a vertex of V is *bad* if it is bad for some $\mathbf{i} \in I$. Since $|I| = r^{k-1}$, the number of bad vertices is at most $\mu_1 n$.

We will ensure that an analogue of condition (i) holds for each $H_{\mathbf{i}}$ by deleting all bad vertices. Since by assumption every vertex of V is contained in at least $dn^{k-1} \geqslant k\mu_1 n^{k-1}$ edges $e \in H$ with $\mathbf{i}_{\mathcal{P}}(e) \in I$, we may greedily choose a matching M of size at most $\mu_1 n$ which covers all of the bad vertices and such that $\mathbf{i}_{\mathcal{P}}(e) \in I$ for each $e \in M$. Let

$V' := V \setminus V(M')$.

Next we wish to define a partition $\mathcal{Q}^\cap$ of V' to be the 'common refinement' of the partitions $\mathcal{Q}_\mathbf{i}$ for $\mathbf{i} \in I$, so that two vertices will be in the same part of $\mathcal{Q}^\cup$ if and only if they are in the same part of $\mathcal{Q}_\mathbf{i}$ for every $\mathbf{i} \in I$. We will also ensure that every part of $\mathcal{Q}^\cap$ is not too small by first deleting some further vertices. Let $\sim$ be the relation on V' in which $x \sim y$ if and only if x and y lie in the same part of $\mathcal{Q}^0_\mathbf{i}$ for *every* $\mathbf{i} \in I$, and observe that $\sim$ is an equivalence relation. Since $\mathcal{P}$ has at most k^2 parts, each of which is partitioned into at most k parts by each $\mathcal{Q}_\mathbf{i}$, the relation $\sim$ has at most $K := k^2 \cdot k^{|I|} = k^{2+r^{k-1}}$ equivalence classes. This means that there is some $j \in [K+1]$ such that no part of $\mathcal{Q}^\cap$ has size between $(j-1)(kK)^{j-1}D/\eta$ and $j(kK)^jD/\eta$. Fix such a j, and let B be the union of all equivalence classes of $\sim$ of size at most $j(kK)^jD/\eta$. Then $|B| \leqslant K \cdot (j-1)(kK)^{j-1}D/\eta$. Similarly as before we may greedily choose a matching M' in H of size at most $|B|$ which covers every vertex of $|B|$ so that every edge $e \in M'$ has $\mathbf{i}(e) \in I$.

We now delete all of the at most $2k\mu_1 n$ vertices covered by $M \cup M'$. To avoid using more complicated notation, we continue to write V for $V \setminus (M \cup M')$, $S_\mathbf{i}$ for $S_\mathbf{i} \setminus (M \cup M')$, V_j for $V_j \setminus (M \cup M')$ and H and $H_\mathbf{i}$ for the restrictions of these k-graphs to the undeleted vertices. Following these deletions we define $\mathcal{Q}^\cap$ to be the partition of V generated by the equivalence relation $\sim$ (now restricted to the undeleted vertices). We will also need to consider the 'common coarsening' $\mathcal{Q}^\cup$ of the partitions $\mathcal{Q}_\mathbf{i}$. Indeed, let $\sim_\cup$ be the relation on V defined by $x \sim_\cup y$ if and only if there is some $\mathbf{i} \in I$ such that $x, y \in S_\mathbf{i}$ and x and y lie in the same part of $\mathcal{Q}_\mathbf{i}$. Then $\sim_\cup$ may not be an equivalence relation, but the transitive closure $\sim^*_\cup$ of $\sim_\cup$ is an equivalence relation. Let $\mathcal{Q}^\cup$ be the partition of V generated by $\sim^*_\cup$. For clarity we now state together the properties of the structures we have obtained so far.

(A1) $\mathcal{P}' \prec \mathcal{P} \prec \mathcal{Q}^\cup \prec \mathcal{Q}^0_\mathbf{i} \prec \mathcal{Q}^\cap$ for any $\mathbf{i} \in I$. Indeed, $\mathcal{P}' \prec \mathcal{P}$ by assumption, and the other relations follow from the definitions of $\mathcal{Q}^0_\mathbf{i}$, $\mathcal{Q}^\cup$ and $\mathcal{Q}^\cap$.

(A2) $\mathcal{Q}^\cap$ has at most K parts, each of which has size at least $2D/\eta$. Indeed, the vertices of any equivalence class of $\sim$ with size less than $j(kK)^j D/\eta$ were included in M' before defining $\mathcal{Q}^\cap$; since $|M'| \leqslant |B|$ any part of $\mathcal{Q}^\cap$ has size at least

$$\begin{aligned} \frac{j(kK)^j D}{\eta} - k|B| &\geqslant \frac{j(kK)^j D}{\eta} - \frac{(j-1)(kK)^j D}{\eta} \\ &= \frac{(kK)^j D}{\eta} \geqslant \frac{2D}{\eta}. \end{aligned}$$

(A3) Every part of $\mathcal{P}$, $\mathcal{Q}^\cup$ and each $\mathcal{Q}_\mathbf{i}$ has size at least $(1/\ell + \gamma/2)n$. Indeed, each part of any $\mathcal{Q}_\mathbf{i}$ had size at least $(1/\ell + 2\gamma/3)n$ before the deletions, and at most $2k\mu_1 n \leqslant \gamma n/6$ vertices were deleted. The bounds for $\mathcal{Q}^\cap$ and $\mathcal{P}$ follow by (A1).

(A4) For any $\mathbf{i} \in I$, $\mathcal{Q}_\mathbf{i}$ is (c, μ_2, μ_8)-robustly maximal with respect to $H_\mathbf{i}$, and every vertex in $S_\mathbf{i}$ lies in at least $4dn^{k-1}$ edges of $H_\mathbf{i}$ with $\mathbf{i} \in I_\mathbf{i}$. Indeed, since $\mathcal{Q}_\mathbf{i}$ was chosen to be $(c/2, \mu_0, \mu_9)$-robustly maximal with respect to $H_\mathbf{i}$ (before any vertices were deleted) the first statement holds by Proposition 4.5.3. The second part holds since M covered all vertices which were contained in fewer than $5dn^{k-1}$ edges of $H_\mathbf{i}$ with $\mathbf{i} \in I_\mathbf{i}$ and at most $2k\mu_1 n \leqslant dn$ vertices were deleted.

(A5) $\mathbf{i}_\mathcal{P}(V) \in L$. Indeed, this was true before any vertices were deleted by assumption, and $\mathbf{i}_\mathcal{P}(V(M \cup M')) \in L$ since this was true of each edge of M.

(A6) $I^{\mu'}_{\mathcal{Q}_\mathbf{i}}(H_\mathbf{i}) = I_\mathbf{i}$. Indeed, we observed before any deletions that $I^{1/2\ell^k}_{\mathcal{Q}_\mathbf{i}}(H_\mathbf{i}) = I^{\mu_0}_{\mathcal{Q}_\mathbf{i}}(H_\mathbf{i}) = I_\mathbf{i}$. So if $\mathbf{i} \in I_\mathbf{i}$ then there were at least $(kn)^k/2\ell^k$ edges of $H_\mathbf{i}$ of index $\mathbf{i}$, at most $2k\mu_1 n^k$ of which were deleted. If $\mathbf{i} \notin I_\mathbf{i}$ then there were at most $\mu_0(kn)^k < \mu'(kn - 2k\mu_1 n)^k$ edges of $H_\mathbf{i}$ of index $\mathbf{i}$.

(A7) If V' is formed from V by deleting another $A \leqslant \gamma n/2$ vertices from each part of $\mathcal{P}$, then we have $\mathbf{i}_\mathcal{P}(V') \in PC(I)$. Indeed, let V'_j be the undeleted vertices from vertex class V_j; then each V_j' has equal size $n' := n - |M \cup M'| - A$, and from (A3) we find that at least n'/k vertices remain undeleted in each part of $\mathcal{P}$. Recall our earlier observations that $d_{\mathrm{Aux}}(S) = 1$ for any $\mathcal{P}_{\mathrm{Aux}}$-partite set S of $k-1$ vertices of Aux and $I(\mathrm{Aux}) = I$; using these we may apply Proposition 4.4.10 to deduce that $\mathbf{i}_{\mathcal{P}}(V') \in PC(I)$.

(A8) For any $\mathbf{i} \in I$ and vector $\mathbf{n}$ with respect to $\mathcal{Q}_{\mathbf{i}}$ such that $(\mathbf{n} \mid \mathcal{P}_{\mathbf{i}})$ is an integer multiple of $\mathbf{1}$, there exist parts X, X' of $\mathcal{Q}_{\mathbf{i}}$ so that $\mathbf{n}_{\mathbf{i}} - \mathbf{u}_X + \mathbf{u}_{X'} \in L_{\mathbf{i}}$. Indeed, this follows by applying Proposition 4.4.5(vii) to $H_{\mathbf{i}}$ before the deletion of any vertices.

(A9) The number of $\mathcal{P}_{\mathbf{i}}$-partite sets $S \subseteq S_{\mathbf{i}}$ of size $k-1$ with $d_{H_{\mathbf{i}}}(S) < (1/\ell + \gamma/2)n$ is at most $\mu_1 n^{k-1}$. Indeed, before any deletions this was true with $2\gamma/3$ in place of $\gamma/2$ by Proposition 4.4.5(v), and at most $2k\mu_1 n \leqslant \gamma n/6$ vertices were deleted in total.

Step 3: Determining the size of each random subgraph. In Step 4, the final step of the proof, we will randomly choose, for each $\mathbf{i} \in I$, a set $T_{\mathbf{i}}$ containing $\rho_{\mathbf{i}}$ vertices from each part of $\mathcal{P}_{\mathbf{i}}$ (the restriction of $\mathcal{P}'$ to $S_{\mathbf{i}}$). Claim 4.6.7 will then show that if $T_{\mathbf{i}}$ is chosen in this manner then the restriction $H[T_{\mathbf{i}}]$ meets all the conditions of the statement of the lemma, with weaker constants and $\ell - 1$ in place of ℓ, except for the requirement that $\mathbf{i}_{\mathcal{Q}_{\mathbf{i}}}(T_{\mathbf{i}}) \in L^{\mu_*}_{\mathcal{Q}_{\mathbf{i}}}(H[T_{\mathbf{i}}])$ (for an appropriate value of μ_*). In this step we determine how many vertices of $T_{\mathbf{i}}$ should be chosen from each part of $\mathcal{Q}^{\cap}$ so that firstly, this final requirement will also be satisfied for each $\mathbf{i} \in I$, and secondly, the sets $T_{\mathbf{i}}$ for $\mathbf{i} \in I$ partition V. This will allow us to apply the inductive hypothesis in Step 4 to obtain a perfect matching in each $H[T_{\mathbf{i}}]$, completing the proof.

To determine how many vertices of $T_{\mathbf{i}}$ should be chosen from each part of $\mathcal{Q}^{\cap}$, we proceed through three claims. The first determines the value $\rho_{\mathbf{i}}$, namely how many vertices are to be chosen from each part of $\mathcal{P}_{\mathbf{i}}$. The second determines how many vertices are to be chosen from each part of $\mathcal{Q}^{\cup}$, whilst the third determines how many vertices are to be chosen from $\mathcal{Q}^\cap$, as required.

Claim 4.6.2 *There exist integers $\rho_{\mathbf{i}}$ for each $\mathbf{i} \in I$ which satisfy*

(B1) $\sum_{\mathbf{i}\in I} \rho_{\mathbf{i}}\mathbf{i} = \mathbf{i}_{\mathcal{P}}(V)$,

(B2) $\rho_{\mathbf{i}} \geqslant \eta n$ *for each* $\mathbf{i} \in I$.

Proof. Let $N := \lceil 2\eta n \rceil$, and for each $\mathbf{i} \in I$ delete N vertices from the parts indexed by $\mathbf{i}$. Recall that each part of $\mathcal{P}$ is indexed by precisely r^{k-2} members of I, so exactly $r^{k-2}N \leqslant \gamma n/2$ vertices are deleted from each part of $\mathcal{P}$. Let V' consist of all undeleted vertices of V; then by (A7) we have $\mathbf{i}_{\mathcal{P}}(V') \in PC(I)$. That is, we may fix reals $\lambda_{\mathbf{i}} \geqslant 0$ for each $\mathbf{i} \in I$ so that $\sum_{\mathbf{i}\in I} \lambda_{\mathbf{i}}\mathbf{i} = \mathbf{i}_{\mathcal{P}}(V')$. Let $a_{\mathbf{i}} = \lfloor \lambda_{\mathbf{i}} \rfloor$ for each $\mathbf{i} \in I$. Then $\mathbf{i}_{\mathcal{P}}(V') - \sum_{\mathbf{i}\in I} a_{\mathbf{i}}\mathbf{i} \in B(\mathbf{0}, kr^{k-1})$. Since $\mathbf{i}_{\mathcal{P}}(V')$ was obtained by subtracting a linear combination of vectors $\mathbf{i} \in I$ from $\mathbf{i}_{\mathcal{P}}(V)$, and $\mathbf{i}_{\mathcal{P}}(V) \in L$ by assumption, we also have have $\mathbf{i}_{\mathcal{P}}(V') - \sum_{\mathbf{i}\in I} a_{\mathbf{i}}\mathbf{i} \in L$. So by Proposition 4.4.11 we may choose integers $b_{\mathbf{i}}$ with $|b_{\mathbf{i}}| \leqslant \eta n$ for each $\mathbf{i} \in I$ so that $\sum_{\mathbf{i}\in I} b_{\mathbf{i}}\mathbf{i} = \mathbf{i}_{\mathcal{P}}(V') - \sum_{\mathbf{i}\in I} a_{\mathbf{i}}\mathbf{i}$. Let $\rho_{\mathbf{i}} = a_{\mathbf{i}} + b_{\mathbf{i}} + N$ for each $\mathbf{i}$. Then

$$\sum_{\mathbf{i}\in I} \rho_{\mathbf{i}}\mathbf{i} = \sum_{\mathbf{i}\in I} a_{\mathbf{i}}\mathbf{i} + \sum_{\mathbf{i}\in I} b_{\mathbf{i}}\mathbf{i} + \sum_{\mathbf{i}\in I} N\mathbf{i} = \mathbf{i}_{\mathcal{P}}(V') + \sum_{\mathbf{i}\in I} N\mathbf{i} = \mathbf{i}_{\mathcal{P}}(V),$$

so (i) holds, and (ii) holds since $a_{\mathbf{i}} \geqslant 0$ and $|b_{\mathbf{i}}| \leqslant \eta n$ for each $\mathbf{i} \in I$. □

The vector $\mathbf{n}_{\mathbf{i}}^\cup$ with respect to $\mathcal{Q}^\cup$ obtained in the next claim tells us how many vertices should be taken from each part of $\mathcal{Q}^\cup$. Indeed, for each $\mathbf{i} \in I$ and each part $X \in \mathcal{Q}^\cup$, the set $T_{\mathbf{i}}$ will include $(n_{\mathbf{i}}^\cup)_X$ vertices from X. Note that condition (C1) says that if the sets $T_{\mathbf{i}}$ are disjoint then they will form a partition of V, whilst condition (C3) ensures that $T_{\mathbf{i}}$ takes roughly the same proportion of vertices from each part of $\mathcal{Q}^\cup$ which is a subset of a part of $\mathcal{P}$ indexed by $\mathbf{i}$, and no vertices from any other part of $\mathcal{Q}^\cup$.

Claim 4.6.3 *There exist vectors $\mathbf{n}_{\mathbf{i}}^\cup$ with respect to $\mathcal{Q}^\cup$ for $\mathbf{i} \in I$ which satisfy*

(C1) $\sum_{\mathbf{i}\in I} \mathbf{n}^{\cup}_{\mathbf{i}} = \mathbf{i}_{\mathcal{Q}^{\cup}}(V)$,

(C2) $\mathbf{n}^{\cup}_{\mathbf{i}} \in L^{\mu_9}_{\mathcal{Q}^{\cup}}(H_{\mathbf{i}})$ *for every* $\mathbf{i} \in I$*, and*

(C3) For any part W *of* $\mathcal{P}$ *and any part* $X \subseteq W$ *of* $\mathcal{Q}^{\cup}$ *we have*

$$(n^{\cup}_{\mathbf{i}})_X = \begin{cases} \rho_{\mathbf{i}}\frac{|X|}{|W|} \pm D & \textit{if } \ i_W = 1 \\ 0 & \textit{otherwise.} \end{cases}$$

Proof. Choose $\rho_{\mathbf{i}}$ for each $\mathbf{i} \in I$ as in Claim 4.6.2. Now fix some $\mathbf{i} \in I$, and choose a vector $\mathbf{n}_{\mathbf{i}}$ with respect to $\mathcal{Q}_{\mathbf{i}}$ by taking $(\mathbf{n}_{\mathbf{i}})_X$ to be either $\lfloor \rho_{\mathbf{i}}|X|/|W| \rfloor$ or $\lceil \rho_{\mathbf{i}}|X|/|W| \rceil$ for each part W of $\mathcal{P}_{\mathbf{i}}$ and each part X of $\mathcal{Q}^{\cup}$ with $X \subseteq W$. Make these choices so that $\sum_{X \subseteq W} (\mathbf{n}_{\mathbf{i}})_X = \rho_{\mathbf{i}}$ for each part W of $\mathcal{P}$ indexed by $\mathbf{i}$; this is possible since

$$\sum_{X \in \mathcal{Q}^{\cup}, X \subseteq W} \frac{\rho_{\mathbf{i}}|X|}{|W|} = \frac{\rho_{\mathbf{i}}|W|}{|W|} = \rho_{\mathbf{i}}.$$

Now, for any $\mathbf{i} \in I$, by (A8) we may choose parts Y, Y' of $\mathcal{Q}_{\mathbf{i}}$ so that $\mathbf{n}^1_{\mathbf{i}} := \mathbf{n}_{\mathbf{i}} - \mathbf{u}_Y + \mathbf{u}_{Y'} \in L_{\mathbf{i}} = L^{\mu'}_{\mathcal{Q}_{\mathbf{i}}}(H_{\mathbf{i}})$. Let $\mathbf{n}^0_{\mathbf{i}}$ be the vector with respect to $\mathcal{Q}^0_{\mathbf{i}}$ corresponding to $\mathbf{n}^1_{\mathbf{i}}$, that is, with additional zero co-ordinates corresponding to the parts of $\mathcal{P}$ not indexed by $\mathbf{i}$. So $\mathbf{n}^0_{\mathbf{i}} \in L^{\mu_9}_{\mathcal{Q}^0_{\mathbf{i}}}(H_{\mathbf{i}})$. Finally, let $\mathbf{n}'_{\mathbf{i}} := (\mathbf{n}^0_{\mathbf{i}} \mid \mathcal{Q}^{\cup})$, so $\mathbf{n}'_{\mathbf{i}} \in L^{\mu_9}_{\mathcal{Q}^{\cup}}(H_{\mathbf{i}})$ by Proposition 4.4.3. The key observation is that for any part W of $\mathcal{P}$ and any part $X \subseteq W$ of $\mathcal{Q}^{\cup}$ we have

$$(n'_{\mathbf{i}})_X = \begin{cases} \rho_{\mathbf{i}}\frac{|X|}{|W|} \pm 2k & \text{if } \ i_W = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{4.6.4}$$

Choose $\mathbf{n}'_{\mathbf{i}}$ as above for every $\mathbf{i} \in I$; then $\sum_{\mathbf{i}\in I} \mathbf{n}'_{\mathbf{i}} \in L^{\mu_9}_{\mathcal{Q}^{\cup}}(H)$. Now, since $\mathcal{Q}^{\cup}$ is a refinement of $\mathcal{P}$ with parts of size at least ckn, and $\mathcal{P}$ is (c, μ, μ')-robustly maximal by assumption (and therefore (c, μ_9, μ')-robustly maximal), we have $\mathbf{i}_{\mathcal{Q}^{\cup}}(V) \in L^{\mu_9}_{\mathcal{Q}^{\cup}}(H)$ by

Proposition 4.5.4. So

$$\mathbf{i}_{\mathcal{Q}^{\cup}}(V) - \sum_{\mathbf{i}\in I} \mathbf{n}'_{\mathbf{i}} \in L^{\mu_9}_{\mathcal{Q}^{\cup}}(H) \cap B(\mathbf{0}, 2k|I|).$$

This means we may apply Proposition 4.4.11 to obtain integers $a_{\mathbf{i}'}$ with $|a_{\mathbf{i}'}| \leqslant (D-2k)/k^k$ for each $\mathbf{i}' \in I^{\mu_9}_{\mathcal{Q}^{\cup}}(H)$ such that

$$\mathbf{i}_{\mathcal{Q}^{\cup}}(V) - \sum_{\mathbf{i}\in I} \mathbf{n}'_{\mathbf{i}} = \sum_{\mathbf{i}'\in I^{\mu_9}_{\mathcal{Q}^{\cup}}(H)} a_{\mathbf{i}'}\mathbf{i}'.$$

For each $\mathbf{i} \in I$, define $\mathbf{n}^{\cup}_{\mathbf{i}} := \mathbf{n}'_{\mathbf{i}} + \sum a_{\mathbf{i}'}\mathbf{i}'$, where the sum is taken over all $\mathbf{i}' \in I^{\mu_9}_{\mathcal{Q}^{\cup}}(H)$ with $(\mathbf{i}' \mid \mathcal{P}) = \mathbf{i}$. Then $\mathbf{n}^{\cup}_{\mathbf{i}} \in L^{\mu_9}_{\mathcal{Q}^{\cup}}(H_{\mathbf{i}})$, since both $\mathbf{n}'_{\mathbf{i}}$ and $\sum a_{\mathbf{i}'}\mathbf{i}'$ are members of $L^{\mu_9}_{\mathcal{Q}^{\cup}}(H_{\mathbf{i}})$. So (C2) is satisfied. Furthermore,

$$\mathbf{i}_{\mathcal{Q}^{\cup}}(V) - \sum_{\mathbf{i}\in I} \mathbf{n}^{\cup}_{\mathbf{i}} = \mathbf{i}_{\mathcal{Q}^{\cup}}(V) - \sum_{\mathbf{i}\in I} \mathbf{n}'_{\mathbf{i}} - \sum_{\mathbf{i}'\in I^{\mu_9}_{\mathcal{Q}^{\cup}}(H)} a_{\mathbf{i}'}\mathbf{i}' = \mathbf{0}$$

so (C1) is satisfied.

Finally, fix some $\mathbf{i} \in I$, a part W of $\mathcal{P}$ indexed by $\mathbf{i}$ and a part $X \subseteq W$ of $\mathcal{Q}^{\cup}$. Then by definition of $\mathbf{n}^{\cup}$ we have $|(n^{\cup}_{\mathbf{i}})_X - (n'_{\mathbf{i}})_X| \leqslant \sum a_{\mathbf{i}'} \leqslant D - 2k$; together with (4.6.4) we have (C3) also. □

Claim 4.6.5 *There exist vectors $\mathbf{n}_{\mathbf{i}}$ with respect to $\mathcal{Q}^{\cap}$ for $\mathbf{i} \in I$ which satisfy*

(D1) $\sum_{\mathbf{i}\in I} \mathbf{n}_{\mathbf{i}} = \mathbf{i}_{\mathcal{Q}^{\cap}}(V)$,

(D2) $(\mathbf{n}_{\mathbf{i}} \mid \mathcal{Q}_{\mathbf{i}}) \in L_{\mathbf{i}}$ *for every* $\mathbf{i} \in I$*, and*

(D3) For any part W of $\mathcal{P}$ and any part $Z \subseteq W$ of $\mathcal{Q}^\cap$ we have

$$(n_{\mathbf{i}})_Z = \begin{cases} \rho_{\mathbf{i}} \frac{|Z|}{|W|} \pm 2D & \text{if } i_W = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Proof. Fix vectors $\mathbf{n}_{\mathbf{i}}^\cup$ which satisfy (C1)-(C3). Now, for each $\mathbf{i} \in I$, each part X of $\mathcal{Q}^\cup$ and each part $Z \subseteq X$ of $\mathcal{Q}^\cap$, take $(n_{\mathbf{i}})_Z$ to be either $\lfloor (n_{\mathbf{i}}^\cup)_X |Z|/|X| \rfloor$ or $\lceil (n_{\mathbf{i}}^\cup)_X |Z|/|X| \rceil$, with choices made so that $\sum_{\mathbf{i} \in I} (n_{\mathbf{i}})_Z = |Z|$ for every $\mathbf{i} \in I$ and $\sum_{Z \subseteq X} (n_{\mathbf{i}})_Z = (n_{\mathbf{i}})_X$ for each part X of $\mathcal{Q}^\cup$. Then the vectors $\mathbf{n}_{\mathbf{i}}$ satisfy (D1). Note also that the latter condition may be reformulated as $(\mathbf{n}_{\mathbf{i}} \mid \mathcal{Q}^\cup) = \mathbf{n}_{\mathbf{i}}$. Furthermore, for any part W of $\mathcal{P}$ and any part $Z \subseteq W$ of $\mathcal{Q}^\cap$ it follows from (C3) that $(n_{\mathbf{i}})_Z$ is equal to $\rho_{\mathbf{i}}|Z|/|W| \pm (D+1)$ if $i_W = 1$ and is zero otherwise.

The claim follows from repeated use of the following observation. Fix $\mathbf{i}_1, \mathbf{i}_2 \in I$, and suppose that Z and Z' are parts of $\mathcal{Q}^\cap$ which are subsets of distinct parts Y_1 and Y_1' respectively of $\mathcal{Q}_{\mathbf{i}_1}$, but that Z and Z' are subsets of the same part Y_2 of $\mathcal{Q}_{\mathbf{i}_2}$. Now suppose that we increment both $(n_{\mathbf{i}_1})_Z$ and $(n_{\mathbf{i}_2})_{Z'}$ by one, and decrement both $(n_{\mathbf{i}_2})_Z$ and $(n_{\mathbf{i}_1})_{Z'}$ by one (all other co-ordinates of $\mathbf{n}_{\mathbf{i}_1}$ and $\mathbf{n}_{\mathbf{i}_2}$ remain unchanged, as do all other vectors $\mathbf{n}_{\mathbf{i}}$). Then clearly $\sum_{\mathbf{i} \in I} \mathbf{n}_{\mathbf{i}}$ is unchanged by this operation, as is $(\mathbf{n}_{\mathbf{i}} \mid \mathcal{Q}_{\mathbf{i}})$ for any $\mathbf{i} \neq \mathbf{i}_1, \mathbf{i}_2$. Also, since Z and Z' are contained in the same part of $\mathcal{Q}_{\mathbf{i}_2}$, the vector $(\mathbf{n}_{\mathbf{i}_2} \mid \mathcal{Q}_{\mathbf{i}_2})$ is also unchanged. However, $(\mathbf{n}_{\mathbf{i}_2} \mid \mathcal{Q}_{\mathbf{i}_2})$ has changed; specifically $(\mathbf{n}_{\mathbf{i}_1} \mid \mathcal{Q}_{\mathbf{i}_1})_{Y_1}$ has increased by one and $(\mathbf{n}_{\mathbf{i}_1} \mid \mathcal{Q}_{\mathbf{i}_1})_{Y_1'}$ has decreased by one. We shall repeat this technique to satisfy (D2) for each $\mathbf{i} \in I$ in turn.

Now fix some $\mathbf{i} \in I$. By (A8) we may choose parts Y, Y' of $\mathcal{Q}_{\mathbf{i}}$ such that $(\mathbf{n}_{\mathbf{i}} \mid \mathcal{Q}_{\mathbf{i}}) - \mathbf{u}_Y + \mathbf{u}_{Y'} \in L_{\mathbf{i}}$. Since $(\mathbf{n}_{\mathbf{i}} \mid \mathcal{Q}^\cup) = \mathbf{n}_{\mathbf{i}}^\cup \in L^{\mu_9}_{\mathcal{Q}^\cup}(H_{\mathbf{i}})$, and $L^{\mu_0}_{\mathcal{Q}^\cup}(H_{\mathbf{i}})$ is $(1,-1)$-free, Y and Y' must be subsets of the same part X of $\mathcal{Q}^\cup$. Choose parts Z and Z' of $\mathcal{Q}^\cap$ with $Z \subseteq Y$ and $Z' \subseteq Y'$. Then by the definition of $\mathcal{Q}^\cup$ there is a sequence $Z = Z_0, \dots, Z_s = Z'$ of parts of $\mathcal{Q}^\cap$ and a sequence $Y_1, \dots, Y_s$ such that

1. for each j there exists $\mathbf{i}_j \in I$ such that Y_j is a part of $\mathcal{Q}_{\mathbf{i}_j}$,

2. for each j we have $Z_{j-1}, Z_j \subseteq Y_j$, and

3. $s \leqslant |\mathcal{Q}^\cap| \leqslant k^{2+k^k}$.

For each $0 \leqslant j \leqslant s$ let Y_j^* be the part of $\mathcal{Q}_\mathbf{i}$ which contains Z_j. So $Y_0^* = Y$ and $Y_s^* = Y'$.

For each $j \in [s]$ in turn, we proceed as follows. If $Y_{j-1}^* = Y_j^*$, then we do nothing. Otherwise, we apply our observation above with $\mathbf{i}$, $\mathbf{i}_j$, Z_{j-1} and Z_j in place of $\mathbf{i}_1$, $\mathbf{i}_2$, Z and Z' respectively. As noted above, the effect of this is that $(\mathbf{n_i} \mid \mathcal{Q}_\mathbf{i})_{Y_j^*}$ is increased by one, and $(\mathbf{n_i} \mid \mathcal{Q}_\mathbf{i})_{Y_{j-1}^*}$ is decreased by one, whilst every other co-ordinate of $(\mathbf{n_i} \mid \mathcal{Q}_\mathbf{i})$ and every $\mathbf{n}_{\mathbf{i}'}$ for $\mathbf{i}' \neq \mathbf{i}$ is unchanged. So after doing this for each $j \in [s]$ in turn, the net effect is that $(\mathbf{n_i} \mid \mathcal{Q}_\mathbf{i})_{Y_0^*}$ has increased by one, and $(\mathbf{n_i} \mid \mathcal{Q}_\mathbf{i})_{Y_s^*}$ has decreased by one. That is, we have added $\mathbf{u}_Y - \mathbf{u}_{Y'}$ to$(\mathbf{n_i} \mid \mathcal{Q}_\mathbf{i})$. By choice of Y and Y' we conclude that after these modifications we have $(\mathbf{n_i} \mid \mathcal{Q}_\mathbf{i}) \in L_\mathbf{i}$, and crucially, that$(\mathbf{n}_{\mathbf{i}'} \mid \mathcal{Q}_{\mathbf{i}'})$ is unchanged for any $\mathbf{i}' \neq \mathbf{i}$ and $\sum_{\mathbf{i} \in I} \mathbf{n_i}$ is unchanged.

Proceed in this manner for every $\mathbf{i} \in I$; then the vectors $\mathbf{n_i}$ obtained at the end of this process must satisfy (D2). Since (D1) held before we made any modifications, and $\sum_{\mathbf{i} \in I} \mathbf{n_i}$ is preserved by each modification, we conclude that (D1) still holds. So it remains to prove (D3). For this observe that we applied our observation at most $s \leqslant k^{2+k^k}$ times for each $|I|$, and so at most $k^{2+k^k} \cdot r^{k-2} \leqslant D - 1$ times in total. On each occasion no co-ordinate of any $\mathbf{n_i}$ is increased or decreased by more than one. So we conclude that for any $\mathbf{i} \in I$, any part Z of $\mathcal{Q}^\cap$ and any part W of $\mathcal{P}$ the co-ordinate $(n_\mathbf{i})_Z$ is equal to $\rho_\mathbf{i}|Z|/|W| \pm 2D$ if $i_W = 1$ and is zero otherwise.

This completes the proof of Claim 4.6.5. □

Step 4: The random selection. Fix vectors $\mathbf{n_i}$ as in Claim 4.6.5. We now partition V into disjoint sets $T_\mathbf{i}$ for $\mathbf{i} \in I$, where for each $Z \in \mathcal{Q}^\cap$ the number of vertices of $T_\mathbf{i}$ taken from Z is $(n_\mathbf{i})_Z$. To ensure that this is possible, we require that $\sum_{\mathbf{i} \in I} \mathbf{n_i} = \mathbf{i}_{\mathcal{Q}^\cap}(V)$ and that each co-ordinate of each $\mathbf{n_i}$ is non-negative. The first of these conditions holds by (D1), and the second by (D3), since for any $\mathbf{i} \in I$, any part W of $\mathcal{P}$ and any part $Z \subseteq W$ of $\mathcal{Q}^\cap$ we have $\rho_\mathbf{i} \geqslant \eta n$ by (B2), $|Z| \geqslant 2D/\eta$ by (A2) and $|W| \leqslant n$. So $\rho_\mathbf{i}|Z|/|W| - 2D \geqslant 0$ as required.

Choose such a partition uniformly at random. That is, for each part Z of $\mathcal{Q}^\cap$, choose uniformly at random a partition of Z into sets $Z_\mathbf{i}$ for $\mathbf{i} \in I$ so that $|Z_\mathbf{i}| = (n_\mathbf{i})_Z$ for each $\mathbf{i} \in I$. For each $\mathbf{i} \in I$ let $T_\mathbf{i} := \bigcup_{Z \in \mathcal{Q}^\cap} Z_\mathbf{i}$, and let $H^*_\mathbf{i} = H[T_\mathbf{i}]$. Since $(\mathbf{n_i})_W = 0$ for any part W of $\mathcal{P}$ not indexed by $\mathbf{i}$, we have $T_\mathbf{i} \subseteq S_\mathbf{i}$, and so $H^*_\mathbf{i}$ is a $\mathcal{P}_\mathbf{i}$-partite k-graph on the vertex set $T_\mathbf{i}$. Also note that since $(\mathbf{n_i} \mid \mathcal{Q}_\mathbf{i}) \in L^{\mu_9}_{\mathcal{Q}^0_\mathbf{i}}(H)$, and every $\mathbf{i}' \in I^{\mu_9}_{\mathcal{Q}^0_\mathbf{i}}(H)$ has $(\mathbf{i}' \mid \mathcal{P}) = \mathbf{i}$, we have $(\mathbf{n_i} \mid \mathcal{P}) = t_\mathbf{i}\mathbf{i}$ for some integer $t_\mathbf{i}$. Note that for any part W of $\mathcal{P}$ which is indexed by $\mathbf{i}$ we have

$$t_\mathbf{i} = (\mathbf{n_i} \mid \mathcal{P})_W = \sum_Z (n_\mathbf{i})_Z \overset{\text{(D3)}}{=} \sum_Z \left(\frac{\rho_\mathbf{i}|Z|}{|W|} \pm 2D \right) \overset{\text{(A2)}}{=} \rho_\mathbf{i} \pm 2DK, \tag{4.6.6}$$

where both sums are taken over all parts Z of $\mathcal{Q}^\cap$ with $Z \subseteq W$. So in particular we have $\eta n/2 \leqslant t_\mathbf{i} \leqslant n$ for any $\mathbf{i} \in I$ by (B2).

Let $\mu_* = \mu_3$, $\mu'_* = \mu_7$, $\gamma_* = \gamma/4$ and $\varepsilon_* = \mu_2$.

Claim 4.6.7 *Fix $\mathbf{i} \in I$. Then the following properties each hold with probability at least $1 - 1/k^k$.*

(E1) At most $\varepsilon_* t_\mathbf{i}^{k-1}$ $\mathcal{P}_\mathbf{i}$-partite subsets $S \subseteq T_\mathbf{i}$ of size $k-1$ have $d_{H^*_\mathbf{i}}(S) < (1/(\ell-1)+\gamma_*)t_\mathbf{i}$.

(E2) $\mathcal{Q}_\mathbf{i}$ is (c, μ_*, μ'_*)-robustly maximal with respect to $H^*_\mathbf{i}$.

(E3) For any vertex $x \in T_\mathbf{i}$ there are at least $d(t_\mathbf{i})^{k-1}$ edges $e \in H^*_\mathbf{i}$ with $x \in e$ and $\mathbf{i}_{\mathcal{Q}_\mathbf{i}}(e) \in L^{\mu_*}_{\mathcal{Q}_\mathbf{i}}(H^*_\mathbf{i})$.

Proof. First observe that (E2) holds with probability $1 - 1/2k^k$ by Lemma 4.5.6. For (E1), recall from (A9) that the number of $\mathcal{P}_\mathbf{i}$-partite sets $S \subseteq S_\mathbf{i}$ of size $k-1$ with $d_{H_\mathbf{i}}(S) <$

$(1/\ell + \gamma/2)n$ is at most $\mu_1 n^{k-1} \leqslant \varepsilon_* t_{\mathbf{i}}^{k-1}$. So it suffices to show that with probability at least $1-1/n$ any $\mathcal{P}_{\mathbf{i}}$-partite subset $S \subseteq T_{\mathbf{i}}$ of size $k-1$ with $d_{H_{\mathbf{i}}}(S) \geqslant (1/\ell+\gamma/2)n$ satisfies $d_{H^*_{\mathbf{i}}}(S) \geqslant (1/(\ell-1)+\gamma_*)t_{\mathbf{i}}$. Fix some such S, and let W be the part of $\mathcal{P}$ such that $N_{T_{\mathbf{i}}}(S) \subseteq W$, so $|N(S) \cap W| \geqslant (1/\ell + \gamma/2)n$. Then

$$\begin{aligned}\mathbb{E}[d_{H^*_{\mathbf{i}}}(S)] &= \sum_{Z \subseteq W} \mathbb{E}[|N(S) \cap Z_{\mathbf{i}}|] = \sum_{Z \subseteq W} \frac{(n_{\mathbf{i}})_Z |N(S) \cap Z|}{|Z|} \\ &\overset{\text{(D3)}}{\geqslant} \frac{\rho_{\mathbf{i}} |N(S) \cap W|}{|W|} - 2KD \\ &\geqslant \frac{t_{\mathbf{i}}(1/\ell + \gamma/2)}{1 - 1/\ell - \gamma/2} - 4KD \\ &\geqslant (1+\varepsilon)\left(\frac{1}{\ell-1} + \gamma_*\right) t_{\mathbf{i}},\end{aligned}$$

Since $d_{H^*_{\mathbf{i}}}(S)$ is a sum of independent hypergeometric random variables we may apply Corollary 1.8.2 to obtain

$$\mathbb{P}\left(d_{H^*_{\mathbf{i}}}(S) < \left(\frac{1}{\ell-1} + \gamma_*\right) t_{\mathbf{i}}\right) < \frac{1}{kn^k}.$$

Taking a union bound over the at most kn^{k-1} such sets S we find that (E1) holds with probability at least $1 - 1/n$.

To prove (E3), fix a vertex $x \in T_{\mathbf{i}}$, and let $E(x)$ denote the number of edges $e \in H^*_{\mathbf{i}}$ with $\mathbf{i}_{\mathcal{Q}_{\mathbf{i}}}(e) \in L_{\mathbf{i}}$ which contain x. To estimate $E(x)$, recall from (A4) that x is contained in at least $4dn^{k-1}$ edges $e \in H_{\mathbf{i}}$ with $\mathbf{i}_{\mathcal{Q}_{\mathbf{i}}}(e) \in L_{\mathbf{i}}$. At most dn^{k-1} of these edges contain a vertex from a part of $\mathcal{Q}^\cap$ of size at most dn/K. Fix one of the remaining $3dn^{k-1}$ edges $e = \{x_1, \dots, x_{k-1}, x\} \in H$, and for each $j \in [k-1]$ let Z_j and W_j be the parts of $\mathcal{Q}^\cap$ and $\mathcal{P}$ containing x_j respectively. So $|Z_j| \geqslant dn/K$ for each $j \in [k-1]$. Then the probability that $e \in H^*_{\mathbf{i}}$ is equal to

$$\prod_{j\in[k-1]} \frac{(n_{\mathbf{i}})_{Z_j}}{|Z_j|} \overset{\text{(D3)}}{=} \prod_{j\in[k-1]} \frac{\rho_{\mathbf{i}}}{|W_j|} \pm \frac{2DK}{dn} \overset{(4.6.6)}{>} \frac{t_{\mathbf{i}}^{k-1}}{n^{k-1}} - \frac{2(\ell+1)DK}{dn}.$$

So $\mathbb{E}(E(x)) \geqslant 2dt_{\mathbf{i}}^{k-1}$. Now fix an order $v_1, \dots, v_{|V|}$ of the vertices of V. We define a martingale $A_j := \mathbb{E}(E(x)|v_1, \dots, v_j)$ for $0 \leqslant j \leqslant |V|$, where the 'uncovered event' is which part of $Z_{\mathbf{i}}$ contains v_j. Then $A_{|V|} = E(x)$ and $A_0 = \mathbb{E}(E(x))$, and for any $j \in [|V|]$ we have $|A_j - A_{j-1}| \leqslant 2(kn)^{k-2}$. Now applying Theorem 1.8.3 with $t = dt_{\mathbf{i}}^{k-1}/2(kn)^{k-2} = \Omega(n)$, we obtain

$$\mathbb{P}(E(x) < dt_{\mathbf{i}}^{k-1}) \leqslant 2\exp\left(-\frac{t^2}{2n}\right) = e^{-\Omega(n)} \leqslant 1/kn^2.$$

Taking a union bound over the at most kn vertices $x \in V$ gives (E3), completing the proof of the claim. $\square$

Since $|I| = r^{k-2}$ we may fix an outcome of the random partition of V so that the properties (E1)-(E3) of Claim 4.6.7 hold for every $\mathbf{i} \in I$. Then for each $\mathbf{i} \in I$ we have the following: $T_{\mathbf{i}}$ is a vertex set partitioned into parts of size $t_{\mathbf{i}}$ by $\mathcal{P}_{\mathbf{i}}$, and $H^*_{\mathbf{i}}$ is a $\mathcal{P}_{\mathbf{i}}$-partite k-graph on $T_{\mathbf{i}}$ in which by (E1) at most $\varepsilon_* t_{\mathbf{i}}^{k-1}$ $\mathcal{P}'$-partite subsets $S \subseteq T_{\mathbf{i}}$ of size $k-1$ have $d(S) < (1/(\ell-1)+\gamma_*)n$. $\mathcal{Q}_{\mathbf{i}}$ is a partition of $T_{\mathbf{i}}$ which refines $\mathcal{P}_{\mathbf{i}}$ into parts of size at least cn and by (E2) is (c, μ_*, μ'_*)-robustly maximal with respect to $H^*_{\mathbf{i}}$. Also, for any vertex $x \in T_{\mathbf{i}}$ there are at least dn^{k-1} edges $e \in H^*_{\mathbf{i}}$ with $x \in e$ and $\mathbf{i}_{\mathcal{Q}_{\mathbf{i}}}(e) \in L^{\mu_*}_{\mathcal{Q}_{\mathbf{i}}}(H^*_{\mathbf{i}})$ by (E3), and $\mathbf{i}_{\mathcal{Q}_{\mathbf{i}}}(T_{\mathbf{i}}) = (\mathbf{n}_{\mathbf{i}} \mid \mathcal{Q}_{\mathbf{i}}) \in L^{\mu_*}_{\mathcal{Q}_{\mathbf{i}}}(H^*_{\mathbf{i}})$ by (D2). Since

$$1/t_{\mathbf{i}} \ll \varepsilon_* \ll \mu_* \ll \mu'_* \ll c, d, \ll \gamma_*, 1/k,$$

we conclude that $H^*_{\mathbf{i}}$ contains a perfect matching $M_{\mathbf{i}}$ by our inductive hypothesis. Fix a perfect matching $M_{\mathbf{i}}$ in $H^*_{\mathbf{i}}$ for each $\mathbf{i} \in I$; then $M \cup M' \cup \bigcup_{\mathbf{i}\in I} M_{\mathbf{i}}$ is a perfect matching in (the original) H. $\square$

The next lemma is a non-partite form of Lemma 4.6.1. Essentially the proof of this lemma is to take a random k-partition of H, and apply Lemma 4.6.1 to obtain a perfect matching. However, showing that the conditions on H transfer to the k-partite subgraph formed by the random partition is technical and non-trivial.

Lemma 4.6.8 *Let $k \geqslant 2$, and suppose that constants $n_0, \varepsilon, \mu, \mu', c, d, \gamma$ satisfy $1/n_0 \ll \varepsilon \ll \mu \ll \mu' \ll c, d, \gamma \ll 1/k$. Let $n \geqslant n_0$, let V be a vertex set of size kn and let H be a k-graph on V such that $d_k(S) \geqslant (1+\gamma)n$ for all but at most εn^{k-1} subsets S of V of size $k-1$.*

Let $\mathcal{P}$ be a partition of V with parts of size at least $(1+\gamma)n$ which is (c, μ, μ')-robustly maximal with respect to H. Suppose that

(i) for any vertex $x \in V$ there are at least dn^{k-1} edges $e \in H$ with $x \in e$ and $\mathbf{i}_{\mathcal{P}}(e) \in L^{\mu}_{\mathcal{P}}(H)$, and

(ii) $\mathbf{i}_{\mathcal{P}}(V) \in L^{\mu}_{\mathcal{P}}(H)$.

Then H contains a perfect matching.

Proof. In this proof we omit floor and ceiling symbols where they do not affect the argument. We choose a random partition $\mathcal{P}' = (V_1, \ldots, V_k)$ of V into k parts each of size n, let H' be the $\mathcal{P}'$-partite subgraph of H and let $\widehat{\mathcal{P}}$ be the common refinement of $\mathcal{P}$ and $\mathcal{P}'$.

Let $\mu'_* = (\mu')^3$, $d_* = d/2k!$ and $\gamma_* = \gamma/2k$ and note that $\mu \ll \mu'_* \ll c, d_*, \gamma_* \ll 1/k$. We will show that with positive probability the following conditions hold for H':

(A1) Every part of $\widehat{\mathcal{P}}$ has size at least $(1/k + \gamma_*)n$,

(A2) All but at most εn^{k-1} $\mathcal{P}'$-partite $(k-1)$-sets in V have codegree at least $(1/k+\gamma_*)n$,

(A3) $\widehat{\mathcal{P}}$ is (c, μ, μ'_*)-robustly maximal with respect to H', (A4) For any vertex $x \in V$ there are at least $d_* n^{k-1}$ edges $e \in H'$ with $x \in e$ and $\mathbf{i}_{\widehat{\mathcal{P}}}(e) \in L^{\mu}_{\widehat{\mathcal{P}}}(H)$,

(A5) $\mathbf{i}_{\widehat{\mathcal{P}}}(V) \in L^{\mu}_{\widehat{\mathcal{P}}}(H')$.

We then choose partitions $\mathcal{P}'$ and $\widehat{\mathcal{P}}$ which satisfy (A1)-(A5), choose an integer D such that $1/n_0 \ll 1/D \ll \varepsilon$ and apply Lemma 4.6.1 to H' with $\widehat{\mathcal{P}}$ in place of $\mathcal{P}$ and μ'_*, d_*, γ_* in place of μ', d, γ to obtain a perfect matching in H', which is also a perfect matching in H.

To prove (A1), observe that $\mathbb{E}(|X|) = |W|/k \geqslant (1/k + \gamma/k)n$ for each part $W \in \mathcal{P}$ and each part $X \subseteq W$ of $\widehat{\mathcal{P}}$. Since $|X|$ is distributed hypergeometrically, Corollary 1.8.2 implies that

$$\mathbb{P}\left[|X| < \left(\frac{1}{k} + \frac{\gamma}{2k}\right)n\right] \leqslant 2\exp\left(-\frac{\gamma^2 n}{12k^3}\right) \leqslant \frac{1}{k^2 n},$$

and a union bound implies that with high probability $|X| \geqslant (1/k + \gamma/2k)n$ for every $X \in \widehat{\mathcal{P}}$.

To prove (A2), first note that for all but at most εn^{k-1} subsets $S \subseteq V$ of size $k-1$ and any $1 \leqslant i \leqslant k$ we have $\mathbb{E}[d_{V_i}(S)] = d_V(S)/k \geqslant (1+\gamma)n/k$. Since $d_{V_i}(S)$ is distributed hypergeometrically, Corollary 1.8.2 implies that

$$\mathbb{P}\left[d_{V_i}(S) < \left(1 + \frac{\gamma}{2}\right)\frac{n}{k}\right] \leqslant 2\exp\left[-\frac{\gamma^2 n}{12k}\right] \leqslant \frac{1}{n^k}.$$

Now a union bound implies that with high probability, $d_{V_i}(S) \geqslant (1/k + \gamma/2k)n$ for all but at most εn^{k-1} such subsets.

Next we prove (A5). By Proposition 4.4.3 (i), $(\mathbf{i} \mid \mathcal{P}) \in L^{\mu}_{\mathcal{P}}(H)$ for every $\mathbf{i} \in L^{\mu}_{\widehat{\mathcal{P}}}(H') \subseteq L^{\mu}_{\widehat{\mathcal{P}}}(H)$, and hence $L^{\mu}_{\widehat{\mathcal{P}}}(H')$ is $(1,-1)$-free. We now show that $\mathbf{i} \in L^{\mu}_{\widehat{\mathcal{P}}}(H')$ for every index vector $\mathbf{i}$ with respect to $\widehat{\mathcal{P}}$ such that $(\mathbf{i} \mid \mathcal{P}')$ is a multiple of $\mathbf{1}$ and $(\mathbf{i} \mid \mathcal{P}) \in L^{\mu}_{\mathcal{P}}(H)$. To see this, let X be a part of $\widehat{\mathcal{P}}$ such that $X \subseteq V_k$. By Proposition 4.4.5 (vii) there exists a part X' of $\widehat{\mathcal{P}}$ such that $\mathbf{i} - \mathbf{u}_X + \mathbf{u}_{X'} \in L^{\mu}_{\widehat{\mathcal{P}}}(H')$. Since H' is $\mathcal{P}'$-partite, X' must also be contained in V_k. Now $(\mathbf{u}_X - \mathbf{u}_{X'} \mid \mathcal{P}) \in L^{\mu}_{\mathcal{P}}(H)$. But $L^{\mu}_{\mathcal{P}}(H)$ is $(1,-1)$-free and hence X and X' must be contained in the same part of $\mathcal{P}$. So $X' = X$ and $\mathbf{i} \in L^{\mu}_{\widehat{\mathcal{P}}}(H')$. (A5) follows immediately, since $(\mathbf{i}_{\widehat{\mathcal{P}}}(V) \mid \mathcal{P}') = n \cdot \mathbf{1}$ and $(\mathbf{i}_{\widehat{\mathcal{P}}}(V) \mid \mathcal{P}) = \mathbf{i}_{\mathcal{P}}(V) \in L^{\mu}_{\mathcal{P}}(H)$ by (ii).

We now use a similar method to that of Lemma 4.5.6 to prove (A3). Again we use Theorem 4.5.5 to obtain an integer n', an ε-regular refinement $\mathcal{Q}$ of $\mathcal{P}$ with parts of size n' and a reduced graph R with n_R vertices where edges correspond to (d,ε)-regular k-graphs for $d \geqslant \mu'/4$, where $n_R \geqslant k^k/c$. Suppose there exists a refinement $\widehat{\mathcal{P}}^*$ of $\widehat{\mathcal{P}}$ with parts of size at least ckn such that $L^{\mu'_*}_{\widehat{\mathcal{P}}^*}$ is $(1,-1)$-free. Let ℓ be the number of parts into which $\widehat{\mathcal{P}}^*$ splits each part of $\mathcal{P}'$ (note that Proposition 4.4.5(ii) implies that ℓ must be the same for each part of $\mathcal{P}'$). Choose $\mu' \ll \psi \ll c$. Call an index vector $\mathbf{i}$ *simple* if $(\mathbf{i} \mid \mathcal{P}') = \mathbf{1}$.

Claim 4.6.9 *There exists a function $\mathbf{i}^* : V(R) \to \mathbb{Z}^{\widehat{\mathcal{P}}^*}$ such that:*

(B1) $\mathbf{i}^*(Y)$ is a simple index vector for each $Y \in \mathcal{Q}$,

(B2) Whenever $Z \in \widehat{\mathcal{P}}^*$ and $i^*(Y)_Z = 1$, $|Y \cap Z| \geqslant \psi n'$, and

(B3) For every $Z \in \widehat{\mathcal{P}}^*$ there are at least n_R/k parts Y such that $i^*(Y)_Z = 1$.

Proof. Choose $\mathbf{i}^*$ randomly as follows: For each $Y \in V(R)$ and each $V_i \in \mathcal{P}'$, we randomly select $Z \in \widehat{\mathcal{P}}^*$ such that $Z \subseteq V_i$ and set $i^*(Y)_Z = 1$. For all other $Z \in \widehat{\mathcal{P}}^*$ with $Z \subseteq V_i$ we set $i^*(Y)_Z = 0$. The probability of selecting $Z \in \widehat{\mathcal{P}}^*$ is zero if $|Y \cap Z| \leqslant \psi n'$ and otherwise is proprtional to $|Y \cap Z|$. All random choices are made independently.

Note that $\mathbb{P}[i^*(Y)_Z = 1] \geqslant |Y \cap Z|/|Y \cap V_i|$ for every Y and Z such that $|Y \cap Z| \geqslant \psi n'$. (B1) and (B2) are then satisfied automatically, so it remains to prove that (B3) holds with high probability, at which point we may choose $\mathbf{i}^*$ such that all three properties hold. To show this, we split $V(R)$ into the clusters Y such that $|Y \cap Z| < \psi n'$ and those such that $|Y \cap Z| \geqslant \psi n'$. The small intersection classes can only eat up a small fraction of the vertices, leaving the rest to contribute to the expected number of Y with $i^*(Y)_Z = 1$.

More precisely, let R_Z be the number of $Y \in V(R)$ with $i^*(Y)_Z = 1$. Firstly note that $\mathbb{E}[|Y \cap V_i|] = n'/k$ for each $Y \in \mathcal{Q}$ and each $V_i \in \mathcal{P}'$. Since $|Y \cap V_i|$ is distributed hypergeometrically, Corollary 1.8.2 implies that with high probability $|Y \cap V_i| = (1 \pm \varepsilon)n'/k$ for every $Y \in \mathcal{Q}$ and $V_i \in \mathcal{P}'$. For any $V_i \in \mathcal{P}'$ and any $Z \in \widehat{\mathcal{P}}^*$ with $Z \subseteq V_i$ we have

$$\sum_{Y \in V(R)} \frac{|Y \cap Z|}{|Y \cap V_i|} \geqslant \frac{k}{(1+\varepsilon)n'} \sum_{Y \in V(R)} |Y \cap Z| = \frac{k|Z|}{(1+\varepsilon)n'}$$

and hence

$$\mathbb{E}[R_Z] \geqslant \sum_{Y \in V(R), |Y \cap Z| \geqslant \psi n'} \frac{|Y \cap Z|}{|Y \cap V_i|} \geqslant \frac{k|Z|}{(1+\varepsilon)n'} - 2k\psi n_R \geqslant (1 + \gamma/10)n_R/k.$$

Now we apply Lemma 1.8.1(i) to the independent random variables $i^*(Y)_Z$ for a fixed $Z \in \widehat{\mathcal{P}}^*$ to obtain

$$\mathbb{P}\left[R_Z < \frac{n_R}{k}\right] \leqslant \exp\left[-\frac{2(\gamma n_R)^2}{(10k)^2 n_R}\right] \leqslant \frac{1}{k^2 n_R} \leqslant \frac{c}{k^k},$$

and a union bound implies that with probability at least $1 - 1/k^k$, $R_Z \geqslant n_R/k$ for every $Z \in \widehat{\mathcal{P}}^*$. $\square$

Call a cluster $Y \in V(R)$ *bad* if there exists $Z \in \widehat{\mathcal{P}}^*$ such that $\mathbf{i}^*(Y)_Z = 0$, but $|Y \cap Z| \geqslant \mu^{1/4}n'$, and *good* otherwise. Note that every good cluster in $V(R)$ contains fewer than $\mu^{1/4}n'/c$ vertices in parts $Z \in \widehat{\mathcal{P}}^*$ with $\mathbf{i}^*(Y)_Z = 0$. Hence in total there are at most $k\mu^{1/4}n/c$ vertices of H which are contained in $Y \cap Z$ for some good cluster Y and some $Z \in \widehat{\mathcal{P}}^*$ such that $\mathbf{i}^*(Y)_Z = 0$. Call these vertices *bad* vertices.

Call an edge e of R *bad* if it contains either a bad cluster Y_1, or two clusters Y_1, Y_2 such that $\mathbf{i}^*(Y_1), \mathbf{i}^*(Y_2)$ are neither identical nor orthogonal to each other. We will show that for any bad edge e, the k-partite subgraph of H corresponding to e contains at least

$(\mu'/4-\varepsilon)\psi^{k-1}\mu^{1/4}n'^k$ edges $e' \in H$ for which $\mathbf{i}_{\widehat{\mathcal{P}}^*}(e') \notin L^{\mu}_{\widehat{\mathcal{P}}^*}(H)$. Noting that there are at most $\mu n^k k^{2k}$ edges of H in total with $\mathbf{i}_{\widehat{\mathcal{P}}^*}(e) \notin L^{\mu}_{\widehat{\mathcal{P}}^*}(H)$, it follows that R contains at most $\mu(nk)^k/(\mu'/4-\varepsilon)\psi^{k-1}\mu^{1/4}n'^k = (\mu k^k/(\mu'/4-\varepsilon)\psi^{k-1}\mu^{1/4})n_R^k \leqslant \sqrt{\mu}n_R^k$ bad edges, where the inequality follows from $\mu \ll \mu', \psi$.

Suppose that e contains a bad cluster Y_1 and that $Z \in \widehat{\mathcal{P}}^*$ is such that $\mathbf{i}^*(Y_1)_Z = 0$, but $|Y_1 \cap Z| \geqslant \mu^{1/4}n'$. Without loss of generality we may suppose that $Z \subseteq V_1$. Then let $Y_2, \dots, Y_k$ be the remaining clusters of e and select $Z_j \in \widehat{\mathcal{P}}^*$ such that $Z_j \subseteq Y_j$ and $\mathbf{i}^*(Y_j)_{Z_j} = 1$ for each Y_j with $j \in [k]$. We set F_1 to be the induced k-partite subgraph of H on $(Y_j \cap Z_j)_{j\in[k]}$ and F_2 the induced k-partite subgraph on $Y_1 \cap Z$ and $(Y_j \cap Z_j)_{2\leqslant j\leqslant k}$. Let F be the induced k-partite subgraph of H on $(Y_i)_{i\in[k]}$. Since $e \in R$, F is (d,ε)-regular for some $d \geqslant \mu'/4$. But F_1 and F_2 are subgraphs of F which contain at least $\psi n' \geqslant \varepsilon n'$ vertices from $k-1$ vertex classes and at least $\mu^{1/4}n' \geqslant \varepsilon n'$ vertices from the remaining vertex class, and hence F_1 and F_2 have density at least $\mu'/4-\varepsilon$, i.e., F_1 and F_2 each have at least $(\mu'/4-\varepsilon)\psi^{k-1}\mu^{1/4}n'^k$ edges, all of which are contained in F.

Note that every edge of F_1 has the same index vector $\mathbf{i}_1$ with respect to $\widehat{\mathcal{P}}^*$, and that every edge of F_2 has the same index vector $\mathbf{i}_2$. Note further that $\mathbf{i}_1 - \mathbf{i}_2 = \mathbf{u}_{Z_1} - \mathbf{u}_Z$, and hence at least one of $\mathbf{i}_1$ and $\mathbf{i}_2$ does not lie in $L^{\mu}_{\widehat{\mathcal{P}}^*}(H)$, since $L^{\mu}_{\widehat{\mathcal{P}}^*}(H)$ is $(1,-1)$-free. So $\mathbf{i}_{\widehat{\mathcal{P}}^*}(e') \notin L^{\mu}_{\widehat{\mathcal{P}}^*}(H)$ for each of the at least $(\mu'/4-\varepsilon)\psi^{k-1}\mu^{1/4}n'^k$ edges $e' \in F_1$ or $e' \in F_2$, respectively.

Alternatively, suppose that e contains two clusters Y_1, Y_2 such that $\mathbf{i}^*(Y_1), \mathbf{i}^*(Y_2)$ are neither identical nor orthogonal to each other. Then we may suppose without loss of generality that there exist $Z_1, Z_1', Z_2 \in \widehat{\mathcal{P}}^*$ with $Z_1, Z_1' \subseteq Y_1$ and $Z_2 \subseteq Y_2$, such that $\mathbf{i}^*(Y_1)_{Z_1} = \mathbf{i}^*(Y_2)_{Z_1'} = \mathbf{i}^*(Y_1)_{Z_2} = \mathbf{i}^*(Y_2)_{Z_2} = 1$. Let $Y_3, \dots, Y_k$ be the remaining clusters of e and select $Z_j \in \widehat{\mathcal{P}}^*$ such that $Z_j \subseteq Y_j$ and $\mathbf{i}^*(Y_j)_{Z_j} = 1$ for each Y_j with $3 \leqslant j \leqslant k$.

Now we define F_1 to be the induced k-partite subgraph of H whose vertex classes are $(Y_j \cap Z_j)_{j\in[k]}$, and F_2 to be the induced k-partite subgraph on vertex classes $Y_2 \cap Z_1', Y_1 \cap Z_2$

and $(Y_i \cap Z_i)_{3 \leqslant j \leqslant k}$. Similarly to the previous case, either F_1 or F_2 will contain at least $(\mu'/4 - \varepsilon)\psi^k n'^k$ edges $e' \in H$ for which $\mathbf{i}_{\widehat{\mathcal{P}}^*}(e') \notin L^{\mu}_{\widehat{\mathcal{P}}^*}(H)$.

For each simple index vector $\mathbf{i}$ with respect to $\widehat{\mathcal{P}}^*$ we define $R_{\mathbf{i}} = \{Y \in V(R) \mid \mathbf{i}^*(Y) = \mathbf{i}\}$. We will show that for any $\mathbf{i}_1 \neq \mathbf{i}_2$ which are not orthogonal we have either $|R_{\mathbf{i}_1}| \leqslant \mu^{1/5} n_R$ or $|R_{\mathbf{i}_2}| \leqslant \mu^{1/5} n_R$. Indeed, if neither of these holds then we have at least $(n_R)^{k-3}|R_{\mathbf{i}_1}||R_{\mathbf{i}_2}|/k! \geqslant \mu^{2/5}(n_R)^{k-1}$ $(k-1)$-tuples of clusters in $V(R)$ with at least one cluster in $R_{\mathbf{i}_1}$ and at least one in $R_{\mathbf{i}_2}$. It follows by a simple argument from the almost-codegree condition that all but at most $2\varepsilon n_R^{k-1}/\gamma$ of these $(k-1)$-tuples are contained in at least n_R/k edges of R each. Hence R contains at least $(\mu^{2/5} - 2\varepsilon/\gamma)n_R^{k-1}(n_R/k) > \sqrt{\mu} n_R^k$ bad edges, which establishes a contradiction.

Let $T = \{\mathbf{i} \in \mathbb{Z}^{\widehat{\mathcal{P}}^*} \mid |R_{\mathbf{i}}| \geqslant \mu^{1/5} n_R\}$. Clearly $U = \bigcup_{\mathbf{i} \in T} R_{\mathbf{i}}$ contains all but at most $\mu^{1/6} n_R$ vertices of R; further, we have shown that T consists of pairwise orthogonal index vectors. Hence $|T| \leqslant \ell$. Further, for each $Z \in \widehat{\mathcal{P}}^*$ there must exist $\mathbf{i} \in T$ such that $i_Z = 1$. Indeed, if there exists Z such that $i_Z = 0$ for every $\mathbf{i} \in T$ then the number of $Y \in V(R)$ with $i^*(Y)_Z = 1$ is at most $\mu^{1/6} n_R$, which contradicts (B3). From this we deduce that $|T| = \ell$ and that for each $Z \in \widehat{\mathcal{P}}^*$ there is precisely one $\mathbf{i} \in T$ with $i_Z = 1$. Hence $|R_{\mathbf{i}}| \geqslant n_R/k - \mu^{1/6} n_R$ for each $\mathbf{i} \in T$.

We define a partition $\mathcal{P}^* = \{X^*_{\mathbf{i}}\}_{\mathbf{i} \in T}$ of V by setting $X^*_{\mathbf{i}} = \bigcup_{Z \in \widehat{\mathcal{P}}^*, i_Z = 1} Z$ for each $\mathbf{i} \in T$. Note that every part of $\mathcal{P}^*$ has size at least $ck^2 n$. Condition (B2) implies that either $R_{\mathbf{i}} = \emptyset$ or $(\mathbf{i} \mid \mathcal{P}) = k\mathbf{u}_W$ for some part $W \in \mathcal{P}$ for any index vector $\mathbf{i}$. Hence $\mathcal{P}^*$ refines $\mathcal{P}$. We will show that $\mathbf{i} \in I^{\mu'_*}_{\widehat{\mathcal{P}}^*}(H)$ for any index vector $\mathbf{i}$ with respect to $\widehat{\mathcal{P}}^*$ such that $(\mathbf{i} \mid \mathcal{P}^*) \in I^{\mu'}_{\mathcal{P}^*}(H)$. Let Z_j be the part of $\widehat{\mathcal{P}}^*$ contained in V_j on which $\mathbf{i}$ is positive and let $\mathbf{i}_j \in T$ be such that $(\mathbf{i}_j)_{Z_j} = 1$, for each $j \in [k]$.

Let U' be the set of good clusters in U. Now there are at least $\mu'(kn)^k$ edges $e' \in H$ such that $\mathbf{i}_{\mathcal{P}^*}(e') = (\mathbf{i} \mid \mathcal{P}^*)$, and at least $\mu'(kn)^k - \varepsilon n^k - \mu'(kn)^k/4 - k^{k+2}\mu^{1/4} n^k \geqslant 2\mu'(kn)^k/3$ of these edges are contained in k-partite k-graphs corresponding to edges of

R and contain no bad vertices. Call these edges *good* edges. Now the number of edges $e \in R$ containing at least one good edge $e' \in H$ is at least $2\mu' n_R^k/3$. At most $\sqrt{\mu} n_R^k$ of these edges are bad and at most $\mu^{1/6} n_R^k$ contain a cluster in $V(R)\backslash U$, and hence at least $2\mu' n_R^k/3 - \sqrt{\mu} n_R^k - \mu^{1/6} n_R^k \geqslant \mu' n_R^k/2$ of the edges e are contained in $R[U']$. Observe that every such e must contain exactly $(\mathbf{i} \mid \mathcal{P}^*)_{X_\mathbf{i}^*}$ clusters in $R_\mathbf{i}$ for each $\mathbf{i} \in T$; otherwise, the k-partite k-graph corresponding to e could not contain any good edges. Order the clusters of e as $(Y_1, \ldots, Y_k)$ so that $\mathbf{i}^*(Y_j) = \mathbf{i}_j$, and hence $\mathbf{i}^*(Y_j)_{Z_j} = 1$, for every $j \in [k]$. Now the sets $(Y_j \cap Z_j)_{j \in [k]}$ each have size at least $\psi n' \geqslant \varepsilon n'$ by (B2), and so by ε-regularity they form a k-partite subgraph of H which contains at least $(\mu'/4 - \varepsilon)\psi^k n'^k$ edges $e \in H$ with $\mathbf{i}_{\widehat{\mathcal{P}}^*}(e) = \mathbf{i}$. Hence in total there are at least $(\mu'/4 - \varepsilon)(\psi^k n'^k)(\mu' n_R^k/2) \geqslant \mu'_*(kn)^k$ edges $e \in H$ with $\mathbf{i}_{\widehat{\mathcal{P}}^*}(e) = \mathbf{i}$, where the inequality follows from $\mu' \ll \psi$. So $\mathbf{i} \in I_{\widehat{\mathcal{P}}^*}^{\mu'_*}(H)$, as desired.

It follows immediately that $\mathbf{i} \in L_{\widehat{\mathcal{P}}^*}^{\mu'_*}(H)$ for any index vector $\mathbf{i}$ with respect to $\widehat{\mathcal{P}}^*$ such that $(\mathbf{i} \mid \mathcal{P}^*) \in L_{\mathcal{P}^*}^{\mu'}(H)$ and $(\mathbf{i} \mid \mathcal{P}')$ is a multiple of $\mathbf{1}$. Recall that each part of $\mathcal{P}^*$ had size at least ck^2n. So the robust maximality of $\mathcal{P}$ implies that $L_{\mathcal{P}^*}^{\mu'}(H)$ contains a $(1,-1)$-vector $\mathbf{u}_{X_1^*} - \mathbf{u}_{X_2^*}$. Let Z_1, Z_2 be the intersections of X_1^* and X_2^* respectively with V_1; now $\mathbf{u}_{Z_1} - \mathbf{u}_{Z_2} \in L_{\widehat{\mathcal{P}}^*}^{\mu'_*}(H)$, contradicting our assumption that $L_{\widehat{\mathcal{P}}^*}^{\mu'_*}(H)$ was $(1,-1)$-free. Hence $\widehat{\mathcal{P}}$ is (c, μ, μ'_*)-robustly maximal with respect to H, proving (A3).

Finally, (A4) can be proved by using Azuma's inequality, similarly to the proof of (E3) of Claim 4.6.7. Fix a vertex $x \in V(H)$ and recall that x is contained in at least dn^{k-1} edges $e \in H$ with $\mathbf{i}_\mathcal{P}(e) \in L_\mathcal{P}^\mu(H)$. Let $E(x)$ denote the number of edges $e \in H'$ with $\mathbf{i}_{\widehat{\mathcal{P}}}^\mu(e) \in L_{\widehat{\mathcal{P}}}^\mu(H)$. To estimate $E(x)$, note that every edge of H containing x is $\mathcal{P}'$-partite (and so belongs to H') with probability at least $1/k!$. Hence $\mathbb{E}(E(x)) \geqslant dn^{k-1}/k!$.

Now fix an order $v_1, \ldots, v_{|V|}$ of the vertices of V. We define a martingale $A_j := \mathbb{E}(E(x)|v_1, \ldots, v_j)$ for $0 \leqslant j \leqslant |V|$, where the 'uncovered event' is which part of $\mathcal{P}'$ contains v_j. Then $A_{|V|} = E(x)$ and $A_0 = \mathbb{E}(E(x))$, and for any $j \in [|V|]$ we have

$|A_j - A_{j-1}| \leqslant (kn)^{k-2}$. So applying Theorem 1.8.3 with $t = dn^{k-1}/2k!(kn)^{k-2} = \Omega(n)$, we obtain

$$\mathbb{P}(E(x) < dn^{k-1}/2k!) \leqslant 2\exp\left(-\frac{t^2}{n}\right) = e^{-\Omega(n)} \leqslant 1/kn^2,$$

and a union bound implies (A4). □

4.7 Proof of Theorem 1.5.2

The main goal of this section is to complete the proof of Theorem 1.5.2, with our main tool being Lemma 4.6.8. We first need some simple lemmas which establish some algebraic properties of the lattices we have been using so far, beginning with a simple application of the pigeonhole principle.

Proposition 4.7.1 *Let $G = (X, +)$ be an abelian group of order m, and suppose that elements $x_i \in X$ for $i \in [r]$ are such that $\sum_{i\in[r]} x_i = x'$. Then $\sum_{i\in I} x_i = x'$ for some $I \subseteq [r]$ with $|I| \leqslant m-1$.*

Proof. It suffices to show that if $r > m-1$ then $\sum_{i\in[r]} x_i = \sum_{i\in I} x_i$ for some $I \subseteq [r]$ with $|I| < r$. To see this, note that there are $r+1 > m$ partial sums $\sum_{i\in[j]} x_i$ for $j \in [r]$, so by the pigeonhole principle some two must be equal, that is, there exist $j_1 < j_2$ so that $\sum_{i\in[j_1]} x_i = \sum_{i\in[j_2]} x_i$. Then

$$\sum_{i\in[r]} x_i = \sum_{i\in[r]\setminus\{j_1+1,\dots,j_2\}} x_i + \sum_{i\in[j_2]} x_i - \sum_{i\in[j_1]} x_i = \sum_{i\in[r]\setminus\{j_1+1,\dots,j_2\}} x_i,$$

as required. □

Lemma 4.7.2 *Let $1/n \ll \mu \ll \gamma$.*

(i) *Let H be a k-graph on n vertices with $\delta(H) \geqslant (1/k+\gamma)n$. Let $\mathcal{P}$ be a partition of $V(H)$ into d parts of size at least $(1/k+\gamma)n$ and such that $L^\mu_{\mathcal{P}}(H)$ is $(1,-1)$-free.*

Let $L = L^{\mu}_{\mathcal{P}}(H)$, and let L_{max} be the lattice of vectors in $\mathbb{Z}^d$ whose coordinates sum to a multiple of k. Then $|L_{max}/L| \leqslant d$.

(ii) *Let $\mathcal{P}'$ be a partition of V into sets $V_1, \dots, V_k$ of size n. Let H be a $\mathcal{P}'$-partite k-graph on V, and let $\mathcal{P}$ be a refinement of $\mathcal{P}'$ which splits each part of $\mathcal{P}'$ into d parts of size at least $n/k + \gamma n$ and such that $L^{\mu}_{\mathcal{P}}(H)$ is $(1, -1)$-free. Let $L = L^{\mu}_{\mathcal{P}}(H)$, and let L_{max} be the lattice of vectors $\mathbf{i} \in \mathbb{Z}^{kd}$ such that $(\mathbf{i} \mid \mathcal{P}')$ is a multiple of $\mathbf{1}$. Then $|L_{max}/L| \leqslant d$.*

Proof. (i) Let $\mathbf{i}'$ be a non-negative index vector with respect to $\mathcal{P}$ such that $\sum_{X \in \mathcal{P}} i_X = k - 1$. We will show that every coset of L/L_{max} contains an index vector $\mathbf{i} = \mathbf{i}' + \mathbf{u}_{X'}$ for some $X' \in \mathcal{P}$. Indeed, by Proposition 4.4.6(i) there exists $X \in \mathcal{P}$ such that $\mathbf{i}' + \mathbf{u}_X \in L$. Now let $N = L + \mathbf{v}$ be a coset of L/L_{max}. By Proposition 4.4.6(v) we obtain $X' \in \mathcal{P}$ such that $-\mathbf{v} - \mathbf{u}_X + \mathbf{u}_{X'} \in L$. But now $\mathbf{i}' + \mathbf{u}_{X'} = (\mathbf{i}' + \mathbf{u}_X) + (-\mathbf{v} - \mathbf{u}_X + \mathbf{u}_{X'}) + \mathbf{v} \in N$.

It follows immediately that there are at most d cosets. □

(ii) Choose a part $W \in \mathcal{P}'$ and let $\mathbf{i}'$ be a non-negative index vector with respect to $\mathcal{P}$ such that $(\mathbf{i}' \mid \mathcal{P}') = \mathbf{1} - \mathbf{u}_W$. We will show that every coset of L/L_{max} contains an index vector $\mathbf{i} = \mathbf{i}' + \mathbf{u}_{X'}$ for some $X' \in \mathcal{P}$ with $X' \subseteq W$. Indeed, by Proposition 4.4.5 (i) there exists $X \in \mathcal{P}$ such that $\mathbf{i}' + \mathbf{u}_X \in L$. Now let $N = L + \mathbf{v}$ be a coset of L/L_{max}. By Proposition 4.4.5 (vii) we obtain $X' \in \mathcal{P}$ such that $-\mathbf{v} - \mathbf{u}_X + \mathbf{u}_{X'} \in L$. But now $\mathbf{i}' + \mathbf{u}_{X'} = (\mathbf{i}' + \mathbf{u}_X) + (-\mathbf{v} - \mathbf{u}_X + \mathbf{u}_{X'}) + \mathbf{v} \in N$.

It follows immediately that there are at most d cosets, since there are only d such sets X'. □

The aim of the following lemma is to allow us to remove inconvenient edges of a k-graph H, so that we may obtain a correspondence between small index vectors in $L_{\mathcal{P}}(H)$ and the index vectors which can actually be realised as the index vectors of matchings in

H. We do this by eliminating any edges e for which H contains few other (disjoint) edges of index vector $\mathbf{i}_{\mathcal{P}}(e)$ with respect to $\mathcal{P}$.

Lemma 4.7.3 *Let H be a k-graph on n vertices and let $H_1, H_2, \ldots, H_m$ be edge-disjoint sub-k-graphs of H. Let $C \geqslant k+2$ and $D \geqslant k+1$ be integers. Then there exists $D \leqslant f \leqslant C^m D$, a set $S \subseteq [m]$ and a matching $M^* \subseteq H$ such that the following conditions hold:*

(i) *M^* contains exactly Cf edges in H_i for every $i \in S$,*

(ii) *Any matching in $\bigcup_{i' \notin S} H_{i'}$ has size at most mf.*

Proof. For each $S \subseteq [m]$, let $g(S)$ be the largest integer such that there exists a matching $M \subseteq H$ with $|H_i \cap M| \geqslant g(S)$ for every $i \in S$ (if $S = \emptyset$ then we set $g(S) = \infty$). We will choose $D \leqslant f \leqslant C^m D$ and $S \subseteq [m]$ such that $g(S) \geqslant Cf$ and $g(S') \leqslant f$ for every strict superset S' of S. Let $S^{(1)} = \emptyset$; we then proceed iteratively by setting $S^{(j+1)}$ to be a superset of $S^{(j)}$ such that $g(S') \geqslant C^{m-j+2} D$ for each $1 \leqslant j \leqslant m+1$. If no such S' exists, then the procedure terminates and we set $f = C^{m-j+1} D$ and $S = S^{(j)}$, so that $g(S) \geqslant Cf$. Note that $|S^{(j+1)}| \geqslant |S^{(j)}| + 1$ for each j; since $|S^{(j)}| \leqslant m$ for each j, it follows that the procedure must terminate with $j \leqslant m+1$. Now since $g(S) \geqslant Cf$ there exists a matching M^* which contains exactly Cf edges in H_i for every $i \in S$. Clearly M^* satisfies (i), and so it suffices to show that it also satisfies (ii). Note that by our choice of f and S,

$$g(S \cup \{i'\}) \leqslant f \tag{4.7.4}$$

for any $i' \notin S$.

Suppose for a contradiction that $\bigcup_{i' \notin S} H_{i'}$ contains a matching M' of size greater than mf. Then by the pigeonhole principle there exists $i_0 \notin S$ such that $|M' \cap H_{i_0}| \geqslant f+1$. Let M'_{i_0} be a submatching of $M' \cap H_{i_0}$ of size $f+1$ and note that $M_S - V(M'_{i_0})$ still contains at least $Cf - k(f+1) \geqslant f+1$ edges in H_i for every $i \in S$. Hence $(M_S - V(M'_{i_0})) \cup M'_{i_0}$ is a matching in H which contains $f+1$ edges in H_i for every $i \in S \cup \{i_0\}$, which contradicts (4.7.4). So $\bigcup_{i' \notin S} H_{i'}$ contains at most mf disjoint edges, which proves (ii).

$\square$

We also need to be able to ensure that every vertex in $V(H)$ is contained in many edges of H whose index vectors lie in $L^{\mu}_{\mathcal{P}}(H)$. We use this technique, rather than finding a matching which covers the offending vertices, since the latter would be inconvenient in the proof.

Lemma 4.7.5 *Let $1/n \ll \mu \ll \mu' \ll c \ll 1/k$. Let H be a k-graph such that $\delta_{k-1}(H) \geqslant n/k$ and let $\mathcal{P}$ be a partition of $V(H)$ with parts of size at least n/k which is (c, μ, μ')-robustly maximal with respect to H. Then there exists a partition $\mathcal{P}'$ of $V(H)$ such that:*

(i) $\mathcal{P}'$ is $(2c, \sqrt{\mu}, \mu'/2)$-robustly maximal with respect to H,

(ii) $L^{\sqrt{\mu}}_{\mathcal{P}'}(H) = L^{\mu}_{\mathcal{P}}(H)$ and

(iii) Every vertex $v \in V(H)$ is contained in at least $n^{k-1}/2k^k$ edges $e \in H$ with $\mathbf{i}_{\mathcal{P}'}(e) \in L^{\sqrt{\mu}}_{\mathcal{P}'}(H)$.

Proof. Note that the minimum vertex degree of H is at least $(n/k)^{k-1}$. Set $\psi = k^{2k}\mu$. We first show that all but at most ψn vertices of H are contained in at least n^{k-1}/k^k edges $e \in H$ with $\mathbf{i}_{\mathcal{P}}(e) \in L^{\mu}_{\mathcal{P}}(H)$. To see this, suppose for a contradiction that there exists a set $B \subseteq V(H)$ of ψn vertices which are each contained in fewer than n^{k-1}/k^k such edges. By the Inclusion-Exclusion principle there are at least $\psi(1/k^{k-1} - 1/k^k)n^k - \binom{\psi n}{2}\binom{n-\psi n}{k-2} \geqslant \psi n^k/k^k = k^k \mu n^k$ edges containing a vertex of B such that $\mathbf{i}_{\mathcal{P}}(e) \notin L^{\mu}_{\mathcal{P}}(H)$. But this is a contradiction since there are fewer than $k^k \mu n^k$ such edges in total.

For each vertex v which is contained in fewer than n^{k-1}/k^k edges $e \in H$ with $\mathbf{i}_{\mathcal{P}}(e) \in L^{\mu}_{\mathcal{P}}(H)$, let X be the part of $\mathcal{P}$ containing v and define a function $g_X : H \to \mathcal{P}$ as follows: For each edge $e \in H$, let X' be any part of $\mathcal{P}$ such that $\mathbf{i}_{\mathcal{P}}(e) - \mathbf{u}_X + \mathbf{u}_{X'} \in L^{\mu}_{\mathcal{P}}(H)$. Note that g_X is well-defined by Proposition 4.4.6(v). Now by the pigeonhole principle there exists $X' \in \mathcal{P}$ such that $g_X(e) = X'$ for at least n^{k-1}/k^k edges $e \in H$ containing v. We move v into X' and note that now $\mathbf{i}_{\mathcal{P}}(e) \in L^{\mu}_{\mathcal{P}}(H)$ for all of these edges. By Proposition 4.4.6 (ii), $L^{\sqrt{\mu}}_{\mathcal{P}'}(H) = L^{\mu}_{\mathcal{P}}(H)$. By Lemma 4.5.3, $\mathcal{P}'$ is still $(2c, \sqrt{\mu}, \mu'/2)$-robustly maximal after this modification; further, after the modification is complete every vertex is contained in at least $n^{k-1}/k^k - \psi n^{k-1} \geqslant n^{k-1}/2k^k$ edges $e \in H$ with $\mathbf{i}_{\mathcal{P}'}(e) \in L^{\sqrt{\mu}}_{\mathcal{P}'}(H)$. $\square$

Finally, the following lemma essentially covers the forward implication in Theorem 1.5.2. Once we have established this, we can proceed to prove the theorem in full.

Lemma 4.7.6 *Let $1/n_0 \ll \mu \ll \gamma$. Let H be a k-graph on $n \geqslant n_0$ vertices with $\delta(H) \geqslant (1/k + \gamma)n$ and which contains a perfect matching M. Let $\mathcal{P}$ be a partition of $V(H)$ into parts of size at least $(1/k + \gamma)n$ and such that $L^{\mu}_{\mathcal{P}}(H)$ is $(1, -1)$-free. Then there exists a submatching $M' \subseteq M$ of size at most $k - 2$ such that $\mathbf{i}(V(H) \setminus V(M')) \in L^{\mu}_{\mathcal{P}}(H)$*

Proof. Let $d \leqslant k - 1$ be the number of parts into which $\mathcal{P}$ partitions $V(H)$. Let $L = L^{\mu}_{\mathcal{P}}(H)$. Since $V(M) = V(H)$ we have $\mathbf{i}(V(H) \setminus V(M)) = \mathbf{0} \in L$. Define L_{max} as in Lemma 4.7.2, let $G = L_{max}/L$ and note that by Lemma 4.7.2 (i) $|G| \leqslant d \leqslant k - 1$. Now let $M = \{e_1, e_2, \dots, e_{n/k}\}$ and let $x_j = \mathbf{i}_{\mathcal{P}}(e_j) + L$ for each $j \in [n/k]$. Now $\sum_{j=1}^{n/k} x_j = 0$, and hence by Proposition 4.7.1 there exists a set $S \subseteq [n/k]$ with $|S| \leqslant k - 2$ such that $\sum_{j \in S} x_j = 0$. Now taking $M' = \{e_j \mid j \in S\}$ we have $\mathbf{i}(V(H) \setminus V(M')) \in L^{\mu}_{\mathcal{P}}(H)$. $\square$

Proof of Theorem 1.5.2. Choose n_0 so that $1/n_0 \ll 1/k, \gamma, 1/C_0$ (and hence $1/n \ll 1/k, \gamma, 1/C_0$). It suffices to show that (i) implies (ii) when $C = C_0$ and that (ii) implies (i) when $C = k^{k^k}$. To prove the first implication, suppose that H contains a perfect matching M. Let $\mathcal{P}$ be a partition of $V(H)$ into parts of size at least n/k and let L be a $(1, -1)$-free lattice such that H contains fewer than C disjoint edges e with $\mathbf{i}_{\mathcal{P}}(e) \notin L$. Choose $1/n \ll \mu \ll \gamma$. It follows immediately that $L^{\mu}_{\mathcal{P}}(H) \subseteq L$ and hence that $L^{\mu}_{\mathcal{P}}(H)$ is

$(1,-1)$-free. Now by Lemma 4.7.6 there exists a submatching $M' \subseteq M$ of size at most $k-2$ such that $\mathbf{i}_{\mathcal{P}}(V \setminus V(M')) \in L^{\mu}_{\mathcal{P}}(H)$ and hence $\mathbf{i}_{\mathcal{P}}(V \setminus V(M')) \in L$.

We now prove the second implication. Choose new constants $\mu_1, \mu_2, \ldots, \mu_k, c, d$ such that $1/n \ll \mu_1 \ll \mu_2 \ll \ldots \ll \mu_k \ll c, d, \gamma$ and apply Proposition 4.5.2 to obtain $j \in [k-1]$ and a partition $\mathcal{P}$ of $V(H)$ into parts of size at least cn which is $(c/4, \mu_j, \mu_{j+1})$-robustly maximal with respect to H. Proposition 4.4.6 implies that every part of $\mathcal{P}$ has size at least n/k. Let $\mu = 2\sqrt{\mu_j}$ and $\mu' = \mu_{j+1}/4$, so that $\mathcal{P}$ is $(c/4, \mu^2/4, 4\mu')$-robustly maximal with respect to H. Let $L = L^{\mu^2/4}_{\mathcal{P}}(H)$ and note that L is $(1,-1)$-free. We now apply Lemma 4.7.5 to obtain a partition $\mathcal{P}'$ of $V(H)$ which is $(c/2, \mu/2, 2\mu')$-robustly maximal with respect to H, and such that $L^{\mu/2}_{\mathcal{P}'}(H) = L$ and every vertex of H is contained in at least $2dn^{k-1}$ edges $e \in H$ with $\mathbf{i}_{\mathcal{P}'}(e) \in L$.

We will now define a new partition $\mathcal{P}''$ and a lattice L' such that $\mathcal{P}'$ refines $\mathcal{P}''$, L' is $(1,-1)$-free and H contains fewer than k^{k^k} disjoint edges e with $\mathbf{i}_{\mathcal{P}''}(e) \notin L'$; this will allow us to apply property (iii). For each non-negative index vector $\mathbf{i}$ with respect to $\mathcal{P}'$, let $H_{\mathbf{i}} = \{e \in H \mid \mathbf{i}_{\mathcal{P}'}(e) = \mathbf{i}\}$. Note that the number of such $\mathbf{i}$ is at most $(k-1)^k$. We apply Lemma 4.7.3 to the collection $(H_{\mathbf{i}})_{\mathbf{i} \notin L}$ with $C = k+2$ and $D = k+1$ to obtain an integer $k+1 \leqslant f \leqslant (k+2)^{(k-1)^k}(k+1)$, a set S of index vectors with respect to $\mathcal{P}'$ and a matching M^* which contains exactly $(k+2)f$ edges e, such that $\mathbf{i}_{\mathcal{P}'}(e) = \mathbf{i}$ for every $\mathbf{i} \in S$, and such that there are at most $k^k f \leqslant k^{k^k}$ edges $e \in H$ with $\mathbf{i}_{\mathcal{P}'}(e) \notin L \cup S$ (the inequality follows from $k^k - k \geqslant 2(k-1)^k + 2$ and $k^2 \geqslant k+2$).

Now let L^* be the lattice generated by $L \cup S$. Define an equivalence relation $\sim$ on $\mathcal{P}'$ by setting $X_1 \sim X_2$ whenever L^* contains $\mathbf{u}_{X_1} - \mathbf{u}_{X_2}$, and let $\mathcal{P}''$ be the partition formed by taking unions of equivalence classes under $\sim$. Let $L' \subseteq \mathbb{Z}^{|\mathcal{P}''|}$ be the natural restriction $L' = \{(\mathbf{i} \mid \mathcal{P}'') \mid \mathbf{i} \in L^*\}$ of L^* under $\sim$, and note that L' is $(1,-1)$-free.

For any edge $e \in H$ such that $\mathbf{i}_{\mathcal{P}'}(e) \in L^*$, we have $\mathbf{i}_{\mathcal{P}''}(e) = (\mathbf{i}_{\mathcal{P}'}(e) \mid \mathcal{P}'') \in L'$. Suppose for a contradiction that are at least k^{k^k} disjoint edges $e \in H$ such that $\mathbf{i}_{\mathcal{P}''}(e) \notin L'$. Then for each such e we have $\mathbf{i}_{\mathcal{P}'}(e) \notin L^*$ and hence $\mathbf{i}_{\mathcal{P}'}(e) \notin L \cup S$, contradicting our choice of S.

By (ii), there exists a matching M' in H of size at most $k-2$ such that $\mathbf{i}_{\mathcal{P}''}(V \backslash V(M')) \in L'$. Now let X be any part of $\mathcal{P}'$. By Proposition 4.4.6 (v) there exists $X' \in \mathcal{P}'$ such that $\mathbf{i}_{\mathcal{P}'}(V \backslash V(M')) - \mathbf{u}_X + \mathbf{u}_{X'} \in L \subseteq L^*$. Note that $(\mathbf{i}_{\mathcal{P}'}(V \backslash V(M')) - \mathbf{u}_X + \mathbf{u}_{X'} \mid \mathcal{P}'') \in L'$ and hence $(\mathbf{u}_X - \mathbf{u}_{X'} \mid \mathcal{P}'') \in L'$. Since L' is $(1,-1)$-free, this implies that $(\mathbf{u}_X - \mathbf{u}_{X'} \mid \mathcal{P}'') = \mathbf{0}$, i.e., that X and X' are contained in the same part of $\mathcal{P}''$. Hence $\mathbf{u}_X - \mathbf{u}_{X'} \in L^*$ and thus $\mathbf{i}_{\mathcal{P}}(V \backslash V(M')) \in L^*$. We will show that further there exists a disjoint matching M'' in H such that $\mathbf{i}_{\mathcal{P}'}(V \backslash V(M' \cup M'')) \in L$. Let $M^{**} = M^* - V(M')$ and note that M^{**} contains at least $(k+2)f - k(k-2) \geqslant k-2$ edges in $H_{\mathbf{i}}$ for every $\mathbf{i} \in S$.

Let $G = L^*/L$. It follows immediately from Lemma 4.7.2 (i) that $|G| \leqslant k-1$. Now write $\mathbf{i}_{\mathcal{P}'}(V \backslash V(M'))$ as a sum $\sum_{j=1}^{m} \mathbf{i}_j$, where $\mathbf{i}_j \in S$ for each $j \in [m]$. For each $j \in [m]$, let $\mathbf{i}_j + L$ be the coset of L containing $\mathbf{i}_j$. Note that $\sum_{j=1}^{m}(\mathbf{i}_j + L) = \mathbf{i}_{\mathcal{P}'}(V \backslash V(M')) + L$. Now by Lemma 4.7.1 we can write $\mathbf{i}_{\mathcal{P}}(V \backslash V(M')) + L = \sum_{j \in T} \mathbf{i}_j + L$ for some $T \subseteq [m]$ with $|T| \leqslant k-2$. Choose edges $(e_j)_{j \in T}$ of M^{**} such that $\mathbf{i}_{\mathcal{P}}(e_j) = \mathbf{i}_j$ for each $j \in T$, and let M'' be the submatching consisting of these edges. Let $V' = V \backslash V(M' \cup M'')$; now $\mathbf{i}_{\mathcal{P}'}(V') \in L$.

Now consider the k-graph $H' = H[V']$. By Lemma 4.5.3 $\mathcal{P}'$ is still (c, μ, μ')-robustly maximal with respect to H', and by Lemma 4.4.1 $L^{\mu}_{\mathcal{P}'}(H') = L$ and hence $\mathbf{i}_{\mathcal{P}'}(V') \in L^{\mu}_{\mathcal{P}'}(H')$. Further, every vertex is clearly contained in at least dn^{k-1} edges $e \in H'$ with $\mathbf{i}_{\mathcal{P}'}(e) \in L$. We choose additional constants D and ε such that $1/n \ll 1/D \ll \varepsilon \ll \mu$ and apply Lemma 4.6.8 to obtain a perfect matching in H'. Now the union of this matching with M' and M'' is a perfect matching in H. $\square$

Chapter 5

A note on Maker-Breaker games

In this chapter we construct a 4-graph G_4 of maximum degree 3, on which Maker wins the Maker-Breaker game. Since G_4 is rather large and intricate, we first give a 3-graph Γ which contains the main essential structure which allows Maker to win. We then derive a further hypergraph Γ' from Γ and finally derive G_4 from Γ'. Each of these transformations is relatively straightforward; in effect, we successively replace 3-edges of Γ by collections of 4-edges (which in general involves adding extra vertices to Γ), while preserving the low maximum degree and the property that Maker wins the Maker-Breaker game on each graph.

5.1 Construction of Γ

The 3-graph Γ is constructed as follows: Let $W = \{w_i \mid i \in [5]\}$, $X = \{x_{ij} \mid i \in [5], j \in [3]\}$ and $T = \{t_{ij} \mid i \in [5], j \in [3]\}$. Then

$$V(\Gamma) = W \cup X \cup T.$$

Let $e_i = x_{i1}x_{(i+2)2}x_{(i+3)3}$ for each $i \in [5]$, where x_{ij} denotes $x_{i(j-5)}$ for $j \geqslant 6$. Then the edges of Γ are

$$\{w_i x_{ij} t_{ij} \mid i \in [5], j \in [3]\} \cup \{e_1, \dots, e_5\}.$$

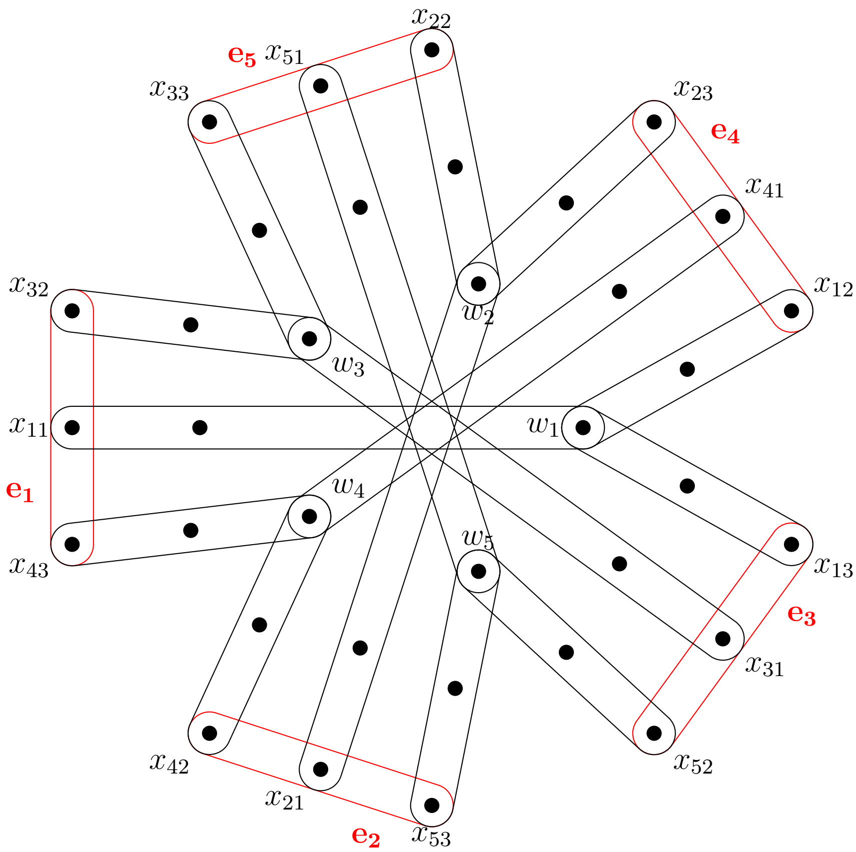

Figure 5.1: The graph Γ

(See Figure 5.1; vertices in T are not labelled.) In total Γ has 35 vertices and 20 edges. Note that the vertices w_i have degree 3 for each i, the vertices t_{ij} have degree 1 and the remaining vertices have degree 2.

Lemma 5.1.1 *Maker has a winning strategy for the Maker-Breaker game on Γ, where Breaker goes first.*

Proof. Roughly speaking, Maker's strategy is to gain an advantage by playing on the vertices $\{w_i \mid i \in [5]\}$ (the 'inner' vertices) and then to use this advantage to claim one of the edges $e_1, \ldots, e_5$. By claiming a vertex w_i when x_{ij} and t_{ij} are unclaimed for each $j \in [3]$, Maker effectively gets two 'outer' vertices in a single turn: Breaker can claim one of the x_{ij}, but Maker will then claim the other two and force a response at the corresponding t_{ij} each time. The effect is even greater if Maker has already claimed one of the x_{ij}; in this case, Breaker is forced to respond immediately at t_{ij} and Maker gets yet more vertices effectively for free. By contrast, if Breaker claims an inner vertex w_i he accomplishes comparatively little. Maker was not planning to win on the edges incident to w_i anyway, and he can simply use the above tactic with a different inner vertex.

We now give a rigorous description of Maker's winning strategy for the Maker-Breaker game on Γ, where Breaker goes first. Whenever Breaker plays at a vertex t_{ij} and Maker has not claimed both w_i and x_{ij}, Maker plays as if Breaker had played at w_i or x_{ij}, whichever is free (or arbitrarily if neither is free). In other respects the winning strategy is as follows:

Case 1: Breaker plays at w_i for some i. Without loss of generality Breaker plays at w_1. Maker then plays at w_2. The strategy then branches depending on Breaker's next move:

Case 1.1: Breaker plays at a vertex of e_4 or at w_4. Then Maker plays at x_{21} (forcing t_{21}), then at x_{22} (forcing t_{22}) and then at x_{51}. Breaker is forced to play at x_{33} or Maker wins immediately by playing there and claiming e_5. Then Maker plays at w_5 (forcing t_{51}), x_{53} (forcing t_{53}), and then at x_{42}, claiming the edge e_2 and winning.

Case 1.2: Breaker plays at a vertex of e_2. Then Maker plays at x_{22} (forcing t_{22}), x_{23} (forcing t_{23}) and then at x_{41}. Breaker is forced to play at x_{12} or Maker wins immediately by playing there and claiming e_4. Then Maker plays at w_4 (forcing t_{41}), x_{43} (forcing t_{43}) and then at x_{32}. Breaker is forced to play at x_{11} or Maker wins immediately by playing there and claiming e_1. Then Maker plays at w_3 (forcing t_{32}), x_{33} (forcing t_{33}), and then at x_{51}, claiming the edge e_5 and winning.

Case 1.3: Breaker plays at any other unclaimed vertex of Γ. Then Maker plays at x_{21} (forcing Breaker to play at t_{21}), then at x_{23} (forcing t_{23}) and then at x_{41}. Breaker is forced to play at x_{12} or Maker wins immediately by playing there and claiming e_4. Then

Maker plays at w_4 (forcing t_{41}), x_{42} (forcing t_{42}), and then at x_{53}, claiming the edge e_2 and winning.

Case 2: Breaker plays at x_{i1} for some i. Without loss of generality Breaker plays at x_{11}. Maker then plays at w_2. The strategy then branches depending on Breaker's next move:

Case 2.1: Breaker plays at a vertex of e_4 or at w_4. Then Maker wins as in Case 1.1, which is possible since Breaker's first move at x_{11} does not prevent this.

Case 2.2: Breaker plays at a vertex of e_2. Then Maker plays at x_{22} (forcing t_{22}), x_{23} (forcing t_{23}) and then at x_{12}. Breaker is forced to play at x_{41} or Maker wins immediately by playing there and claiming e_4. Then Maker plays at w_1 (forcing t_{12}), x_{13} (forcing t_{13}) and then at x_{52}. Breaker is forced to play at x_{31} or Maker wins immediately by playing there and claiming e_3. Then Maker plays at w_5 (forcing t_{52}), x_{51} (forcing t_{51}) and then at x_{33}, claiming the edge e_5 and winning.

Case 2.3: Breaker plays at any other unclaimed vertex. Then Maker wins as in Case 1.3.

Case 3: Breaker plays at x_{i2} or x_{i3} for some i. Without loss of generality Breaker plays at x_{32}. Maker then plays at w_2. The strategy then branches depending on Breaker's next move:

Case 3.1: Breaker plays at a vertex of e_4 or at w_4. Then Maker wins as in Case 1.1.

Case 3.2: Breaker plays at a vertex of e_2. Then Maker wins as in Case 2.2.

Case 3.3: Breaker plays at any other unclaimed vertex. Then Maker wins as in Case 1.3. $\square$

5.2 Construction of G_4 from Γ

We derive the (non-uniform) hypergraph Γ' from Γ as follows: Add new vertices $(y_{ijk})_{k\in[6]}$ and $(z_{ijk})_{k\in[4]}$ to Γ for every $i \in [5]$ and $j \in [3]$. Then for every $i \in [5]$ and $j \in [3]$, writing $w = w_i$, $x = x_{ij}$, $t = t_{ij}$, $y_k = y_{ijk}$ for $k \in [6]$ and $z_k = z_{ijk}$ for $k \in [4]$, we replace the edge wxt by the edges

$$wty_1y_2,\ xty_3y_4,\ wty_5y_6$$

and

$$y_1y_3y_5z_1,\ y_1y_3y_5z_2,\ y_2y_4y_6z_3,\ y_2y_4y_6z_4.$$

(See Figure 5.2.)

In total Γ' has 155 vertices and 110 edges. Note that all of the edges of Γ', apart from the edges $e_1, \ldots, e_5$, have size 4, since all of the other original edges of Γ have been replaced and every edge we added had size 4. Further, Γ' has maximum degree 3. Indeed, at each of the vertices of W we replaced each 3-edge by a single 4-edge, at the vertices of X we left one 3-edge as it is and replaced the other by two 4-edges, and Figure 5.2 illustrates that the vertices z_{ijk} have degree 1 and the remaining vertices degree 3.

Lemma 5.2.1 *Maker has a winning strategy for the Maker-Breaker game on Γ', where Breaker goes first.*

Proof. Initially Maker plays only on the vertices $\{w_i \mid i \in [5]\}$ and $\{x_{ij} \mid i \in [5], j \in [3]\}$ and plays according to the winning strategy for Γ from Lemma 5.1.1. If Breaker plays at y_{ijk} or z_{ijk} for some i, j, k then Maker plays as if Breaker had played at t_{ij}. (If Breaker had previously played at t_{ij} then we choose a free x_{ij} arbitrarily and play as if Breaker had played there instead.)

Since Maker is following the winning strategy for Γ, at some point he will claim an edge of Γ. If this edge is also an edge of Γ' then he wins immediately, so we may assume

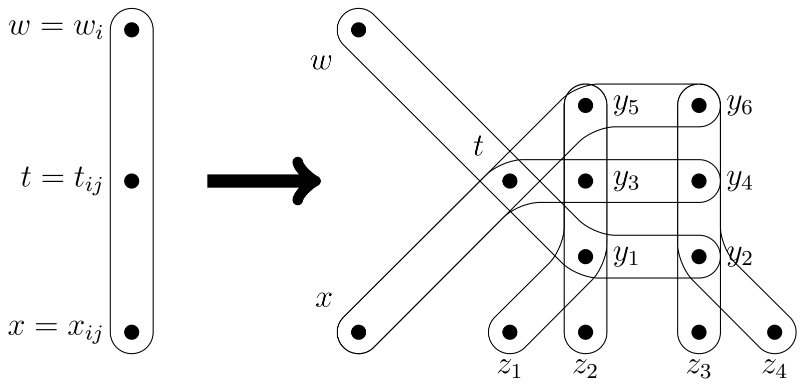

Figure 5.2: Forming Γ' from Γ

that Maker claims $w_i x_{ij} t_{ij}$ for some $i \in [5]$ and $j \in [3]$. Since Maker claimed t_{ij}, Breaker cannot have played at y_{ijk} or z_{ijk} for any k. Since moves by Breaker apart from y_{ijk} or z_{ijk} for some k will not affect what follows, we may assume that Breaker's next move is at one of these vertices. Suppose that Breaker's next move is either y_{ij1}, z_{ij1} or z_{ij2} (other possible moves are covered in a similar way). Now Maker plays at y_{ij4}. Breaker is forced to play at y_{ij3}, or Maker plays there and wins. Maker then plays at y_{ij6}, forcing Breaker to play at y_{ij5}. Finally Maker plays at y_{ij2}, and now takes one of z_{ij3} and z_{ij4} and wins. $\square$

Now we construct a 4-graph G_4 as follows: Form the disjoint union of three copies $\Gamma'_1, \Gamma'_2, \Gamma'_3$ of Γ', and let $e_{i1}, \ldots, e_{i5}$ be the 3-edges of Γ'_i for each $i \in [3]$. Add extra vertices $v_1, v_2, v_3, v_4, s_1, s_2, s_3$. Now replace the edges e_{i1} and e_{i2} by $e_{i1} \cup v_i$ and $e_{i2} \cup v_i$, and replace the edges e_{i3}, e_{i4}, e_{i5} by $e_{i3} \cup s_i, e_{i4} \cup s_i, e_{i5} \cup s_i$ for each $i \in [3]$. Finally add the edge $v_1 v_2 v_3 v_4$. (See Figure 5.3.)

Note that G_4 is indeed a 4-graph, since we replaced each 3-edge e_{ij} by either the 4-edge $e_{ij} \cup v_i$ or the 4-edge $e_{ij} \cup s_i$. Further, it is easily seen that G_4 has maximum degree 3. In total G has 472 vertices and 331 edges.

Proposition 5.2.2 *Maker has a winning strategy for the Maker-Breaker game on G_4,*

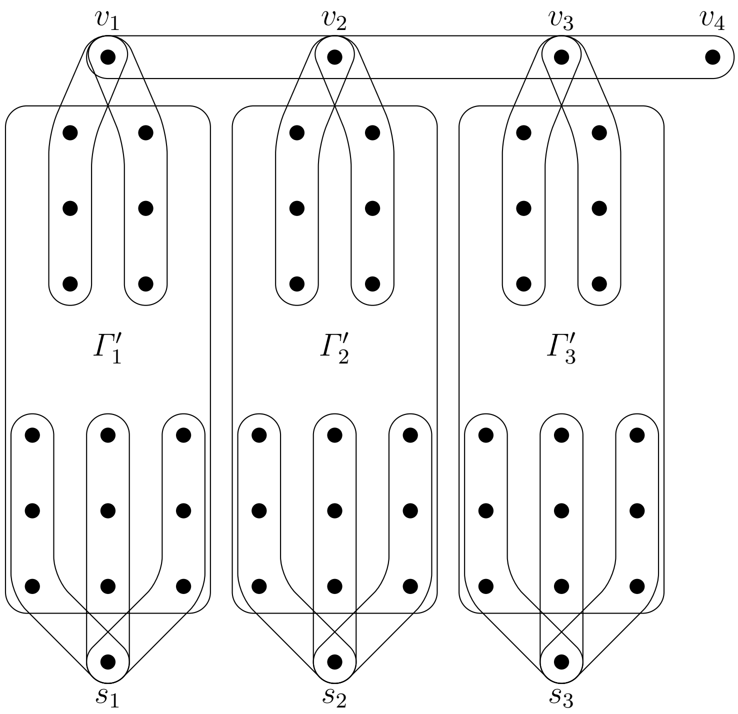

Figure 5.3: Construction of G_4, using three copies of Γ'

where Maker goes first.

Proof. Firstly, suppose that at some point during the game Maker plays at v_i for some $i \in [3]$, such that Breaker has not yet played at s_i or at any vertex of Γ_i'. Suppose further that Breaker does not immediately respond at any of these vertices. Then Maker plays at s_i and can now use the Breaker-first winning strategy on Γ_i' from Lemma 5.2.1. Thus Maker claims an edge of Γ_i'. But any 4-edge of Γ_i' is also an edge of G_4 and if Maker claims a 3-edge e_{ij} then he also claims either the 4-edge $e_{ij} \cup v_i$ or the 4-edge $e_{ij} \cup s_i$ (whichever is an edge of G_4). Hence Maker wins in either case.

So we may assume that every time Maker plays at v_i for some $i \in [3]$, Breaker immediately responds at s_i or at some vertex of Γ_i', assuming he has not done so already. But in this case, Maker simply claims each v_i in turn, ending with v_4, and wins. $\square$

List of References

[1] R. Aharoni, A. Georgakopoulos and P. Sprüssel, Perfect matchings in r-partite r-graphs, *European J. of Combin.* **30** (2009), 39–42.

[2] M. Ajtai, J. Komlós and E. Szemerédi, The first occurrence of Hamilton cycles in random graphs, *Ann. Disc. Math.* **27** (1985), 173–178.

[3] N. Alon, P. Frankl, H. Huang, V. Rödl, A. Ruciński, B. Sudakov, Large matchings in uniform hypergraphs and the conjectures of Erdős and Samuels, *J. Combin. Theory A* **119** (2012), 1200–1215.

[4] N. Alon, P. Frankl and L. Lovasz, The chromatic number of Kneser hypergraphs, *Transactions of the American Mathematical Society* **298** (1986), 359–370.

[5] N. Alon and J. H. Spencer, *The Probabilistic Method (Second Ed.)*, John Wiley & Sons, USA, 2000.

[6] N. Alon and R. Yuster, H-factors in dense graphs, *J. Combin. Theory B* **66** (1996), 269–282.

[7] B. Alspach, J.-C. Bermond and D. Sotteau, Decompositions into cycles. I. Hamilton decompositions, *Cycles and rays (Montreal, PQ, 1987)*, Kluwer Acad. Publ., Dordrecht, 1990, 9–18.

[8] B. Alspach, D. Bryant and D. Dyer, Paley graphs have Hamilton decompositions, preprint.

[9] A. Asadpour, U. Feige and A. Saberi, Santa Claus meets hypergraph matchings, *ACM Transactions on Algorithms* **8** (2012), No. 24.

[10] D. Bal and A. Frieze, Packing tight Hamilton cycles in uniform hypergraphs, preprint.

[11] J. Beck, *Combinatorial Games: Tic-Tac-Toe Theory*, Cambridge University Press, New York, 2008.

[12] J. Balogh, A.V. Kostochka and A. Treglown, On perfect packings in dense graphs, preprint.

[13] S. Ben-Shimon, M. Krivelevich, and B. Sudakov, On the resilience of Hamiltonicity and optimal packing of Hamilton cycles in random graphs, preprint.

[14] B. Bollobás, The evolution of sparse graphs, *Graph theory and combinatorics*, Academic Press, London (1984), 35–57.

[15] B. Bollobás, *Random Graphs*, Academic Press, London, 1985.

[16] B. Bollobás and A. Frieze, On matchings and Hamiltonian cycles in random graphs. *Random graphs '83* (Poznan, 1983), North-Holland Math. Stud., 118, North-Holland, Amsterdam (1985), 23–46.

[17] A. Bondy, Basic graph theory: paths and circuits, *Handbook of Combinatorics* **1**, Elsevier, Amsterdam (1995), 3–110.

[18] J. Böttcher, *Embedding large graphs–The Bollobás–Komlós conjecture and beyond*, PhD thesis, Technische Universität München, 2009.

[19] J. Böttcher and S. Müller, Forcing spanning subgraphs via Ore type conditions, *Electronic Notes in Disc. Math.*, **34** (2009), 255–259.

[20] J. Böttcher, K. Pruessmann, A. Taraz and A. Würfl, Bandwidth, expansion, treewidth, separators and universality for bounded-degree graphs, *European J. Combin.* **31** (2010), 1217–1227.

[21] J. Böttcher, M. Schacht and A. Taraz, Spanning 3-colourable subgraphs of small bandwidth in dense graphs, *J. Combin. Theory B* **98** (2008), 85–94.

[22] J. Böttcher, M. Schacht and A. Taraz, Proof of the bandwidth conjecture of Bollobás and Komlós, *Math. Ann.* **343** (2009), 175–205.

[23] P. Châu, An Ore-type theorem on hamiltonian square cycles, *Graphs Combin.*, published online 10 April 2012, doi:10.1007/s00373-012-1161-3.

[24] D. Christofides, P. Keevash, D. Kühn and D. Osthus, Finding Hamilton cycles in robustly expanding digraphs, *J. Graph Algorithms and Applications* **16** (2012), 337–360.

[25] D. Christofides, D. Kühn and D. Osthus, Edge-disjoint Hamilton cycles in graphs, preprint.

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

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

[28] V. Chvátal, On Hamilton's ideals, *J. Combin. Theory B* **12** (1972), 163–168.

[29] V. Chvátal, V. Rödl, E. Szemerédi and W. T. Trotter Jr., The Ramsey number of a graph with bounded maximum degree, *J. Combin. Theory B* **34** (1983), 239–243.

[30] K. Corrádi and A. Hajnal, On the maximal number of independent circuits in a graph, *Acta Math. Acad. Sci. Hungar.* **14** (1964), 423–439.

[31] B. Csaba, On embedding well-separable graphs, *Disc. Math.* **308** (2008), 4322–4331.

[32] D. E. Daykin, R. Häggkvist, Degrees giving independent edges in a hypergraph, *Bull. Aust. Math. Soc.* **23**, (1981) 103–109.

[33] G.A. Dirac, Some theorems on abstract graphs, *Proc. London Math. Soc.* **2** (1952), 69–81.

[34] J. Edmonds, Paths, trees, and Flowers, *Canad. J. Math.* **17** (1965), 449–467.

[35] G. Egorychev, The solution of the Van der Waerden problem for permanents, *Dokl. Akad. Nauk SSSR* **258** (1981), 1041–1044.

[36] P. Erdős, Some remarks on the theory of graphs, *Bull. Amer. Math. Soc.* **53** (1947), 292–294

[37] P. Erdős and A. Rényi, On random graphs. I., *Publicationes Mathematicae* **6** (1959), 290–297.

[38] P. Erdős and A. Rényi, On the evolution of random graphs, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5** (1960), 17–61.

[39] P. Erdős, Problem 9, in: M. Fieldler (Ed.), *Theory of Graphs and its Applications*, Czech. Acad. Sci. Publ., Prague, 1964, p. 159.

[40] D. Falikman, A proof of the Van der Waerden problem for permanents of a doubly stochastic matrix, *Mat. Zametki* **29** (1981) 931–938.

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

[42] A. Frieze and M. Krivelevich, On packing Hamilton cycles in ε-regular graphs, *J. Combin. Theory B* **94** (2005), 159–172.

[43] A. Frieze and M. Krivelevich, On two Hamilton cycle problems in random graphs, *Israel J. Math.* **166** (2008), 221–234.

[44] A. Frieze and M. Krivelevich, Packing Hamilton cycles in random and pseudo-random hypergraphs, *SIAM J. Discrete Math.* **26**, 435–451.

[45] A. Frieze, M. Krivelevich and P.-S. Loh, Packing tight Hamilton cycles in 3-uniform hypergraphs, *Random Structures and Algorithms* **40** (2012), 269–300.

[46] H. Gebauer, Disproof of the neighbourhood conjecture with implications to SAT, in *Proceedings of 17th European Symposium on Algorithms*, *Lecture Notes in Computer Science* **5757** (2009), 764–775.

[47] H. Gebauer, T. Szabó, G. Tardos, The Local Lemma is Tight for SAT, in *Proceedings of 22nd Symposium on Disc. Algorithms* (2011), 664–674.

[48] E. N. Gilbert, Random Graphs, *Ann. Mathematical Statistics* **30** (1959), 1141–1144.

[49] R. Glebov, M. Krivelevich and T. Szabó, On covering expander graphs by Hamilton cycles, preprint.

[50] W. T. Gowers, Lower bounds of tower type for Szemerdi's uniformity lemma, *Geometric And Functional Analysis* **7** (1997), 322–337.

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

[52] A. Hajnal and E. Szemerédi, Proof of a conjecture of Erdős, *Combinatorial Theory and its Applications vol. II* **4** (1970), 601–623.

[53] A. W. Hales and R. I. Jewett, Regularity and positional games, *Trans. Amer. Math. Soc.* **106** (1963), 222–229.

[54] H. Hàn, *Einbettungen bipartiter Graphen mit kleiner Bandbreite*, Master's thesis, Humboldt-Universität zu Berlin, Institut für Informatik (2006).

[55] H. Hán, Y. Person and M. Schacht, On perfect matchings in uniform hypergraphs with large minimum vertex degree, *SIAM J. Disc. Math.* **23** (2009) 732–748.

[56] S. G. Hartke and T. Seacrest, Random partitions and edge-disjoint Hamilton cycles, preprint.

[57] P. E. Haxell, B. Nagle, and V. Rödl, An algorithmic version of the hypergraph regularity method, *SIAM J. Comput.* **37** (2008), 1728–1776.

[58] D. Hefetz, D. Kühn, J. Lapinskas and D. Osthus, Optimal covers with Hamilton cycles in random graphs, preprint.

[59] H. Huang, C. Lee and B. Sudakov, Bandwidth theorem for random graphs, *J. Combin. Theory B* **102** (2012), 14–37.

[60] S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*, Wiley-Interscience, 2000.

[61] R. M. Karp, Reducibility among combinatorial problems, *Complexity of Computer Computations* (1972), 85–103.

[62] M. Karpiński, A. Ruciński and E. Szymańska, The complexity of perfect matching problems on dense hypergraphs, in *Proceedings of International Symposium on Algorithms and Computation*, *Lecture Notes in Computer Science* **5878** (2009), 626–636.

[63] P. Keevash, A hypergraph blow-up lemma, *Random Structures and Algorithms* **39** (2011), 275–376.

[64] P. Keevash, D. Kühn and D. Osthus, An exact minimum degree condition for Hamilton cycles in oriented graphs, *J. London Math. Soc.* **79** (2009), 144–166.

[65] P. Keevash and R. Mycroft, A geometric theory for hypergraph matchings, preprint.

[66] L. Kelly, D. Kühn and D. Osthus, A Dirac-type result on Hamilton cycles in oriented graphs, *Combin. Probab. Comput.* **17** (2008), 689–709.

[67] I. Khan, Perfect matching in 3-uniform hypergraphs with large vertex degree, preprint.

[68] I. Khan, Perfect Matchings in 4-uniform hypergraphs, preprint.

[69] H. A. Kierstead and A. V. Kostochka, An Ore-type Theorem on Equitable Coloring, *J. Combin. Theory B* **98** (2008), 226–234.

[70] H. A. Kierstead and A. V. Kostochka, A short proof of the Hajnal-Szemerédi theorem on equitable colouring, *Combin., Probability and Comput.* **17** (2008), 265–270.

[71] J.H. Kim and N.C. Wormald, Random matchings which induce Hamilton cycles, and Hamiltonian decompositions of random regular graphs, *J. Combin. Theory B* **81** (2001), 20–44.

[72] F. Knox, D. Kühn and D. Osthus, Approximate Hamilton decompositions of random graphs, *Random Structues and Algorithms* **40** (2012), 133–149.

[73] F. Knox, D. Kühn and D. Osthus, Edge-disjoint Hamilton cycles in random graphs, preprint.

[74] F. Knox and A. Treglown, Embedding spanning bipartite graphs of small bandwidth, to appear in *Combin., Probability and Comput.*

[75] J. Komlós, The Blow-up Lemma, *Combin., Probability and Comput.* **8** (1999), 161–176.

[76] J. Komlós, G.N. Sárközy and E. Szemerédi, Blow-up Lemma, *Combinatorica* **17** (1997), 109–123.

[77] J. Komlós, G.N. Sárközy and E. Szemerédi, Proof of the Seymour conjecture for large graphs, *Ann. of Combin.* **2** (1998), 43–60.

[78] J. Komlós, G.N. Sárközy and E. Szemerédi, Proof of the Alon–Yuster conjecture, *Disc. Math.* **235** (2001), 255-269.

[79] D. Kühn, J. Lapinskas and D. Osthus, Optimal packings of Hamilton cycles in graphs of high minimum degree, preprint.

[80] D. Kühn and D. Osthus, Critical chromatic number and the complexity of perfect packings in graphs, in *Proceedings of 17th Symposium on Disc. Algorithms* (2006), 851–859.

[81] D. Kühn and D. Osthus, Multicoloured Hamilton cycles and perfect matchings in pseudo-random graphs, *SIAM J. Disc. Math.* **20** (2006), 273–286.

[82] D. Kühn and D. Osthus, The minimum degree threshold for perfect graph packings, *Combinatorica* **29** (2009), 65–107.

[83] D. Kühn and D. Osthus, Hamilton decompositions of regular expanders: a proof of Kelly's conjecture for large tournaments, preprint.

[84] D. Kühn and D. Osthus, Hamilton decompositions of regular expanders: Applications, preprint.

[85] D. Kühn, D. Osthus and A. Treglown, Hamilton decompositions of regular tournaments, *Proc. London Math. Soc.* **101** (2010), 303–335.

[86] D. Kühn, D. Osthus and A. Treglown, An Ore-type theorem for perfect packings in graphs, *SIAM J. Disc. Math.* **23** (2009), 1335–1355.

[87] D. Kühn, D. Osthus and A. Treglown, Hamiltonian degree sequences in digraphs, *J. Combin. Theory B* **100** (2010), 367–380.

[88] D. Kühn, D. Osthus and A. Treglown, Matchings in 3-uniform hypergraphs, preprint.

[89] D. Kühn, R. Mycroft and D. Osthus, An approximate version of Sumner's universal tournament conjecture, *J. Combin. Theory B* **101** (2011), 415–447.

[90] M. Krivelevich and W. Samotij, Optimal packings of Hamilton cycles in sparse random graphs, preprint.

[91] M. Krivelevich and B. Sudakov, Sparse pseudo-random graphs are Hamiltonian, *J. Graph Theory* **42** (2003), 17–33.

[92] M. Krivelevich and B. Sudakov, Pseudo-random graphs, *More Sets, Graphs and Numbers*, Bolyai Society Mathematical Studies v.15, Springer (2006), 199–262.

[93] A. Lo and K. Markström, Perfect matchings in 3-partite 3-uniform hypergraphs, preprint.

[94] E. Lucas, *Récréations Mathématiques*, Vol. 2, Gautheir-Villars, 1892.

[95] W. Mantel, Problem 28, *Wiskundige Opgaven*, 10:60–61, 1907.

[96] K. Markström, A. Ruciński, Perfect matchings (and Hamilton cycles) in hypergraphs with large degrees, *European J. Combin.* **32** (2011) 677–687.

[97] C. St. J. A. Nash-Williams, Hamiltonian arcs and circuits, in *Recent Trends in Graph Theory*, Springer 1971, 197–210.

[98] O. Ore, Note on Hamilton circuits, *Amer. Math. Monthly* **67** (1960), 55.

[99] D. Osthus and K. Staden, Approximate Hamilton decompositions of regular robustly expanding digraphs, preprint.

[100] J. Petersen, Die Theorie der regulären Graphen, *Acta Mathematica* **15** (1891) 193–220.

[101] O. Pikhurko, Perfect matchings and K_4^3-tilings in hypergraphs of large codegree, *Graphs Combin.* **24** (2008), 391–404.

[102] L. Pósa, A theorem concerning hamiltonian lines, *Publ. Math. Inst. Hungar. Acad. Sei.* **7** (1962), 225–226.

[103] L. Pósa, Hamiltonian circuits in random graphs, *Disc. Math.* **14** (1976), 359–364.

[104] V. Rödl and A. Ruciński, Dirac-type questions for hypergraphs-a survey (or more problems for Endre to solve), *An Irregular Mind (Szemerédi is 70)* (2010), 561–590.

[105] V. Rödl, A. Ruciński and E. Szemerédi, Perfect matchings in uniform hypergraphs with large minimum degree, *Europ. J. Combin.* **27** (2006), 1333–1349.

[106] V. Rödl, A. Ruciński and E. Szemerédi, Perfect matchings in large uniform hypergraphs with large minimum collective degree, *J. Combin. Theory Ser. A* **113:3** (2009), 613–636.

[107] V. Rödl, J. Skokan, Regularity lemma for uniform hypergraphs, *Random Structures & Algorithms* **25** (2004), 1–42.

[108] P. Seymour, Problem section, in *Combinatorics: Proceedings of the British Combinatorial Conference 1973* (T.P. McDonough and V.C. Mavron eds.), 201–202, Cambridge University Press, 1974.

[109] B. Sudakov and J. Vondrak, A randomized embedding algorithm for trees, *Combinatorica* **30** (2010), 445–470.

[110] E. Szemerédi, Regular partitions of graphs, *Problémes Combinatoires et Théorie des Graphes Colloques Internationaux CNRS* **260** (1978), 399–401.

[111] E. Szymańska, The complexity of almost perfect matchings in uniform hypergraphs with high codegree, in *Proceedings of 20th International Workshop on Combinatorial Algorithms*, *Lecture Notes in Computer Science* **5874** (2009), 438–449.

[112] A. Thomason, Pseudo-random graphs, *Ann. of Disc. Math.* **33** (1987) 307–331.

[113] C. Thomassen, Edge-disjoint Hamiltonian paths and cycles in tournaments, *Proc. London Math. Soc.* **45** (1982), 151–168.

[114] T. W. Tillson, A Hamiltonian decomposition of $K^*_{2m}; 2m \geqslant 8$, *J. Combin. Theory B* **29** (1980), 68–74.

[115] A. Treglown and Y. Zhao, Exact minimum degree thresholds for perfect matchings in uniform hypergraphs, *J. Combin. Theory A* **119** (2012), 1500–1522.

[116] W. T. Tutte, The factorisation of linear graphs, *J. London Math. Soc.* **22** (1947), 107–111.

[117] W. T. Tutte, The factors of graphs, *Canad. J. Math.* **4** (1952), 314–328.

[118] P. Turán, On an extremal problem in graph theory, *Matematikai s Fizikai Lapok* **48** (1941), 436-452.

[119] V. A. Vatutin and V. G. Mikhailov, Limit theorems for the number of empty cells in an equiprobable scheme for group allocation of particles, *Theory Probab. Appl.* **27** (1982), 734–743.

[120] W. D. Wallis, One-factorizations, Kluwer Academic Publishers, Dordrecht, Boston, London, 1997.